Library Hi Tech

A RDF-based approach to metadata crosswalk for semantic interoperability at the data element level

Ya-Ning Chen

A RDF-based approach to metadata crosswalk for semantic interoperability at the data element level

Ya-Ning Chen
Department of Information and Library Science, Tamkang University, New Taipei City, Taiwan

Received 7 August 2014
Revised 20 March 2015
Accepted 5 April 2015

Abstract

Purpose – The purpose of this paper is to propose a Resource Description Framework (RDF)-based approach to transform metadata crosswalking from equivalent lexical element mapping into semantic mapping with various contextual relationships. RDF is used as a crosswalk model to represent the contextual relationships implicitly embedded between described objects and their elements, including semantic, hierarchical, granular, syntactic and multiple object relationships to achieve semantic metadata interoperability at the data element level.

Design/methodology/approach – This paper uses RDF to translate metadata elements and their relationships into semantic expressions, and also as a data model to define the syntax for element mapping. The feasibility of the proposed approach for semantic metadata crosswalking is examined based on two use cases – the Archives of Navy Ships Project and the Digital Artifacts Project of National Palace Museum in Taipei – both from the Taiwan e-Learning and Digital Archives Program.

Findings – As the model developed is based on RDF-based expressions, unsolved issues related to crosswalking, such as sets of shared terms, and contextual relationships embedded between described objects and their metadata elements could be manifested into a semantic representation. Corresponding element mapping and mapping rules can be specified without ambiguity to achieve semantic metadata interoperability.

Research limitations/implications – Five steps were developed to clarify the details of the RDF-based crosswalk. The RDF-based expressions can also serve as a basis from which to develop linked data and Semantic Web applications. More use cases including biodiversity artifacts of natural history museums and literary works of libraries, and conditions, constraints and cardinality of metadata data elements will be required to make revisions to fine tune the proposed RDF-based metadata crosswalk.

Originality/value – In addition to reviving contextual relationships embedded between described objects and their metadata elements, nine types of mapping rules were developed to achieve a semantic metadata crosswalk which will facilitate the design of related mapping software. Furthermore, the proposed approach complements existing crosswalking documents provided by authoritative organizations, and enriches mapping language developed by the CIDOC community.

Keywords Digital libraries, Digital archives, RDF, Metadata, Crosswalk, Semantic interoperability

Paper type Research paper

Introduction

In digital libraries, metadata plays an essential role in resource description and discovery. Many existing generic and domain-specific metadata standards have been developed by various user communities and adopted as schema to build up digital library systems or

The author is grateful for the helpful comments of the anonymous reviewers. Furthermore, the author thanks the National Palace Museum (hereafter NPM) in Taipei for offering metadata instance of Table IV selected from Digital Archiving System of NPM Antiquities and Dr Wei-Lee Lu of ROC Naval Academy for provision of metadata instance listed at Table II.

0737-8831
DOI 10.1108/LHT-08-2014-0078

repositories. Each metadata standard has been developed for a different purpose to serve the needs of a particular user community or domain (Chan and Zeng, 2006). With the use of varying metadata standards, integration or sharing of heterogeneous metadata between digital library systems or repositories, i.e., the interoperability between standards becomes an issue. Metadata interoperability can be divided into schema, record and repository levels (Chan and Zeng, 2006). According to definition proposed by Chan and Zeng (2006), "crosswalks are by far the most commonly used method to enable interoperability between metadata schemas at the schema level." Most current crosswalks are a semantic mapping of elements without a metadata conversion specification (St Pierre and LaPlant, 1998). This reveals that most metadata crosswalks are still focussed on mapping equivalent lexical definitions of metadata elements between source and target metadata standards. However, in addition to lexical semantics, a complete semantic crosswalk has to take into account other issues such as common terminologies, granularity of elements, syntax, semantic and hierarchical relationships embedded between elements, multiple objects and their relationships, and corresponding mapping rules (St Pierre and LaPlant, 1998).

Literature review

In computer science automatic schema and ontology matching approaches have been developed for data integration, e-business, data warehousing and semantic query processing. They can achieve approximate matching between source and target metadata elements (Rahm, 2011; Rahm and Bernstein, 2001) or lexical matching without contextual relationships (Kitamura *et al.*, 2008). The reason lies in each metadata standard is rooted in a distinct domain with heterogeneous semantics (Nogueras-Iso *et al.*, 2005), such as Categories for the Description of Works of Art (hereafter CDWA) from museums and Encoded Archival Description (hereafter EAD) from archives. On the other hand, in recent years numerous projects have been initiated to achieve crosswalks between different metadata standards and their applications, including selected two or more metadata standards to build up a mapping mechanism to share metadata elements between similar or different communities (Cao *et al.*, 2004; Chandrakar, 2002, 2005; Godby *et al.*, 2008; Lagunas and Basurto, 2005; Lightle and Ridgway, 2003; Shepherd and West, 2003). The issues and practices explored by these studies are useful guidelines in practical mapping for similar cases and projects, but current proposed approaches have focussed on mapping equivalent semantic definitions between elements through by representation with a table or chart. In addition to table or chart, Morform and XSL has been used to specify the syntax (Godby *et al.*, 2008), as well as a set of step-by-step procedures (Lim *et al.*, 2012) and a mapping language (Kondylakis *et al.*, 2006) has been proposed for metadata crosswalk. However, to this day no attempts have been made to develop a generic approach to semantic metadata crosswalks at the data element level.

