

A Study on Feature Analysis of Archival Metadata Standards in the Records Lifecycle*

Jae-Eun Baek (백재은)**

Contents

- | | |
|--|--|
| 1. Introduction | 5. Feature Analysis of Metadata Schemas based on Lifecycle |
| 2. Models and Standards for Archiving and Preservation | 5.1 Viewing Differences from Descriptive Elements |
| 2.1 Metadata Schemas for Record Management and Archives - AGLS, EAD, ISAD(G), OAIS, PREMIS | 5.2 Mapping to Determine the Primary Stages in the Lifecycle |
| 2.2 Properties and Characteristics of Archival Metadata | 6. Facet Analysis of Archival Metadata Schemas for Metadata Interoperability |
| 2.3 Selection Decision Tree of Digital Materials for Long-term Retention | 6.1 Task-oriented View of Records Lifecycle-Task Model |
| 2.4 Lifecycle of Records - NARA Lifecycle | 6.2 Task-centric View of Metadata Elements and 5W1H Categories |
| 3. Related Works | 6.3 Analysis Criteria of Classification |
| 4. From Resource-centric to Task-centric View of Metadata Schema | 6.4 Mapping Metadata Schemas in 5W1H Categories and Lifecycle Tasks |
| | 7. Conclusion |

ABSTRACT

Metadata schemas are well recognized as one of the important technological components for archiving and preservation of digital resources. However, a single standard is not enough to cover the whole lifecycle for archiving and preserving digital resources. This means that we need to appropriately select metadata standards and combine them to develop metadata schemas to cover the whole lifecycle of resources (or records). Creating a unified framework to understand the features of metadata standards is necessary in order to improve metadata interoperability that covers the whole resource lifecycle. In this study, the author approached this issue from the task-centric view of metadata, proposing a Task model as a framework and analyzing the feature of archival metadata standards. The proposed model provides a new scheme to create metadata element mappings and to make metadata interoperable. From this study, the author found out that no single metadata standard can cover the whole lifecycle and also that an in-depth analysis of mappings between metadata standards in accordance with the lifecycle stages is required. The author also discovered that most metadata standards are primarily resource-centric and the different tasks in the resource lifecycle are not reflected in the design of metadata standard data models.

Keywords: Archival Metadata Standard, Metadata Interoperability, Preservation, Records Lifecycle, Task Model

* 본 논문은 일본 筑波大學大學院 図書館情報メディア研究科(도서관정보학 전공) 박사학위논문(2013년 11월)을 축약·재구성한 것임.

** 덕성여자대학교 문헌정보학 강사(jaeunb1979@duksung.ac.kr, jaeunbb@yahoo.co.jp)
논문접수일자: 2014년 7월 15일 최초심사일자: 2014년 8월 4일 게재확정일자: 2014년 8월 19일
한국문헌정보학회지, 48(3): 71-111, 2014. [http://dx.doi.org/10.4275/KSLIS.2014.48.3.071]

1. Introduction

Due to the rapid growth of the Internet and WWW (World Wide Web), a quantity of information could be constantly produced and received in various forms. In our modern information environment, we cannot imagine our daily lives without digital resources and ubiquitous networks.

Digital resources have their own problems of management and preservation. 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 well known that archiving and preservation of digital resources is not straightforward. The problem is not only the quantitative, but also the preservation of a digital resource in its original form for the next generation. In other words, we need a number of solutions for long-term preservation and management of non-digital and digital resources for the future.

For the long-term preservation and archiving of digital resources, 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, and metadata schemas for digital preservation.

A metadata standard is well recognized as one of the important components required in the creation, management, recordkeeping, archiving, and preservation of digital resources. Metadata standards are usually designed for a specific purpose and used in different services, e.g., searching resources, rights management, and accessibility control. There are many major metadata standards used for management, recordkeeping, archiving, and preservation of digital resources, e.g. Dublin Core, AGLS, AGRkMS, EAD, ISAD(G), METS, MoReq2, OAIS, PREMIS, and more.

Metadata schema for purposes such as finding aids, rights management, and accessibility descriptions are used in accordance with the requirements of a particular stage of the resource's lifecycle. Metadata schema is also related to different resource tasks throughout the whole resource lifecycle. Metadata elements are created and revised by resource tasks and change according to the content and the purpose of the tasks. Resources perform different tasks according to the stage of their lifecycle, which means that metadata, associated with the resource, needs to change. We need appropriate metadata schemas related to the lifecycle stage. Moreover, we need guidelines to select appropriate metadata standards and to define profiles for the tasks and stages based on the metadata standards. However, most metadata standards do not explicitly mention the resource lifecycles or tasks. In other words, it is not clearly defined when a descriptive element should be assigned or where its value should be revised in the lifecycle.

For example, PREMIS has five types of entities in its data model-intellectual entity, digital object, event, right and agent, and elements. PREMIS is primarily for preservation. However, some elements of an intellectual entity of PREMIS such as title and creator are assigned when the entity is created, which happens in the very early stages of the lifecycle. Thus, the data model of a metadata standard does not explicitly reflect lifecycle stage(s) for which the standard is primarily designed.

A major question is 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 practice, multiple metadata standards are frequently used in a single system. 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 (Baek Jae-eun and Sugimoto 2010).

Metadata vocabulary mapping is not new. There are notable examples such as VMF (The Vocabulary Metadata Framework) (Dunsire 2010). However, these mappings do not explicitly use the lifecycle to identify the semantics of the metadata elements.

VMF is used for the mapping of vocabularies from major metadata standards. Also, VMF is designed as a tool to automate finding the 'best fit' mapping among terms in controlled vocabularies in different metadata schemes (JISC 2009). This means that, on one hand, we need to appropriately choose one or more metadata standard(s) and define a metadata schema for a particular application system. On the other hand, we may need to combine different metadata standards to define an application profile in accordance with the requirements given to the application system. In addition, we may need to define crosswalks between metadata schemas for data exchange.

Based on the observation about metadata schemas for archiving and preservation of digital resources, the author explains and proposes a methodology to analyze metadata schemas in order to help selection and combination of metadata schemas used throughout the whole lifecycle, i.e. from creation to preservation and re-use. Specifically, she analyzed the relationship between a resource task and available metadata schemas for digital archiving and preservation.

A metadata standard is generally focused on resources from the viewpoint of the purpose of description. Mapping metadata standards using each stage of a lifecycle is not a suitable method. In order to analyze the features of archival metadata, she examined the relation between the metadata standards and the stages of a lifecycle. The author proposes a mapping method between

metadata standards in order to link the different metadata standards with the tasks within the stages of a lifecycle. Furthermore, she did a detailed analysis from the viewpoint of the task of a resource. This paper proposes a framework to characterize descriptive elements of metadata vocabularies and to improve mapping among them.

First, she analyzed relationships between the lifecycle stages and the metadata standards by an analysis of patterns based on the lifecycle. From the crosswalk and mapping between metadata and the stage of a lifecycle, she examined and identified a stage, where the value of the element is updated and a particular metadata standard is most frequently used, for every element where an initial value of the element is given.

In the second research, the author proposed a Task model based on a task-centric point of view for more detailed analysis of the element sets. The author clarified the viewpoint of an 'Event' performed within a task, using the 5W1H attribute set (what, why, where, who, when, how) and used it in order to categorize a metadata element in the context of each task where the element is used. The Task model and the 5W1H attribute set are important to narrow the scope of mapping and categorizing in order to perform efficient mapping between descriptive elements focusing on a task.

To show the features of archival metadata standards, the author thinks that an analysis using various metadata standards shows a clearer difference when comparing metadata. So, she have chosen AGRkMS, EAD & ISAD(G), PREMIS form as typical standards in their particular domains. Although the AGLS, OAIS, and DPC attribute sets (a set of attributes extracted from the decision tree for a preservation process defined by the Digital Preservation Coalition) are not designed as metadata schema for archiving or preservation, she has included them as comparable objects in order to show the characteristics of archival metadata standards. Also, in order to analyze the relationships between a resource task and the metadata standards, she used the records lifecycle of NARA.

The author first started her 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, she clarified the features of the standards from the viewpoint of relationships between the elements and the lifecycle stages.

This paper is organized as follows. Section 2 and 3 describes and arranges a fundamental concept - metadata standards for archiving and preservation of digital resources, records lifecycle model, and literature reviews, as the background. Section 4 explains the relation of a task and

the metadata standard. Section 5 shows the feature analysis of archival metadata standards from a viewpoint of a resource lifecycle, according to the first research. Section 6 explains about the feature analysis for interoperability of a metadata standard, and proposes the basic models - the 5WIH categories and the Task model-and, shows several example mappings among the standards, following the second research. In section 7, she has conclusions.

