

Internal Structure of Information Packages in Digital Preservation

Seungmin Lee *

Department of Library and Information Science
Sookmyung Women's University, Republic of Korea
E-mail: ableman@sookmyung.ac.kr

ABSTRACT

The description of preserved resources is one of the requirements in digital preservation. The description is generally created in the format of metadata records, and those records are combined to generate information packages to support the process of digital preservation. However, current strategies or models of digital preservation may not generate information packages in efficient ways. To overcome these problems, this research proposed an internal structure of information packages in digital preservation. In order to construct the internal structure, this research analyzed existing metadata standards and cataloging rules such as Dublin Core, MARC, and FRBR to extract the core elements of resource description. The extracted elements were categorized according to their semantics and functions, which resulted in three categories of core elements. These categories and core elements were manifested by using RDF syntax in order to be substantially applied to combine metadata records in digital preservation. Although the internal structure is not intended to create metadata records, it is expected to provide an alternative approach to enable combining existing metadata records in the context of digital preservation in a more flexible way.

Keywords: Digital preservation, Information package, Metadata record, Internal structure

Open Access

Accepted date: November 24, 2014
Received date: November 12, 2014

*Corresponding Author: Seungmin Lee
Assistant professor
Department of Library and Information Science
Sookmyung Women's University, Republic of Korea
E-mail: ableman@sookmyung.ac.kr

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>). Under this license, authors reserve the copyright for their content; however, they permit anyone to unrestrictedly use, distribute, and reproduce the content in any medium as far as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

1. INTRODUCTION

With the rapid increase of resources in digital formats, there is growing attention to the preservation of those resources. In this changed information environment, the importance of digital preservation is now becoming a hot issue in many communities, including the library and information science field. The preservation of traditional resources in printed formats has long been done in systematic ways. However, that of digital resources seems to be extremely different from the traditional archiving of resources.

In its initial stage, the major object of digital preservation was born-digital resources. With the advantages of digital preservation in preserving resources, however, any type of resources, including printed materials, is now the object of digital preservation. These resources in analog formats are transformed or converted into digital formats and preserved in digital archives.

The process of digital preservation consists of multiple phases that many experts from various fields need to participate in, including acquisition, archiving processes, storage, and dissemination of the resources preserved. In each phase, the target resource of digital preservation needs to be described in order to provide additional information. Therefore, the description of preserved resources is considered as the core of digital preservation. It can support an effective and efficient approach to resource management which is optimized for digital preservation.

Resource description is generally created in the format of metadata records using metadata standards. Digital preservation has also adopted various metadata standards, including rights, technical, administrative, and structural metadata, in order to provide detailed descriptions of preserved resources. In each phase of digital preservation, many metadata records are used to fully support the preservation process, which generates different types of information packages. However, these metadata standards adopt heterogeneous elements and the records generated based on these standards are considered as independent records. This has resulted in inefficiency in generating information packages and may not fully support each phase of the process of digital preservation.

In order to support efficient administration and management of digital preservation, heterogeneous

metadata records need to be combined in a more flexible way during the process of digital preservation. From this perspective, this research proposes an approach to combine metadata records generated by heterogeneous metadata standards and to support the generation of information packages in digital preservation. This approach considers a set of metadata records as an independent unit which can also be used as an information package that can support the process of digital preservation. In order to constitute information packages, an internal structure of the information packages needs to be constructed based on the elements used in a set of metadata records.

To achieve this goal, this research analyzes the core elements used in metadata records in the context of digital preservation. By extracting the core aspects of resource description in the process of digital preservation, the criteria for combining different metadata records can be established. Based on the core aspects, an internal structure for combining metadata records is constructed. This research also adopted Resource Description Framework (RDF) in order to manifest the internal structure of information packages, which can be substantially applied in the combination of metadata records and the generation of information packages. This internal structure can clearly show the components of information packages and establish the relationships between metadata records optimized for digital preservation.

2. CONCEPTUAL FOUNDATIONS

2.1. Concept of Resource Description and Digital Preservation

With the development of Information and Communication Technology (ICT), various types of information resources are rapidly increasing and have been published in digital format on the Web. In addition, ICTs can play an important role in support of the use of those resources. In this changed information environment, the importance of digital preservation gains more weight in the management and re-use of information resources.

Although there are a variety of formats of information resources, most resources in both printed and digital formats are generally preserved in digital format be-

cause digital formatting has advantages in preserving resources from the perspective of retention and re-use of those resources.

