Metadata

for the Long Term Preservation
of Electronic Publications

Catherine Lupovici

Julien Masanès

Bibliothèque nationale de France

Metadata

Metadata

for the Long Term Preservation

of Electronic Publications

Catherine Lupovici
Julien Masanès

This is report 2 of the NEDLIB Report Series.

content

1 Summary

The objective of this report is to define 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.

Preservation of digital documents for the long term requires above all to solve the problem of technological obsolescence. Accessing to digital documents in 20 or 100 years will be impossible if we, or our successor, can't process the bit stream underlying digital documents. We can be sure that the modality of data processing will be different in 20 or 100 years. It is thus our task to collect key information about today's data processing to ensure future access to these documents.

This report focus strictly on this kind of information. Other kind of information, descriptive, administrative or legal one is out of the scope of this work.

This report doesn't intend to cover in detail every type of digital document with its specificity. It is limited to the most generic information about digital objects.

We propose to define 8 metadata elements and 38 sub-elements following the OAIS taxonomy of information objects. A layered information analysis of the digital document is proposed in order to list all information involved in the data processing of the bit-stream.

The last metadata element is what we could call a cross-metadata in that sense that it documents every change made in the other metadata.

These metadata elements are intended to be created, as much as possible, in an automatic way to make it possible to handle large amounts of documents.

2 Introduction

The NEDLIB project defined the concept of a DSEP (Deposit System for Electronic Publications) largely based on the OAIS model.[1] The DSEP functional model is composed of the following six processes in which various metadata types are used:

- Ingest process, which receives the SIPs (Submission Information Package) that are standard packages prepared for archiving from the publications delivered or captured by the Library. The publications are checked and metadata are created. The SIPs are transformed into AIPs (Archival Information Package) and sent to the Archival storage process and the metadata are sent to the Data management process.

- Archival Storage stores the AIPs and delivers them on demand to the Access process. This process deals with the storage of the bit streams.

- Preservation ensures that the bit streams stored remain accessible to the user even if the original environment is obsolete. The Preservation module is added by NEDLIB to the OAIS reference model in order to address additional preservation actions to the single medium migration (refreshing or copying a publication) which is handled in the Archival storage module. This module is configured according to the deposit library preservation policy. The migration and emulation approaches are detailed in the DSEP model. The resulting output is either a new version of a formerly deposited publication that is re-ingested as a new AIP in the system, or it is a set of specifications for building emulators that can render a generation of publications on a future (unknown) platform. In both cases, new preservation metadata will be generated and fed into Data–Management.

- Data Management stores the metadata that are used by the Administration process to monitor the Deposit system.

- Access process makes available an electronic publication and its associated metadata through DIPs (Dissemination Information Package). The Access module is much more restrictive in NEDLIB then what is defined in the OAIS corresponding module, as the functions of managing the users and applying access controls are generally handled by the Digital Library system.

- Administration process monitors and controls the archival procedures within the DSEP system.

- The DSEP itself is connected to the general library environment through two main processes.

- Delivery and capture which receives the electronic publications and their accompanying metadata when available. This process converts the deposited publications into SIPs standard packages to submit them to the deposit system through the ingest process.

- Packaging and Delivery, which translates the user's requests into queries to the Access process. It also receives the DIPs and unpacks them for the user.

These processes are out of the scope of NEDLIB.

The different types of metadata that are used for deposited electronic publications are:

- Descriptive metadata

Descriptive metadata include the bibliographic metadata, which provide a bibliographic description of the publication and are used for retrieval purposes. They also include the structural metadata providing information about the relations between parts of publications such as serial title, issues and articles that are used for electronic collections browsing. Part of those metadata may have been created by the author

the DSEP
functional
model

1 Reference Model for an Open Archival Information System, CCSDS 650.0-R-1, Red Book, May 1999

and by the publisher. The librarians check and improve those metadata by creating links with authority files for the main access points or by organising the items into the electronic collection structure. These metadata are used directly by the users to select the electronic publications in which they are interested. Traditionally some technical metadata are recorded in the descriptive information. They are written in a textual non-standard and non-coded form. They provide a description of the original technical environment of the publication as part of its history. But they do not correspond to the metadata for preservation used in DSEP: they are not precise enough and they cannot be used for automatic processing by the DSEP.

• Administrative metadata recorded for the deposit system management purposes

• Metadata for preservation

Unlike other work on metadata for preservation and access carried out in projects such as CEDARS or PANDORA, or by RLG, this report is focussing strictly on preservation metadata and not metadata that have to be preserved. The objective is to define 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. This excludes for instance all kinds of finding aids or copyright information but also descriptive information such as image resolution or audio duration.

The information involved in long term preservation metadata is information about the data processing of the digital objects that are to be preserved. This kind of information may consist of a format name that describes a certain structure of the bit stream of the digital object (ASCII or Unicode character set for instance) or in the name of an application that can process this structure. The problem is to describe precisely this processing or/and the elements involved in it, because the modality of data processing will be different in 10 or 200 years. That is why the image format (unlike its resolution) is part of the preservation information.

We can then say that the main problem metadata for long term preservation will help to solve is the problem of technological obsolescence.

It is first necessary to document the original data processing environment of the publication exactly in a manner that allows large amounts of data items to be handled. An example of the amount of data we have in mind is the last complete harvest of the Swedish web, from spring 1999, which comprises 15 million files. Such a figure makes it impossible for a human operator to collect information about these files one by one.

