The Preservation Metadata Schema of the Military Academy Digital Library Application Software System of China

Wang Songlin

Nanjing Politics College at Shanghai_Shanghai_ China _200433
Tel. 86-21-51256536
Mail:slwang518@hotmail.com

Liu Yu

Nanjing Politics College at Shanghai_Shanghai_China _200433
Tel. 86-21-51256094
Mail:liuyudj@etang.com

Abstract. The OAIS Reference Model framework has been adopted by many organizations and institutions, including the government departments and digital libraries. This paper, upon the basis of the introduction to the Reference Model of Open Archival Information System, discusses and analyses the preservation metadata schema of the Military Academy Digital Library Application Software System of China, including its elements and sub-elements of Content Information, Preservation Description Information and Descriptive Information.

Keywords. OAIS, Functional model, Information model, MDLS, Preservation metadata schema

OAIS Reference Model,[1] as a standard, is established and regulated by the National Aeronautics and Space Administration (NASA) and Consultative Committee for Space Data System (CCSDS). The OAIS Reference Model aims at providing a reference model framework, which functions to gain access to and preserve information recourses for long-term. The framework concerned was officially defined as the ISO 14721 standard in January 2002. Since the OAIS Reference Model can be applied to any system or organization targeting to provide long-term preservation and utilization of digital resources, the OAIS Reference Model framework has been adopted by many organizations and institutions, including the government departments and digital libraries. For example, the Library of Congress, the British Library, OCLC and many other organizations applied this OAIS Reference Model to their projects. The National Library of China also follows the same approach in establishing its Chinese Metadata Schema.[2] Furthermore, the Military Academy Digital Library Application Software System of China (hereinafter referred to as the "MDLS") follows the OAIS Reference Model upon its general framework structure, including the framework, terms and concepts explored by the said Reference Model. This paper focuses on the preservation metadata scheme of MDLS, which is based upon the introduction to the OAIS Reference Model.

1 OAIS Reference Model

As a Reference Model, OAIS defines a general framework for preserving data, a framework that comprises required functions, the uniformed concepts and terms. The complete English name of OAIS is the "Open Archival Information System", among of which, the term "Open" implies that the OAIS Reference Model and any related recommendation and standard developed upon the said Model in the future, are under an open circumstance. However, this does not mean that the access to OAIS is thoroughly unrestricted. Under the Reference Model, the Archive, like the OAIS, can be explained as an information preservation organization or system, an organization or system that is responsible to preserve information for long-term and make information service function available to a designated community. An archive has the characteristics as follows:

to accept appropriate information from Information Producers;

to preserve the information concerned for long-term;

to determine its service objects;

to assure that the information preserved can be independently understood by its service objects, which means that the designated community can understand such information without the need for assistance requested to those information production specialists;

to strictly follow the policy of and procedure for information preservation;

to guarantee to safely preserve the said information against any contingencies;

to assure that the information to be disseminated is the authenticated copies of the original one, or

the one traceable to the original one; and
to assure that the information preserved is available to the designated community.[3]
The fundamental entities interacted with an OAIS can be identified from the characteristics mentioned above. The Reference Model names those fundamental entities together with their relations to OAIS, as OAIS environment, which includes Information Producer, OAIS, Consumer, and Management.[4] The Reference Model, upon the basis of these entities and relations, defines the Functional Model, the Information Model of OAIS and the respective internal details of the models concerned.

1.1 OAIS Functional Model

As far as the Functional Model is concerned, normally, the OAIS system has seven functional entities, which include Ingest, Archival Storage, Data Management, Administration, Preservation Planning, Access, and Common Services and more, the Common Services entity needs to depend upon the other six entities. Obviously, the OAIS has similar functions to digital libraries. It can be concluded that an OAIS is such a digital library that emphasizes a preservation function. This may explain why the OAIS Function Model can be applied to digital libraries, particularly till now, no "matured" or "universal" systematic structure available to the creation of digital libraries.[5]