2. Models and Standards for Archiving and Preservation

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

In this study, she used widely known metadata standards for recordkeeping, record management, archiving, and preservation. In order to propose a new model to clarify the features of those standards, she have chosen AGRkMS, EAD, ISAD(G), OAIS and PREMIS from these standards as typical standards in their particular domains-i.e., AGRkMS for record keeping, EAD for archives, and OAIS and PREMIS for digital preservation.

OAIS does not define a metadata element set in itself. She used the element set of CEDARS preservation metadata as the CEDARS set was drafted in close consultation with the OAIS reference model, to the extent that the elements borrow the concepts, terminology, and organization embedded within the OAIS framework.

In addition to these standards, the author included AGLS and the DPC attribute set which are not archival metadata standards but contain metadata elements used in the lifecycle-AGLS for resource discovery, DPC for appraisal. The next paragraphs briefly introduce these metadata standards referred to in this study.

2.1.1 AGLS

Australian Government Locator Service (AGLS) Metadata standard is to refer descriptive information about resources, and it is known as resource discovery metadata.

AGLS Metadata was designed to facilitate, discover, and search resources by users online and was used to improve the visibility and discoverability of Australian government resources in the online environment.

AGLS Metadata Standard provides a set of metadata properties, policies and guidelines defined

for a particular application or implementation, and metadata property set consists of 60 properties (National Archives of Australia 2006; 2010; 2011).

2.1.2 AGRkMS

Australian Government Recordkeeping Metadata Standard (AGRkMS) describes the “information about records and the contexts in which they are captured and used” (International Organization for Standardization 2006; 2009).

AGRkMS is based on the AGLS standard and sets out the type of recordkeeping metadata (National Archives of Australia 2008).

AGRkMS differs from the first standard in that it is based on a multiple-entity model, allowing for the description of five separate entities - Record, Agent, Business, Mandate, and Relationship. It defines a basic set of 26 metadata properties and an additional 44 sub-properties that may be used to describe these entities (National Archives of Australia 2010; 2011).

2.1.3 EAD

Encoded Archival Description (EAD) is an XML standard used to encode archival finding aids in a network (online) environment. In addition to the content description of digital resources, EAD has the elements for structural description (Baek, Jae-eun and Sugimoto 2010), (The Library of Congress 2002).

“EAD Elements section of the tag library contains descriptions of 146 elements and the EAD tag set is used both to describe a collection as a whole, and also to encode a detailed multi-level inventory of the collection. EAD is a metadata schema for archiving digital resources, keeping compatibility with ISAD(G) and one of the guiding principles of EAD is to maintain compatibility with ISAD(G)” (The Library of Congress 2002).

2.1.4 ISAD(G)

The General International Standard Archival Description (ISAD(G)) was 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, and it 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 storage period, usage, copy condition, context of resource, etc (Baek Jae-eun and Sugimoto 2007).

ISAD(G) has 26 elements of which six are mandatory and rules. All elements of ISAD(G)

“covered by these general rules are available for use, but only a subset need to be used in any given description” (International Council on Archives 2000).

2.1.5 PREMIS

The Preservation Metadata and Implementation Standard (PREMIS) is a metadata schema for preservation of digital resources and “is designed to be an effective and inexpensive implementable tool that provides the metadata or information needed to preserve digital information assets for the long term.” PREMIS defines a data model of instances which are subject to metadata description for preservation and the data dictionary.

The PREMIS data dictionary is the international standard for metadata to support the preservation of digital objects and it defines preservation metadata as the information a repository uses to support the digital preservation process (PREMIS Editorial Committee 2012).

2.1.6 Preservation metadata elements for the CEDARS project

The CEDARS (CURL Exemplars in Digital ARchiveS) approach adopts the OAIS information model (concepts and terminology) as an underlying framework for their metadata. “The CEDARS metadata also is supplied by the Resource Description element, which for the CEDARS project, is implemented as a Dublin Core record. This record can be supplemented by any other existing metadata records (e.g., MARC) associated with the digital object” (OCLC/RLG Working Group on Preservation Metadata 2001).

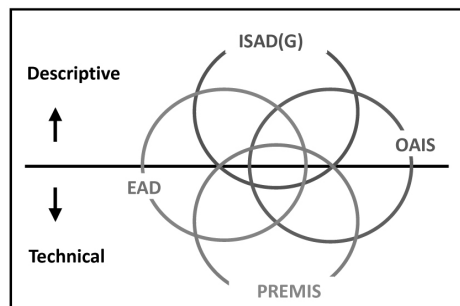
The CEDARS metadata scheme treats Reference Information as metadata for resource discovery and includes descriptive, administrative, technical, and legal information. “The CEDARS metadata element set is intended to enable the long-term preservation of digital resources” (OCLC/RLG Working Group on Preservation Metadata 2001).

2.2 Properties and Characteristics of Archival Metadata

Each metadata standard for archives has its own set of elements and controls vocabularies. A typical metadata description contains elements such as title, creator, related resource, and history of 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 details of the descriptive and technical metadata.

As a typical metadata of archives and preservation, she analyze the features using four meta-

data-EAD, ISAD(G), OAIS, PREMIS. ISAD(G) contains descriptive elements of resources in an appropriate granularity, i.e., fond, sub-fond, series, file, and item. EAD and OAIS have elements to describe intellectual content, structural features, and administrative and technology information. Intellectual content obviously needs descriptive metadata and technology information is in technical metadata. Structural and administrative information have both descriptive and technical features. PREMIS has many elements to describe the technical features and structure of the digital resources. Figure 1 shows the features of these four standards (Baek Jae-eun and Sugimoto 2007).



〈Figure 1〉 Characteristics of Metadata for Archiving and Preservation

In this analysis for the metadata elements, the author has shown that, on one hand, these metadata schemas have common features, but on the other, they have different features determined by their objectives and purposes. This means that it is crucial to select and use appropriate metadata standards and combine them correctly when designing a metadata schema for a specific archival system. In other words, the crucial metadata issues for the archival system are to create mappings between the lifecycle stages and the metadata standards and between the 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 (Baek Jae-eun and Sugimoto 2010).

2.3 Selection Decision Tree of Digital Materials for Long-term Retention

The Digital Preservation Coalition (DPC) promotes 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. Also, DPC provides a selection decision tree for long-term preservation (Digital Preservation Coalition 2006).

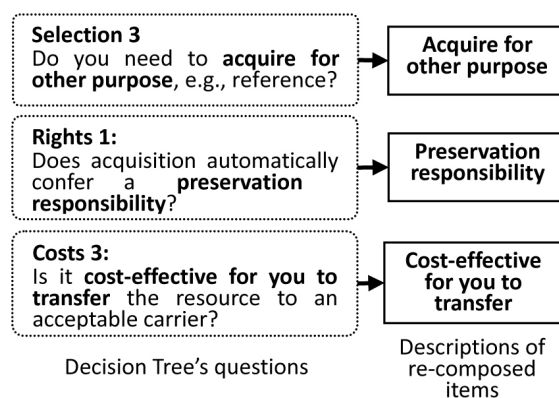
Decision tree of DPC is a decision process for the selection of digital materials for long-term retention. The decision tree is composed of three sections - *Rights & Responsibility, Technology & Metadata, Documents & Costs*. Each section is expressed as a sub-tree of the whole process. And, the decision 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*.

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. The questions contain crucial semantic attributes to help choose an 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 she extracts metadata attributes, is regarded as a metadata standard like other standards described in the previous section. For the conversion of the decision tree into a metadata schema, she extracted phrases from the questions and organized them into descriptive elements. The method of extracting phrases from the question statements 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 from description elements of a metadata schema.

In this way, she got 27 attributes from the set of questions in the decision tree. Figure 2 shows transformation of Questions (of decision tree) to Metadata Attributes.



<Figure 2> Transformation of Decision Questions to Metadata Attributes

2.4 Lifecycle of Records – NARA Lifecycle

Huge numbers of documents and resources are created, used, preserved, and destroyed in accordance with the management processes determined by the individual organizations (National Archives and Records Administration). 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. The NARA's records lifecycle has seven stages defined independently from any resource types, such as digital resources, official documents, archives, and national records, and also from any media types. The paragraphs below explain the stages of the NARA lifecycle.

(1) Creation

Records are created by people 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 reasons to change the media.

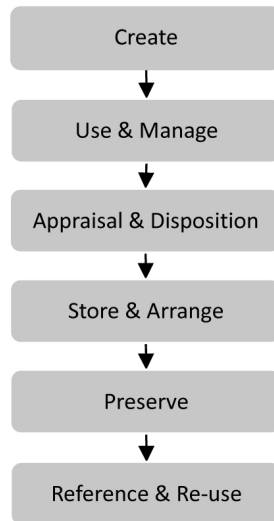
(6) Reference

Supply the preserved records to provide search and reference services.