Digital preservation refers to the series of managed activities necessary to ensure continued access to digital resources for as long as necessary (Digital Preservation Coalition, 2012). It includes all of the actions required to maintain access to digital resources beyond the limits of media failure or technological change (Gomes, Miranda & Costa, 2011).

In order to process digital preservation, it is required to identify and retrieve all related components for the purpose of preserving a digital object. From this perspective, the process of digital preservation is inseparable from accessing the preserved object, which can be supported by resource description (Vardigan & Whiteman, 2007). For these reasons, the object needs to be re-created in a surrogate form, which is mainly generated in the format of metadata records.

In order for successful digital preservation, fundamental principles for resource description need to be established. More specifically, metadata standards that can support the description of the preserved resources may be the core of the process of digital preservation. However, there are many limitations in establishing appropriate disciplines that can be applied to digital preservation. From this perspective, many research efforts have been conducted to address these challenges and to develop digital preservation activities.

According to Heslop, Davis, and Wilson (2002), the object preserved in digital archives is often not an original physical object but a conceptual or virtual object that contains, replicates, or embodies the original object's significant properties or essential qualities. It refers to the characteristics that need to be preserved as descriptive records. For these reasons, many research studies have made standardized conceptual models for supporting the description of preserved objects that can suit the purpose of their communities (Lynch, 2003).

However, it is possible that a preserved resource can change its format during its conversion or transformation process in order to be included in digital preservation. Metadata is required in this process and describes every aspect related to the resource from its origins to the current status in the digital preservation process. At this point, one important thing is that every change or property needs to be described in a standardized and

formal way (Calhoun, 2006). The description of preserved resources is the core of digital preservation. In order to support resource description in digital preservation, many reference models and strategies have been proposed, including the Open Archival Information System (OAIS) Reference Model and DCC Curation Lifecycle Model.

In digital preservation communities, there have been many efforts to efficiently conduct digital preservation by standardizing the process of preserving resources in any format, including both born-digital and digitized resources. OAIS Reference Model is a representative model designed to support the overall process of digital preservation. The process provided by OAIS model consists of six steps. Each step requires different packages of additional information or metadata, and also needs different values for the same elements according to the purpose of each step, although they describe the same resource. From this perspective, the OAIS Reference Model defines long-term preservation in terms of the concept of an information package, for example, digital content and its associated metadata as a single package moving into (Submission Information Package: SIP), through (Archival Information Package: AIP), and out of (Dissemination Information Package: DIP) the archival system (Lane, 2012). Different metadata records may constitute different information packages. However, current approaches lack a well-defined architecture to support the use of preserved resources.

2.2. Challenges in Resource Description for Digital Preservation

Digital preservation is expected to provide a way of utilizing and reusing resources and the efficient long-term management of valuable resources. Because of arbitrary generation of information packages in the process of digital preservation, however, there are problems with resource description and the generation of information packages (IMS, 2003). In addition, because each set of metadata records is generated based on heterogeneous metadata standards and is independent from other records, they may not be efficiently reused when the process goes on. Therefore, each step may generate metadata records for the same resource redundantly. It may increase the complexity of the overall process of digital preservation. For these reasons, practical issues are considered in order to make appropriate use of re-

source description in the context of digital preservation. There is also a need to develop approaches and models for the re-use and connection of metadata records.

Resources for preservation go along with resource description. This description as additional information will have to be linked together in some way and methods of updating the resources and related information will have to be considered (NLII, 2003). From this perspective, a preserved resource is any resource that can be used to facilitate information activities and has been described using metadata. A resource cannot be regarded as a resource in digital preservation until it is made explicit through the addition of metadata. For these reasons, there is a growing awareness that preserving information resources will need to be expanded to include appropriate description through information packages by combining interrelated and relevant metadata records (Beagrie, 2002; Crow, 2002).

To fully support resource description in the context of digital preservation, metadata records that deal with every aspect of the preserved resource should be generated in each phase of digital preservation (Hedstrom, 1998). Because the same resource is described based on different metadata standards, the description can vary according to the unique purpose of each phase in the process of digital preservation. In the current status of digital preservation, however, the adoption of heterogeneous metadata standards has resulted in the inefficiency of generating information packages. This is mainly because the current reference models provide heterogeneous metadata records in each phase. For this reason, each metadata record is considered as an independent unit that may be isolated in each phase and may not be able to be fully reused in other phases

3. METHODOLOGY

To overcome the limitations and problems in resource description when preserving various resources, it is required to have a standardized way for resource description that can be applied to the overall process of digital preservation. The purpose of this research is to manifest a conceptual framework that represents each aspect of a variety of resources and supports the process of digital preservation by providing detailed resource description.

