Feature Analysis of Metadata Schemas for Records Management and Archives from the Viewpoint of Records Lifecycle

기록 생애주기 관점에서 본 기록관리 메타데이터 표준의 특징 분석

Jae-Eun Baek(백재은)*
Shigeo Sugimoto(스기모토 시게오)**

목 차

<ABSTRACT>

Digital resources are widely used in our modern society. However, we are facing fundamental problems to maintain and preserve digital resources over time. Several standard methods for preserving digital resources have been developed and are in use. It is widely recognized that metadata is one of the most important components for digital archiving and preservation. There are many metadata standards for archiving and preservation of digital resources, where each standard has its own feature in accordance with its primary application. This means that each schema has to be appropriately selected and tailored in accordance with a particular application. And, in some cases, those schemas are combined in a larger frame work and container metadata such as the DCMI application framework and METS. There are many metadata standards for archives of digital resources. We used the following metadata standards in this study for the feature analysis me metadata standards - AGLS Metadata which is defined to improve search of both digital resources and non-digital resources, ISAD(G) which is a commonly used standard for archives, EAD which is well used for digital archives, OAIS which defines a metadata framework for preserving digital objects, and PREMIS which is designed primarily for preservation of digital resources. In addition, we extracted attributes from the decision tree defined for digital preservation process by Digital Preservation Coalition (DPC) and compared the set of attributes with these metadata standards. This paper shows the features of these metadata standards obtained through the feature analysis based on the records lifecycle model. The features are shown in a single frame work which makes it easy to relate the tasks in the lifecycle to metadata elements of these standards. As a result of the detailed analysis of the metadata elements, we clarified the features of the standards from the viewpoint of relationships between the elements and the lifecycle stages. Mapping between metadata schemas is often required in the long-term preservation process because different schemes are used in the records lifecycle. Therefore, it is crucial to build a unified framework to enhance interoperability of these schemes. This study presents a basis for the interoperability of different metadata schemas used in digital archiving and preservation.

Keywords: digital resource, lifecycle of records, recordkeeping, metadata for archives, metadata for preservation, metadata schema

* Graduate School of Library, Information and Media Studies, University of Tsukuba, Japan Ph.D Student(jaeeunb@slis.tsukuba.ac.jp) (main author)
** Graduate School of Library, Information and Media Studies, University of Tsukuba, Japan Professor(sugimoto@slis.tsukuba.ac.jp)

■ 접수일자 2010년 11월 24일 ■ 수정일자 2010년 12월 21일 ■ 게재확정일자 2010년 12월 22일

〈초 록〉

인터넷과 컴퓨터의 발전으로 다양한 환경이 끊임없이 제공되고, 이로 인해 대량의 디지털 리소스가 축적, 발신되고 있다. 이는 다양한 문제를 가져왔고, 우리는 디지털 리소스를 미래에 이용가능하도록 지속하고 보존하기 위한 기본적인 문제에 직면하게 되었다. 디지털 리소스를 장기간 보존하기 위해서는 리소스에 적합한 보존 방침과 방법이 필요하고, 따라서 여러 스탠다드가 개발되고 사용되어지고 있다. 메타데이터는 디지털 리소스를 장기간 유지하기 위한 디지털 아카이브에서 가장 중요한 구성요소 중 하나 이다. 디지털 리소스의 아카이빙과 보존을 위해 사용되는 메타데이터는 많이 있다. 그러나 각각의 스탠다드 는 주된 어플리케이션에 따라 각각의 특징을 가지고 있다. 이는 각각의 스키마가 특정한 어플리케이션에 따라 적절하게 선택하고 맞춰지지 않으면 안 되는 것을 의미한다. 경우에 따라서는DCMI의 어플리케이션 프레임워크와 METS와 같이, 스키마는 거대한 프레임워크와 컨테이너 메타데이터로 결합되어 있다. 다양한 메타데이터가 있는 가운데, 본 논문에서는 아카이브를 행하기 위해 용이되어 있는 메타데이터 스키마로, 공문서 혹은 행정문서등의 아카이브를 위해 기술하고 있는 ISAD(G), 디지털 리소스를 위해 작성된 EAD, 보존한 디지털 리소스를 위해 메타데이터 프레임워크를 정의하고 있는 OAIS, 디지털 리소스의 보존을 위한 PREMIS, 그리고 리소스의 관리와 검색을 위해 작성된 AGLS Metadata를 사용하여, '보존해야 되는 리소스에 하나의 메타데이터만을 선택해서 이용한다면 어떠한 문제가 생기는 가'라고 하는 의문을 바탕으로 접근하였다. 본 논문은 기록 생애주기 모델을 기초로, 스탠다드의 특징분석을 통해서 알게 된 메타데이터 스탠다드의 특징을 보여주고 있다. 특징은 이들 스탠다드의 메타데이터 기술요소가 기록 생애주기에서의 작업(task)에 관련하는 것을 간단하게 단일의 프레임워크로 보여줬다. 메타데이터 기술요소의 상세한 분석을 통해서, 우리는 기술 생애주기의 단계와 기술요소 간의 관계의 관점에서부터 스탠다드의 특징을 확실하게 할 수 있었다. 메타데이터 스키마간의 매핑은 다른 스키마가 기록 생애주기에서 사용되기에 장기 보존과정에 있어 자주 요구된다. 따라서 이러한 스키마의 상호운용성을 향상시키기 위해서는 통일된 프레임워크를 구축하는 것이 중요하다. 이 연구에서는 디지털 아카이빙과 보존에 사용되는 다른 메타데이터 스키마의 상호운용성을 기초로 제시한다.

주제어: 기록 생애주기, 디지털 리소스, 레코드 라이프 사이클, 레코드 키핑, 메타데이터 스키마, 보존을 위한 메타데이터, 아카이브를 위한 메타데이터, 아카이빙

1. Introduction

In our modern information environment, our daily lives heavily rely on the digital information resources that have drastically increased since 1990s. The increased usage of digital resources has brought us serious demands to preserve the digital resources over time, even though the media on which information resources are stored is continuously changing and it is widely known that the preservation of digital resources is not straight forward.

Memory organizations such as archives and libraries already have vast amount of digital resources which may be born digital or turned digital from physical objects. Those memory organizations which are responsible for the long-term management and preservation of digital resources are keen to develop systems for digital preservation. Not only the memory organizations but also governments, industries and universities need to preserve highly valuable resources for future. However, it becomes more and more difficult to maintain digital resources as time goes. For long-term preservation of digital resources, proper preservation policies and strategies are necessary. Many factors have to be taken into account to develop the policies and methods - evaluation and prioritization to select resources for preservation, laws and regulations for digital preservation, preservation technologies such as migration and emulation, metadata schemas for digital preservation (National Library of Australia 2001). In general, preservation policies and strategies have to be clearly defined in accordance with the type of resources to be preserved and the purpose of preservation.