(7) Continuing use

Proper management and continuing use of preserved records is promoted.

In this study, she merge the last two stages of NARA’s lifecycle into one and define the resource lifecycle model as shown in Figure 3 because both of the last two stages, Reference and Continuing Use, mean the use of the archived resources. This resource lifecycle model was used for feature analysis and she used this resource lifecycle model to define the Task model of the resource lifecycle.



<Figure 3> Lifecycle of This Research

3. Related Works

(1) Metadata for Digital Preservation: A Review of Recent Developments

Michael Day (2001) 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 (Day 2001).

The author referenced various definitions, descriptions, projects, and metadata for recordkeeping from the paper. Reviewing the digital preservation and researching on the importance of metadata for preservation can help us make it clear for her study background.

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

Joanne Evans et al. (2005) analyses and explores the development of metadata for multiple archival purposes and relevance to future archival systems using the Clever Recordkeeping Metadata Project (CRKMP). The paper offers a good example of metadata use in the whole records lifecycle. This project explains the interoperability and the theory of the Records Continuum as a conceptual framework (Evans, Mckemmish and Bhoday 2005).

Through this paper, she refers to the role, definitions, description of recordkeeping system or record management system, and interoperability. The author learned the importance of record-keeping system for integrated systems, and metadata interoperability through the CRKMP.

(3) Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information

Murtha Baca (2003) focuses on the selection of appropriate metadata schemas for Cultural Heritage Information. And the article focuses on the combination of controlled vocabularies and classification systems (Baca 2003).

Her study used the definition of metadata mapping and crosswalks from the article. She referenced the description about 'Selection of metadata schemas', 'Metadata mapping and crosswalk', through sample mappings of each metadata schema for museum, bibliographic, archival, and Web resources.

(4) Metadata Elements for Object Description and Representation: A Case Report form a Digitized Historical Fashion Collection Project

Marcia Lei Zeng (1999) discusses the application of existing metadata formats to a historical fashion collection and develops a catalog into digitized historical fashion collection objects. The paper describes how to choose, compare, and use the different elements of metadata schema for the creation of catalog (Zeng 1999).

Metadata interoperability is an important aspect in her research. So, she referenced the explanations and concepts about metadata interoperability, and examined metadata mapping methods in this paper. The significant difference is that her study is based on the resource lifecycle which is an essential aspect of metadata for archive and preservation.

(5) Metadata Interoperability and Standardization: A Study of Methodology Part 1

Chan and Zeng (2006) studies interoperability problems with multiple metadata schemas, such

as having the same subject domain and resources of the same type (Chan and Zeng 2006).

Metadata interoperability, Application profiles etc are very important aspect for the mapping and classification in her study. The author referenced definition and description about the metadata interoperability, application profiles, crosswalks, and metadata interoperability projects of different levels in the paper.

(6) A Survey of Techniques for Achieving Metadata Interoperability

The survey by Haslhofer and Klas (2010) 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 (Haslhofer and Klas 2010). According to each of these four blocks, various metadata standards and metadata mappings and their techniques are explained in a study of metadata interoperability from different viewpoints.

The mappings that she has created among the metadata standards improve interoperability of the metadata standards. This survey paper gives hints to compare and mapping between metadata schemas performed in the study described in the paper.

(7) Interdisciplinary Contents Management Using 5W1H Interface for Metadata

Keiko Shimazu et al. (2006) studies a metadata exchange interface for interdisciplinary content-sharing. The paper shows the interface module which converts tag-labels using 5W1H categories. In this paper, the interface for the metadata abstraction module for contents-circulation across various disciplines was designed using the concept of 5W1H, a representative result of communication study in the field of sociology. 5W1H, which stands for each initial letter, was proposed as the standard solution (Shimazu, Arisawa and Saito 2006).

Her study uses the Task model and 5W1H categories to identify the contexts of the resources which are the objectives of metadata descriptions. This is a unique feature of this study in contrast with those works in survey papers and those listed in the paragraphs above. Especially, she examined the usage of 5W1H, and the metadata abstraction module using 5W1H - the metadata labels (of Dublin Core) to 5W1H, the labels of noun types to 5W1H.

(8) A Metadata Lifecycle Model for Digital Libraries: Methodology and Application for an Evidence-based Approach to Library Research.

Chen et al. (2003) describes and proposes the Metadata Lifecycle Model (MLM). The paper introduces

MLM as a methodology of whole process of metadata provision for digital libraries. The purpose of the model is to achieve a consistent method for developing metadata for digital library projects, and to conduct a content-based analysis for digital collections (Chen, Chen and Lin 2003).

In her study, she proposed and used the records lifecycle model and the Task model to carry out a feature analysis of metadata elements. Through the metadata lifecycle model that is provided in this paper, she discovered the various views of lifecycle models and referred to them. She also learned about the metadata analysis which uses the MLM.

4. From Resource-centric to Task-centric View of Metadata Schema

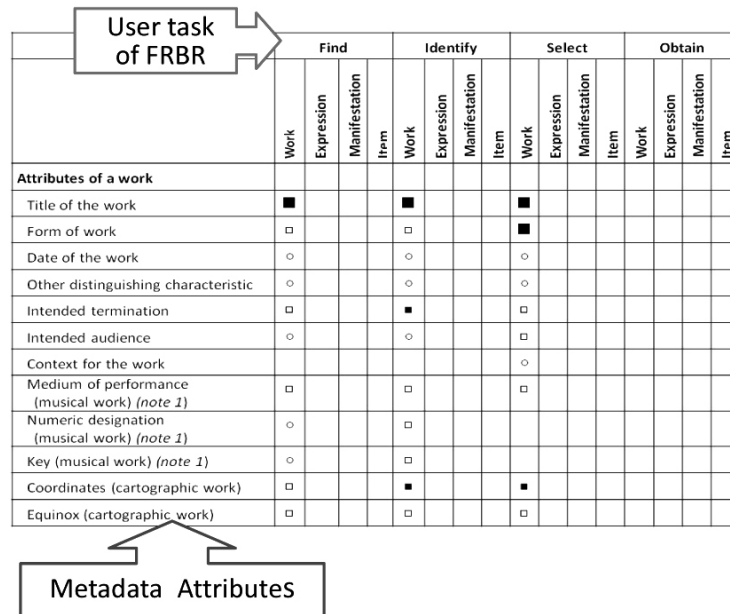
A resource may be affected by a task performed in a lifecycle stage - for example, in the *Appraisal & Disposition* stage, a resource disposed may be revised in the appraisal process in accordance with the preservation policy of the given archive. Metadata should be able to record the change of the resource as the lifecycle stage proceeds. Thus, the metadata elements are assigned values or updated in the lifecycle stages. Most metadata standards are designed in accordance with the lifecycle stages where the metadata standards are applied. However, most metadata standards make no mention about the resource tasks.

FRBR (Functional Requirements for Bibliographic Records) which is defined as a model for bibliographic description includes some generic tasks and metadata elements (i.e., metadata attributes) used in those tasks, e.g. title-of-work is required to find a work.

FRBR shows the four generic tasks - find, identify, select, and obtain-to explain the relationship between the attributes and tasks (International Federation of Library Associations and Institutions 2009). Figure 4 show the mapping of four generic tasks and *Work* (one of four elementary attributes) in the entity-relationship model. “Each task is in turn broken out into four sub-tasks defined in relation to the entity on which the task is focused (i.e., find work, find expression, find manifestation, find item, etc.). The symbols (■=High value, □=Medium value, ○=Low value) used in the tables indicate the relative value of each attribute or relationship in supporting a specific user task focused on a particular entity” (International Federation of Library Associations and Institutions 2009).

FRBR User Tasks are included in a stage between *resource creation* and *use* in the lifecycle because of the nature of bibliographic description. User Task shows the relationship of metadata attributes and a task using the importance of metadata value, by applying to a task the metadata

that describes a resource. FRBR User Task shows the metadata attribute is related to resource task. In addition, this means that the author can shows metadata attributes from task-oriented viewpoint.