This research consists of three phases: (1) analysis of

the core aspects of metadata records in the context of digital preservation, (2) conceptualization of the internal structure of information packages optimized for digital preservation processes, and (3) the manifestation of internal structure to generate information packages.

This research begins with a review of the concept and role of resource description. This may be the foundation of digital preservation because it provides detailed description of preserved resources. In digital preservation, each resource description is not independent but interrelated with each other and provides a set of descriptions to support each phase of digital preservation. By reviewing the core aspects of resource description, the embedded meanings and the substantial functions of the description can be identified. Based on the reviews of resource description, the combination of metadata records can be achieved in a more flexible way and can support the overall process of digital preservation.

Based on the notion of resource description, the essential components of metadata records in digital preservation would be extracted. The existing metadata standards can be used to extract the core elements of metadata records for digital preservation. The identified elements are categorized according to their intended use and the relationships among the elements. These categories constitute the basic framework of the internal structure with multi-layered structure for generating information packages.

A structured framework for combining metadata records is constructed on the basis of the internal structure of information packages. This framework is not for acquiring metadata interoperability, but is a conceptual foundation that can combine existing metadata records in order to satisfy the purpose of each phase of the process of digital preservation. Each of the core elements that constitute the internal structure can be used as a reference point that provides a flexible combination of metadata records and generates various information packages applied in digital preservation.

In order to manifest the internal structure of information packages, this research adopts Resource Description Framework (RDF) so that the existing metadata records can be combined in a flexible way, which is an optimized way for each phase of digital preservation.

The approach proposed in this research might be helpful to clarify the internal structure of information packages, so that the management and description of

resources contained in digital archives can be supported.

4. CONSTRUCTION OF INTERNAL STRUCTURE OF INFORMATION PACKAGE

To provide more efficient ways of utilizing information packages and to support digital preservation, metadata records applied to each phase need to be combined according to the unique purpose of each phase. In order to achieve the combination of metadata records, the internal structure of information packages should be established.

4.1. Layers of Metadata Records

The combination of metadata records requires a set of elements that can uniquely identify each metadata record. The element set can be used as primary keys of connecting metadata records. By using the key elements, heterogeneous metadata records can retain their intrinsic elements that may fully satisfy the purpose of each phase. In addition, each metadata record can be applied in a more flexible way in generating information packages whenever it is required. In setting up a set of elements as primary keys, however, the core elements used in metadata records should be identified, which can be manifested as the internal structure of metadata records.

An element can be related to other elements if they share some meanings. This establishes element relationships. This relationship makes the meaning and functions of each element clear. Each element can have attributes that describe certain aspects of the element and prescribe the range of the meaning of the element. These attributes can be also considered as components of resource description. However, most of these attributes can be applied to only one element, although some attributes can be broadly applied to other elements. The attribute may not be important in constructing the internal structure of information packages because it does not establish relationships with other elements and the range of the meaning is relatively narrow. Therefore, this research does not consider the attribute as the component of information packages.

The elements traditionally used to describe an information resource are linearly enumerated in one-dimensional structure. Although this linear structure can

clearly show the meanings of the elements embedded in information packages, it cannot fully show the relationships among elements if one element has multiple relationships with other elements. In some cases, hierarchical structure can be more efficient to show the relationships among elements. However, not all elements do have a hierarchical relationship and the hierarchical structure may not provide the space for the enumeration of the meaning of each element.

In this research, a hybrid approach was used that combines linear and hierarchical structure. It can have multiple layers with enumeration of elements. Each element in each layer has hierarchical relationships with other elements in other layers. This multi-layered structure can enumerate the meanings of elements in detail and can also show the hierarchical relationships among elements.

4.2. Identification of Core Elements for Information Packages

In order to construct the internal structure of information packages, the components of an information package need to be transformed to a set of descriptive elements that can be substantially applied to the description of preserved resources. The elements of information packages can be transformed from each aspect of the information resources.

The elements of an information package may be similar to the traditional descriptive elements for information resources. However, they may be related to other elements in multiple layers. This is because an information package is a group of metadata records that share the same contents. In order to comprehensively combine different metadata records for the same resource, the elements used in metadata records should be tied in a structured format.