Semantic metadata crosswalks at the data element level are more than equivalent mapping based on lexical definitions of metadata elements. Researchers (Chan and Zeng, 2006; Dunsire *et al.*, 2011; Park, 2002) have regarded this kind of crosswalk as lexical mapping that is based on lexical form, appearance or meanings of metadata elements. Lexical crosswalks still require a set of common or shared terms to reconcile their semantic heterogeneity (Cao *et al.*, 2004; St Pierre and LaPlant, 1998). Most current lexical crosswalking practices lack explicit contextual information and relationships for mapping (Dunsire *et al.*, 2011; St Pierre and LaPlant, 1998; Woodley, 2008). Thus, semantic (St Pierre and LaPlant, 1998), hierarchical (Cao *et al.*, 2004; Lim *et al.*, 2012; St Pierre and LaPlant, 1998; Woodley, 2008), granular (i.e. many-to-one and one-to-many)

(Cao *et al.*, 2004; Chandrakar, 2005; Lightle and Ridgway, 2003; Machovec, 2002; Shepherd and West, 2003; St Pierre and LaPlant, 1998; Woodley, 2008), syntactic relationships (Chan and Zeng, 2006; Woodley, 2008), and multiple objects and their relationships (St Pierre and LaPlant, 1998; Woodley, 2008) that are implicitly embedded between elements must be taken into account in a semantic metadata crosswalk. To achieve semantic metadata interoperability at the data element level, in addition to equivalent lexical mapping based on definitions of metadata elements, the above issues need to be considered as an integral part of crosswalks. In summary, a semantic metadata crosswalk not only includes lexical mapping, but also contextual information and relationships including hierarchical, granular, syntactic and multiple objects are required for matching metadata elements. However, to the best of our knowledge no attempts have developed a generic approach to semantic metadata crosswalks at the data element level.

Methodology

This study adopted Resource Description Framework (RDF) as the data model and examined the feasibility of semantic metadata mapping for interoperability at the data element level. Two projects from the Taiwan e-Learning and Digital Archives Program (hereafter TELDAP) were selected as use cases to illustrate the RDF-based crosswalk. The projects selected were the Archives of Navy Ships Project (hereafter ANSP) and the Digital Artifacts Project of the National Palace Museum in Taipei (hereafter DAPNPM). Three metadata element standards, EAD, CDWA and Dublin Core (hereafter DC), were employed as examples to delineate the RDF-based crosswalk at the data element level. The reasons to select a manual approach rather than an automatic one for this study are as follows: first, a semantic crosswalk between source and target metadata elements requires a deep knowledge of each domain-specific metadata standard. Furthermore, it is difficult to do a mapping from a hierarchical metadata standard (e.g. CDWA and EAD) to a highly abstract metadata standard in a flat structured way such as DC without human intervention. Third, although the proposed approach by this study is manual, however, a semantic crosswalk with RDF-based expression also paves the way for automatic metadata mapping at the data element level.

The remainder of this paper is composed of the following sections: first, a section on a RDF-based approach to semantic metadata crosswalks describes a process of five steps through which RDF can be implemented in a crosswalk model for metadata element mapping. Then, in the section following, the feasibility of the proposed approach is examined in detail by exploring two use cases. Finally, the Discussion presents the benefits of the RDF-based crosswalk in comparison with existing semantic mappings, and the Conclusion clarifies the contributions of the approach proposed.

A RDF-based approach to semantic metadata crosswalks

This section explains how RDF can be used as a crosswalk model to express mapping of equivalent lexical metadata elements with various contextual relationships, including semantic, hierarchical, granular, multiple object and syntactic, for semantic metadata crosswalking at the data element level. The RDF-based crosswalk approach proposed consists of five steps as listed below.

Description of the approach

Step 1: identification of the multiple objects and their relationships embedded in source metadata elements. Owing to widespread adoption of the DC's one-to-one principle, it is a typical metadata norm and practice that a record selects a single object as a subject for description or mapping. In reality, not all metadata standards are restricted to a single object without relationships to others (St Pierre and LaPlant, 1998). Metadata can also be used to indicate how compound objects are put together (Gilliland, 2008; NISO, 2004). For example, the EAD is used to describe and arrange archival materials into compound objects at various levels (i.e. fonds, series, files and items), whereas the CDWA is to describe museum artifacts and their granular relationships (i.e. series, sets and collections, to components and items). Multiple object relationships are therefore often lost and their corresponding elements are mixed together during mapping from domain-specific metadata standards into DC. In terms of RDF expression, multiple objects can be regarded as subject and object, and the relationship between them as predicate (Figure 1(a)). Therefore, the first step in the semantic metadata crosswalk is to adopt RDF to identify multiple object relationships (such as is-Part-of, has-Part, and is-Related-to) implicitly embedded in the metadata standard.

Step 2: selection of the adopted objects and their metadata elements from the source standard. Although metadata elements can be further categorized into different groups such as mandatory, recommended and optional, in practice not all the elements of metadata standards need to be selected for mapping. Those mapped should depend on the types of digitized objects and the requirements of digital library projects (St Pierre and LaPlant, 1998). Similarly, not all multiple objects and their metadata elements need to be included in mapping. There are two possible general mapping rules for multiple objects. One is to include all multiple objects and their metadata elements into mapping and select an element of both the source and the target metadata standard to build up a bi-directional relationship between objects and their records. The other is that one of

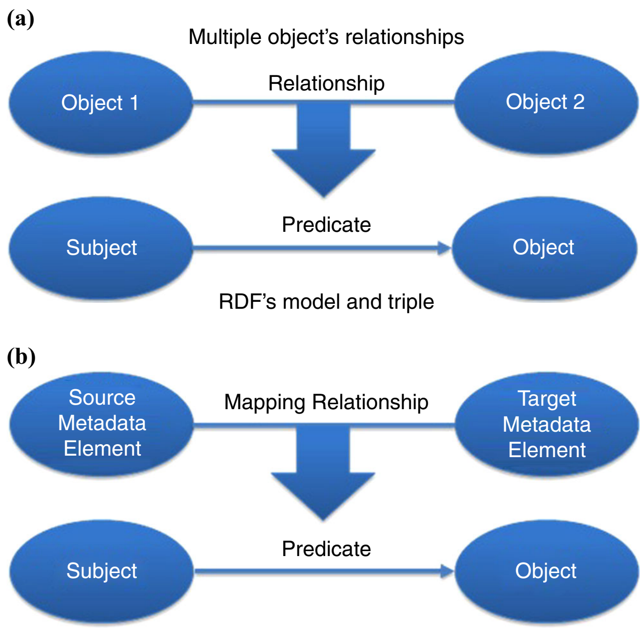

Figure 1.
RDF-based crosswalk approach for semantic metadata mapping

the multiple objects is included and the other objects are excluded from the mapping. In addition, the relationship between the mapping object and non-mapping objects needs to be recorded in the target metadata standard. Thus the second step we propose is to identify adopted metadata elements and their instances from the source standard. This means that metadata professionals have to assure the objects described and their used metadata elements, instead of including all elements defined in the metadata standard for mapping.