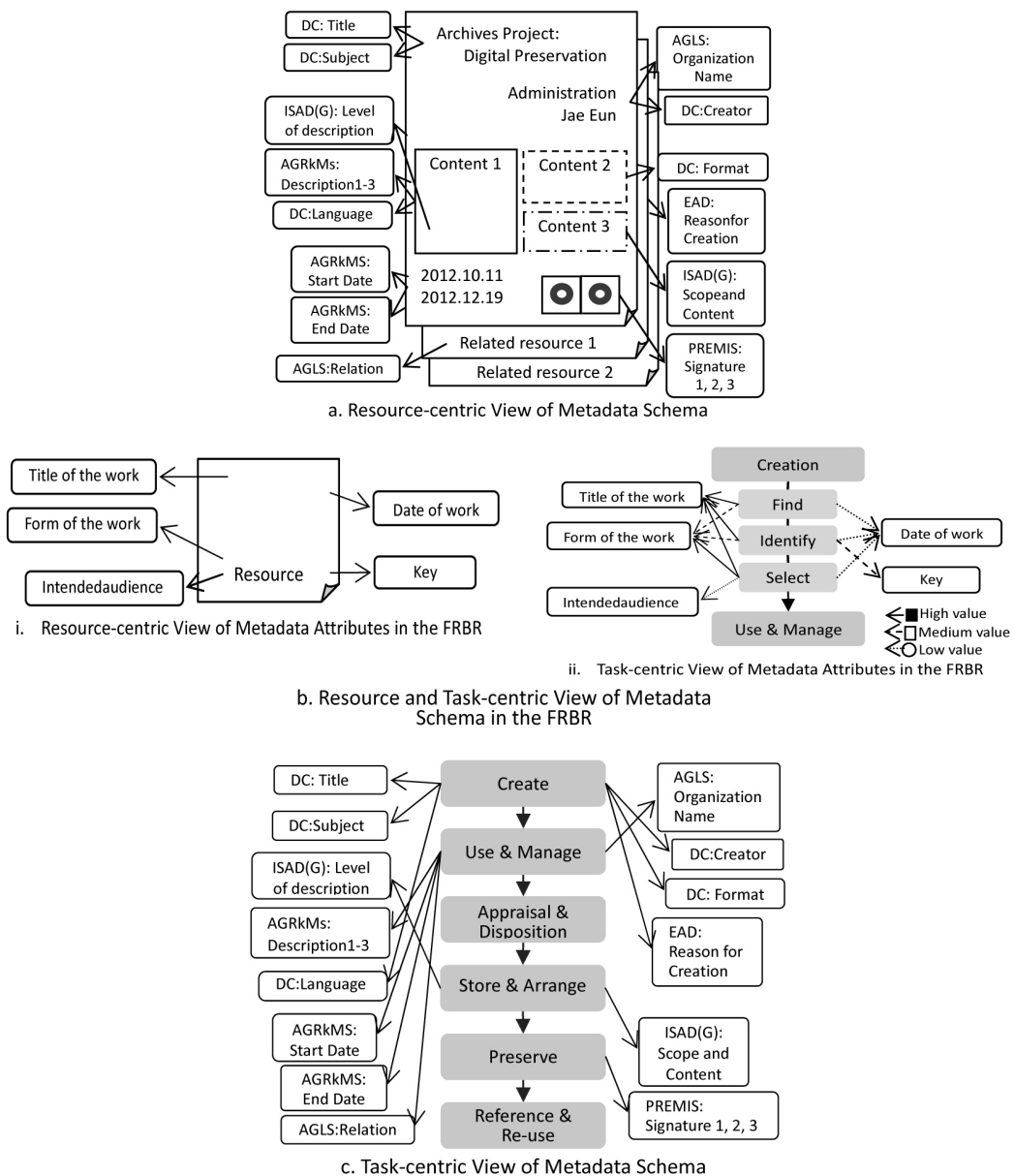
<Figure 4> The Mapping of User Task and Attributes in the FRBR
 (From IFLA Study Group on the Functiona Requirements for Bibliographic Records)

Figure 5a shows the metadata elements (title, creator, language, date, signature, relation and so on). This means that metadata elements are designed from a resource-centric view. As mentioned above, she uses the relation between metadata and tasks to identify features of metadata schemas in this study. So, she examines metadata in each stage of the lifecycle from a task-centric view. Figure 5b shows the FRBR User Tasks in the lifecycle. Figure 5c shows some metadata elements and their related stages in the lifecycle.

For example, *ISAD(G): level of description* is an element that describes the level of a resource for archiving. If this element applied to a stage in the lifecycle, it should be included and used in the storing or archiving stage. For another example, *reason for creation* of EAD expresses the reason why the resource is created. This element applies to the creation stage in the lifecycle. In other words, a resource is examined in every stage of task of lifecycle for the tasks in the stage.

In general, a metadata standard is defined from a resource-oriented viewpoint in accordance with the purpose of the standard. On the other hand, each metadata element is used in a task at a lifecycle

stage. The task-attribute relationship given in FRBR is a well-known example of the relationship. The task-attribute relationship is useful to clarify the feature of a metadata standard from the viewpoint of tasks performed in the resource lifecycle stages. Task oriented view of metadata standards is advantageous to define mappings between metadata standards along with the lifecycle stages.



<Figure 5> From Resource-centric to Task-centric View of Metadata Schema

5. Feature Analysis of Metadata Schemas based on Lifecycle

5.1 Viewing Differences from Descriptive Elements

As the first research, the author performed element mapping and analysis of archival metadata from a viewpoint of lifecycle in order to analyze the feature of metadata standards.

Mapping of metadata standards into the records lifecycle is examined explicitly to 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.

During a workflow that takes place according to a metadata standard, a metadata element is created at some point and used in the whole records lifecycle. Therefore, the author used the viewpoint of 'Creation and revision (update) of metadata'. She defines a Creation Stage and Update Stage of a metadata element in the records lifecycle as the stage where the metadata element is given an initial value and revised, respectively. The creation and update stages are called a primary stage of the metadata element.

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. 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.

In order to analyze descriptive elements into a corresponding stage of the lifecycle, she carried out classification and mapping, using the two analysis methods.

1. Analyze the feature of metadata standard.

For example,

preservationLevelDateAssigned of PREMIS,

Before analyzing the descriptive element, PREMIS is metadata standard for preservation of digital objects and is used in the preservation stage of records lifecycle basically.

2. Find and classify a corresponding keyword or a related meaning from the value of descriptive element.

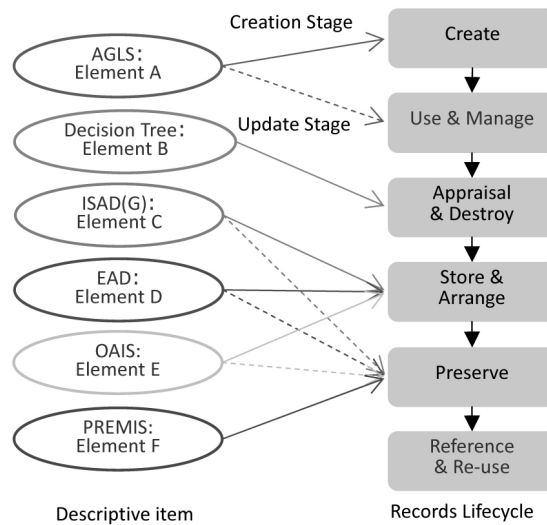
For example,

preservationLevelDateAssigned of PREMIS defined "The date, or date and time, when a

particular preservationLevelValue was assigned to the object”. This element means not only the period which determines a preservation level, but also the period which changes the preservation demand and policy etc of repository. Thus, she decided this element as a preservation stage of lifecycle and classified it.

And then, mapping metadata standards into the records lifecycle is done in two steps:

- Step 1.** Extract every metadata element from each metadata schema standard one by one, and determine the primary stages in the records lifecycle for the element.
- Step 2.** 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 over viewing of the metadata element sets across the stages of the lifecycle.



〈Figure 6〉 Classification Criteria of Metadata Standards into the Records Lifecycle using the Step 1 and 2

5.2 Mapping to Determine the Primary Stages in the Lifecycle

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

This section shows an analysis of a metadata element extracted from each metadata standard.

Because every schema has many elements, this section shows the analysis using examples. Each element shown in the paragraphs below 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 or updated.

(1) AGLS Metadata

Availability is primarily used for non-digital resources, providing information on how the user may acquire physical accesses to a resource. Because this element explains the availability of resources in the real usage environment, she classifies the stage of this element as *Use & Manage*. The value of the element is updated in *Appraisal & Destroy* and *Reference & Re-Use*.

<Table 1> An Example of AGLS Metadata

Element of AGLS Metadata: <i>Availability</i>	
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. This element was classified 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: <i>Acquire for other purpose</i>	
Point of view	Lifecycle Stages
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: <i>archdesc</i>	
Point of view	Lifecycle Stages
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): <i>Level of Description</i>	
Point of view	Lifecycle Stages
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: <i>change history before archiving</i>	
Point of view	Lifecycle Stages
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: <i>creating Application</i>	
Point of view	Lifecycle Stages
Creation	Create
Update	Store & Arrange, Preserve

Based on this investigation, she analyzed the relationship between each metadata standard and the lifecycle stages. The full result is shown Appendix 1, 2, 3 in doctoral dissertation.