In order to extract the core elements for resource description in the context of digital preservation, this research used two existing metadata standards (Dublin Core and MARC) and one cataloging rule (FRBR). Dublin Core is one of the most representative metadata standards across communities and provides core elements in describing resources in any format. Although MARC was originally designed for printed materials in the library community, it is now applied to any format of resources both in printed and digital formats. These metadata standards are expected to identify the

elements that can be essentially used in the resource description for digital preservation. Functional Requirements for Bibliographic Records (FRBR) is not a metadata standard, but a kind of cataloging rule. However, it provides a conceptual structure in describing resources and in supporting the establishment of relationships between elements used in resource description. It can be also used to combine inter-related elements when describing resources. By analyzing these three standards, the essential elements for digital preservation can be extracted (see Appendix 1). The core elements that can be applied to generate information packages appear to be nine elements (see Table 1).

Although these elements may be used as a set of core elements in digital preservation, each element cannot properly identify the preserved resources. Therefore, a set of core elements need to be connected as an independent unit. By using the set of elements, it can properly function as core elements that can support the generation of information packages.

4.3. Multi-Layered Element Relationships for Information Packages

The internal structure of an information package with semi-hierarchical structure proposed in this research consists of three layers with hierarchies: content, carrier, and publication. The layer 'content' can be defined as a set of bibliographic elements that describes specific as-

pects of a resource related to the content of the resource, such as title, creator, and subject. The layer 'carrier' is a set of elements that describes the medium of resource which manifests and communicates its contents. Representative elements contained in the 'carrier' layer include medium, language, date, and identifier. The bibliographic elements that represent the information related to the publication of the content and media are categorized in the 'publication' layer. It includes the elements such as publisher and edition, which are related to the publication of the resource.

These layers were based on the categories of inter-related elements extracted from the three standards. Each layer contains related elements as sub-elements and establishes certain types of relationship with one or more elements. These relationships are established when one element has effect on the meaning of the related elements in constructing an internal structure of the information package.

This research adopted the social network theory when establishing element relationships. The social network theory seems to be useful to identify the relationship between elements in information packages because each element can be seen as a component of element relationships within bibliographic description. The relationship between elements in information packages can be divided into three main types according to strength of relations: strong, medium, and weak ties. The strength

Table 1. Core Elements of Information Packages

Core Element	Definition
Creator	A group of creators responsible for the content of individual information resources that share the same or variation of the original content
Title	A group of names of information resources that share the same or a variation of content
Subject	A set of one or more terms that represent the content contained in an information package
Publication	Information related to the manifestation of the same or a variation of content in individual information resources
Language	Specific languages used to represent the same or a variation of content contained in an information package
Edition	A certain version of individual information resources contained in an information package
Identifier	A set of characters assigned to individual information resources contained in an information package
Date	Information related to the date of individual resources contained in an information package
Medium	Formats of individual information resources contained in an information package

of relationship refers to the closeness between two or more components and the range of the function of each component. This strength of relationship can represent the detailed relationships with related elements and can play an important role in constructing the internal structure of information packages.

4.3.1. Content layer

Among the nine elements provided in Table 1, Creator and Title are connected with strong ties. The element Title also establishes a strong tie with the element Subject (see Fig. 1). These two strong ties establish a weak tie among the elements Creator and Subject. These three elements are the ones that describe the content of the information package. The content of the information package is not changed or affected by other parts of the information package, no matter what forms the information package takes. The three elements that describe the content do not describe or prescribe other elements.

This means that the meaning of these three elements is not changed or affected by other elements. They can be considered to be grouped together because they are all related to the content of the information package.

4.3.2. Carrier layer

The elements Language, Identifier, Date, and Medium specify each aspect of the carrier of information resources. They only describe and prescribe the elements in other layers. The meanings of these elements can be changed according to context and the element value can be also changeable. These elements can be grouped together because they are related to the carrier of the information package. However, there is no direct relationship among these elements. These elements can only be connected through other elements. This type of connection establishes weak ties among these elements (see Fig. 2).