It is also necessary to record the history of the technical migration or emulation strategies applied to the publication as technological environments evolve.

Finally long-term preservation requires checking the integrity of the bit stream. This can be achieved with fixity information about the data object to be preserved.

3 The OAIS Data Model

The NEDLIB Deposit System for Electronic Publications (DSEP) is based on the OAIS information model. The model was developed by the Space community as a general framework for any archive. It is applicable to a large range of different organisations responsible to provide long term access to digital information. The DSEP functional model of NEDLIB is the adaptation of the reference model to the digital libraries and digital archives context where the selection and description of digital works, the creation of access aids and the provision for user access are already provided through the library or the archive automated system. This adaptation is focussing on the storage and preservation functions.

3.1 The OAIS Information Object concept

The OAIS reference model is a functional model that allows a deposit system organisation to be described. It is not of course a metadata model in itself. But for functional modelling it defines different types of conceptual information objects, some of them specifically defined for preservation and access purposes. Each type of information object is built on a single generic model in which a data object can be interpreted by using its associated Representation Information that can be represented as follows:
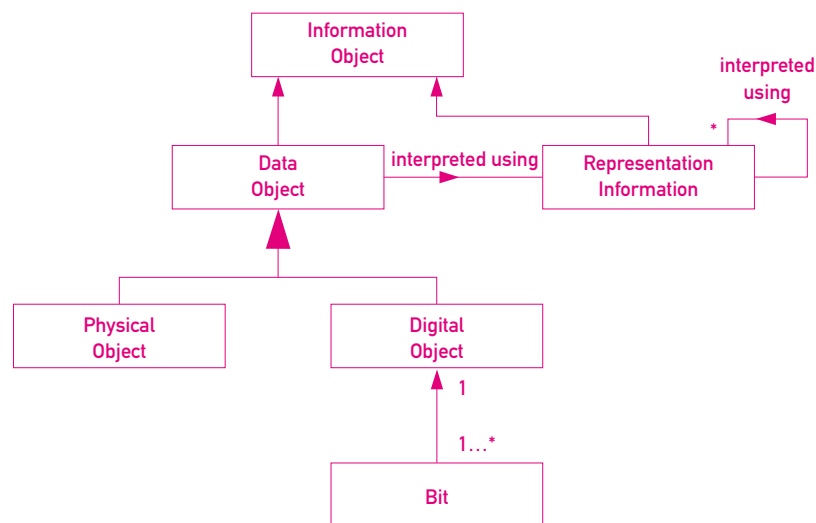
Figure 1   Information object [from OAIS Fig. 4-9, p. 4-16]

The *Data Object* may be expressed as a physical or a digital object. The *Representation Information* accompanying a digital object is used to understand the information carried by the digital object. The representation information itself is an information object that can be in digital form and needs itself representation information to be understood.

A representation information can also be composed of multiple components, each component having its own representation and the whole being described as a *Representation Network*. A Representation network for a multimedia document composed of HTML, JPEG and ASCII files, which, in order to be understood needs the standards and the mapping rules to combine the standards into the concepts required to be expressed by the document, is the following:
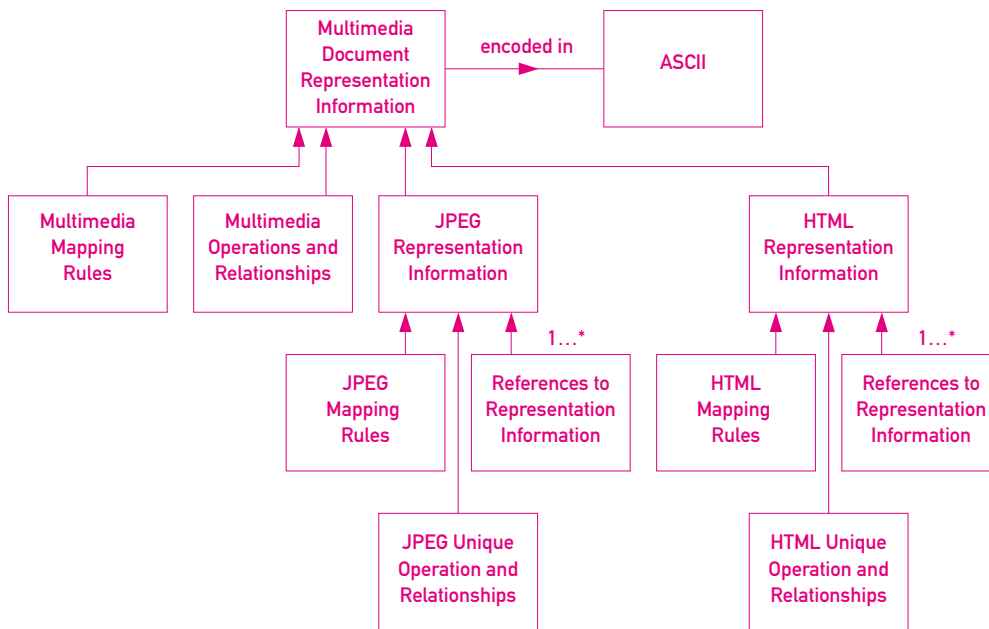
Figure 2   Representation Network Object Model [from OAIS Fig.4-12, p. 4-21]

In this example each underlying standard is supposed to be an ASCII file needing no additional representation information.