5.2.2 Step 2: Determine Primary Stages for Metadata Standards

This section shows lifecycle stage that mainly expresses all metadata elements and each metadata standard.

(1) AGLS Metadata

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

In the first study, the author mapped the lifecycle stages to metadata elements extracted from the metadata standards. In this mapping, for every element extracted from metadata standards, she 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. And, Figure 7 is a figure that expressed based on Table 7.

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.

In order to show the stage with highest percentage in resource lifecycle, the author highlighted

the corresponding stages. In other words, AGLS, DPC, EAD & ISAD(G), OAIS & PREMIS show the stages of highest percentage of *Use & Manage*, *Appraisal & Destroy*, *Store & Arrange*, and *Preserve*, respectively. Thus, these stages show the stage where each metadata is mostly used in resource 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, this mapping shows that there is a clear split between Create stage and Update stage. 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, she can identify the overall features of the metadata standards shown in Figure 7 from the statistics shown in Table 7.

Every metadata schema is related to all stages of the lifecycle except the decision tree. Figure 7 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.

Figure 7 is useful to view the stages where crosswalks between metadata schemas are efficiently performed. This is because it helps us to identify the correspondence between elements of similar meanings by showing the correspondence of elements to lifecycle stages.

The unified framework (new viewpoint) to identify the features of archival metadata standards proposed in the first study 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.

<Table 7> Metadata Standards shown by Figures

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

	AGLS	DPC	EAD	ISAD (G)	OAIS	PREMIS
Create						
Use & Manage						
Appraisal & Destroy						
Store & Arrange						
Preserve						
Reference & Re-use						

〈Figure 7〉 Stage of Lifecycle shown by Metadata Elements

6. Facet Analysis of Archival Metadata Schemas for Metadata Interoperability

From the feature analysis discussed in the previous chapter based on the resource lifecycle, she has learned that we need to use not only the semantic description given in the definition of a metadata element but also the context information of the element which can be obtained from the lifecycle.

In her second study which is presented in this chapter, she introduces a task-centric view of metadata elements in order to create metadata mappings across the lifecycle stages. In the rest of this chapter, she describes a task-based model of the resource lifecycle, which she calls the Task Model. Then, she defines a task-centric view of metadata elements and she introduces 5W1H categories to characterize metadata elements for a task-oriented semantics analysis of the metadata elements (Baek Jae-eun and Sugimoto 2012).

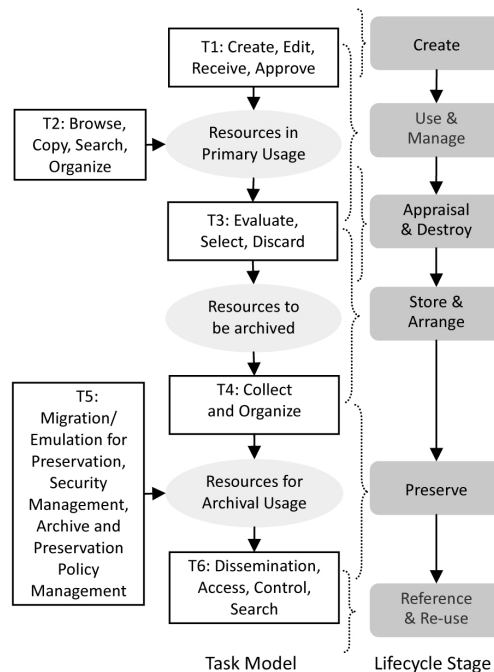
6.1 Task-oriented View of Records Lifecycle-Task Model

The records lifecycle defines stages of records—from creation at offices to preservation in archives. In order to examine in detail of the relationship between resource and each stage in records lifecycle, the author proposes the Task model. The Task model is defined in parallel to the records lifecycle.

The Task Model is a model that is created based on the records lifecycle. The Task model is proposed in this study in order to analyze metadata standards in detail from the viewpoint of the tasks performed in each stage of the records lifecycle. In this model, task groups, which are composed of several tasks and linked to resources, are associated with the lifecycle stages. The records lifecycle briefly describes what tasks are performed in each stage but it is not clear how the resource attributes are used in the stages. On the other hand, the Task model is more descriptive than the lifecycle because each task in the groups indicates attributes of resources used in the task.

As shown in Figure 8, the Task Model defines the tasks performed in each stage of the records lifecycle. The Task Model is composed of six task groups (T1-T6) defined as follows,

- Task 1: Creation tasks: Tasks used for initial creation including those for the approval process,
- Task 2: Primary Usage tasks: Tasks for primary users to find and browse resources,
- Task 3: Appraisal and Retention tasks: Tasks to select and discard resources,
- Task 4: Archival Transformation tasks: Conversion and transformation tasks for archival storage,
- Task 5: Preservation tasks: Maintenance tasks for archival storage, and
- Task 6: Archival Usage tasks: Tasks to find and use archived resources



<Figure 8> Task Model

The lifecycle stages are shown to the right of the Task model in Figure 8. The Task model complements the lifecycle model in the aspects of tasks performed at each stage of the lifecycle and explicitly shows the transition in status of the resources.

6.2 Task-centric View of Metadata Elements and 5W1H Categories

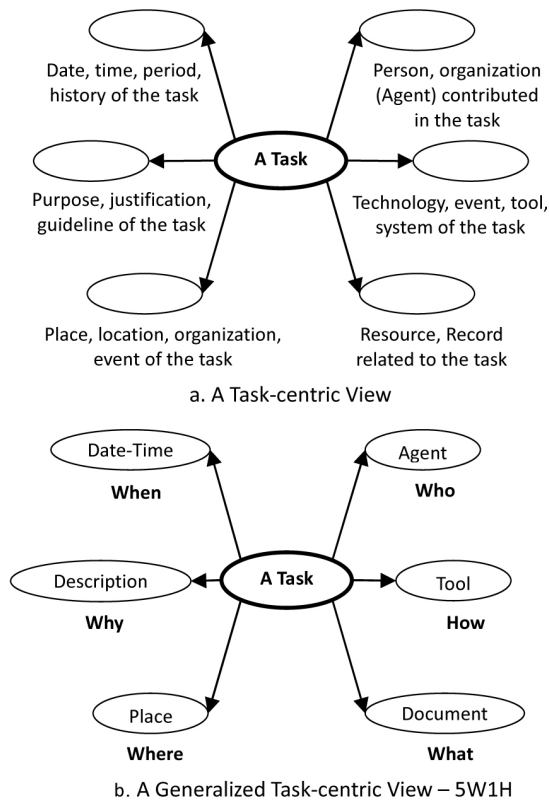
6.2.1 Combination of Task-oriented Model and Metadata Elements

An execution of a task causes an event on resources. The author describes the relationships of the metadata elements and an event (on resources).

Figure 9 shows a task-centric view of metadata. A task-centric view of metadata is to define metadata standards in the context of tasks. In Figure 9a, 'A Task' is linked to values of metadata elements, i.e., an entity, such as right, time, purpose, or person. This is the reverse of Figure 9b. As shown in Figure 9a, every single task is associated with those entities shown as a circle. These entities are agents that play some roles in the task, locations or institution where the task is performed, reasons and guidelines that perform the task, and so forth. Generally, the relationships between a task and its associated entities are determined task-by-task, but we need an appropriate categorization of these tasks.

In Figure 9b, links from the task are labeled using 5W1H categories, i.e., an input link to an entity is reversed as a metadata element of the entity. In this study, we propose to use 5W1H categories - who, where, when, what, why and how - as generalized categories to express the relationship of a related entity and task, as shown in Figure 9b. Figure 9b is derived from Figure 9a by categorizing the relationship from the task to the values. The paragraphs below show detailed explanation of this categorization.

Many, but not all, of the entities associated with a task are recorded as a metadata value in accordance with the schema used in a particular system. However, in general, data models of metadata standards are defined based on data entities but not tasks. This means that the metadata elements are not explicitly related to the tasks, in spite of the correspondence between lifecycle stages and metadata elements, which she found in her previous study. In addition, the difference of data models of metadata standards has to be taken into account to map their metadata elements. The underlying idea of this study is to use the generalized task-centric view of metadata to map metadata schemas instead of the data entity-centric view in conventional mapping.



<Figure 9> Task-centric View of Metadata and 5W1H

6.2.2 5W1H Categories

5W1H attributes are used to identify categories of metadata elements. A metadata element category represented by a 5W1H attribute is called 5W1H categories in the rest of this paper. The paragraphs below show definitions of the 5W1H categories for this study.