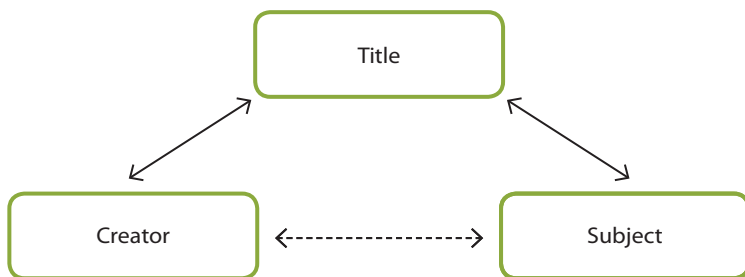


Fig. 1 Element relationships in Content layer

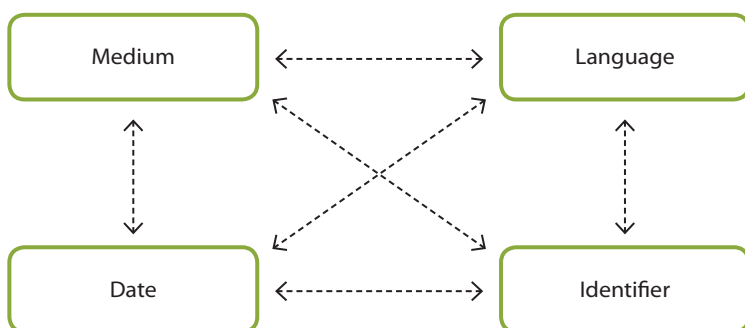


Fig. 2 Element relationship in Carrier layer

4.3.3. Publication layer

The elements Publisher and Edition share many characteristics and can be considered as similar elements in the context of digital preservation. This is because Edition is also a kind of publication of an information resource. In terms of bibliographic information, however, these two elements describe different aspects of information resources. Therefore, Publisher and Edition should be considered as different elements in the internal structure of information packages. These two elements are described by other elements, and simultaneously describe and prescribe the meaning of other elements. These elements represent the process of publishing information resources in order to transmit and communicate information. Therefore, these elements can be considered as a manifestation of the carrier of information. They are also related to the content of an information package. The elements Publisher and Edition establish a strong tie and can be grouped together because they are related to the publication process of information packages (see Fig. 3).

4.4 Construction of Internal Structure

The internal structure of information packages can be divided into two parts: layer-level and element-level. The relationships in layer-level construct the framework of the internal structure of an information package. The layer-level relationship concentrates on creating layers by combining elements that have the same or similar meanings and representing the relationships between the layers. The layer relationship adopts a hierarchical structure that can clearly represent hierarchical relationships among elements in different layers. This hierarchical structure is also useful to represent multiple relationships if an element establishes relationships with more than one element.

Linear structure is used to represent element relationships in each layer in element-level. Each layer enumerates elements that have the same or similar meaning and

functions. The linear structure is efficient to describe the detailed meanings and characteristics of each element in element-centered aspects which hierarchical structure cannot clearly represent. In addition, the element relationships that do not establish hierarchical relationship can be clarified by using linear structure.

By using the structure that combines hierarchical and linear structure, the meanings and functions of the elements in information packages can be clearly represented. The proposed internal structure of information packages consists of three layers: Content layer, Publication layer, and Carrier layer. Each layer adopts linear structure to enumerate the embedded elements. The enumerated elements are connected with other elements by establishing specific relationships (see Fig. 4).

The Content layer is placed on the top of the internal structure. The elements embedded in the Content layer are connected with each other by establishing strong ties with the element Title. Although the elements Creator and Subject are not directly connected, the weak ties are established through the element Title. The elements in the Content layer are described and prescribed by the elements in other layers.

The Publication layer is placed in the middle of the internal structure. It functions as a mediator that connects the Content layer and the Carrier layer. The two elements embedded in the Publication layer establish strong relationships and describe the element Title in the Content layer. The two elements in the Publication layer do not establish any direct relationship with the elements Creator and Subject in the Content layer, thus weak relationships are established with those elements through the element Title. Therefore, the Publication layer can be considered to describe the Content layer. The Publication layer is also described by the Carrier layer.

The Carrier layer is placed at the bottom of the internal structure. The elements in this layer describe all elements in the Publication layer. This element relationship

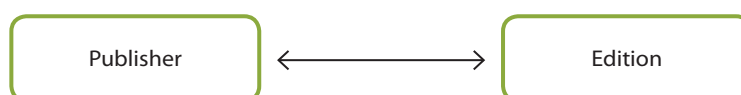


Fig. 3 Element relationship in Publication layer

is established by medium ties between the Publication and the Carrier layers. The elements in the Carrier layer also describe the elements in the Content layer. Therefore, the Carrier layer establishes medium ties with both the Publication layer and weak ties with the Content layers. However, not all of the elements in the Carrier layer describe all of the elements in the Content layer. Therefore, the relationship between the Content and Carrier layers is not a layer-level relationship. The relationship can be considered as an element-level relationship. Although the elements embedded in the Carrier layer are mainly related to represent the characteristics of the carrier of information, these elements are also used to describe the elements in the Publication and the Content layers. The elements in the Carrier layer play important roles in connecting all components of the information package through certain types of relationships.