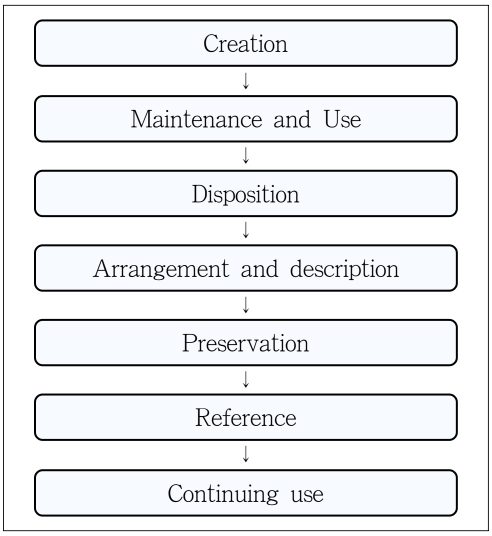
On one hand, a number of factors in different aspects have to be examined in order to preserve digital resources. On the other hand, it is too complicated to examine all of the factors at the same time. In this paper, we study metadata for preservation and archiving, which is widely recognized as a very important issue for digital preservation. There are several classes of metadata schema standards for preservation of digital resources, and every class is defined based on a purpose, feature, role, and so forth. It is a straightforward question whether a single standard is sufficient for digital resource preservation. If we have to use multiple metadata schemas we have to have an appropriate framework to enhance the interoperability between the schemas. In a practical environment, more than one metadata standard are frequently used in a single system, e.g. descriptive metadata, administrative metadata and technical metadata. From another viewpoint, it is crucial to record information about a resource from the moment when the resource is created and to maintain the information in accordance with tasks required in every stage of the lifecycle of the resource. Thus, we naturally use more than one metadata schema in the record management and archiving process. This paper aims to clearly identify the relationships among metadata standards for archiving and record management by comparing the metadata standards in detail from the viewpoint of the "Lifecycle of Records".

2. Background - Metadata and Records Lifecycle

Metadata for preservation is defined as the information to describe, manage and identify the structure of digital resources in order to preserve the resources over time (後藤敏行 2007, 74). This information is used to support the process of digital preservation. There are many kinds of metadata standards for resources. AGLS Metadata is defined to improve search of both digital resources and non-digital resources. ISAD(G) is a commonly used standard for archives. EAD is well used for digital resources. OAIS, which defines a reference model for archival systems, defines a metadata framework for preserving digital objects. PREMIS, which is a new standard, is designed primarily for preservation of digital resources. The Decision Tree which is developed by the Digital Preservation Coalition (DPC)[1] for the digital preservation process provides a set of attributes that have to be examined for preservation. In this study, the set of attributes extracted from the decision tree are used as a metadata schema to help preservation tasks, although the Decision Tree is not designed as a metadata standard.

Records lifecycle defines the stages of a lifecycle of a record for records management and the tasks in each stage. Governments and archival organizations use the records lifecycle model to appropriately keep track of the resources. When resources come to the appraisal stage for archiving, the organizations need to decide whether to retain or destroy the resources (Government of South Australia). The lifecycle model of the US National Archives and Records Administration (NARA) is used as the base lifecycle model in this paper because it is widely know and applied to various official and historic resources in the USA.

1) The Digital Preservation Coalition was established in 2001 to foster joint action to address the urgent challenges of securing the preservation of digital resources in the UK. The DPC offers a generic advice service and provide guidance (e.g. Preservation Handbook, Technology Watch Reports, DPC Annual Reports etc).



〈Figure 1〉 Lifecycle of NARA

2.1 Metadata Schemas for Record Management and Archives - AGLS, EAD, ISAD(G), OAIS, PREMIS

(1) AGLS Metadata (Australian Government Locator Service metadata)

AGLS Metadata defined by the Australian Government based on Dublin Core, contains 19 descriptive elements. It was designed to facilitate, discover and search resources by users online (National Archives of Australia 2006).

(2) EAD (Encoded Archival Description)

EAD (The Library of Congress 2002) is a metadata schema for archiving digital resources, keeping compatibility with ISAD(G) (The Library of Congress 2002). In addition to the content description of digital resources, EAD has the elements for structural description.

(3) ISAD(G) (General International Standard Archival Description)

ISAD(G) (International Council on Archives 2000) is originally designed for archived resources in traditional archives and is not specific to digital resources. ISAD(G) is applied to descriptions of all kinds of resources in archives. ISAD(G) expresses the type of a resource, the source organization of the resource, storage information of the resource and the history of the resource. ISAD(G) also describes information about collection, storage period, usage, copy condition, description element for context of resource, etc (白才恩 et al. 2007, 17).

(4) OAIS (Open Archival Information System)

OAIS (Consultative Committee for Space Data Systems 2002) is an international standard for preservation of digital resources. OAIS is a reference model for archive systems to guarantee to access (Harvard University Library 2008). The OAIS reference model outlines the functions required to access the information objects and guarantee efficient long-term preservation (Digital Curation Centre 2009). The most fundamental features of metadata scheme in OAIS are that OAIS clearly splits digital object and representation information and that it defines four categories of metadata required as preservation description information - provenance, context, reference, and fixity (JuhaHakala 2001).

(5) PREMIS (Preservation Metadata and Implementation Standard)

PREMIS (Online Computer Library Center 2008) is a new metadata schema for preservation of digital resources. PREMIS defines a data model of instances which are subject to metadata description for preservation. The description elements are defined in its data dictionary. The PREMIS data model consists of five entities – intellectual entity, digital object, agent, rights and event (Online Computer Library Center 2008). Unbundling of intellectual entity from a digital object is a crucial feature for digital resources because a digital resource in a particular format is frequently converted into another format without changing its intellectual contents. In addition to intellectual entities and digital objects, the PREMIS data dictionary defines the elements for the rights, agents and events.

2.2 Decision Tree for Long-term Retention as a Metadata Scheme for Digital Preservation

Digital Preservation Coalition (DPC) is an organization to promote information sharing and activities for long-term access of digital resources to reduce the obstacles in the way of preservation of resources. DPC has been working for preservation of digital resources from various viewpoints, and has suggested the guidelines for digital preservation in the Digital Preservation Handbook (DPH).[2] DPH shows a *decision process* for the selection of digital

2) The handbook provides an internationally authoritative and practical guide to the subject of managing digital resources over time and the issues in sustaining access to them. It has been developed and maintained the Digital Preservation Coalition.

materials for long-term retention, which is called Decision Tree. The decision process shows an evaluation process for the resources in the form of *Questions and Choices* (Digital Preservation Coalition 2006). The questions and choices assist in the ultimate decision to accept or reject long-term preservation responsibility.