- (1) What: Information about preservation processes and tasks such as resources used for reservation, rights, and rules for preservation.
- (2) Why: A reason for an operation on a resource, e.g., purpose of creation or criteria for preservation.
- (3) When: Time, date, period, and era when the task was performed, e.g. date of creation or expiration.
- (4) Where: Place, location, organization, or institution where the task was performed.
- (5) Who: An agent related to a resource, e.g., a person or an organization that has made a contribution to the task.

- (6) How: Operations performed on a resource and related information, e.g., file formats, software tools, rights management, and so forth.

6.3 Analysis Criteria of Classification

6.3.1 Vocabulary of Systematic Classification by 5W1H Categories and Lifecycle Tasks

In the second study, she classifies every metadata element - AGLS, AGRkMS, EAD, OAIS, PREMIS, and the attribute sets of DPC - using the 5W1H categories and tasks in the lifecycle. She used explanation texts of each metadata element to find keywords. And keywords are used to classify all the metadata elements into the 5W1H categories and the lifecycle tasks. This classification was carried out manually because she had to interpret the meanings and intention of the explanations. The author prepared a set of keywords for each task group and 5W1H categories and used the keywords to classify every element into a task and map it to 5W1H categories.

Table 8 and Table 9 show the keywords for the 5W1H categories and the Task model, respectively. The keywords are manually extracted as typical words to express a category and a task, respectively. More than one keyword may appear in the definition of a metadata element.

<Table 8> Classification Vocabulary with 5W1H Categories

5W1H Categories	Keywords (Example)
Who	Agent, Author, Creator, Institution, Name, Organization, People, Person etc
When	Date(s), Period, Time, Month, Day, Year etc
Where	Agent, Country, Institution, Location, Name, Organization, Place etc
What	Administration, Bibliography, Description, History, Policy, Relationship, Right etc
How	Action, Event, File format, Hardware/Software, Metadata scheme, Technique, Tool, etc
Why	Purpose, Reason etc

<Table 9> Classification Vocabulary with Task Model

Task Group	Keywords(Example)
T1: Create, Receive, Approve	Create, Make, Produce etc
T2: Browse, Copy, Search, Organize	Access, Manage, Use etc
T3: Evaluate, Select, Discard	Accept, Appraise, Destruct, Select etc
T4: Collect and Organize	Archive, Collect, Manage, Store etc
T5: Migration/Emulation for Preservation, Archive/ Preservation Policy Management	Archive, Manage, Store, Preserve etc
T6: Dissemination, Access, Control, Search	Access, Search, Use etc

The paragraphs below show the classification guideline,

1. Find keywords in the title, definition and guideline texts of a metadata element,
2. If no keyword is found, find a term (or terms) whose meaning is similar to a keyword,
3. If matching by 1 or 2 does not succeed find a keyword (-s) in a use-case example of the element.

6.3.2 Classification Procedure

The classification workflow has two steps, classification by 5W1H categories (step 1) and classification by 5W1H categories in the lifecycle tasks (step 2).

Step 1 Metadata Mapping by 5W1H categories

- 1-1 Classification of descriptive elements: For every element of each metadata standard, examine whether the definition of the element includes one or more keywords listed in Table 8 and, if found, classify the element to the corresponding category (-ies).
- 1-2 A mapping among metadata standards: In each 5W1H category, compare elements among the standards and create mappings. If a mapping table for any of the standards exists, it is also used to determine the mapping.

Step 2 Classification by 5W1H categories in the Lifecycle Tasks

- 2-1 Classification of descriptive elements by tasks: For every element of each metadata standard, examine whether the definition text of the element includes one or more keywords listed in Table 9 to classify the element to the corresponding task(s).
- 2-2 Classification of descriptive elements by 5W1H categories: For every element classified to a task, apply Step 1-1 to classify the element by 5W1H in each task.
- 2-3 A mapping among metadata standards: In each 5W1H category of each task, create mappings.

6.4 Mapping Metadata Schemas in 5W1H Categories and Lifecycle Tasks

This section shows 5W1H categories and task groups using example mappings among the elements of metadata standards chosen for the comparison.

6.4.1 Classification of Descriptive Elements in 5W1H Categories

(1) *Publisher* of AGLS Metadata

The *Publisher* element of AGLS means an entity is responsible to make a resource available.

AGLS says that this element may be used to provide details of the organization that provides access to the service. As shown in Table 8, vocabulary of 5W1H categories and agents such as organizations, and institutions are often used as a location. Therefore, agents in Table 8 includes both *Who* and *Where*.

Corresponding elements of EAD and OAIS in these categories are shown in the table. These elements have similar keywords and meaning, like AGLS. So, she classified them both *Who* and *Where*. Elements of other standards have not corresponding elements to this element. It means that other standard elements have no corresponding vocabularies of this element.

Metadata which have no corresponding elements to *Publisher*, AGRkMS is the minimum metadata standard for record management. AGRkMS use general metadata element that describes resource, from AGLS. In PREMIS, the element which has relevance to intellectual entity is premised on using from other metadata standard. And, DPC does not contain element about intellectual contents that AGLS express, because of the attribute which is extracted from the evaluation process for preservation.

<Table 10> AGLS: *Publisher*

5W1H	Metadata Standards		
	AGLS	EAD	OAIS
Who	Publisher	Publication Statement	Name of publisher
Where	Publisher	Publication Statement	Place of Publication
		Publisher	Name of publisher

(2) *Date Range* of Australian Government Recordkeeping Metadata Standard

Date Range element of AGRkMS means date and time associated with an entity. It has *Start Date* and *End Date* as its sub-elements. The category of these elements is obviously *When*. Corresponding elements with *Date Range* element of AGRkMS includes AGLS, EAD, OAIS and PREMIS.

<Table 11> AGRkMS: *Date Range*, *Start Date* and *End Date*

5W1H	Metadata Standards				
	AGLS	AGRkMS	EAD	OAIS	PREMIS
When	Date	Date Range		Date of Publication	dateCreatedByApplication
		Start Date		Change History Before Archiving	
					PreservationLevelDateAssigned
		End Date			

(3) *Multiple media formats* of DPC Decision Tree Attributes

Multiple media formats element of the DPC attributes means that a resource could have more than one media format regardless of digital or non-digital.

Here, *format* means a type of media of a resource and also a technology required to render a resource. Therefore, the former is categorized in *What* and the latter in *How*. Corresponding elements of the DPC attributes in these categories have AGLS, AGRkMS.

<Table 12> DPC Attribute Set: Multiple media formats

5W1H	Metadata Standards			
	AGLS	AGRkMS	DPC	PREMIS
What	Format	Format	Multiple media formats	
How	Format	Format	Multiple media formats	Format

(4) *Title of the Unit* of EAD

Title of the Unit element of EAD means the name of the described materials. Since *Title of the Unit* expresses a name of a resource handled in a task, it is categorized in *What*.

<Table 13> EAD: *Title of the Unit*

5W1H	Metadata Standards			
	AGLS	AGRkMS	EAD	OAIS
What	Title	Name	Title of the Unit	Resource description

(5) *Reason for Creation* of OAIS

Reason for Creation element of OAIS is used to specify a reason(s) of creation of a resource. Table 8 shows that reasons or purposes which create, manage, destroy, and preserve resource are included in *Why*. Thus, this element is categorized in *Why*. Corresponding elements are shown in AGLS. *Description* of AGLS is included here as an element of a broader meaning.

<Table 14> OAIS: Reason for Creation

5W1H	Metadata Standards	
	AGLS	OAIS
Why	Description	Reason for Creation

(6) *Size* of PREMIS

The *Size* element of PREMIS expresses a technical value such as file size. Elements that express technical values are primarily categorized in *How*. It is mapped to *Description* of AGLS which has a broader meaning than *Format* of AGRkMS and *Extent* of EAD as well.

<Table 15> PREMIS: *Size*

5W1H	Metadata Standards			
	AGLS	AGRkMS	EAD	PREMIS
How	Format	Format	Extent	Size

6.4.2 Mapping in the Task Groups

Followed by the mapping of descriptive elements in 5W1H categories, this section shows the classification obtained by an application of Step 2. This section describes an example, which shows ‘T3: Evaluate, Select, Discard’ of the Task model. A part of the whole classified table shows Table 16.

T3 is associated with “Appraisal and Disposition” in the lifecycle (Figure 3) where the resource is selected and evaluated for archiving. She has classified elements of all metadata standards

By the keywords discussed in section 6.3 and shown in Table 9, T3 includes the keywords, such as appraisal, selection, destruction,

approval etc. The result of classification that performed using these keywords express that no element of PREMIS is included T3. PREMIS has no element directly related to T3, because PREMIS is primarily designed for the ‘Preservation’ stage in the records lifecycle. Table 16 shows a part of the all mappings among the elements classified to the task group T3. This mapping table shows the relationships between the elements classified in the 5W1H category in each task group. Format of *How*, which expresses the format/environment (a technology/format that has a technical meaning) for performing a resource in T3, is mapped *Format* of AGLS, *Format* of AGRkMS, many elements of DPC and *Table Column Specification** of EAD etc. This description is a part of examples and Format corresponding to more metadata elements.