Synthetically, the layer-level relationship constructs the framework of the internal structure of the information package. This layer relationship is established by the relationships between the elements embedded in each layer. Therefore, the element relationship among elements embedded in each layer provides the foundation of constructing the internal structure of the information package. The elements function as internal nodes that

connect each layer. They can also function as external nodes to connect other information packages that contain the same element values.

5. MANIFESTATION OF INTERNAL STRUCTURE

The proposed internal structure of information packages identifies the core elements that constitute an information package and the element relationships that connect the components of the information package. The internal structure can clearly identify the meanings and functions of the elements of information packages.

Once the categories of core elements are established, they can be used to combine a set of metadata records in digital preservation. However, these categories and elements are not designed for substantially describing resources. They are intended to be used to combine metadata records that describe preserved resources in a standardized way. In order to achieve the combination of records, the categories and core elements need to be manifested in a formal format of metadata. This research applied Resource Description Framework (RDF) based on eXtensible Markup Language (XML) and con-

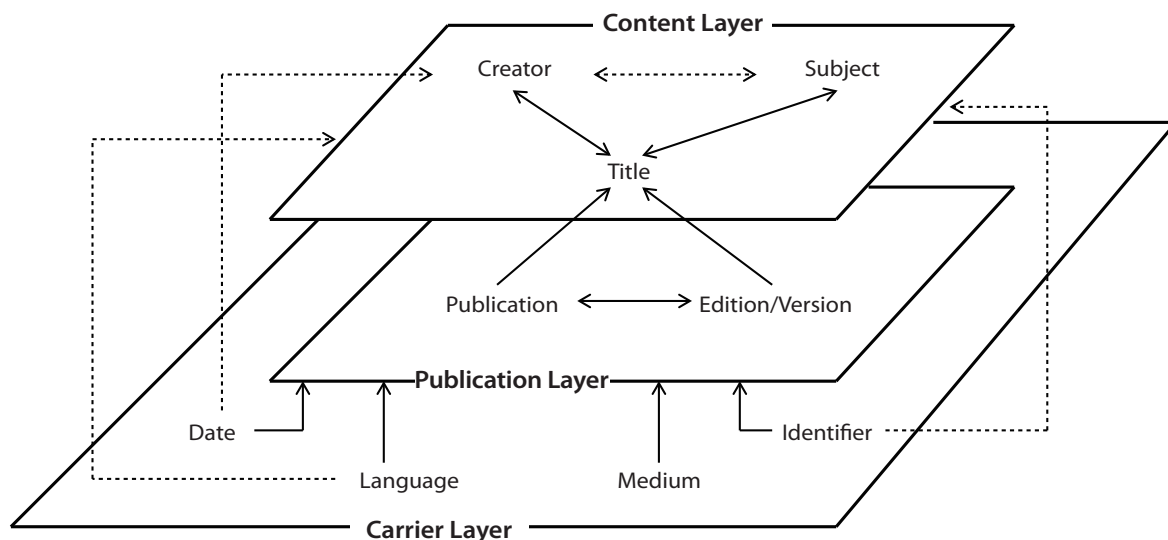


Fig. 4 Proposed internal structure of information package

verted the internal structure into a metadata framework.

In the RDF syntax, each category of core elements functions as an RDF Class. Because the core elements that are expected to be used to combine metadata records are categorized and set up as element categories, each Class needs to be represented as a category, not an element. Class Category defines the categories as Classes in the proposed RDF schema. Class Category functions to indicate each layer in the internal structure of information packages and contains Content, Carrier, and Publication as sub-Class (see Fig. 5).

However, these Classes are nothing but the representation of categories in RDF syntax if they are used to categorize the core elements. Therefore, the core el-

ements need to be manifested as a set of attributes that specify each aspect of the information package. In order to manifest the attributes in a formal format, the relationships between core elements should be syntactically represented. This is manifested as external properties (see Fig. 6).

Each core element functions as a property of the categories converted into RDF Classes. By using these Classes and Properties, the internal structure of information packages can be manifested as a metadata framework in RDF. It can be used to combine existing metadata records in digital preservation in a flexible way. The RDF syntax for these categories and core elements are manifested in Fig. 7.

```
<rdfs:Class rdf:about="#Category">
  <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
</rdfs:Class>
<rdfs:Class rdf:about="#Content">
  <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
  <rdfs:subClassOf rdf:about="#Category"/>
</rdfs:Class>
<rdfs:Class rdf:about="#Carrier">
  <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
  <rdfs:subClassOf rdf:about="#Category"/>
</rdfs:Class>
<rdfs:Class rdf:about="#Publication">
  <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
  <rdfs:subClassOf rdf:about="#Category"/>
</rdfs:Class>
```

Fig. 5 Class definition of information package

```
<rdf:Property rdf:ID="Relationship">
  <rdfs:range rdf:resource="#Resource"/>
  <rdfs:domain rdf:resource="#Resource"/>
</rdf:Property>
```

Fig. 6 Property definition of information package

As shown in Fig. 7, the nine core elements function as properties of the Class Category. Each property is expected to function as a mediator that combines the external elements used in the overall process of digital preservation. Based on the core elements as properties, the elements in information packages can be inter-related with each other and can provide more detailed resource description according to the phases of the overall process of digital preservation.