Step 3: identification of the semantic and hierarchical relationships between source and target elements. In addition to semantic mapping based on lexical definitions, semantic and hierarchical relationships implicitly embedded in metadata elements between source and target standards also need to be taken into account in semantic metadata crosswalks. Semantic relationships between source and target metadata elements can be of various types, such as same, synonymous, abbreviation, acronym, language, variation and so forth (Lim *et al.*, 2012). Each metadata standard has developed a different structure to serve the different needs of different user communities or domains (Cao *et al.*, 2004) because of the rooting of their elements in distinctive documentation requirements. Some have a hierarchical structure and some are flat between elements (St Pierre and LaPlant, 1998; Woodley, 2008). In terms of RDF expression, the source element is regarded as subject, the target element as object, and mapping relationship between the two as predicate (Figure 1 (b)). These RDF triples can be used to illustrate various semantic and hierarchical relationships, such as is-Synonymous-with[1] (Figure 2(a)), as well as is-Combined-as and is-Divided-into (Figure 2(b)). Therefore, the third step we propose is to adopt RDF to identify the embedded semantic and hierarchical relationships between source and target elements and specify related mapping rules.

Step 4: identification of the granular and syntactic relationships between source and target elements. In addition to semantic and hierarchical relationships existing between source and target metadata elements, the issue of granular relationships embedded between source and target metadata elements also impacts the precision of a semantic crosswalk. One-to-many and many-to-one are two general types of granular relationship (St Pierre and LaPlant, 1998). In terms of many-to-one, some have suggested appended together (Cao *et al.*, 2004), and some have proposed to use a qualifier approach (Dunsire *et al.*, 2011) to specify semantic mapping rules. On the other hand, there is a need to encode the syntax of metadata elements as a mapping wrapper (Woodley, 2008). If all granular and syntactic relationships can be clearly illustrated in a structured way, such as with RDF-based representations like is-Generalized-by and is-Specialized-by (Figure 2(c)) as well as is-Composed-by and is-Decomposed-into (Figure 2(d)), then metadata professionals can specify rules for mapping elements from source to target without ambiguity. Therefore, the fourth step we propose is to employ the RDF triples as a basis to specify granular and syntactic relationships between metadata elements and their mapping rules from source standard to the target.

Step 5: mapping of all equivalent lexical elements from source metadata into the target with reference to RDF-based relationships. Existing official metadata crosswalks released by authoritative organizations are a useful starting point, especially for direct and no mappings. Such as DC to EAD (www.loc.gov/ead/ag/agappb.html) maintained by the Network Development and MARC Standards Office of the Library of Congress (hereafter LC) and the Society of American Archivists, and Metadata Standards Crosswalk (www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html) released by the Getty Research Institute (hereafter

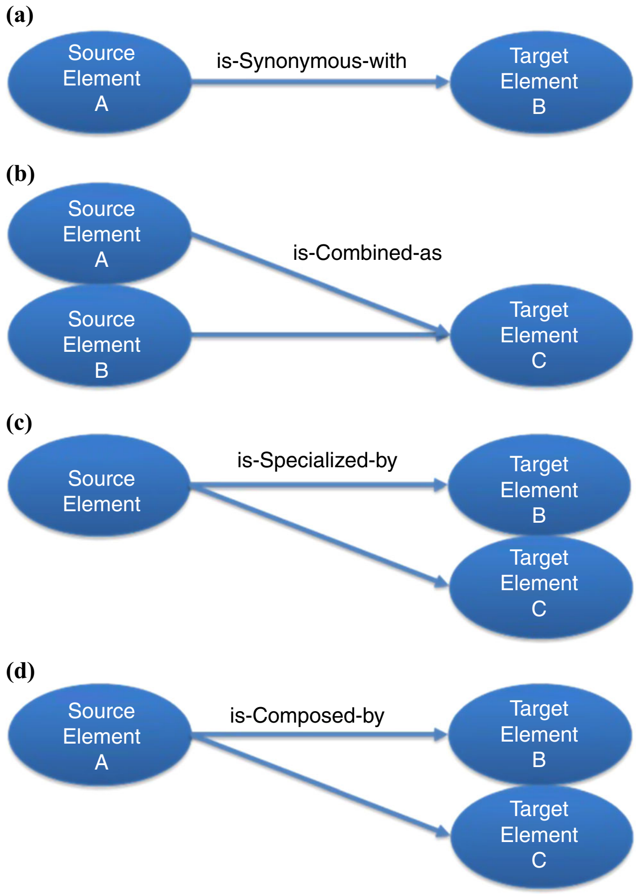

Figure 2.
Types of RDF-based semantic metadata mapping

GRI). Therefore, as the fifth step we propose to combine the RDF-based contextual associations with existing authoritative documents to achieve semantic metadata crosswalk at the data element level.

Use cases

Two individual projects were selected from TELDAP to examine the feasibility of the approach proposed above. The first project is part of a digital archival project where EAD was used as a metadata standard to document the archival description. The other project is part of a digital museum project where CDWA was adopted for description of museum artifacts.

Use case 1: digital archives project – ANSP

Description of the application. The ANSP is a digital archival project to digitize various materials such as photos, blueprints, logs, captain's diaries and chronological events of navy ships, with metadata descriptions. EAD was adopted by the ANSP as the metadata standard to build up a digital repository, and as a crosswalk mechanism to export the metadata records into the DC-based union catalog of TELDAP (hereafter

TELDAP UC). The result is that several identical EAD tags, such as < unittitle >, are repeatedly used at various archival levels from fonds to series, files and items; and provenance relationships also diminished when EAD tags were exported into the TELDAP UC. Issues such as syntactic, hierarchical, identical element and granularity relationships could not be illustrated clearly by the above official crosswalking document and therefore remain unsolved.