Figure 1: OAIS Functional Entities
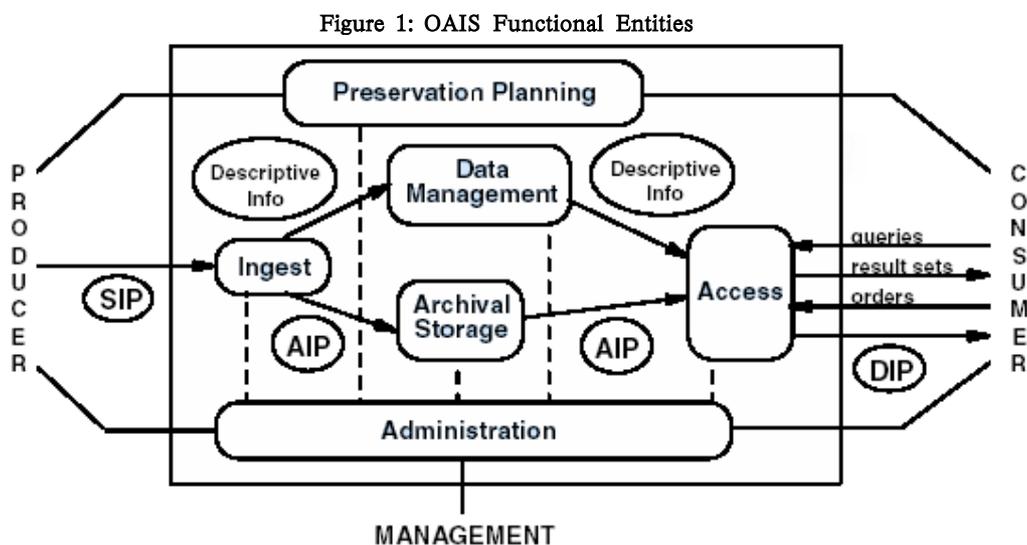
In Figure 1, the real lines describe the information communication paths, which feature a flow in both directions among the functional entities, and the dashed to Administration and Preservation Planning drawn hereunder are for clarity only.

1.2 OAIS Information Model

In Figure 1, the "IP" in SIP, AIP and DIP is the abbreviation of Information Package, a significant concept in OAIS Information Model. An Information Package consists of three types of Information as follows:
(1) the Content Information, the Physical Object or Digital Object, i.e., bits and its associated Representation Information needed for assuring the Content Data Object understandable to the designated community;
(2) the Preservation Description Information, the specific information applied to describe the featuring of the Content Information, which is divided into Reference Information, Context Information, Provenance Information, and Fixity Information; and
(3) The Packaging Information, which is to, identify, bind and connect the Content Information and the Preservation Description Information to constitute complete Information Package. For instance, supposed, the information described in the Content Information and the Preservation Description Information being saved in the same CD-ROM, the Packaging Information therein will include the volume and file structure executed by the ISO 9660 plus the information relevant to the file names and directory information in such CD-ROM disk.[6]

Figure 2: OAIS Information Package Model
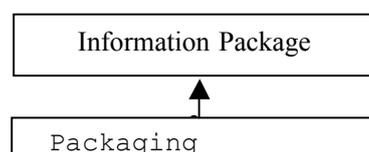
The Figure 2 shows that:

(1) an Information Package, in an OAIS, includes two types of information named as the Content Information and the Preservation Description Information accordingly and further, the Content Information and the Preservation Description Information are viewed as being encapsulated and are identified through the Packaging Information;

(2) supposed, the Information Package searched and identified by the designated community in an OAIS, such search and identification shall have to be processed by and relied upon the Descriptive Information's description to the characteristics and nature concerned. The OAIS Information Model, hereby, divides the Information Objects into the types as follows:

Figure 3: OAIS Information Object Taxonomy
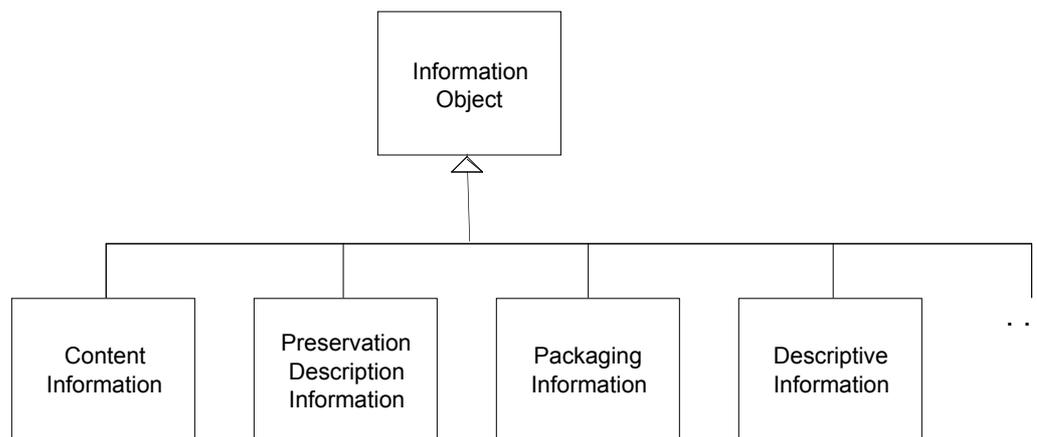
The Preservation Description Information, the Packaging Information and the Descriptive Information showed in Figure 3 are not defined as metadata in the OAIS Information Model, however, it can be stated that except for the Content Information which preserves the original object, the Preservation Description Information, the Packaging Information and the Descriptive Information describe the properties or attributes of the Content Information and the Information Package respectively, provided that we rely upon the analysis presented above. The purpose of the Preservation Description Information and the Packaging Information is to assure long-term preservation of the Content Information and as to the Descriptive Information, it aims at assisting

3

Consumers to discover the Information Package efficiently. Thus, in view of the classification of metadata, the Descriptive Information shall be the descriptive metadata and even, the Preservation Description Information and the Packaging Information shall fall within the category of metadata through the reference to the definition of metadata. A common sense concerning such opinion has been reached in the metadata theory research circle home and abroad.[7]

2 Preservation Metadata Schema of MDLS

Whereas:

(1) the OAIS Reference Model can be applicable to any system or organization that applies itself to providing long-term preservation and access to digital resources, and

(2) the OAIS Reference Model has extensively been adopted by organizations and institutions, including but not limited to, those government departments and digital libraries,

the MDLS thereby determines to design its own functional model and information model by the reference to the OAIS Functional Model and Information Model. This paper, due to a limit to the length, emphasizes the preservation metadata schema (core elements set) of the MDLS only and accordingly, such MDLS metadata schema institutes the elements and sub-elements of the Content Information or more accurately speaking, the Representation Information concerned, together with the Preservation Description Information and the Descriptive Information, which constitute the material parts of the OAIS Information Package.[8]

2.1 Elements and Sub-element of the Content Information in MDLS

As to the Content Information of the Information Package, the MDLS totally creates two elements, namely the Representation-Information and the Primary-Digital-Object. The element Representa- tion-Information further creates and establishes the Structure-Information and the Semantic- Information as its two sub-elements and the sub-element Structure-Information creates six secondary sub-elements, such as the Image, Audio and the Video, etc. These elements, sub-elements and secondary sub-elements hereof are listed in Table 1:

Table 1: Elements and Sub-elements of Content Information

Elements	Sub-elements	Secondary Sub-elements
Representation-Information	Structure-Information	Image | Audio | Video | Text | Database | Executable
	Semantic-Information	
Primary-Digital-Object		

Since the digital object is structured by one or more sequences of bits, the element Representation-Information in Table 1 is to convert such sequences into applicable and valuable information. Surely, to achieve such target, the element Representation-Information shall have to describe the Structure-Information and Semantic-Information, which applies itself to those bits sequences. In other words, the Structure-Information organizes the bits sequence into a data structure characteristic of specific meaning, such as the image, audio and the video and so forth. The Semantic-Information is to provide interpretation and explanation for the relations and specific meaning of these data structure, so as to assure that the said data structure can be converted into applicable and valuable information.

Originally, the Primary-Digital-Object in Table 1 is used for storing the bits data flow of the original digital recourses. It, however, in actual application, is replaced by the pointer that points to the external files which are used for saving primary digital object due to the separate preservation of the metadata files from the primary-digital-object in the form of external files.