<Table 16> Example of Mapping of Metadata Schemas for Archiving and Preservation
(T3: Evaluate, Select, Discard)

Task model	5W1H	AGLS	AGRkMS	Decision Tree of DPC	EAD	OAIS	PREMIS	
T3	Who		Position		Sponsor Publisher Publication Statement Author			
		When	Date	Date Range Start Date		Date Publication Statement Date of Unit		
				End Date				
	Where		Identifier Scheme Position		Sponsor Publisher Author			
		What	Mandate	Permission				
			Right	Security Caveat Right	Long term value Justify preservation	Ingest Process History	Processing Information Appraisal Information	
			Identifier Jurisdiction					
	Format		Format					
	How	Mandate	Permission	Negotiate for the source to supply				
		Right	Right					
		Format	Format	Digital version be selected for preservation			Table ColumnSpecification	
				Manageable file format				
			Document Form		Ingest Process History			
			Change history				Revision Description	
	Why			Long term value justify preservation Acquire for other purposes				
			Document Form					

6.4.3 Analysis of Metadata Schemas using 5W1H Categories and Task Group

In this section, she analyzes the result of classification and mappings shown in the previous sections. She created tables using the classification presented in the previous section. Table 17 and Table 18 show statistics of the classification of the elements into the 5W1H categories and task groups, respectively.

In Table 17, a number in a column shows how many descriptive elements of each standard are classified into each of the 5W1H categories. For example, the numbers of AGLS elements classified into *Who*, *When*, *Where*, *What*, *How* and *Why* are 4, 1, 4, 15, 7, 1, respectively.

The bottom row shows the total number of elements for every standard. Because an element can be classified into one or more 5W1H categories, the sum of the 5W1H rows may not be the same as the number of elements shown at the bottom of each column.

Table 17 shows that the most common portion of the AGLS elements is *What*, but in the case of PREMIS the most common portion is *How*. This means the descriptive element of AGLS expresses the meaning of ‘descriptive information about a resource’ or has the relating elements. And PREMIS mean there are many elements that express the meaning about ‘a technical feature about resource’.

<Table 17> Metadata Standards in 5W1H categories shown by Figures

Metadata \ 5W1H	AGLS	AGRkMS	DPC	OAIS	EAD	PREMIS
	19	20	27	53	146	95
Who	4	0	0	4	18	2
When	1	1	0	2	9	7
Where	4	2	0	8	23	7
What	15	15	16	24	99	21
How	7	13	12	43	47	125
Why	1	1	2	3	0	2

She arranged corresponding metadata standard to 5W1H categories in each task group and, expressed by figures. Table 18 shows similar statistics sorted according to the task groups. This shows a feature of the metadata standards discussed in section 5. This table is sorted by the task groups but not by the lifecycle stages used in her first study.

Table 19 shows the overall distribution of elements in the task groups. Each row of this table shows values for each task group. A column shows values for a standard. Each value in a box contains a percentage of elements classified to a corresponding task group. This table shows

a feature of the metadata standards analyzed from the viewpoint of the task groups. It shows a feature similar to but more refined than in her first study shown in section 5.

<Table 18> Metadata Standards in the Task Groups shown by Figures

Task	5W1H	AGLS	AGRkMS	DPC	EAD	OAIS	PREMIS
Task1	Who	3			15		
	When	1	1		5		1
	Where	3	1		16		
	What	6	3		26		
	How	3	1		11		2
	Why	1				1	
Task2	Who	2			15	1	
	When	1	1		5		
	Where	1	2		18	1	
	What	11	14		29	3	
	How	5	11		9	2	11
	Why		1				
Task3	Who				16		
	When	1	1		5		
	Where		1		15		
	What	10	7	5	19	1	
	How	6	8	11	10	1	
	Why	1	1	2			
Task4	Who				17	3	
	When	1	1		9	2	1
	Where		2		12	7	1
	What	3	14		91	23	2
	How	3	11		46	42	31
	Why		1			2	
Task5	Who				17	2	2
	When	1	1		8	2	5
	Where		1		21	7	5
	What	3	1	11	61	21	18
	How	3	1	1	32	42	119
	Why					3	2
Task6	Who	2			16	1	
	When	1	1		5		
	Where	1			17	1	
	What	4			24	3	
	How	3			10	3	11
	Why						

<Table 19> Metadata Standards in the Task Groups by Percentage

a. The highlighted metadata in task group (from each row)

Task Group	AGLS	AGRkMS	DPC	OAIS	EAD	PREMIS
Task 1	21	7		12	1	1
Task 2	25	33		13	4	5
Task 3	22	21	60	11	1	
Task 4	9	33	40	29	45	17
Task 5	9	5		23	44	72
Task 6	14	1		12	5	5

b. The highlighted task (from each column)

Task Group	AGLS	AGRkMS	DPC	OAIS	EAD	PREMIS
Task 1	21	7		12	1	1
Task 2	25	33		13	4	5
Task 3	22	21	60	11	1	
Task 4	9	33	40	29	45	17
Task 5	9	5		23	44	72
Task 6	14	1		12	5	5

Each column shows the distribution of elements in the lifecycle. For example, AGLS could be used well in task 1, 2, 3 and 6, and PREMIS could be used in task 5. The boxes surrounded by bold lines show the highest value for each standard, and can be interpreted to imply a main task to which the standard is well suited.

Each row of the table shows the different weightings of a task for each standard. The highlighted boxes show the highest values in a row, which would mean the most suitable standard for each task.

Table 19a shows the highlighted metadata in the task groups, from each row (view of task). For example, Task 2 shows the highest value in AGRkMS, Task 4 shows the highest value in EAD. Table 19b shows the highlighted task from each column (viewpoint of metadata standard). For example, AGLS is high-lighted for task 2 and PREMIS shows the highest value for task 5. Percentage is rounded. The highlighted boxes have the highest number in each row.

In this research, she proposed the 5WIH categories and the Task models to analyze the features of descriptive elements of archival and preservation metadata standards, and also to create mappings among the standards.

The author considers that the two models are useful because they provide simple semantics which help to identify meanings of descriptive elements from the viewpoint of tasks in the lifecycle and aspects required to identify the tasks, respectively. The task-centric view proposed in this

paper helps with access to archived information resources across repositories and over time.

Contextual semantics are implicit in the definition of metadata elements, which is one of the major barriers to creating mappings between metadata standards. A shift in the viewpoint of metadata elements, i.e. from resource-centric to task-centric, helps us find and use the contextual information in metadata mappings.

Thus, she defined the Task model using the 5W1H categories for metadata mappings to improve metadata interoperability over the whole lifecycle. The author learned that it is crucial to combine metadata standards for archiving and preservation of digital resources.

7. Conclusion

Digital resources are widely used in our modern society. The rapid growth of digital resources has not only the popularization of digital resource but also some major problems. One of these problems is to manage and maintain digital resources for future generations. Thus, we are facing fundamental problems of how to manage and preserve digital resources over time.

For archiving and long-term preservation of digital resources, proper policies and strategies (developing systems, guidelines, metadata schemas and so on) are necessary. Several standard methods for preserving digital resources have been developed and are in use. It is widely recognized that metadata schema is one of the most important components of archiving and preservation of digital resources.

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. In addition, metadata standards have a base data model, and a metadata element is defined as a property (or an attribute) of an entity included in the data model.

On the other hand, metadata standards are affected by tasks performed in the records lifecycle. Metadata has to be used in accordance with the tasks. However, in general, the data model is not explicitly linked to the records lifecycle or tasks, which means that users have to find appropriate metadata standards in accordance with the lifecycle stages. It is crucial to select and combine metadata standards in accordance with requirements in an application domain and in the records lifecycle.

This study identified and analyzed features of archival metadata standards to select, combine, and use them appropriately throughout the whole resource lifecycle, for archiving and preservation

of digital resource.

In order to analyze the features of the metadata standards, the author identified the primary records lifecycle stage(s) where a standard would be applied. As a result of this analysis, she clarified the features of the standards from the viewpoint of relationships between the elements and the lifecycle stages. In addition, she found that a metadata standard element is related to a task.

Based on this feature analysis, this study has proposed the Task Model to clarify tasks in the records lifecycle and to categorize metadata elements from the viewpoint of the tasks. Based on this, the author has proposed to categorize metadata elements using 5WIH categories coupled with the Task model derived from the resource lifecycle. In this study, metadata elements of the chosen standards are categorized using the 5WIH categories and mapped to each other. The mappings are grouped and sorted in accordance with the Task model.