Application of the RDF as a crosswalk model

Step 1: identification of the multiple objects and their relationships embedded in source metadata elements. According to the ISAD(G) definition, fond, series, file and item have a whole-part relation (International Council on Archives, 2000). By adoption of RDF, archival relationships can be expressed by a set of RDF triples. For instance, fonds and series can be regarded as subject and object, respectively, and the predicate has-Part can connect and illustrate their relationship. Therefore, the top-down archival relationship can be represented in RDF structured expressions. Conversely the bottom-up relationship can be expressed with the is-Part-of predicate to connect from items, to files, series and fonds.

Step 2: selection of the adopted object and metadata elements from the source standard. Owing to loss of most of the provenance information from the archives, archival items rather than archival fonds as are commonly used in archival practices were selected for detailed description for the ANSP. Logs of a navy ship are used as an example to illustrate the deployment of EAD tags used by the ANSP (Table I). As can be seen, several identical EAD tags, such as < unititle >, < unitid > and < persname > were repeatedly used by the ANSP at various archival levels. This shows that the same EAD tag can be utilized to describe characteristics of archives at various levels. In practice, an arrangement (i.e. exclusion) was specified for mapping elements of fonds, series and files into the archival item-level setting. Exclusion means that elements of fonds, series and files are not included in mapping from EAD to DC, because the target objects for crosswalk are archival items. Therefore elements of fonds, series and files that were either identical to or distinct from items were excluded from crosswalk for the ANSP. Furthermore, the value of item was added into DC's Type element to indicate that the target objects belong to the archival item level. On the other hand, a semantic provenance relationship exists between fonds, series, files and items and the arrangement is transformation. This means that some EAD tags that belong to fonds, series and files are not suitable to be mapped directly into DC elements. Instead, their semantics must be transformed to fit into appropriate DC elements at the archival item setting. Therefore, < unittitle > of fonds, series and files are concatenated together into DC's Relation element with attribute of is-Part-of, in order to illustrate the object's whole-part relationship from fonds, series and files to items.

Step 3: identification of the semantic and hierarchical relationships between source and target elements. Hierarchical relationships exist between EAD tags used by the ANSP case as follows:

- < origination > → < persname >
- < langmaterial > → < language >
- < physdesc > → < dimensions >, < physdesc > → < extent >, < physdesc > → < genreform >, and < physdesc > → < physfacet >
- < controllaccess > → < geoname >
- < daogrp > → < daodesc >, and < daogrp > → < daoloc >

EAD elements used by the ANSP	Fonds level	Series level	Files level	Items level
< accessrestrict >			X	△
< acquinfo >				△
< archdesc >	X			
< bibliography > → < bibref >				△
< bioghist >	X	X	X	△
< bioghist > → < chronlist >		X	X	
< bioghist > → < chronitem >		X	X	
< bioghist > → < event >		X	X	
< container >			X	
< controlaccess > → < geoname >				△
< daogrp > → < daodesc >			X	△
< daogrp > → < daoloc >			X	△
< did >	X	X	X	△
< langmaterial > → < language >			X	△
< note >		X	X	△
< origination > → < persname >			X	△
< physdesc > → < dimensions >				△
< physdesc > → < genreform >				△
< physdesc > → < extent >			X	△
< physdesc > → < physfacet >				△
< physloc >		X	X	
< processinfo >			X	△
< relatedmaterial >		X		
< repository >		X	X	
< repository > → < corpname >		X		△
< scopecontent >	X	X	X	
< unitdate >			X	△
< unitid >	X	X	X	△
< unittitle >	X	X	X	△
< userestrict >			X	△

Notes: X and △ stand for elements used by ANSP, and only those marked △ were selected for metadata conversion into TELDAP UC

Table I.
EAD elements at various archival hierarchies used by the ANSP

During mapping from EAD to DC, only one set (i.e. < daogrp > → < daodesc > and < daogrp > → < daoloc >) was excluded from the crosswalk. The other four sets were included in the crosswalk; however, this study specified two different mapping rules: map directly, and separate with display label. The former maps sub-hierarchical EAD tags to equivalent DC elements directly according to the contextual semantic meanings, such as map < geoname > to Coverage-Spatial, and < language > to Language in DC. The latter maps sub-hierarchical EAD tags into different DC elements with a display label. These sub-hierarchical EAD tags are also hierarchical EAD tags, but they are given with different contextual semantic meanings in DC. Thus < genreform > was mapped to the Type element in DC with the < genreform > display label. The remaining sub-hierarchical EAD tags under the same < physdesc > tag hierarchy were combined into the Format element of DC with display labels, such as < dimension >, < extent > and < physicalfacet >.

Step 4: identification of the granular and syntactic relationships between source and target elements. In terms of granular relationship, < accessrestrict > and < userestrict > EAD tags were appended together into the Rights of DC with display labels in order to indicate many-to-one mapping. During this step, the RDF is also useful to express the syntax for mapping. For instance, the <persname> EAD tag can be composed of first name and last name in a syntactic way. The syntactic crosswalk can be processed to map encoded <persname> into Creator of DC with structured syntax. Thus, the approach proposed in this study can map elements from many fine-grained EAD elements to one coarse-grained DC element with specified display labels and syntax.

Step 5: mapping of all equivalent lexical elements from source metadata into the target with reference to RDF-based relationships. In this study RDF was used as a crosswalk model to visualize and illustrate EAD tags and their various embedded relationships in the ANSP's log archives. It also transforms the lexical metadata mapping into semantic metadata mapping that includes various contextual relationships. A crosswalk between EAD and DC for the ANSP use case is shown in Table II.