2.2 Elements and Sub-elements of Preservation Description Information in MDLS

To realize long-term preservation and utilization of the Content Information, other relevant information, here, referring to the Preservation Description Information, is needed for describing the characteristic of the Content Information. As to the Preservation Description Information of the Information Package, the MDLS totally creates and establishes four elements, namely the Reference-Information, the Context-Information, the Provenance-Information and the Fixity-Information, each of which includes some sub-elements. These elements and sub-elements hereof are listed in Table 2:

Table 2: Elements and Sub-elements of Preservation Description Information

Elements	Sub-elements
Reference-Information	MUD

4

Context-Information	Relation | Existing-Metadata | Reason-for-Creation
Provenance-Information	Reason-for-Preservation | Original-Technical-Environments | Custody-History | Change-History-before-Archiving | Ingest-Process-History | Maintain-History | Negotiation-History | Digital-Publisher-Name | Digital-Copyright-Statement
Fixity-Information	Validation-Key

MUD, as the exclusive sub-element of the element Reference-Information in Table 2, refers to the Military Academic Union Digital Resources Code. The Military Academic Union Digital Resources Code is the exclusive identifier of any independently indexed resources in the Military Drilling Information Web and also, the basis of implementing and achieving series of high-level application, such as the open right and interest managements, the open linkage, and the database registration mechanism etc. Due to the special status of MUD in the entire system, MDLS does not require it to fall under the element Identifier, which falls within the scope of the Descriptive Information set below.

Since the Relation, the sub-element of the element Context-Information in Table 2, is equal to the element Relation in DC elements, details of analysis and discussion relating to the Relation shall not be necessary hereunder.

The element Provenance-Information in Table 2 makes MDLS element set distinguished from other metadata formats, such as DC and MARC and even, can be used for recording the history of the Content Information, including but not limited to its origin or source, the reason for preservation, the change history before archiving, the custody history, the ingest and process history and the maintain history. Any other information relevant to the storage, handling, and migration of the Content information, such as the original technical environments, the digital resources producer, producing date and copyright statement, also comes under the element Provenance-Information.

The Validation-Key, as the exclusive sub-element of the element Fixity-Information in Table 2, is the Message Digest value, or namely, the Digital Fingerprint of the digital resource object. The Validation-Key is an output value acquired under the process exerting MD5 algorithms with the digital resource object regarded as the input.

2.3 Elements and Sub-elements of Descriptive Information in MDLS

An Information Package, consisting of the aforesaid Content Information, Preservation Description Information and the Packaging Information, requires the Descriptive Information to describe its characteristics and nature as this will guarantee that the Consumer is able to search out the Information Package and to assess the value of that information. As to the Descriptive Information of an Information Package, the MDLS creates fifteen elements in total, including the Identifier, Title, Subject, Description, Type, Publisher, Creator, Contributor, Source, Date, Language, Coverage, Format, Support and the Security and only some of them include sub-elements. The elements and sub-elements are listed in Table 3.

Table 3: Elements and Sub-elements of Descriptive Information

Elements	Sub-elements
Identifier	ISBN | ISSN | ISMN | ISRC | ISRN | CSSN | CODEN | SPN | URI | URL | DOI | Article-Identifier | U-D-I
Title	Chief | Parallel | Uniform | Other-Title
Subject	Keyword | MSC | CLC | LCCAS | DDC | U-D-C
Description	Abstract | Contents | Version | Audience | Award | Common
Type	
Publisher	
Creator	
Contributor	
Source	Source-Title | Year | Volume | Number
Date	Available | Created | Issued | Modified | Valid
Language	Text-Language | Caption-Language | Dub-Language
Coverage	
Format	Extend | Media-Type | System-Req
Support	Project | Branch | Fund
Security	Secret-Level | Authorization

In Table 3, since the first 13 elements, also namely the Identifier, Title, Subject, Description, Type, Publisher, Creator, Contributor, Source, Date, Language, Coverage and the Format, are corresponding with the related elements in DC, details of analysis and discussion is unnecessary hereunder. In addition, the element Support records the background supporting information that is closely connected with the creation of resources, including but not limited to, the project the resources affiliated to, the branch that provide support to the content producer, and the fund sponsor granted for that content producer to produce resources.

The element Security in Table 3 documents the information required by exercising security control upon the resource usage, which includes the secret-level and the authorization. The former one identifies the safety level of the resources, while the latter, in the form of code, defines the operation permission extent of the resources against a designated user. The complete control information provided for the digital right and interest management system shall be available by the marriage of the two elements.

Further illumination needs to be presented hereunder:

(1)The Chinese metadata Schema adopts such an approach as exploring the Descriptive Information of MDLS to fall under the element Reference-Information of the Preservation Description Information. This approach is in accordance with the opinion that the Descriptive Information is generally derived from the Content Information and the Preservation Description Information;[9] and

(2) The MDLS concerns the Descriptive Information together with the Preservation Description Information and Content Information as the sub-elements of the Information Package. This means that the Information Package is the root element of all the aforesaid information (elements & sub-elements included) in MDLS.