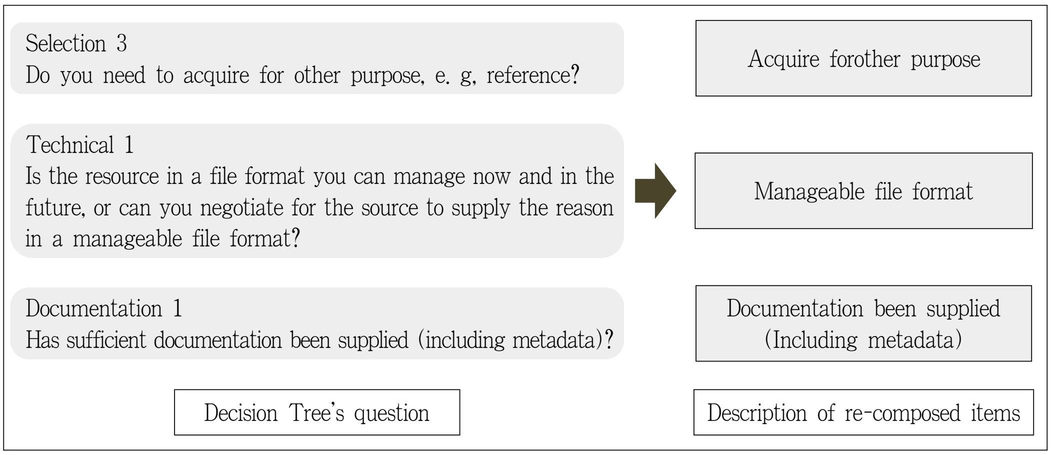
The decision tree is divided into the four sections. Each section is expressed as a sub-tree of the whole process. The decision tree is composed of three sections - *Rights & Responsibility, Technology & Metadata, Documents & Costs.*

The decision tree is composed of questions and answers – a question is a node and an answer is an edge coming from the node. An advice may be attached to a node as an answer to the question. We can evaluate resources and find appropriate technology and strategies for preservation. The decision tree does not have attributes as a metadata schema because it is not designed as a metadata standard but it has a set of questions as a tool to help choose a preservation strategy. The questions contain crucial semantic attributes to choose appropriate technology or method for preservation at every decision point. Therefore, a semantic attribute in a question can be transformed into a metadata attribute. Thus, the answers to a question are the value of an attribute or a class of values for the attribute.

In this paper, the DPC decision tree, from which we extract metadata attributes, is regarded as a metadata standard as well as other standards described in the previous section. For the conversion of the decision tree into a metadata schema, we extracted phrases from question statements of the decision tree, and then, we organized them into descriptive elements. The method of extracting phrase of question statements of the decision tree is as follows:

1. Identify the semantic feature in each question that is a node of the decision tree one at a time.
2. Extract a key word or a phrase from the question.
3. Reorganize the extracted key words and phrases into description elements of a metadata schema.

In this way, we got 27 attributes from the set of questions in the whole decision tree. For example, from a question in selection 3 of the decision tree "Do you need to acquire for other purpose?", we extracted the metadata element "Acquire for other purpose". The extraction and reorganization process are explained in detail in chapter 5.2.



〈Figure 2〉 Digitized Decision Tree's question

2.3 Lifecycle of Records - Lifecycle of NARA

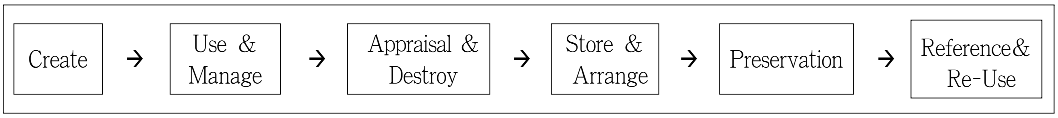
Huge amounts of documents and records are created and disseminated everyday by various organizations and institutions. All of those resources are created, used, preserved and destroyed in accordance with the management process determined by the organizations (The National Archives and Records Administration). Thus, each resource has a lifetime composed of a set of stages known as a records lifecycle.

The model of the records lifecycle used in this paper is based on that of the National Archives and Record Administration (NARA) of the US government. As shown below, the NARA's records lifecycle has seven stages defined independently from any resource types, e.g. digital resources, official documents, archives and national records, and also from any media types, e.g. pictures, maps, photos, and videos. The paragraphs below explain the stages of the NARA lifecycle.

1) Creation
   Records are created by persons or departments that belong to various organizations and institutions.
2) Maintenance and use
   While in use, the record is collected, arranged and stored with similar records.
3) Disposition
   Records are kept according to the record schedule in the organization. And a record is evaluated at this stage. The records appraised are permanently preserved in the National Archives.

4) Arrangement and description
   Administrative information (metadata) is given to the records according to the management policies of the National Archives.
5) Preservation
   Records should be preserved without losing anything. Meanwhile, there are additional cases to change the media.
6) Reference
   Supply the records preserved to provide search and reference services.
7) Continuing use
   Proper management and continuing use of preserved records is promoted.

In our study, we revised the NARA lifecycle into six stages as shown in figure 3.



〈Figure 3〉 Lifecycle of this research

3. Related works

1) Create Once, Use Many Times: The Clever Use of Recordkeeping Metadata for Multiple Archival Purposes

In this research, development of metadata for multiple archival purposes and relevance to future archival systems are analyzed and explored using the Clever Recordkeeping Metadata (CRKM) Project. This project explains the interoperability, and the theory of the Records Continuum as a conceptual frame work. The Monash CRKM Project explains about the challenges of automating metadata creation and sharing metadata between business systems, current recordkeeping system and archival systems (Joanne Evans et al. 2005, 17).

The theory of the Record continuum is used as a conceptual explanation. And recordkeeping metadata, ISAD(G), EAD and Australian Recordkeeping Metadata Schema, etc are also referred to. The relation of the records continuum and metadata for recordkeeping and archives is not mentioned in this research.

2) A survey of Techniques for Achieving Metadata Interoperability

This survey first describes the metadata used in current information systems and its concepts. And then, metadata interoperability and its problems are explained. Especially, the metadata is divided into four blocks using four viewpoints - metadata, model, meta model, meta-meta model (Bernhard Haslhofer et al. 2010, 30). According to each of these blocks, various metadata standards and metadata mappings and their techniques are explained to discuss is a study of metadata interoperability from different viewpoint. This survey paper gives hints to compare and mapping between metadata schemas performed in the study described in this paper.

3) Metadata for Preservation: A Review of Recent Developments

This research describes recent developments relating to digital preservation metadata, and introduces Digital preservation problems, and the importance of metadata for preservation strategies. Specially, the paper explains features of "Library-Based Projects", and projects that relate to preservation, archives and metadata formats for recordkeeping. It also describes the taxonomy of Information object class defined by 'The OAIS Reference Model' and some developments in the records domain and archives (Michael Day 2001, 165). To Review the digital preservation and research on the important of metadata for preservations can help us make it clear for our research background.