Use case 2: digital museum project – digital artifacts project of National Palace Museum
Description of the application. Another example (the DAPNPM) from the National Palace Museum (hereafter NPM) in Taipei was chosen from TELDAP to illustrate the feasibility of RDF as a model for semantic metadata crosswalking. The NPM owns a huge number of Chinese historical artifacts, most of which originated from the imperial court of the Ch'ing Dynasty. Since 2002, projects have been underway to build up a digital repository by digitizing these Chinese cultural artifacts with metadata. The NPM adopted CDWA as the metadata standard for resource description and discovery, and as a crosswalk to export metadata into TELDAP UC by converting CDWA into DC. Initially, the NPM employed an official crosswalk from the GRI as a basis to map metadata between CDWA and DC. However, the complicated relationships embedded between artifacts (e.g. set and item) and their element mappings were not handled properly in the process of the crosswalk.

Applying the RDF as a crosswalk model
Step 1: identification of the multiple objects and their relationships embedded in source metadata elements. In DAPNPM, the whole-part relationship is implicitly embedded in the Chinese cultural artifacts. For example, the grid trellis of Round Treasure Box with Indian Lotus Décor (Ch'ing dynasty, name in Chinese: 清竹絲纏枝番蓮圓盒多寶格) is a set composed of several items, such as Jade in the Shape of an Auspicious Animal (Ch'ing dynasty, name in Chinese: 清舊玉異獸) and Jade Goose (Ch'ing dynasty, name in Chinese: 清舊玉鵝). Therefore, two relationships exist between the above artifacts as follows: is-Part-of and is-Related-to. The former is employed to indicate the whole-part relationship between the set and the item, and the latter is used to show the relationships between items deposited to the same set. Furthermore, all relationships are reciprocal relationships, such as is-Part-of and has-Part which is utilized to represent the relationship between set and item, and is-Related-to which is used to represent the relationship between items.

Step 2: selection of the adopted objects and their metadata elements from the source standard. In terms of mapping, the DAPNPM is different from the ANSP in that it exports metadata elements both of sets and items into the DC-based TELDAP UC. Therefore, two specific instructions for mapping from CDWA to DC were required. First, the granularity of the object had to be indicated to inform users whether artifacts are sets or items. Thus, a default value (either set or item) was added to the Type element of DC. In the case of the DAPNPM, two kinds of relationships embedded in museum's artifacts had to be added into the DC's Relation elements with different attributes: one is has-Part and is-Part-of for sets and items, and the other is is-Related-to for individual

Table II. Crosswalk between EAD and DC for the ANSP at the archival item level

Subject (i.e. EAD)	Predicate	Object (i.e. DC)	Mapping type	Mapping rules and notes	Instance
< accessrestrict > < userestrict >	is-Generalized-by (reverse: is-Specialized-by)	Rights	Append	Map directly with EAD tag as a display label in the DC element	Access restrict: internal use Use restrict: internal use
< controlaccess > → < geoname >	is-Synonymous-with	Coverage-Spatial	Direct	Map directly from sub-hierarchical EAD tag to DC element	東經 121° 44' 8"121 (i.e. 121° 44' 8"121 East) 北緯 25° 8' 2"(i.e. 25° 8' 2" North)
< languagematerial > → < language >	is-Synonymous-with	Language	Direct	Map directly from sub-hierarchical EAD tag to DC element	中文 (i.e. Chinese)
< origination > → < persname >	is-Composed-by (reverse: is-Decomposed-into)	Creator	Direct	Map directly from sub-hierarchical EAD tag to DC element with specified syntax, such as: LastName, FirstName	周, 峻義
< physdesc > → < dimensions > < physdesc > → < extent > < physdesc > → < physfacet >	Is-Combined-as (reverse: is-Divided-into)	Format	Separate	Map directly from sub-hierarchical EAD tag to DC element with display labels, such as: dimension, extent and physical-facet	Dimensions: 38.6×26.3 cm Extent: 2 頁 (i.e. pages) Physical-facet: 紙 (i.e. paper)
< physdesc > → < genreform >	is-Synonymous-with	Type		Map directly from sub-hierarchical EAD tag to DC element with a display label: genreform	Genreform: 件 (i.e. item)

(continued)

Table II.

Subject (i.e. EAD)	Predicate	Object (i.e. DC)	Mapping type	Mapping rules and notes	Instance
< repository > → < corpname >	is-Synonymous-with	Publisher	Direct	Map directly from sub-hierarchical EAD tag to DC element	Navy Memorial Digital Archives
< unitdate >	is-Synonymous-with	Date	Direct	Map directly	1995-06-01
< unitid >	is-Synonymous-with	Identifier	Direct	Map directly	005
< unittitle > of items	is-Synonymous-with	Title	Direct	Map directly	航泊日誌單日本 - 資料頁
< unittitle > of fonds, series and files	is-Part-of (reverse: has-Part)	Relation	Transformation	1. Mapping syntax: relation type (default value: is-Part-of) plus title of Fonds-Serie-Files 2. Concatenate titles of Fonds, Series and Files together with "is-Part-of" default value	海軍總司令部 (i.e. Navy Command Headquarters) – 海軍艦隊司令部 (i.e. Naval Fleet Command) – 洛陽軍艦 (i.e. ROCS Lo-Yang) – 單日本航泊日誌 (i.e. log book)

Notes: Definitions of EAD and DC metadata elements can be referred in the following URLs: EAD is available at: www.loc.gov/ead/tglib/index.html; DC is available at: http://dublincore.org/documents/usageguide/elements.shtml items. Table III is an application of CDWA elements adopted by the DAPNPM. In the case of the DAPNPM, objects both of sets and items share similar CDWA elements.

Step 3: identification of the semantic and hierarchical relationships between the source and target elements. There are five sets of hierarchical relationships embedded in the CDWA elements used by the DAPNPM in the following: Creation-Creation Date, Current Location, Measurements, Object/Work and Ownership/Collecting History. During mapping from CDWA to DC, mapping arrangements for the DAPNPM were mostly different from those for ANSP as follows:

- Append together: some hierarchical CDWA elements were combined together to map to a DC element with a display label. For example, Creation-Creation Date-Earliest Date and Creation-Creation Date-Latest Date were combined together into the DC element "Date."
- Separate into different elements: some CDWA elements belonging to the same hierarchical category were separated into different DC elements. For instance, Current Location-Current Repository Numbers was mapped to DC's Identifier and the Current Location-Current Repository/Geographic Location was mapped to DC's Description. This kind of arrangement can also be applied to Object/Work and Ownership/Collecting History (Table IV).
- Hybrid: some hierarchical CDWA elements were first appended together to map into the same DC element, and were then separated with different display labels. All CDWA's Measurements elements were mapped to DC's Format first. Then Measurements-Dimensions Description was separated from the other CDWA Measurement elements as an independent entry to DC's Format, and the other sub-elements of Measurements were combined together into DC's Format element with different display labels.