3.2 The OAIS taxonomy of Information Object Classes

The types of information objects enabling preservation and access to information within an OAIS are the following:
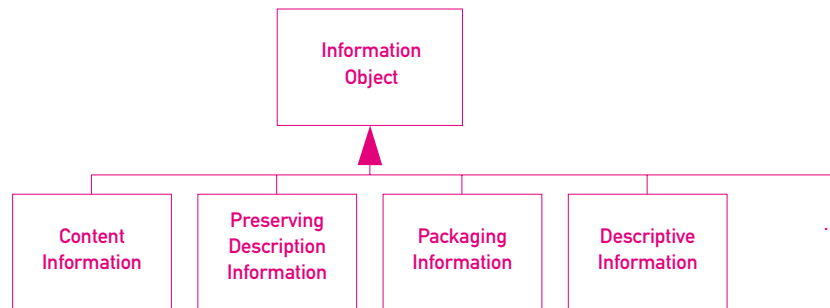
Figure 3   Information Object Taxonomy [From OAIS Fig. 4-13, p. 4-22]

3.2.1 Content Information

The Content Information is the primary information that an Archive has to preserve. It consists of the primary digital objects and all the Representation Information objects needed to provide meaningful information.

3.2.2 Preservation Description Information (PDI)

The Preservation Description Information (PDI) includes a set of Information Objects, which are necessary to preserve the Content Information with which the PDI is associated.

- **Reference Information** It identifies and if necessary describes one or more mechanisms to provide identifiers for the Content Information. The identification must allow outside systems to refer unambiguously to this particular Content Information. Part of it is replicated inside the Package Descriptions in order to enable the Consumers to access Content Information.
- **Context Information** Information documenting the relationship of the Content Information with its environment. This includes how the Content Information relates to other Content Information objects.
- **Provenance Information** Information documenting the history of the Content Information. It is used to give future users some assurance on the reliability of the Content Information. It can be viewed as a special type of Context Information.
- **Fixity Information** Information documenting the authentication mechanisms that will ensure that the Content Information object has not been altered in a non documented manner.

A PDI can only be defined once the related Content Information has been clearly determined.

3.2.3 Packaging Information

This is information that binds logically or physically the components of the package, according to the choice made by the archive to retain or not the original submission exactly as it was received.

3.2.4 Descriptive Information

This information is gathered into Access Aids allowing for search, location and ordering capabilities for the users.

3.3 OAIS archival data packaging

The OAIS model also provides the conceptual information structures for the information objects that are required to accomplish the functions of long-term preservation of information and access by the designated community of users.

3.3.1 Information Package

All those Archival Information objects could be gathered into the concept of an Information Package container for the Content Information and the associated Preservation Description Information. The packaged is viewed and accessed through its associated Descriptive Information.

The Information Package is also associated with two other types of Information Objects: the Packaging information and the Package Descriptions.
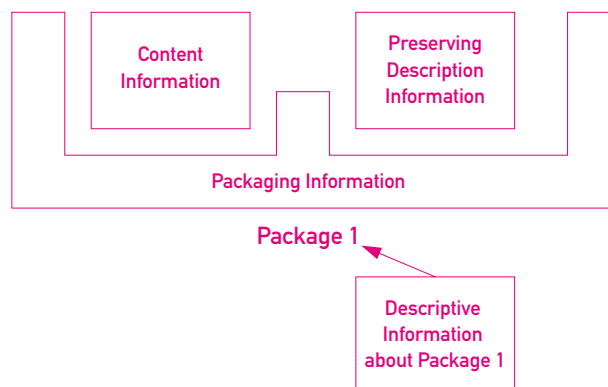

Figure 4   Information Package concepts relationships [From OAIS Fig. 2-3, p. 2-5]

3.3.2 Types of Information Packages

The model defines three subtypes of the Information Package structure.

- the Submission Information Package (SIP) which is sent to the Archival System by the Producer. In the Archival system one or more SIPs are transformed into one or more Archival Information Packages (AIP). The Preservation Description Information associated to the content could be partially provided by the producer.
- the Archival Information Package (AIP) which is the set of information having all the qualities needed for indefinite long term preservation of a designated Content Information Object and its Preservation Description Information.
- the Dissemination Information Package (DIP) which is disseminated from the archival system to a consumer. It contains generally less detailed PDI than the AIP.