The elements can also integrate all the aspects of preserved resources that most of the types of the current

resources contain. Therefore, the proposed structure can support the description of preserved resources optimized for each phase of the process of digital preservation. It can also clearly represent the details of preserved resources and establish relationships across metadata records created during the preservation process.

Although the internal structure of information packages are manifested using RDF syntax, this structure can be used to connect external metadata records created in other communities. Because the core elements are extracted from the existing metadata standards that

```

<rdf:Property rdf:ID="creator">
  <rdfs:range rdf:resource="#Content"/>
</rdf:Property>
<rdf:Property rdf:ID="title">
  <rdfs:range rdf:resource="#Content"/>
</rdf:Property>
<rdf:Property rdf:ID="subject">
  <rdfs:range rdf:resource="#Content"/>
</rdf:Property>
<rdf:Property rdf:ID="publication">
  <rdfs:range rdf:resource="#Publication"/>
</rdf:Property>
<rdf:Property rdf:ID="language">
  <rdfs:range rdf:resource="#Carrier"/>
</rdf:Property>
<rdf:Property rdf:ID="edition">
  <rdfs:range rdf:resource="#Publication"/>
</rdf:Property>
<rdf:Property rdf:ID="identifier">
  <rdfs:range rdf:resource="#Carrier"/>
</rdf:Property>
<rdf:Property rdf:ID="date">
  <rdfs:range rdf:resource="#Carrier"/>
</rdf:Property>
<rdf:Property rdf:ID="medium">
  <rdfs:range rdf:resource="#Carrier"/>
</rdf:Property>

```

Fig. 7 RDF definitions of information package

are commonly adopted in a variety of communities, the metadata records used in those communities can be applied to the generation of information packages via the core elements. From this perspective, the core elements and element categories can be considered as reference points to combine metadata records generated in communities other than digital preservation.

The Classes and Properties in the internal structure of information packages can also have a specific set of metadata elements from other metadata standards as a subset of each element. By adopting those external elements, the internal structure can provide more detailed descriptions for preserved resources. Therefore, the internal structure can be considered as an umbrella structure for both resource description in digital preservation and a conceptual structure for integrating metadata standards. With these strengths, the internal structure proposed in this research is expected to be used as a conceptual foundation for generating information packages that can be flexibly applied to satisfy specific purposes of resource description across communities.

6. CONCLUSION

With the rapid increase of information resources in digital format, the importance of digital preservation gains more weight across communities. Although many research efforts have been conducted to support the process of digital preservation and have provided standardized models and strategies, they have limitations because of inefficient and inconsistent description of preserved resources. In addition, the use of information packages in each phase of digital preservation may not fully support the unique purpose of digital preservation and/or resource description. To overcome these limitations and to address the current problems in digital preservation, this research proposed an internal structure of information packages generated in the process of digital preservation.

The internal structure is based on the core elements that are commonly used in creating metadata records across communities. The set of core elements describes the common characteristics of a variety of information resources and functions as a reference point in generating information packages in the context of digital preservation. The core elements were extracted from

the existing metadata standards such as Dublin Core and MARC. In addition to these standards, a kind of cataloging rule, which is FRBR, is also analyzed to provide conceptual relationships across the extracted core elements. These extracted elements were categorized according to their semantics and functions in describing information resources. The categories of the core elements consist of three groups of elements, which are content-related, publication-related, and carrier-related elements. Each category constitutes a layer that can provide a space for enumeration of the core elements and can provide detailed description of preserved resources. The three layers also constitute a hierarchical structure in resource description for generating information packages.

Based on the constructed internal structure of information packages, this research manifested each category and core element in RDF syntax. This can provide a standardized format of converting the categories and core elements in a metadata format. Through the RDF syntax, the internal structure of information packages can be applied to substantially generate information packages by combining metadata records created during the process of digital preservation.