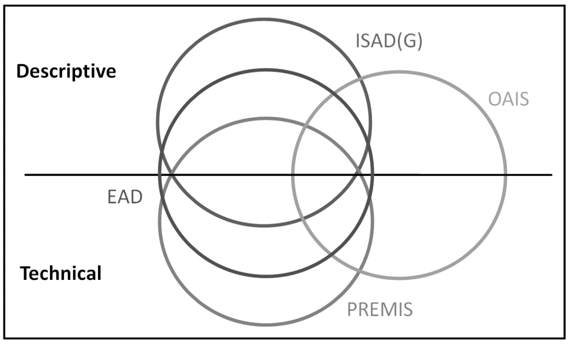
4) Metadata Interoperability and Standardization - A Study of Methodology Part 1

This research explains interoperability problems with multiple metadata, such as having the same subject domain and the resources of same type. It then explains three levels - Schema level, Record level, Repository level - from the same interoperability viewpoint. The six methods - derivation, application profiles, crosswalks, switching-across, framework and registry - are explained to show metadata interoperability with examples (Lois Mai Chan et al. 2006).

4. Properties and Characteristics of Metadata for Archives

Every metadata standard for archives has its own set of elements and controlled vocabularies. We first present the types of metadata and then explain the characteristics of metadata. In this study, we split the metadata schemas into descriptive metadata and technical metadata.

Metadata description contains elements such as title, creator, related resource, history of a resource. Technical metadata explains the technical features of a resource, such as data for management, format, media, hardware and so on. The paragraphs below show the details about descriptive and technical metadata.



〈Figure 4〉 Characteristics of metadata for archives

First, ISAD(G) contains descriptive elements of resources in an appropriate granule, i.e., fond, sub-fond, series, file, and item. EAD and OAIS have elements to describe intellectual contents, structural features, administrative and technology information. Intellectual contents are obviously descriptive metadata and technology information is technical metadata. Structural and administrative information have both descriptive and technical features. PREMIS have many elements to describe technical features and the structure of digital resources. Figure 4 shows the features of these four standards (白才恩 et al. 2007, 18).

By the brief analysis above for the metadata elements, we have shown that, on one hand, these metadata schemas have common features among each other, but on the other hand, they have different features determined by their objectives and purposes of description in the scope of archival metadata. This means that it is crucial to select and use appropriate metadata standards and combine them appropriately for designing a metadata schema for a specific archival system. In other words, the crucial metadata issues for the archival system are to create mapping between the lifecycle stages and metadata standards and to create mapping between metadata elements of different schemas used in the system. Therefore, a unified framework to enhance interoperability of metadata standards is crucial for digital preservation and archiving.

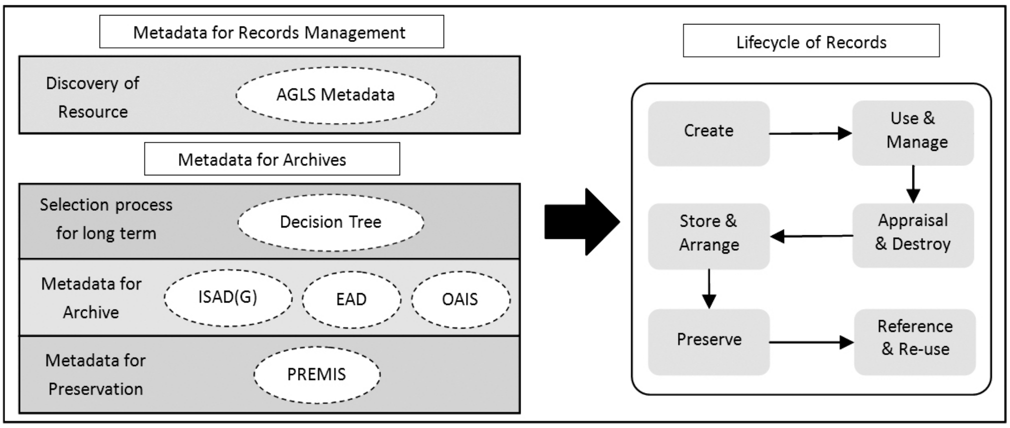
5. Using the Records Lifecycle as an Approach to Feature Analysis

There are many common key aspects in archival and preservation metadata standard, e.g., recording management, preservation of digital resources and so on. On the other hand, every metadata standard has its own features in accordance with its purpose - descriptive elements, structural constraints, base data models, usage guidelines and so forth.

5.1 Identifying Primary Stages of Metadata Element

Mapping of metadata standards into the records lifecycle is examined in this study to explicitly extract and compare the features of metadata schemas used in digital archives and preservation. For the mapping, it is necessary to extract descriptive elements from a metadata schema, and then to examine in which stage of the lifecycle the value of each element is determined.

A metadata record is created at some point and used in the whole records lifecycle. In this study, we define a *primary stage* of a metadata element in the records lifecycle as the stage where the metadata element is given an initial value or revised. A metadata element may have one or more primary stages. For example, as the value of *creator* element of a resource is determined when the resource is created, the primary stage of the creator element is the first stage of the lifecycle, i.e. "create" in Figure 5. Even if the *creator* element is very frequently used in the later stages, the primary stage is "create". If the value is revised (or updated) in a later stage in the lifecycle, the stage is also a primary stage of the element.



〈Figure 5〉 Lifecycle and Metadata for archives

5.2 Mapping Archival Metadata Standards into Lifecycle Model - Analysis of Features of Archival Metadata Standards

"Mapping metadata standards into the records lifecycle" is done in two steps in this study:

Step1. Extract every metadata element from each metadata schema standard one by one, and determine the *primary stages* in the records lifecycle for the element.

Step2. For each metadata schema, determine its *primary stage set* in the lifecycle where the *primary stage set* means a set of stages in which the majority of the metadata elements are given their values or revised. This step requires overviewing of the metadata element sets across the stages of the lifecycle.

In this study, we examined AGLS, ISAD(G), EAD, OAIS, PREMIS and the DPC's Decision Trees. We examined AGLS as a well known schema for resource discovery of governmental documents and services in order to compare its features with the metadata schemas for archives and preservation. Following the steps shown above, we examined all of the six schemas. The full result is shown in the appendix and the following sections explain the Step 1 and 2 in detail.

5.2.1 Step 1: Extract Descriptive Elements of Metadata Standards for Records Management and Archives

This section shows analysis of a metadata element extracted from each metadata standard. Because every schema has many elements, this section shows the analysis by an example. Each element is given its primary stages in two aspects - Creation and Update. Creation shows a stage where initial value of the element is given and Update shows a stage(s) where the element value is changed.

(1) AGLS Metadata

We select an element named *Availability*. *Availability* is primarily used for non-digital resources, provided as the information that is acquired through the user access resource physically. Because this element explains the availability of resources in the real usage environment, we classify the stage of this element as *Use & Manage*. The value of the element is updated in *Appraisal & Destroy* and *Reference & Re-Use*. Table 1 shows the summary of the primary stages for *Availability*.