3.3.2.1 The Archival Information Package

The AIP is the core logical structure for the archival system and all the Preservation Description Information Objects it contains are mandatory.
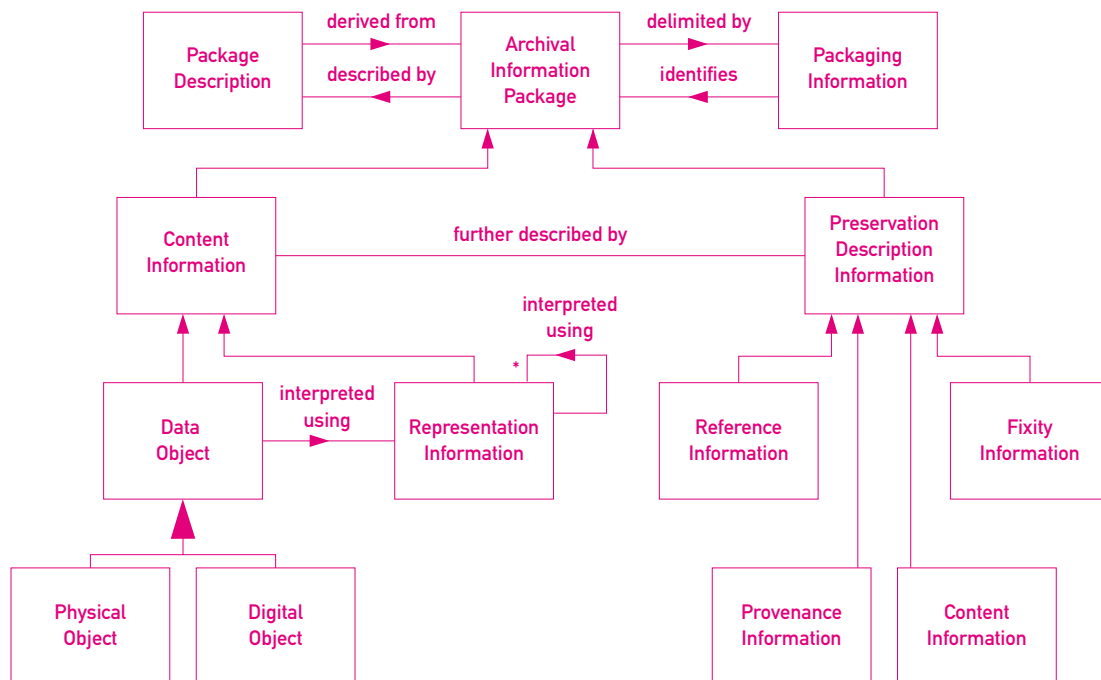
Figure 5   Archival Information Package (Detailed view) [From OAIS Fig. 4-19, p. 4-34]

3.3.2.2 Specialization of the AIP and Package Descriptions

The AIP can be an Archival Information Unit (AIU) or an Archive Information Collection (AIC) according to the complexity of its Content Information and their Package Descriptions and Packaging Information. The AIU has a single content information object described by one set of PDI. An AIC is a collection of other AICs and AIUs, each of which has its own PDI.

There are also two corresponding specializations of the Package Description into the Unit Description and the Collection Description.

4 Metadata for long term preservation

The Deposit System for Electronic Publications needs information metadata to process and monitor the actions needed for the long term preservation of the content information carried by the deposited electronic publications.

The OAIS reference model covers all the functions involved in the whole processing chain for collecting the documents, for routine storage processing, for long term preservation and for access by the target Community of users that the Archival system defines.

NEDLIB is focussed on the long-term preservation processes that have to be undertaken by the Deposit System for Electronic Publications (DSEP) in order to Archive, Preserve and Deliver the Information. As far as National Libraries are concerned, no specific community of users can be defined. National Libraries must preserve each document for anyone and any time.

The long term preservation metadata to be used within the DSEP are presented 1) according to the concept of Representation Information defined by the reference model to go along with the Information Objects and 2) according to the concept of Preservation Description Information defined along with the Content Information Objects.

4.1 Metadata for the Representation Information (RI)

A key issue for the long-term preservation of electronic publications is to maintain access to the binary information stored on a medium.

interpreting
the bit stream
The metadata for Representation Information of digital information objects are metadata related to technical means of rendering the electronic publication content and manipulation facilities for the user.

The Representation Information in the OAIS model is composed of two types of information. The structure information describes the format, the data structure of the bit stream. The semantic information is additional information that is absolutely necessary to interpret the Content Information (CI) such as the date or conditions of a scientific observation. This information is different from the context information of the PDI in this sense that the CI is meaningless without it.

For digital documents, this kind of information is in fact limited to the language in which the document is expressed. Thus for libraries, the RI consists mainly of the structure information necessary to render the content of a document.

4.1.1 Analysis of the components of the Representation Information

The following 5 layers model of information, adapted from annex E of the OAIS model, provides a basis for analysing the components of the Representation Information that must be handled for long term preservation of electronic resources.
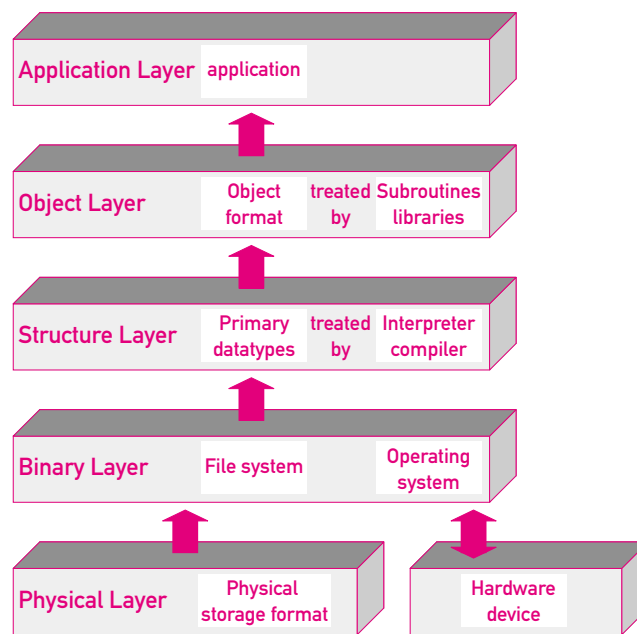
Figure 6   The layered information model

- Physical layer A digital document is first of all presented on a physical or communication medium through a physical storage format, generally standardised (for instance ISO 9660 for the CD-ROM). Device drivers and chips built into the physical storage interface manage this format in order to deliver the bit stream to the next layer.
  We have added a second part to the physical layer as stated by the OAIS annex, concerning the computer. It is not a layered view of the document itself (that is why it is separated on the schema) but of its hardware dependency.