3 Example and Conclusions

3.1 Example

To give a deeper impression of preservation metadata schema of MDLS, part of the Information Package represented using Extensible Markup Language (XML) is given as an example as follows.

```
<?xml version="1.0"encoding="UTF-8">
<!DOCTYPE Information-Package SYSTEM"_____.dtd">
<Information-Package>
  <Description-Information>
    <Identifier>
      <ISSN Field="ISSN">1000-3428</ISSN>
      <CODEN Field="CODEN">JISGEV</CODEN>
      <CSSN Field="CSSN">CN-31-1289/TP</CSSN>
    </Identifier>
    <Title>
      <Chief Field="__">
        _____
      </Chief>
      <Parallel Field="_____">
        Study of Compression Algorithm in Nuclear Fusion Experiment Data Acquisition
              System
      </Parallel>
    </Title>
    <Subject>
      <Keyword Field="_____">
        ____    _____             LZO
      </Keyword>
      <Keyword Field="_____">
        Nuclear fusion; Data acquisition and issuance; LZO
      </Keyword>
      <CLC Field="_____">TP301.6</CLC>
    </Subject>
    <Description>
      <Abstract Field="_____">
```

```
                    _____
                    _____LZO_____
                    _____
                    _____
```
</Abstract>
<Abstract Field=”_____”>

This paper analyzes some common compressing algorithms in application; a kind of compressing algorithm, LZO, which fits the nuclear fusion experiment acquisition and issuance, is accepted. The performance evaluation of LZO application with HT7 case is given in the paper.

</Abstract>
</Description>
…………
</Description-Information>
<Preservation-Description-Information>
  <Reference-Information>
   <MUD Field=”MUD”>80200300000786</MUD>
  </Reference-Information>
…………
</Preservation-Description-Information >
<Content-Information>
…………
</Content-Information >
</Information-Package>

3.2 Conclusions

In the previous practices, metadata had always been defined as supporting resource discovery while OAIS has broken through such orientation, namely that metadata supports the discovery of the resources and simultaneously, the long-term preservation of resources. Moreover, the extension of such function of metadata needs only, based upon the original metadata set, such as the DC, to supply and add relevant elements of the Content Information and the Preservation Description Information. The MDLS just does some attempts in this aspect. The OCLC/RLG also has formally explored the concept of preservation metadata and intends to execute a standard framework of preservation metadata upon the basis of OAIS Reference Model.[10] It is concerned that those researches and practices, home and aboard, force us to attach more importance to the issue of the digital resource long-term preservation in the future creation and construction of digital libraries. In other words, the data long-term preservation function should be given the same emphasis and address as the data discovery function in designing a metadata set.

References

1 Reference Model for an Open Archive Information System. CCSDS 650.0-B-1. Blue Book.http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf

2 _____. _____. 2001-06
 National Library of China. Chinese Metadata Schema. 2001-06

3 Brian Lavoie. Meeting the challenges of digital preservation : The OAIS reference model.
http://www.oclc.org/research/publications/newsletter/repubs/lavoie243

4 ＿＿＿, ＿＿＿, ＿＿＿. OAIS_____. _____, 2003(4)
Xu, Zhouya … [et al.] OAIS Reference Model and Chinese Metadata Schema. New Technology of Library and Information Service, 2003(4)

5 __. _____ : _____. _____, 2002(3)
Zhen, Zhen. Digital Libraries : Their Architecture and Issues and Analysis. Library Journal, 2002(3)

6 Reference Model for an Open Archive Information System. CCSDS 650.0-R-2. Red Book, July 2001. http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html

7 ____, ____, ____. OAIS_____. _____, 2002(3)
Niu, Jinfang … [et al.] OAIS and Digital Library. Library and Information Knowledge, 2002(3)

8 MDLS_____. _____MDC_____, 2003-11
MDLS Working Group. Description Guide for Military Academy Digital Objects. 2003-11

9____, ____. _____. _____, 2002(2)
Niu Jinfang … [et al.] On the Preservation Metadata for Digital Objects. Journal of Academic Libraries, 2002(2)

10 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. http://www.oclc.org/digitalpreservation/presmeta_wp.pdf