〈Table 1〉 An example of AGLS Metadata

Element of AGLS Metadata : Availability	
Point of view	Lifecycle Stages
Creation	Use & Manage
Update	Appraisal & Destroy, Reference & Re-Use

(2) Decision Tree

Acquire for other purpose is used as an example element of the DPC Decision Tree. As mentioned before, the descriptive element of the Decision Tree is re-composed by re-phrasing a question at a node. *Acquire for other purpose* explains appraisal for other purpose in resource selection in the appraisal stage, i.e., *Appraisal & Destroy*. As the Decision Tree is not a metadata scheme, Decision Tree does not include a revision of the element value.

〈Table 2〉 An example of Decision Tree

Element of Decision Tree : Acquire for other purpose	
Point of view	Lifecycle
Creation	Appraisal & Destroy
Update	Not Applicable

(3) EAD

Archdesc gives a description about a resource - contents, contexts, scopes and so forth. The element value is determined in *Create*. Then, it is to be updated in *Appraisal & Destroy*, *Store & Arrange* and *Preserve*. This is because each time a resource is processed in an archival system the description of the resource may be subject to change.

〈Table 3〉 An example of EAD

Element of EAD : archdesc	
Point of view	Lifecycle
Creation	Create
Update	Appraisal & Destroy, Store & Arrange, Preserve

(4) ISAD(G)

Level of Description is an element that expresses units of resource, which is divided into

Fond, File, Item and so on. A unit of the resource may be changed if related resource(s) are added or removed.

A value for *Level of Description* is set in the *Create* stage of the Lifecycle, and updated in the step of *Use & Management* that confirms the related or subordinate resources, while using the resource. The value is updated in the steps in archival phases -*Appraisal & Destroy, Store & Arrange, Preserve* and *Reference* - where archives may change the values in accordance with their policy and changes in the time line.

〈Table 4〉 An example of ISAD(G)

Element of ISAD(G) : Level of Description	
Point of view	Lifecycle
Creation	Create
Update	Use & Management, Store & Arrange, Appraisal & Destroy, Preserve, Reference & Re-use

(5) OAIS

Change history before archiving describes the change history of a resource before it is deposited in an archive. The value of this element should be set in *Store & Arrange* and may be updated in *Preserve*.

〈Table 5〉 An example of OAIS

Element of OAIS : change history before archiving	
Point of view	Lifecycle
Creation	Store & Arrange
Update	Preserve

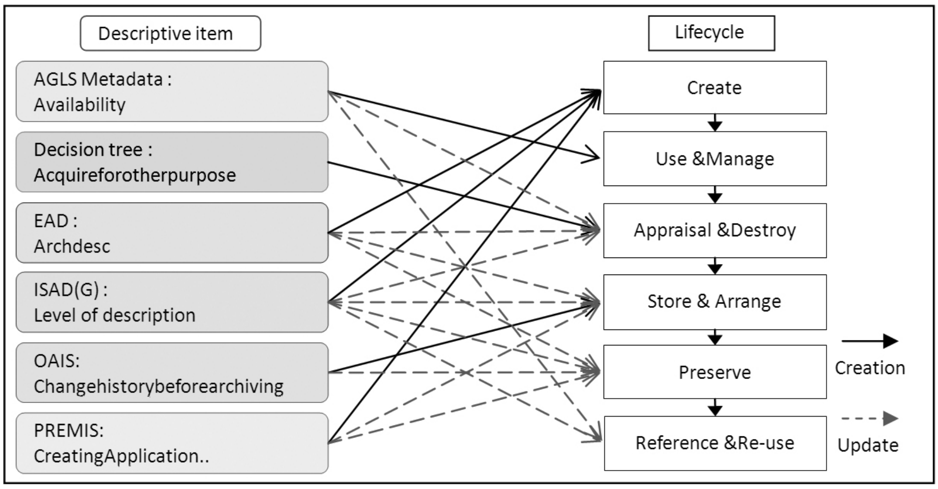
(6) PREMIS

Creating Application describes the applications used when a digital object was created. For this reason, the value of this element is determined in *Create*, and then, updated in *Store & Arrange* and *Preserve* where the digital object may be migrated to a new environment.

〈Table 6〉 An example of PREMIS

Element of PREMIS : creating Application	
Point of view	Lifecycle
Creation	Create
Update	Store & Arrange, Preserve

Figure 6 shows a summary of the analysis presented above. We have applied the analysis method above to all elements of the six schemas and summarized the result in the schema, mentioned in the next section.



〈Figure 6〉 An example of using the lifecycle of records for the description elements

5.2.2 Step 2: Determine Primary Stages for Metadata Standards

We took out every descriptive element from the metadata schemas, and mapped them to the records lifecycle stages in order to determine the primary stages of each element. Based on this investigation, we analyzed the relationship between each metadata standard and the lifecycle stages. Tables A, B and C in the appendix show the relationships between elements and the primary stages of the schemas. In these three tables, all elements of the metadata are shown where Roman and Italic fonts mean Creation and Update, respectively.

Every metadata schema is related to all stages of the lifecycle except the decision tree. Figure 6 shows the overall relationship between the schemas and the records lifecycle. The figure shows the high-density parts where many elements are connected to a specific stage. For example, AGLS has many elements connected to *Create, Use & Manage*, and *Reference & Re-use*. The paragraphs below show the analysis of each standard.

(1) AGLS Metadata

AGLS Metadata is composed of a description about resources according to their contents and for searching. In the lifecycle, we found that AGLS Metadata mainly expresses *Create, Use & Manage* and *Reference & Re-Use*. This is a very natural result because the first two stages are not necessarily related to long-term archiving but to general resource discovery and management, and the last stage is for users who want to find and use resources in the archives. Also, archival metadata schemas have a small set of general descriptive metadata like the ones on AGLS.

(2) Decision Tree

The element set created from the DPC's decision tree is composed of descriptive elements about the evaluation of the resources. Therefore, these elements are used only in *Appraisal & Destroy* and *Store & Arrange*. This crispness is the feature of the decision tree compared with other metadata schema standards.

(3) EAD

EAD mainly has descriptive elements that express the appraisal of the resources, history, origin of resources, and relative information. As elements of EAD are mainly for evaluation and basic description for archives, many elements for *Appraisal & Destroy* and *Store & Arrange* and some elements for *Preservation* are included.

(4) ISAD(G)

ISAD(G) is similar to EAD, but it does not have so many elements for *Preservation* as EAD has. ISAD(G) has elements that express bibliographic information and administrative information for archives such as management, use of resources, history information, and so forth. Thus, ISAD(G) is linked to *Appraisal & Destroy*, and particularly to *Store & Arrange*. On the other hand, the first two stages of the lifecycle are also connected.

(5) OAIS

OAIS has elements to express collection and history of digital objects. On the other hand, it has many elements to express technological and structural contents. OAIS has many elements for re-using resources. This is because dissemination of archived resources is a part of the OAIS reference model. Thus, OAIS covers *Appraisal & Destroy*, *Store & Arrange*, *Preservation*, and *Reference & Re-Use*.