CDWA elements	Set	Item
Condition/Examination History-Condition/Examination Description	△	△
Creation-Creation Date-Earliest Date	△	△
Creation-Creation Date-Latest Date	△	△
Current Location-Current Repository Numbers	△	△
Current Location-Current Repository/Geographic Location	△	△
Measurements-Dimensions Description	△	△
Measurements-Dimensions Extent	△	△
Measurements-Dimensions Type	△	△
Measurements-Dimensions Unit	△	△
Measurements-Dimensions Value	△	△
Object/Work-Catalog Level	△	△
Object/Work-Components/Parts-Components Quantity	△	△
Object/Work-Remarks	△	△
Ownership/Collecting History-Legal Status	△	△
Ownership/Collecting History-Owner's Credit Line	△	△
Ownership/Collecting History-Ownership Date	△	△
Ownership/Collecting History-Transfer Mode	△	△
Related Works-Related Work Label/Identification-Work Relationship Type	△	△
Titles or Names-Title Text	△	△
Titles or Names-Title Type	△	△

Note: △ stands for elements used by ANSP for metadata conversion into TELDAP UC

Table III. CDWA elements at set and item levels used by the DAPNPM

Subject (i.e. CDWA)	Predicate	Object (i.e. DC)	Mapping type	Mapping rules and notes	Instance
Condition/Examination History-Condition/ Examination Description	is-Synonymous-with	Description	Exclusion	Do not export	
Creation-Creation Date-Earliest Date	is-Combined-as	Date	Append	Append together and then map directly from third hierarchical CDWA category to DC element with display label, such as: Creation Date-Earliest Date and Creation Date-Latest Date	Creation Date-Earliest Date: 清 (i.e. Ch'iang dynasty) Creation Date-Latest Date: 清 (i.e. Ch'iang dynasty)
Creation-Creation Date-Latest Date					
Current Location-Repository/Geographic Location-Repository Number	is-Synonymous-with	Identifier	Separate	Map directly	現貯箱號 院 2020 箱
Current Location-Repository/Geographic Location		Description		Map directly from sub-CDWA category to DC element with display label such as: Geographic Location	Geographic location: 雜項庫房
Measurements-Dimensions Description	is-Synonymous-with	Format	Hybrid	Map directly into an independent entry from others	尺寸高 24.5 公分 徑 18.5 公分
Measurements-Dimensions Type	is-Combined-as (reverse: is-Divided-into)	Format-Extent		Append together and then map directly from second hierarchical CDWA category to DC element with display label, such as: Dimension Type, Value and Unit	Dimension type: 高 (i.e. height) Dimension type: 徑 (i.e. diameter)
Measurements-Dimensions Value		Format-Extent			Dimension value: 24.5 Dimension value: 18.5
Measurements-Dimensions Unit		Format-Extent			Dimension unit: 公分 (i.e. cm) Dimension unit: 公分 (i.e. cm)

(*continued*)

Table IV. Crosswalk between CDWA and DC for the DAPNPM

Table IV.

Subject (i.e. CDWA)	Predicate	Object (i.e. DC)	Mapping type		Mapping rules and notes	Instance	
Measurements-Dimensions Extent		Format-Extent				Dimension extent: 全器 (i.e. whole artifact)	Dimension extent: 全器 (i.e. whole artifact)
Object/Work-Catalog Level	is-Synonymous-with	Type	Separate		Default value is either Set or Item	單件 (i.e. item)	
Object/Work-Components/ Parts-Components Quantity		Format			Map directly from 3rd hierarchical CDWA category to DC element	27	
Object/Work-Remarks		Description			Map directly	相關藏品 (i.e. related works) 故玉 005629N000000000 故玉 005618N000000000	
Ownership/Collecting History-Owner's Credit Line Ownership/Collecting History-Transfer Mode	is-Synonymous-with	Description	Separate	Exclusion	Do not export		
Ownership/Collecting History-Legal Status	is-Synonymous-with	Description		Direct	Map directly	古物 (i.e. antique)	
Ownership/Collecting History-Ownership Date	is-Generalized-by (reverse: is-Specialized-by)	Date		Direct	Map directly from second hierarchical CDWA category to DC element with display label such as: Ownership Date	Ownership Date: Unknown	
Titles or Names-Title Text	is-Synonymous-with	Title		Direct	Map directly	清 竹絲殘纏枝番蓮圓盒多寶格	
Titles or Names-Title Type	is-Generalized-by (reverse: is-Specialized-by)	Title	Direct		Map directly from second hierarchical CDWA category to DC element with display label such as: original and English	English: Round treasure box with Indian lotus décor, Ch'ing dynasty	

(*continued*)

Subject (i.e. CDWA)	Predicate	Object (i.e. DC)	Mapping type	Mapping rules and notes	Instance
Related Works-Related Work Label/Identification-Work Relationship Type	is-Part-of (reverse: has-part) or is-Related-to	Relation	Direct	Use "has-Part-of" and "is-Part-of" for collection and its items records, respectively. Use "is-Related-to" for items belonging to the same collection	HasPart: 清 舊玉鵝 (i.e. Jade goose, Ch'ing dynasty)
		Publisher	Addition	Default value is NPM	國立故宮博物院 (i.e. NPM)

Notes: Definitions of EAD and DC metadata elements can be referred in the following URLs: CDWA is available at: www.getty.edu/research/publications/electronic_publications/cdwa/; DC is available at: http://dublincore.org/documents/usageguide/elements.shtml; Metadata instance was retrieved from Digital Archiving System of National Palace Museum Antiquities

Table IV.