Mapping between metadata schemas is often required throughout the preservation process because different schemes are used in different stages of the records lifecycle. Therefore, it is crucial to build a unified framework to enhance the interoperability of metadata schemas. 5WIH categories and the Task model are used as a unified viewpoint in this study. The author thinks that the proposed models help identify the contexts of descriptive elements and define crosswalks among standards. This study presents a basis for the interoperability of different metadata schemas used in digital archiving and preservation.

A major achievement of this study is the feature analysis of archival metadata schemas from the two viewpoints, a records lifecycle-view and a Task model-view. And the core contribution of this study is a shift from a conventional resource-centric view to a task-centric and lifecycle-centric view. Through this study, the author has learned that a metadata standard is related to a task in the records lifecycle. She also has learned that any single metadata standard for archiving and preservation does not cover the whole resource lifecycle.

One of the most difficult aspects of this research was the manual mapping and classification of the metadata vocabularies. The author has not yet applied the mapping table to test metadata interoperability in a practical environment due to a limitation of the resources available for her research. Therefore, the author has not included an evaluation of mapping in this paper. The author believes that evaluation of the semantic mappings between different metadata elements is necessary and important.

As the goal of this study is to propose a unified framework that improves the interoperability between metadata elements, creation of the mappings that cover several major standards and are carried out by manual but semi-formalized process, is sufficient to show the feasibility of the

framework as the goal of this study. Evaluation of the mappings based on real metadata done by machines is left for future work. In addition, the author has left the development of software tools for task groups as an object of future study.

Another issue reserved for future study is to introduce the concept of application profiles into the task-centric model. This is because the metadata schemas expressed as application profiles are primarily resource-centric and task-oriented information is not explicitly described as a part of metadata schema. She thinks that a task-centric application profile for archival metadata may help with metadata interoperability and may help to select necessary metadata elements for each task.

References

- [1] Baca, M. 2003. "Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information." *Cataloging & Classification Quarterly*, 36(3-4): 47-55.
- [2] Baek, Jae-eun and Sugimoto, S. 2007. Selection Guidelines for Preservation Method of Digital Resources. [online] [cited 2013. 2. 15.]
<http://www.tulips.tsukuba.ac.jp/limedio/dla_m/M93/M932902/9.pdf>
- [3] Baek, Jae-eun and Sugimoto, S. 2010. "Feature Analysis of Metadata Schemas for Records Management and Archives from the Viewpoint of Records Lifecycle." *Journal of Korean Society of Archives and Records Management*, 10(2): 75-100.
- [4] Baek, Jae-eun and Sugimoto, S. 2012. "A task-centric model for archival metadata schema mapping based on the records lifecycle." *International Journal of Metadata, Semantics and Ontologies archive*, 7(4): 269-282.
- [5] Chan, L.M. and Zeng, M.L. 2006. Metadata Interoperability and Standardization-A Study of Methodology Part 1. *D-Lib Magazine*, 12(6). [online] [cited 2013. 2. 15.]
<<http://www.dlib.org/dlib/june06/chan/06chan.html>>
- [6] Chen, Y.N., Chen, S.J. and Lin, S.C. 2003. A metadata lifecycle model for digital libraries: methodology and application for an evidence-based approach to library research. In: *Proceedings of the World Library and Information Congress: 69th IFLA General Conference and Council*, Berlin. [online] [cited 2013. 1. 20.] <<http://www.citi.sinica.edu.tw/papers/sophy/117-F.pdf>>
- [7] Day, M. 2001. Metadata for Digital Preservation: A Review of Recent Developments. In: *Proceedings of the Research and Advanced Technology for Digital Libraries: 5th European Conference (ECDL 2001)*, Darmstadt, Germany, 161-172. [online] [cited 2013. 2. 15.]

- <http://link.springer.com/content/pdf/10.1007%2F3-540-44796-2_15.pdf>
- [8] Digital Preservation Coalition. 2006. Decision Tree for Selection of Digital Materials for Long-term Retention. [online] [cited 2013. 2. 15.]
<<http://www.dpconline.org/advice/preservationhandbook/decision-tree>>
- [9] Dunsire, G. 2010. The Vocabulary Mapping Framework and its potential for improving metadata interoperability in the Semantic Web. In: *Proceeding of the EUROVOOC Conference, November 2010*, Luxembourg.
- [10] Evans, J., Mckemmish, S. and Bhoday, K. 2005. "Create Once, Use Many Times: The Clever Use of Recordkeeping Metadata for Multiple Archival Purposes." *Archival Science*, 2005(5): 17-42.
- [11] Haslhofer, B. and Klas, W. 2010. "A Survey of Techniques for Achieving Metadata Interoperability." *ACM Computing Surveys*, 42(7): 1-37.
- [12] International Council on Archives. 2000. ISAD(G): General International Standard Archival Description - Second edition. [online] [cited 2013. 4. 30.]
<<http://www.ica.org/10207/standards/isadg-general-international-standard-archival-description-second-edition.html>>
- [13] International Federation of Library Associations and Institutions. 2009. Functional Requirements for Bibliographic Records-Final Report. [online] [cited 2013. 2. 15.]
<http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf>
- [14] International Organization for Standardization. 2006. *ISO 23081-1:2006: Information and documentation - Records management processes - Metadata for records - Part 1: Principles*.
- [15] International Organization for Standardization. 2009. *ISO 23081-2:2009: Information and documentation - managing metadata for records - Part 2: Conceptual and implementation issues*.
- [16] JISC. 2009. The Vocabulary Mapping Framework (VMF): an introduction. [online] [cited 2013. 2. 15.]
<[http://www.doi.org/VMF/documents/VocabularyMappingFrameworkIntroductionV1.0\(091212\).pdf](http://www.doi.org/VMF/documents/VocabularyMappingFrameworkIntroductionV1.0(091212).pdf)>
- [17] National Archives and Records Administration. [n.d.]. Preservation Programs at the National Archives. [online] [cited 2013. 1. 20.] <<http://www.archives.gov/preservation/internal/>>
- [18] National Archives of Australia. 2006; 2010; 2011. AGLS Metadata: Australian Government Implementation Manual (Version 3.0). [online] [cited 2013. 2. 15.]
<http://www.naa.gov.au/Images/AGLS%20Manual%203.0%20-%20most%20recent%20revision%2030%20September%202011_tcm16-49605.pdf>
- [19] National Archives of Australia. 2008. Australian Government Recordkeeping Metadata Standard

- (Version 2.0). [online] [cited 2013. 5. 1.]
<http://www.naa.gov.au/Images/AGRkMS_Final%20Edit_16%2007%2008_Revised_tcm16-47131.pdf>
- [20] National Archives of Australia. 2010. AGLS Metadata Standard Part 1-Reference Description (Version 2.0). [online] [cited 2013. 5. 1.]
<<http://www.agls.gov.au/pdf/AGLS%20Metadata%20Standard%20Part%201%20Reference%20Description.PDF>>
- [21] National Archives of Australia. 2011. Australian Government Recordkeeping Metadata Standard Implementation Guidelines (Version 2.0). [online] [cited 2013. 5. 1.]
<http://www.naa.gov.au/Images/AGRkMS%20Implementation%20Guidelines_tcm16-50156.pdf>
- [22] Nilsson, M., Baker, T., Johnston, P. and DCMI. 2008. The Singapore Framework for Dublin Core Application Profiles. [online] [cited 2013. 1. 20.]
<<http://dublincore.org/documents/singapore-framework>>
- [23] OCLC/RLG Working Group on Preservation Metadata. 2001. Preservation Metadata for Digital Objects: A Review of the State of the Art. [online] [cited 2013. 3. 9.]
<http://www.oclc.org/resources/research/activities/pmwg/presmeta_wp.pdf>
- [24] PREMIS Editorial Committee. 2012. PREMIS Data Dictionary for Preservation Metadata, version 2.2. [online] [cited 2013. 3. 9.]
<<http://www.loc.gov/standards/premis/v2/premis-2-2.pdf>>
- [25] Shimazu, K., Arisawa, T. and Saito, I. 2006. Interdisciplinary contents management using 5W1H interface for metadata. In: *Proceeding of IEEE/WIC/ACM International Conference on Web Intelligence (WI'06)*, 909-912.
- [26] The Library of Congress. 2002. Encoded Archival Description (EAD). [online] [cited 2013. 4. 30.] <<http://www.loc.gov/ead>>
- [27] Zeng, M.L. 1999. "Matadata Elements for Object Description and Representation: A Case Report from a Digitized Historical Fashion Collection Project." *American Society for Information Science*, 50(13): 1193-1208.