(6) PREMIS

PREMIS has many elements that express technological features for preservation of digital resources. Significant difference from other metadata schemas that are connected to more than one stage in the lifecycle is that PREMIS is concentrated into *Preservation*.

In this study, we mapped the lifecycle stages to metadata elements extracted from the metadata standards. For the mapping, for every element extracted from metadata standards, we determined the primary stages where the element value is initially given or revised. Table 7 shows the statistics of the mapping. In Table 7, the numbers show the percentage of elements of each standard whose values are initially given or revised in a corresponding stage of the lifecycle. For example, in the case of EAD, *Appraisal & Destroy*, *Store & Arrange* and *Preserve* stages are the primary stage for 14%, 33% and 20% of the elements, respectively.

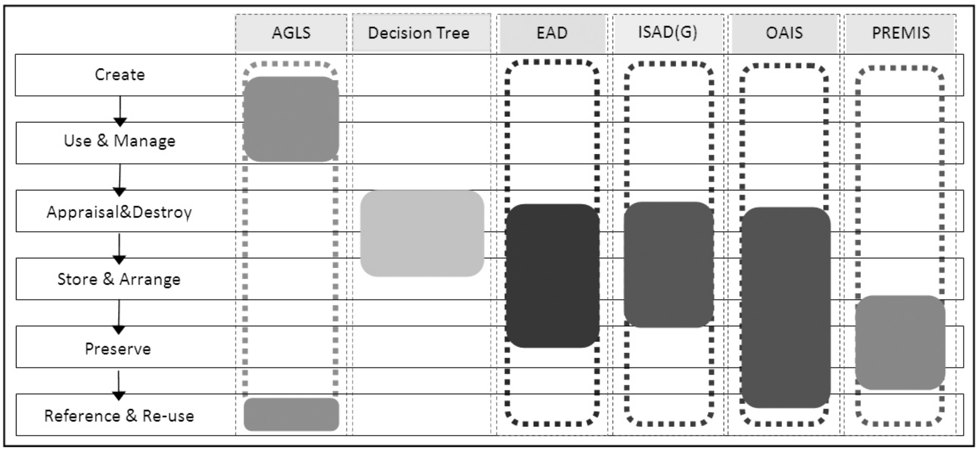
On the other hand, 24% elements are determined their values in the first two stages. This shows that EAD is oriented to resource organization in the archival storages rather than resource discovery and management in live resource repositories used in the early stages of the lifecycle. AGLS is primarily designed for resource discovery and access, which correspond to the first two stages of the lifecycle. In this study, however, the table shows AGLS is used in the whole lifecycle as a finding aid throughout the records lifecycle. As shown in Table 7, the primary stages are spread over the lifecycle but there is a peak in the *Use & Manage* stage. More importantly, Table C in the appendix shows that there is a clear split between Create stages and Update stages. This shows that the values initially given, are used for discovery in the first two stages of the lifecycle and the values may be revised for maintenance at archives. Thus, we can identify the overall features of the metadata standards shown in Figure 7 from the statistics shown in Table 7.

Metadata standards for archiving and preservation of digital resources are various. However, each metadata standards has its own feature in accordance with its primary application. The unified framework to identify the features of archival metadata standards proposed in this paper is useful to combine different archival metadata schemes in a single system because it is straight forward to find stages where mappings between different standards are heavily required. Thus, this unified framework is advantageous to enhance interoperability between the archival metadata standards.

Figure 7 is useful to overview the stages where crosswalking between metadata schemas are efficiently performed. This is because it helps us identify the correspondence between elements of similar meanings by showing the correspondence of elements to lifecycle stages. Thus, new viewpoint to enhance interoperability of the archival metadata schemas are given.

〈Table 7〉 Metadata standard shown by figures (%)

Lifecycle \ Metadata	AGLS	Decision tree	EAD	ISAD(G)	OAIS	PREMIS
Create	16		11	11	1	5
Use & Manage	28		13	6	2	22
Appraisal& Destroy	5	61	14	15	13	
Store & Arrange	18	39	33	43	30	21
Preserve	15		20	19	39	45
Reference & Re-use	18		9	6	15	7



〈Figure 7〉 Stage of lifecycle shown by metadata description elements

6. Conclusion

Because of the rapid increase of digital resources, long-term preservation of digital resources has been recognized as an important but difficult issue. For long-term preservation of digital resources, preservation policies and strategies to cope with various resources are necessary.

Metadata is one of the most important components for archiving and preservation of digital resources. Metadata schemas for digital archiving and preservation should be designed in accordance with the policies and strategies of archives.

We have examined the metadata for preservation and archives of digital resources in this study from the viewpoint of mapping between the metadata standards and the records lifecycle. In our research, we first started our study with a simple question "Is it possible to preserve resources long-term only by one metadata schema?" and another question "Is it possible to design a unified framework for metadata standards for archiving and preservation?" As a result the detailed examination of the metadata elements, we clarified the features of the standards from the viewpoint of relationships between the elements and the lifecycle stages.

Mapping between metadata schemas is a crucial issue because we are frequently required to unify metadata databases. Metadata mapping is required in the long-term preservation process. However, on the other hand, we know that metadata schema mapping is an expensive task. Our next step is to define a framework to help systematically map metadata elements for preservation.

Reference

後藤敏行. 2007. デジタル情報保存のためのメタデータ : 現状と課題.『情報管理』, 50(2): 74.

栗山正光. ディジタル情報保存のためのメタデータ(1). [cited 2010. 5. 12].
〈http://www.tulips.tsukuba.ac.jp/memb/mtkuri/reports/metadata1.html〉.

難波忠清外. 核融合アーカイブズデータベースの共有化. [cited 2010. 5. 11].
〈http://www.nifs.ac.jp/archives/01aA31P.pdf#search='eadの特徴'〉.

日本図書館協會, 栗山正光譯. 『PREMIS保存メタデータのためのデータ辭書第2.0版』. [cited 2010. 9. 11].
〈http://www.jla.or.jp/publish/PREMIS.pdf〉.

白才恩, 杉本重雄. 2007. ディジタルリソースの保存方式選擇のためのガイドライン.『情報處理學會研究報告: 情報學基礎研究會報告』, 2007(109): 15-22. [cited 2010. 9. 27].
〈http://www.tulips.tsukuba.ac.jp/limedio/dlam/M93/M932902/9.pdf〉.

Chan, Lois Mai, and Marcia Lei Zeng. 2006. Metadata Interoperability and Standardization: A Study of Methodology Part 1. *D-Lib Magazine*, 12(6). [cited 2010. 8. 11].

〈http://www.dlib.org/dlib/june06/chan/06chan.html〉.

Consultative Committee for Space Data Systems. 2002. *Reference Model for an Open Archival Information System(OAIS)*. [cited 2010. 8. 11]. 〈http://public.ccsds.org/publications/archive/650x0b1.pdf#search='OAIS'〉.