- Binary layer At this layer the bit stream is organised into labelled blocks in a medium independent manner. The operating system, which is managing the file system, provides the functionality of this level. A file system is not necessarily specific to an operating system but generally each operating system has a favourite file system.

- Structured layer The bits are aggregated into primitive data structures to be manipulated by programmes. The compilers and interpreters of programming languages are providing the functionality at this level.

- Object layer Data are structured into objects meaningful for the application and through it to the user. The object formats could be open ones or proprietary formats.

- Application layer The application software manipulates the objects of the previous layer and presents them to the user.

From this analysis we can define the Representation Information classes necessary for long term preservation in a library context. See part 7, the list we propose for a core preservation metadata set for electronic publications. The metadata's reference will be indicated in square brackets.

technical
information

- Physical layer
  1. The physical storage format is an internal information used by the Archival Storage functional entity and is not actually part of the Representation Information. If the archive has to provide the document in its original format, the information concerning this format should be preserved as part of the PDI's provenance information [2-3].

2. As far as hardware dependency is concerned, this information is redundant with information about the operating system. Of course, in traditional cataloguing we are used to specifying minimum hardware requirements (for instance 'at least Pentium II 200 MHz and 64 Mb memory') but this is only useful in a context in which we have lower class machines that would not be powerful enough to run a program. This information will be useless in 20 or 50 years (think about this specification about a Sinclair ZX81: 'at least 16Ko of memory'). However, a specific metadata might be useful for programs using a specific set of instructions (for instance Intel's MMX instruction set) or in the case of a multimedia application using extra devices (for instance MIDI audio application). We will see more on this in the conclusion.

- Binary layer In the context of the library, information about the type of file system is useless most of the time as operating systems 'know' how to handle them. The information needed at this layer therefore is information about the Operating System (name and version) [1-2].

- Structured layer As long as the DSEP does not preserve Primary datatype nor source code, no information is needed at this level. However, in case of software deposit, if these pieces of software are not compiled, information about the interpreter or the compiler version needed is necessary [1-3].

- Object layer A DSEP may have to preserve a digital object (for instance images) and the information about the format of this object is necessary. The type of subroutine libraries is transparent for a DSEP as this object is rendered by an application using these libraries.
The information necessary at this level is therefore the object format [1-4].

- Application layer The name and version of the application needed to access the digital document may seem to be redundant with the information about the object format. In fact in some cases we can draw a direct correspondence between a specific format and a specific application (for instance the PDF format and Acrobat Reader). But in many cases there are many applications that can view a specific format (for instance JPEG viewers are numerous). And there are also cases where the format is unknown by the DSEP, which has only information about the application that can render it. This often happens in CD-ROM publications in which a specific application manipulates unknown formats. The only possible access to the content of the publication is in this case the proprietary application (for instance cdu.exe in the Universalis Encyclopaedia). In the first case, specifying the application may be useful to be sure that the version of this application is the good one. In the second case (one format-many applications), the metadata may be repeatable and this could also avoid future problems about versions of formats. In the last case (unknown format), information about the application is mandatory.
In any case, information at this level is important and should be carefully collected [1-5].

4.1.2 Two types of electronic publications and documents

This analysis makes it possible to distinguish between two types of electronic publications and electronic documents regarding possible preservation actions.

two types of electronic publications

One type corresponds to the *specific system dependent applications* where metadata on the application layer (Application [1-5]) and the binary layer (Operating System [1-2]) are mandatory to allow the long term preservation of the documents.

Among this kind of applications we can find applications using standard platform configurations (for instance Intel Pentium, SVGA, Sound-Blaster compatibility) but also applications using specific platform configurations or hardware requirements that need more detailed description (for instance MMX instructions set). In this case, and in this case only, we propose to define a hardware dependency metadata, which can concern either the microprocessor, multimedia devices or other peripheral devices [1-1-1], [1-1-2], [1-1-3].

The other type of electronic publications corresponds to *specific system independent formats* (ex. XML, JPEG) where metadata on the object layer (Object Format [1-4]) and possibly the application layer [1-5] (depending on the deposit contract) are mandatory for long term preservation.

The first category corresponds mainly to off-line publications and the second one mainly to online Web applications.

In addition we can distinguish two types of information objects that are documented by a Representation Information: a simple homogeneous object (more or less a single file) and a complex information object with several types of files inter-linked documented by a Representation Network as defined by the OAIS model.

4.2 Metadata for the Preservation Descriptive Information (PDI)

In addition to the content itself, the Task Force on the Archiving of Digital Information (ref. 9) has defined four features (reference, context, provenance and fixity) that determine information integrity. These four features form the basis of the Preservation Descriptive Information in the OAIS model. This additional information 'is specifically focused on describing the past and present states of the Content Information, ensuring it is uniquely identifiable, and ensuring it has not been unknowingly altered'.

• Reference information
In today's digital environment, there is no unique identifier usable for all kind of digital documents. In the library context, traditional identifiers such as title, name of the author, publisher and date of creation of the document might be collected. An assigned identifier, such as an ISSN, may also be part of this reference information. As long as URNs or DOIs are not widely used, the URL might be collected failing anything better. In this case the date of harvesting must be collected to have the possibility to refer to an Internet archiving snapshot.

Given this fluctuating environment, it is necessary to collect not only the value of the reference information but also its meaning or the construction methodology for the assigned identifier.
Note that a unique identifier is given to each item archived in the DSEP, its AIP identifier, but this is just an internal identifier, part of the packaging description.