Step 4: identification of the granular and syntactic relationships between source and target elements. In the use case of the DAPNPM, there was a need to specify mapping rules for the granular relationships, including many-to-one and one-to-many. In terms of many-to-one, two CDWA elements (i.e. Creation-Creation Date and Ownership/Collecting History-Ownership Date) shared the same Date element of DC. Two display labels (i.e. Creation and Ownership) and is-Generalized-by predicate were needed to refine the Date element more specifically. In terms of one-to-many, two types of titles embedded in Title or Names-Title Type of CDWA for the same Chinese artifact were requested by the DAPNPM: original and English title. Thus, Title or Names-Title Type of CWDA's category were split into the DC's Title element with two different display labels (i.e. original and English title), and is-Generalized and is-Specialized-by predicates were employed to illustrate bi-directional mapping between CDWA and DC.

Step 5: mapping of all equivalent lexical elements from source metadata into target with reference to RDF-based relationships. However, this study still used RDF as a crosswalk model to visualize and illustrate CDWA elements and their various embedded relationships of DAPNPM's museum artifacts. It also transformed the lexical metadata mapping into semantic metadata mapping with various contextual relationships. Furthermore, an additional DC element was selected as an additional default element to show which institution was responsible for making this artifact available and the default value was NPM. Results of crosswalk between CDWA and DC for the use case of the DAPNPM are listed in Table IV.

Discussion

Typically crosswalks constitute a chart or table that maps semantic elements from one metadata standard to another (Baca, 2003). In practice, many implicit contextual relationships are not illuminated clearly in such tables resulting in mapping errors in crosswalks. This study developed a semantic crosswalk model using RDF that illuminates the various contexts of metadata elements and their relationships. This model was then tested on two use cases and was proved to be feasible to obtain metadata interoperability. This RDF-based mapping approach illuminated the issues and challenges of semantic metadata crosswalk at the data element level as follows.

One-to-one principle and multiple object mapping

DC is a widely used standard for metadata description and crosswalking. The one-to-one principle defined by DC is also one of most important principles in crosswalking for sharing and exchanging metadata elements between various metadata standards in digital libraries. As seen by the use cases in this study, the issue of information granularity may make crosswalking complicated. As seen with the ANSP, digital archival projects often have compound archival objects and provenance relationships. When standards such as the EAD are used, specific characteristics of archival materials are always implicitly embedded in metadata elements. And metadata standards such as CDWA represent various associative relationships, ranging from collections, groups, series and sets to items and components. Without RDF-based semantic representation, metadata professionals focus their attention on lexical mapping of elements, and lose or ignore the contextual relationships embedded in multiple objects and their corresponding metadata elements. Then all metadata elements are often merged into a single object in DC, which inevitably results in mapping errors. The RDF-based representation of information objects and their contextual relationships between multiple objects is therefore useful for metadata professionals in decision making for mapping. First, it allows clarification of the relationships between multiple objects, from which a table of element distribution with their corresponding objects can be built up, as shown as Tables I and III. Second, selected multiple objects and elements can be further identified into a mapping table from source standard to the target as shown in Tables II and IV. If not all the granular objects are selected for mapping, as in the ANSP case, then metadata professionals have to transform source elements related to multiple object relationships into target elements to indicate the provenance or whole-part relationship. With the RDF representation, professionals can take the context of various multiple objects and their relationships into account for metadata crosswalk and specify appropriate mapping rules.

Complexity of mapping issues and types

Indirect mapping or relative crosswalk is one of the most difficult tasks of semantic metadata mapping. The reason is that it is not only dependent on lexical form, appearance or meanings, but also requires various mapping rules to represent the implicit contextual relationships between source and target metadata elements. Although approaches such as append (Cao *et al.*, 2004) and qualifier (Dunsire *et al.*, 2011) have been proposed as part of metadata conversion specification, the ANSP and DAPNPM use cases show that issues of semantic metadata crosswalk are much more complicated than previous studies have addressed. In addition to lexical equivalence mapping, a complete semantic metadata crosswalk at the data element level has to solve more complicated mapping issues, such as semantic, hierarchical, syntactic, multiple object and granular (i.e. many-to-one and one-to-many) relationships to achieve metadata interoperability. This study has expanded mapping types from existing approaches of direct and no mappings, append and qualifier into addition, exclusion, separate, transformation and hybrid, and has proven the feasibility with definitions and instances. The nine mapping types developed do not only illuminate the complexity of the issues surrounding crosswalking between source and target metadata elements, but can also be regarded as a classification framework of mapping types for in-depth examination based on more case studies in the future. On the other hand, the adoption of display labels in this study borrowed the concept of display constant from MARC, rather than the qualifier approach, can be considered to be consistent with the dumb-down principle of DC to keep the semantics of metadata elements intact, clearly preserve and deliver unambiguous meanings to end users, and facilitate greater global interoperability between heterogeneous metadata standards.

RDF-based visual representation

As a crosswalk is generally a chart or table, many implicit semantic relationships between elements are not illustrated clearly resulting in mapping errors in practice. The proposed RDF-based crosswalk approach consists of five steps, clarified in detail by use cases, which identify selected objects and revive their contextual semantics and relationships in mapping from source elements into target elements. The proposed approach does not only map elements based on semantic definitions, but is also a graph-based, encoded example of a semantic metadata crosswalk. Thus issues such as identical elements, semantic transformation, syntax, hierarchy, many-to-one, one-to-many and multiple objects can be solved to harmonize semantic metadata interoperability. Therefore, the proposed approach is complementary to existing crosswalks offered by authoritative organizations such as the LC and GRI, rather than mutually exclusive. Furthermore, both the proposed

RDF-based approach by this study and the mapping language provided by CIDOC community (Kondylakis *et al.*, 2006) share in including contextual relationships for semantic metadata interoperability. However, the former is different from the latter in elaborating contextual relationships with unambiguous semantics and expressions. Finally based on RDF expression, all detailed expressions can be converted into encoded RDF triples and stored in registry such as Open Metadata Registry to liberate all metadata elements from locally flat vocabularies with an eye toward future linked data and the Semantic Web applications in the open domain.