Day, Michael. 2001. Metadata for Digital Preservation: A Review of Recent Developments. *ECDL 2001, LNCS* 2163: 161-172.

Digital Curation Centre. *All Standards for Any Lifecycle Action*. [cited 2010. 8. 15]. 〈http://www.dcc.ac.uk/resources/standards/diffuse/standards?framework_id=0&lifecycle_id=0&sort=type#Digital%20Preservation%20Standards〉.

Digital Preservation Coalition. 2006. *Decision Tree for Selection of Digital Materials for Long-term Retention*. [cited 2010. 11. 11]. 〈http://www.dpconline.org/advice/preservationhandbook/decision-tree〉.

Dürr, E. H., R. Dekker, and K. van der Meer. 2006. *Archival Metadata for Durable Data Sets*. The Netherlands: DIR'06 Dutch-Belgian Information Retrieval Workshop. [cited 2010. 7. 10]. 〈http://hmi.ewi.utwente.nl/dir2006/abstracts/durr_paper.pdf〉.

Evans, Joanne, Sue Mckemmish, and KarunaBhoday. 2005. Create Once, Use Many Times: The Clever Use of Recordkeeping Metadata for Multiple Archival Purposes. *Archival Science*, 2005(5): 17-42.

Government of South Australia. *Records Life Cycle*. [cited 2010. 8. 18]. 〈http://www.decs.sa.gov.au/rmp/pages/cg0000941/lifecycle/?reFlag=1〉.

Hakala, Juha. 2001. *Metadata for Referencing and Archival Usage*. [cited 2010. 9. 15]. 〈http://eos.wdcb.ru/eps2/eps02009/eps02009.pdf〉.

Harvard University Library. *Metadata Standards*. [cited 2010. 9. 13]. 〈http://hul.harvard.edu/ois/digproj/metadata-standards.html〉.

Haslhofer, Bernhard and Wolfgang Klas. 2010. A Survey of Techniques for Achieving Metadata Interoperability. *ACM Computing Surveys*, 42(2) Article 7: 1-37.

International Council on Archives. 2000. *ISAD(G): General International Standard Archival Description, Second edition*. [cited 2010. 7. 11]. 〈http://www.ica.org/sites/default/files/isad_g_2e.pdf〉.

McKemmish, Sue, Glenda Acland, Nigel Ward, and Barbara Reed. *Describing Records in Context in the Continuum: the Australian Recordkeeping Metadata Schema*. [cited 2010. 8. 18].

〈http://infotech.monash.edu/research/groups/rcrg/publications/archiv01.html〉.

National Archives of Australia. 2006. *Australian Government Implementation Manual : AGLS Metadata*. [cited 2010. 8. 5]. 〈http://www.naa.gov.au/Images/AGLS_Manual_v2_tcm2-914.pdf〉.

National Archives of Australia. 2008. *Australian Government Recordkeeping Metadata Standard Version 2.0*. [cited 2010. 8. 11]. 〈http://www.naa.gov.au/Images/AGRkMS_Final%20Edit_16%2007%2008_Revised_tcm2-12630.pdf〉.

Online Computer Library Center. 2008. *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*. [cited 2010. 10. 1]. 〈http://www.loc.gov/standards/premis/v2/premis-2-0.pdf〉.

The Library of Congress. 2002. *Encoded Archival Description(EAD)*. [cited 2010. 9. 3]. 〈http://www.loc.gov/ead〉.

The National Archives and Records Administration. *Records have Lifecycle*. [cited 2010. 9. 5]. 〈http://www.archives.gov/about/info/whats-a-record.html〉.

Appendix

〈Table A〉 Using the lifecycle of records for the metadata standard (ISAD(G) & Decision Tree)

Metadata Lifecycle	ISAD(G)	Decision Tree
Create	Date(s) Language, scripts of material Level of description Name of creator(s) Title	
Use & Manage	Date(s) *Level of description* Scope and content	
Appraisal & Destory	Appraisal, destruction and scheduling information Archivist`s note *Date(s)* Date(s) of description Immediate source of acquisition or transfer Physical characteristics and technical requirements Rule or conventions	Long term value justify preservation Other purposes Multiple media formats Digital version be selected for preservation Documentation been supplied Negotiate for the source to supply Technically feasible for you to construct Material so valuable that you will accept Accept the costs and risks of trying to manage Cost effective for you to develop Cost-effective for you to transfer Ccept the costs and risks of trying to manage Commit adequate staff Manageable file format Technically feasible for you to transfer the material Available to you online or on a physical carrier Able to collect or receive the resource via a Enough available storage space Carrier that is acceptable for transfer and/or storage Transfer the resource to an acceptable carrier
Store & Arrange	Accruals Administrative, Biographical history Archival history *Archivist`s note* Conditions governing access Conditions governing reproduction *Date(s)* *Date(s) of description* Existence and location of copies Existence and location of originals *Extent and medium of the unit of description* Findings aids *Language, scripts of material* *Level of description* Note Publication note Related units of description Reference code *Rule or conventions* System of arrangement	Institutional remit/collection development policy Preservation responsibility Preservation responsibility been accepted elsewhere Higher degree of preservation commitment or access Acceptable arrangements for acquisition and/or transfer Re-evaluate acquisition The rights to transfer *Technically feasible for you to transfer the material* *Available to you online or on a physical* *Enough available storage space* *Documentation been supplied* *Negotiate for the source to supply* *Cost effective for you to develop*
Preserve	*Accurals* *Archival history* *Archivist`s note* *Date(s)* *Date(s) of description* *Level of description* *Note* *Reference code* *Rule or conventions*	
Reference & Re-use	*Date(s)* *Level of description* *Scope and content*	

〈Table B〉 Using the lifecycle of records for the metadata standard (EAD & OAIS)