• Context and provenance information
In its final report, the Task Force on the Archiving of Digital Information tries to settle means for ensuring the authenticity of digital information. To that purpose they propose to document the context of this information which includes 'a technical dimension, a dimension of linkage to other objects, a communication dimension and a wider social dimension' (op. cit. p.18). In the OAIS model, this category of information is part of the PDI but in a quite different and more limited sense.

First, the 'technical context' is excluded. Part of this information, concerning the association of logical information with physical media, is moved to the package description (see annex B of the OAIS model) but there is nothing about the other information related to hardware and software dependencies. We consider this information to be part of the representation information as stated above.

The OAIS model tries then to build an information category from the other dimensions of the Context as defined by the Task Force which is not that easy. The result is quite unsatisfactory from our point of view.

The distinction between Provenance Information and Context Information does not seem to us to be convenient (why should the indication of the publisher be considered as context information rather than provenance information?). As the OAIS says 'provenance can be viewed as a special type of context information'. It would seem more relevant to make just one information category concerning the History and the Context of the digital object, that is, all the information that does not concern the object itself as it is now (unlike the rest of the PDI, reference and fixity).

We can see that a significant part of this information, as defined by the Task Force, can only be collected by future bibliographic studies and is not relevant for strict archiving purposes. As we've already said, our purpose in this report is to define a core metadata set for preservation purpose. Additional information, particularly descriptive metadata, forms the main part of the PDI but is outside the scope of this work.

For preservation purposes, the information needed is information about the past and present states of the Content Information. Unlike other preservation metadata sets, we propose to reduce and specify the scope of this kind of information. We set aside information about administrative decisions such as the reasons for preservation (in Administrative Metadata), to focus on information concerning any change in the digital object that might imply an update in any preservation metadata already mentioned.

The information needed about this change is at least its date, its description and any existing means to reverse the change. For instance, if a migration has been carried out, it is necessary to have the date and a description of this migration (previous format, tool used for the migration) and a link to the AIP of the original document, if it has been archived.

This metadata is of course repeatable as a change (like a migration) may have an influence on several metadata. In this case, only old and new values of the secondary metadata implied need to be archived.

This metadata can be constructed as follows (with a format migration example):

Change History [2-3]					
Metadata concerned	**Date [2-3-1-1]**	**Old value [2-3-1-2] and [2-3-2-2]**	**New value [2-3-1-3] and [2-3-2-2]**	**Tool [2-3-1-4]**	**Reverse [2-3-1-5]**
Object Format [1-4]	06-06-2000	.doc	.rtf	MS-Word	MS-Word
Application [1-5]		MSWord	MSWord WordPerfect StarOffice...		
Operating System [1-2]		Windows 9X, NT, 20x MAC OS	Idem plus UNIX		
Fixity Information [2-2]		x	y		

The reverse label can refer either to the original version if it has been archived or to a tool (mostly software) that allows the backwards transformation to operate. If this transformation is irreversible, the original version of the digital object should be archived.

This metadata must at least contain the first change operated on the digital object, the migration from its original physical storage format to an internal DSEP format.

5 Metadata Management Issues

Even if this report's focus is not implementation but functional modelling, a few aspects of data management should be addressed at this point.

5.1 WHERE?

There are two main possibilities regarding the location of metadata in an archiving system. The first is to build a separate database containing information about the items that are preserved in the archive. This is the way of working in traditional libraries in which catalogues are made. At least the title, the author and some other information are indicated in the book itself, allowing the catalogue to be reconstructed if a disaster occurs. In fact, only digital information is separated from the printed one. The other possibility is to encapsulate metadata in the archive's items themselves, which is safer but harder to manage. The best solution for archiving purposes is to use both of these possibilities by duplicating metadata from the archive's item to more practical databases.

In the OAIS model, according to the concept of Archival Information Package the long-term preservation metadata and the content information should be put together as seen above. It is subject to the same preservation policy as the content itself.

For everyday use and especially retrieval of the data objects, each AIP is associated with an Associated Description. This Associated Description, which is not part of the AIP itself, provides data for different Access Aids (Finding Aids, Ordering Aids, Retrieval Aids etc.). These Access Aids can be managed in separate databases by the Data Management functional entity.

Among these Access Aids, the Retrieval Aid is of particular interest for long term preservation. According to the OAIS model, it 'translates from the unique identifier assigned by the OAIS to identify the AIP into the set of operations and filenames needed to retrieve the AIP from the file management system used in Archival Storage and returns the Content Information and PDI for the requested AIP' (see OAIS 4.2.2.3, p. 4-32).

From NEDLIB's point of view these operations may involve more than simple file management. This is the reason why we consider that the Preservation functional module is also involved at this level while in the OAIS this Retrieval Aid is only considered to be part of the Archival Storage functional area. Specific operations such as a concomitant retrieval of the Operating System and/or Application and/or Object Format specification may be necessary at this point to allow complete access to the Content Information. If we are in an emulation context, all emulation specifications are required [2] and the emulation process must be released.

This means that the Associated Description of the AIP must contain this information, at least in the form of linkage to specific AIPs containing the bit stream of software or specification. Metadata for long-term preservation as specified above [5.2] should thus be duplicated in this Associated Description. From the PDI, at least the fixity information should also be duplicated for verification purposes provided by the storage entity.