In digital preservation, information packages play an important role because they can satisfy the unique purpose of each phase of digital preservation and provide detailed description of preserved resources. The internal structure of information packages proposed in this research is designed to support the combination of metadata records for the same resource created by heterogeneous metadata standards. Although this approach is not intended to substantially describe actual information resources, it is expected to provide an alternative approach to combine existing metadata records in the context of digital preservation, and functions as a conceptual foundation of generating information packages in a more flexible way.

REFERENCES

- Beagrie, N. (2002). *A continuing access and digital preservation strategy for the Joint Information Systems Committee (JISC), 2002-2005*. Retrieved September 18, 2014, from http://www.jisc.ac.uk/uploaded_documents/dpstrategy2002b.rtf

- Calhoun, K. (2006). *The changing nature of the catalog and its integration with other discovery tools*. Washington, D.C.: Library of Congress. Retrieved September 29, 2014, from <http://www.loc.gov/catdir/calhoun-report-final.pdf>
- Crow, R. (2004). *The case for institutional repositories: A SPARC position paper*. The Scholarly Publishing and Academic Resources Coalition, 2002. Retrieved October 15, 2014, from <http://www.arl.org/sparc/IR/ir.html>
- Digital Preservation Coalition (2012). *Digital preservation handbook*. Retrieved October 27, 2014, from <http://www.dpconline.org/advice/preservationhandbook/topic/168-introduction-definitions-and-concepts>
- Gomes, D., Miranda, J., & Costa, M. (2011). A survey on Web archiving initiatives. In *Proceedings of the 15th International Conference on Theory and Practice of Digital Libraries (TPDL '11): Research and Advanced Technology for Digital Libraries*, 408-420. Berlin: Springer-Verlag.
- Hedstrom, M. (1998). Digital preservation: A time bomb for digital libraries. *Computers and the Humanities*, 31, 189-202.
- Heslop, H., Davis, S., & Wilson, A. (2002). An approach to the preservation of digital records. *National Archives of Australia*. Retrieved October 22, 2004, from http://www.naa.gov.au/recordkeeping/er/digital_preservation/Green_Paper.pdf
- IMS (2003). *IMS Digital Repositories Specification*. Retrieved October 15, 2014, from <http://www.ims-global.org/digitalrepositories/index.cfm>
- Lane, H. (2012). Models for the EU Arctic Information Centre: Engaging the polar libraries colloquy. In *Proceedings of the 24th polar libraries colloquy: Cold regions: Pivot points, focal points*, 169-170. Boulder, CO, USA, June 11-14.
- Lynch, C. A. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. *ARL Bimonthly Report*, 226. Retrieved September 15, 2004, from <http://www.arl.org/newsltr/226/ir.html>
- NLII (National Learning Infrastructure Initiative) (2003). *Learning objects (NLII 2002-2003 Key Themes)*. Retrieved September 20, 2014, from http://www.educause.edu/nlii/keythemes/LearningObjects.asp#nlii_research
- Vardigan, M., & Whiteman, Cole. ICPSR meets OAIS: Applying the OAIS reference model to the social science archive context. *Archival Science*, 7, 73-87.

Appendix 1

Brief Comparison of Elements in Existing Bibliographic Standards: FRBR, Dublin Core, and MARC

	FRBR	Dublin Core	MARC
Creator	Name of person	Creator	1XX Author
	Dates of person	Contributor	7XX Contributor
	Title of person	-	-
	Name of corporate body	-	-
	Place of corporate body	-	-
	Date associated with the corporate body	-	-
Subject	Summarization	Subject	6XX Subject
	Term for the concept	Description	5XX Notes
Edition	Edition/issue	-	245 Title
Publication	Publication place/date	-	260 Publication
	Publisher/distributor	Publisher	-
Identifier	Identifier	Identifier	856 Identifier
Language	Language	Language	041 Language
Title	Title	Title	440 Series statement
	-	-	245 Title
Medium	Form	Format	246 Form of title
	Physical medium	Type	-
Date	Date	Date	-
	-	Relation	-
	-	Coverage	-
	-	Source	-
	-	Rights	-
	Intended termination	-	-
	Extensibility	-	-
	Extent	-	-
	Revisability	-	-
	Restriction	-	-
	Statement	-	-
	Responsibility	-	-
	Manufacturer	-	-
	Access authorization/restriction	-	-
	Availability	-	-
	Intended audience	-	-
	Context	-	-
	Term for the object	-	-
	Term for the event	-	-
Term for the place	-	-	

* This table does not show the entire MARC field set. It only provides the MARC elements that are matched with Dublin Core and FRBR elements.