Metadata / Lifecycle	EAD	OAIS
Create	Address, Addressline, Creation, Corpname Date, Famname, Geogname, Imprint Language, Language, Name, Namegrp Origination, Otherfidaid, P, Persname Publicationstm, Publisher, Ptr, Sponsor Subtitle, Title, Titlepage, Unitdate Unittitle	Reason for Creation
Use & Manage	*Address*, *Addressline*, *Corpname*, *Date* *Dimensions*, Extent, Extref, Extrefloc *Famname*, Genreform, *Geogname*, *Imprint* *Language*, *Langusage*, *Name*, *Namegrp* Occupation, *Otherfidaid*, *P*, *Persname* Prefercite, *Publicationstm*, *Publisher*, Physdesc *Ptr*, Scopeconten, *Sponsor*, *Subtitle* *Unitdate*, *Origination*	Existing Metadata Existing Records
Appraisal & Destory	Abstract, Accruals, Acqinfo, *Address* *Addressline*, Appraisal, Archdesc, *Archref* Author, C, C01 - c12, Container *Corpname*, *Date*, Descrules, *Famname* Frontmatter, *Geogname*, *Imprint*, *Language* *Langusage*, *Name*, *Namegrp*, Note *P*, *Persname*, Processinfo, *Ptr* Repository, *Sponsor*, *Subtitle*, *Unitdate*	Context Information, Custody History Documentation, *Existing Metada* History of Origin, Legislation Text Pointer Permitted by License Original Technical Environments Prerequisites, Procedures *Reason for Creation*, Reason for Preservation Related Information Objects Representation Information
Store & Arrange	*Abstract*, Accessrestrict, *Accruals*, *Address* *Addressline*, Altformavail, *Appraisal*, Archdesc *Archref*, Arrangement, *Author*, Bibliography Bibref, Bibseries, Bioghist, *C* *C01 - C12*, Chronitem, Chronlist, *Container* *Corpname*, Custodhist, Dao, Daodesc Daogrp, Daoloc, *Date*, *Descrules* *Dimensions*, Event, Eventgrp, *Extent* Extptr, Extptrloc, *Extref*, *Extrefloc* *Famname*, *Frontmatter*, Genreform, *Geogname* *Imprint*, Langmaterial, *Langusage*, Legalstatus Materialspec, *Name*, *Note*, *Occupation* Originalsloc, *Otherfindaid*, *P*, *Persname* *Physdesc*, Physfacet, Physloc, Phystech *Prefercite*, *Processinfo*, *Ptr*, *Publisher* *Publicationstm*, *Scopecontent*, *Sponsor*, *Language* Separatedmaterial, *Repository*, *Namegrp* Relatedmateri, Ref, Refloc *Unitdate*, Userestrict, Subject	Actions, Action History Actors, Administration History Change History Before Archivir, Contacts or Rights Holders *Context Information*, Copyright Statement *Custody History*, Date of Publication *Existing Metadata*, *Existing Records* Fixity Information, *History of Origin* Ingest Process History, *Legislation Text Pointer* Licence Text Pointer, Management History Name of Publisher, Negotiation History *Original Technical Environmen*, Permitted by License Place of Publication, *Prerequisites* *Procedures* Provenance Information *Reason for Creation* Reference Information *Related Information Objec* *Representation Informatic* Resource Description Rights Information Rights Management
Preserve	*Address*, *Archdesc*, *Archref*, *Author* *Bibliography*, *Bibref*, *C*, *C01 - C12* *Chronlist*, *Chronitem*, *Container*, *Corpname* *Custodhist*, *Date*, *Dao*, *Daodesc* *Daogrp*, *Daoloc*, *Descrules*, *Event* *Eventgrp*, Extref, *Extrefloc*, *Frontmatter* *Famname*, *Geogname*, *Imprint*, *Language* *Langusage*, *Materialspec*, *Name*, *Namegrp* *Note*, *Occupation*, *Persname*, *Processinfo* *Phystech*, *P*, *Ptr*, *Ref* *Refloc*, *Repository*, *Sponsor*, *Subject* *Unitdate*, *Origination*	*Actions*, *Action History* *Actors*, *Administration History* Authentication Indicator *Change History Before Archiving* *Contacts or Rights Holders*, Content Information *Context Information*, *Copyright Statement* *Custody History*, *Date of Publication* *Existing Metadata*, *Existing Records* *Fixity Information*, *History of Origin* *Ingest Process History*, Input Format *Legislation Text Pointer*, *Licence Text Pointer* *Management History*, *Name of Publisher* *Negotiation History*, Output Format Parameters, Permitted by Statute *Place of Publication*, Platform Policy History Preservation Description Information *Provenance Information*, *Reason for Preservation* *Reference Information*, *Related Information Objects* Render/Analyse Engines, *Representation Information* *Resource Description*, *Rights Information* *Rights Management*, Rights Warning Structure Information, Transformer Objects (TOs) Underlying Abstract Form Description
Reference & Re-use	*Address*, *Corpname*, *Date* *Famname*, *Geogname*, *Name* *Namegrp*, *Persname*, *Prefercite* *Sponsor*, *Unitdate*, Extref *Language*, *Imprint*, *Extrefloc* *Langusage*, *Occupation*, *P* *Ptr*, *Ref*, *Refloc*	*Actions*, *Actors* *Contacts or Rights Holders*, *Existing Metadata* *Input Format*, *Legislation Text Pointer* *Licence Text Pointer*, *Negotiation History* *Output Format*, *Parameters* *Permitted by License*, *Permitted by Statute* *Platform*, *Render/Analyse Engines* *Rights Information*, *Rights Management* Rights warning

〈Table C〉 Using the lifecycle of records for the metadata standard (AGLS & PREMIS)

Metadata / Lifecycle	AGLS Metadata	PREMIS
Create	Contributor Creator Date Format Identifier Language Publisher Rights Title	ObjectCharacteristics CreatingApplication Environment OriginalName
Use & Manage	Audience Availability Coverage *Date* Description *Format* Funcion *Identifier* Mandate Relation *Rights* Source Subject Type	*ObjectCharacteristics* *Environment* Relationship LinkingEventIdentifier LinkingIntellectualEntityIdentifier LinkingRightsStatementIdentifier EventType EventDateTime EventDetail LinkingAgentIdentifier LinkingObjectIdentifier CopyrightInformation LicenseInformation StatuteInformation LinkingObjectIdentifier LinkingAgentIdentifier
Appraisal & Destory	*Availability* *Date* *Rights*	
Store & Arrange	*Date* *Description* *Funcion* *Format* *Mandate* *Mandate* *Relation* *Rights* *Subject* *Type*	*ObjectCharacteristics* *OriginalName* Storage *Environment* *Relationship* *LinkingEventIdentifier* *LinkingIntellectualEntityIdentifier* *LinkingRightsStatementIdentifier* *EventDateTime* *LinkingAgentIdentifier* *LinkingObjectIdentifier* *CopyrightInformation* *LicenseInformation* *LinkingObjectIdentifier* *LinkingAgentIdentifier*
Preserve	*Date* *Description* *Format* *Identifier* *Mandate* *Relation* *Rights* *Type*	ObjectIdentifier ObjectCategory PreservationLevel SignificantProperties *ObjectCharacteristics* *OriginalName* *Storage* *Environment* SignatureInformation *Relationship* *LinkingEventIdentifier* *LinkingIntellectualEntityIdentifier* *LinkingRightsStatementIdentifier* EventIdentifier *EventType* *EventDateTime* *EventDetail* EventOutcomeInformation *LinkingAgentIdentifier* *LinkingObjectIdentifier* AgentIdentifier AgentName AgentType RightsStatement RightsStatementIdentifier RightsBasis *CopyrightInformation* *LicenseInformation* *StatuteInformation* RightsGranted LinkingObjectIdentifier *LinkingAgentIdentifier* RightsExtension
Reference & Re-use	*Availability* *Audience* *Date* *Format* *Function* *Identifier* *Rights* *Source* *Relation* *Type*	*ObjectCharacteristics* *Environment* *EventDateTime* *EventDetail* *LinkingAgentIdentifier*