5.2 HOW?

We have seen that metadata for long term preservation are included in the AIP (as RI and PDI) and that part of them are duplicated in the package's Associated Description and are used by the Retrieval Aid. It must be clearly stated that the use of these metadata in a DSEP is an automatic one. The DSEP must be able to handle a huge amount of documents and to settle all problems due to their digital form.

---

2 In his report on 'An experiment in using Emulation to preserve Digital Publications', Jeff Rothenberg describes a specific AIP, the Emulator Specification AIP for obsolete hardware platform written in an emulator specification language. To interpret this language, an Emulator Specification Interpreter AIP is required, to run on the current version of the virtual machine (p.22 sq.).

This means that metadata are often used to release automatic processes. Given this aim, they must be in machine-readable form rather than in human-readable form. This is particularly the case for metadata for long term preservation as we have seen.

automatic processing of documents

That is the reason why Representation Information should include the indication of the AIP containing either the format description or the load image of the software. For instance, rather than only indicating 'HTML 4', the metadata should also contain a pointer to the AIP containing specifications for HTML 4 (as well as the HTML source code which indicates the path of the W3C's site containing the DTD of this version of HTML). For software, instead of only stating 'Netscape Navigator version 4.7', the metadata should also contain the unique identifier of the AIP containing this piece of software.

In this form, metadata will have the great advantage to be directly and automatically usable in the DSEP itself, without using a table of correspondence. Of course, when human intervention is needed to complete or to check metadata for instance, an automatic translation to more understandable names could be implemented.

6 Proposal For A Core Preservation Metadata Set

The following list doesn't contain a table for each metadata sub-element. When no specific comment or information is necessary (ex: Name of the Application) the name and identifier of the sub-element only appears in the table of the upper level element.

6.1 Metadata for Representation Information

NEDLIB Name	Specific Hardware requirements
Identifier	1-1
Definition	Description of non-standard platform configuration or hardware requirements
Obligatory	Mandatory only for a specific system dependent application when the content can only be viewed through an application that needs specific hardware devices
Repeatable	No
Comment	
Sub-element	1-1-1 Microprocessor 1-1-2 Multimedia device 1-1-3 Peripheral device
CEDARS relationship	
NLA relationship	

NEDLIB Name	Specific microprocessor requirements
Identifier	1-1-1
Definition	Description of specific microprocessor instructions set (for instance MMX instructions set) or co-processor
Obligatory	Mandatory only for a specific system dependent application when the content can only be viewed through an application that needs a specific microprocessor instructions set or co-processor
Repeatable	Yes
Comment	
Sub-element	
CEDARS relationship	
NLA relationship	

NEDLIB Name	Specific multimedia requirements
Identifier	1-1-2
Definition	Description of non-standard multimedia hardware requirements
Obligatory	Mandatory only for a specific system dependent application when the content can only be viewed through an application that needs non-standard multimedia devices (for instance MIDI audio applications)
Repeatable	Yes
Comment	
Sub-element	
CEDARS relationship	
NLA relationship	

NEDLIB Name	Specific peripheral requirements
Identifier	1-1-3
Definition	Description of non-standard peripheral devices (for instance a ZIP storage device)
Obligatory	Mandatory only for a specific system dependent application when the content can only be used through an application that needs non-standard peripheral devices
Repeatable	Yes
Comment	
Sub-element	
CEDARS relationship	
NLA relationship	

NEDLIB Name	Operating system
Identifier	1-2
Definition	The operating system on which the publication can run
Obligatory	Mandatory for a specific system dependent application, where the content can only be viewed through an application that needs this OS
Repeatable	Yes
Comment	
Sub-element	1-2-1 Name 1-2-2 Version
CEDARS relationship	
NLA relationship	Partially 6. Known System Requirements

NEDLIB Name	Interpreter and compiler
Identifier	1-3
Definition	Piece of programme allowing to analyse and execute each statement in a source programme or allowing to translate a programme expressed in a high level language into machine language
Obligatory	Yes only for digital documents which are source code programmes or written in a high level language
Repeatable	Yes
Comment	
Sub-element	1-3-1 Name 1-3-2 Version 1-3-3 Instruction
CEDARS relationship	
NLA relationship	

NEDLIB Name	Object format
Identifier	1-4
Definition	Name of the object format
Obligatory	Yes
Repeatable	Yes in case of a multimedia complex object or for XML's or SGML's DTD
Comment	Can be an open format or a proprietary format which is at the same time specific application dependent. For mark up languages, the DTD is part of Object format information. This metadata should then be duplicated (Object format)
Sub-element	1-4-1 Name 1-4-2 Version
CEDARS relationship	1.2.1.1.2.5 Input format and also 1.2.1.1.2.4 Output Format
NLA relationship	5. File description

NEDLIB Name	Application
Identifier	1-5
Definition	Name and version of the application
Obligatory	Yes
Repeatable	
Comment	The application can be either system dependent or system independent
Sub-element	1-5-1 Name 1-5-2 Version
CEDARS relationship	
NLA relationship	Partially 6. Known System Requirements

6.2 Metadata for Preservation and Description Information

NEDLIB Name	Reference Information
Identifier	2-1
Definition	Information that identifies the Content Information
Obligatory	Yes
Repeatable	
Comment	Part of this information is duplicated in the Package Description to enable access to the Content
Sub-element	2-1-1 Creator 2-1-2 Title 2-1-3 Date of creation 2-1-4 Publisher 2-1-5 Assigned Identifier 2-1-6 URL
CEDARS relationship	
NLA relationship	2 Date of creation

NEDLIB Name	Assigned Identifier
Identifier	2-1-5
Definition	Unique identifier that identifies the Content Information
Obligatory	Optional
Repeatable	Yes
Comment	
Sub-element	2-1-5-1 Value 2-1-5-2 Construction method 2-1-5-3 Responsible agency
CEDARS relationship	
NLA relationship	

NEDLIB Name	URL
Identifier	2-1-6
Definition	Location of the document on the World Wide Web
Obligatory	Optional
Repeatable	Yes
Comment	As the URL may change over time or even disappear, it is necessary to specify a date at which it was valid
Sub-element	2-1-6-1 Value 2-1-6-2 Date of validation
CEDARS relationship	
NLA relationship	

NEDLIB Name	Fixity Information
Identifier	2-2
Definition	Data used to prove the authenticity of an AIP
Obligatory	Yes
Repeatable	Yes
Comment	
Sub-element	2-2-1 Checksum
	2-2-2 Digital signature
CEDARS relationship	1.1.4 Fixity Information
NLA relationship	12. Validation

NEDLIB Name	Checksum
Identifier	2-2-1
Definition	Information about the use of a checksum
Obligatory	Optional
Repeatable	
Comment	
Sub-element	2-2-1-1 Value
	2-2-1-2 Algorithm
CEDARS relationship	
NLA relationship	

NEDLIB Name	Change History
Identifier	2-3
Definition	Information about every change that has occurred in the digital object and which has implied a change in any metadata for long term preservation
Obligatory	Yes
Repeatable	Yes
Comment	Part of this information is duplicated in the Package Description to be used by the Data Management entity
Sub-element	2-3-1 Main metadata concerned
	2-3-2 Other metadata concerned
CEDARS relationship	Partially 1.1.3.2.2.1 Action History
NLA relationship	Partially 13. Relationships and 23.10 Changes made by the process applied to the digital object

NEDLIB Name	Main metadata concerned
Identifier	2-3-1
Definition	Information about a change that occurred in the digital object
Obligatory	Yes
Repeatable	Yes
Comment	The main metadata concerned is the one concerning the aspect of the digital object that was intended to be changed (Object Format for a migration)
Sub-element	2-3-1-1 Date
	2-3-1-2 Old Value
	2-3-1-3 New value
	2-3-1-4 Tool
	2-3-1-5 Reverse
CEDARS relationship	
NLA relationship	

NEDLIB Name	Tool
Identifier	2-3-1-4
Definition	Tool that has been used to operate the transformation on the digital object
Obligatory	Yes
Repeatable	Yes
Comment	The DSEP must of course archive the software that has been used to operate any transformation
Sub-element	2-3-1-4-1 Name
	2-3-1-4-2 Version
CEDARS relationship	
NLA relationship	Partially 23.4 Critical software used in process applied to the digital object

NEDLIB Name	Reverse
Identifier	2-3-1-5
Definition	Designate either the content or the previous version of the digital object if it has been archived, or the tool to operate the backwards transformation
Obligatory	Yes
Repeatable	No
Comment	If the transformation is irreversible, the previous version of the digital objet should be archived
Sub-element	
CEDARS relationship	
NLA relationship	

NEDLIB Name	Other metadata concerned
Identifier	2-3-2
Definition	Information about other aspects of a change that occurred in the digital object
Obligatory	Yes
Repeatable	Yes
Comment	It concerns other aspects of the digital object that have changed (Checksum for a migration). For this metadata, only the old and new values are required (information about date and tools is already present in [2-3-1])
Sub-element	2-3-2-1 Old Value
	2-3-2-2 New value
CEDARS relationship	
NLA relationship	

references

1. Consultative Committee for Space Data Systems, *Reference Model for an Open Archival information System* (OAIS). *Red book,* 1999.
   URL: http://www.ccsds.org/RP9905/650x0r1.pdf (Last visited: 13-Jul-2000).
2. Dollar, C.M., *Archival theory and information technologies: the impact of information technologies on archival principles and methods.* Macerata: University of Macerata Press, 1992.
3. National Library of Australia, *Preservation Metadata for Digital Collections (draft)*, 1999.
   URL: http://www.nla.gov.au/preserve/pmeta.html (Last visited: 13-Jul-2000).
4. RLG Working Group on Preservation Issues of Metadata, *Final Report May,* 1998.
   URL: http://www.rlg.org/preserv/presmeta.html (Last visited: 13-Jul-2000).
5. Rothenberg J., *An Experiment in Using Emulation to Preserve Digital Publications,* The Koninklijke Bibliotheek, Den Haag, April 2000.
   URL: http://www.kb.nl/coop/nedlib/results/emulationpreservationreport.pdf
   (Last visited: 13-Jul-2000).
6. Rothenberg J., *Avoiding technological quicksand: finding a viable technical foundation for digital preservation.* Washington, D.C.: Council on Library and Information Resources, 1999.
   URL: http://www.clir.org/pubs/reports/rothenberg/contents.html (Last visited: 13-Jul-2000).
7. Russel K, Sergeant D., Stone A., Weinberger R., Day M., *Metadata for digital preservation: the Cedars project outline specification,* March 2000.
   URL: http://www.leeds.ac.uk/cedars/MD-STR~5.pdf (Last visited: 20-Jul-2000).
8. Task Force on the Archiving of Digital Information, *Preserving digital information: report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group.* Commission on Preservation and Access. Washington, D.C., 1996.
   URL: http://www.rlg.org/ArchTF/index.html (Last visited: 13-Jul-2000).