Conclusion

The use cases illustrated in this study (the ANSP and DAPNPM), show that the RDF-based approach to semantic crosswalking is distinctive from existing approaches as it solves complicated mapping issues at the data element level, including semantic, hierarchical, granular, syntactic and multiple object relationships. With visualized RDF semantic representation reviving contextual relationships embedded between metadata elements, the RDF-based crosswalk approach not only manifests the hidden relationships, but also provides more detailed contextual information as a basis for the specification of mapping rules. Furthermore, RDF-based mappings can be transformed into RDF triples smoothly to allow extension of applications into linked data and the Semantic Web in the open domain.

This study has also expanded the metadata crosswalk at the data element level from append and qualifier into a set of nine mapping types. These nine mapping types have further illustrated the complicated issues surrounding semantic metadata crosswalking. The proposed nine mapping types can be used as a classification framework and inspire more studies to examine its feasibility and wider applications. The proposed mapping types also form a useful basis from which to develop mapping software for automatic processing.

One limitation of this study may be conditions, cardinality and constraints of metadata elements, as well as the target subjects and their cases of use. Conditions (e.g. mandatory, optional and recommendation), constraints (e.g. free text and numeric range) and cardinality (i.e. repeatable and non-repeatable) of metadata elements have not included and well elaborated as part of the proposed RDF-based crosswalking approach. Furthermore, only two use cases were selected from archives and museums to illustrate how to achieve semantic metadata crosswalking at the data element level. If the approach used in this study can be expanded into more diverse cases of use and types of materials, such as biodiversity artifacts of natural history museums, and literary works of libraries, then more revisions may be made to produce a more robust RDF-based crosswalk for semantic metadata interoperability in cultural heritage.

Note

1. A owl:sameAs link is defined to indicate that two concepts are the same individual (i.e. the same source). A skos:exactMatch link is used to indicate that two concepts can be used interchangeably for the same individual. As each metadata standard is rooted in a distinct domain with heterogeneous semantics, EAD and CDWA are more domain-specific standards than DC, and they are defined with contextual semantics for museums and archives communities. This study used is-Synonymous-with to indicate that two terms and their concepts are synonymous only under mapping conditions for semantic metadata crosswalk. In fact most synonymous terms between EAD and DC, and CDWA and DC listed in Tables II and IV are not used interchangeably nor do they refer to the same individual in the real world

because the concepts of DCs are too broad and abstract to result in a semantic gap or confusion. The instances include EAD's geoname and DC's Coverage-Spatial, and CDWA's Measurement and DC's Format. Halpin *et al.* (2010) reported five types of errors for incorrect use of owl:sameAs. Therefore, this study uses is-Synonymous-with rather than owl:sameAs and skos:exactMatch for semantic metadata mapping.

References

Baca, M. (2003), "Practical issues in applying metadata schemas and controlled vocabularies to cultural heritage information", *Cataloging & Classification*, Vol. 36 Nos 3/4, pp. 47-55.

Cao, Y., Lin, F., McGreal, R., Schafer, S., Friesen, N., Tin, T., Anderson, T., Kariel, D., Powell, B. and Anderson, M. (2004), "Facilitating e-learning with a MARC to IEEE LOM metadata crosswalk application", *Lecture Notes in Artificial Intelligence*, Vol. 3029, pp. 739-748.

Chan, L. and Zeng, M. (2006), "Metadata interoperability and standardization: a study of methodology part I: achieving interoperability at the schema level", *D-Lib Magazine*, Vol. 12 No. 6, available at: www.dlib.org/dlib/june06/chan/06chan.html (accessed March 19, 2009).

Chandrakar, R. (2002), "CCF and INFLIBNET standard metadata to dublin core crosswalk", *Journal of Internet Cataloging*, Vol. 5 No. 3, pp. 33-57.

Chandrakar, R. (2005), "An approach to mapping CCF to dublin core", *The Electronic Library*, Vol. 23 No. 5, pp. 577-590.

Dunsire, G., Hillmann, D.I., Phipps, J. and Coyle, K. (2011), "A reconsideration of mapping in a semantic world", *Proceedings of International Conference on Dublin Core and Metadata Applications 2011*, *September 21-23, Hague*, available at: http://dcpapers.dublincore.org/index.php/pubs/article/view/3622/1848 (accessed March 20, 2009).

Gilliland, A.J. (2008), "Setting the stage", in Baca, M. (Ed.), *Introduction to Metadata* (online edition, ver. 3.0), Getty Research Institute, Los Angeles, CA, pp. 1-19, available at: www.getty.edu/research/publications/electronic_publications/intrometadata/setting.pdf (accessed December 28, 2012).

Godby, C.J., Smith, D. and Childress, E. (2008), "Toward element-level interoperability in bibliographic metadata", *Code4Lib Journal*, No. 2, available at: http://journal.code4lib.org/articles/54 (accessed March 20, 2009).

Halpin, H., Hayes, P.J., McCusker, J.P., McGuinness, D.L. and Thompson, H.S. (2010), "When owl:sameAs Isn't the same: an analysis of identity in linked data", *Lecture Notes in Computer Science*, Vol. 6496, pp. 305-320.

International Council on Archives (2000), "ISAD(G): general international standard archival description", 2nd ed., available at: www.icacds.org.uk/eng/ISAD(G).pdf (accessed December 28, 2014).

Kitamura, Y., Segawa, S., Sasajima, M., Tarumi, S. and Mizoguchi, R. (2008), "Deep semantic mapping between functional taxonomies for interoperable semantic search", *Lecture Notes on Computer Science*, Vol. 5367, pp. 137-151.

Kondylakis, H., Doerr, M. and Plexousakis, D. (2006), "Mapping language for information integration", ICS-FORTH Technical Report No. 385, December 6, available at: www.cidoc-crm.org/docs/Mapping_TR385_December06.pdf (accessed March 12, 2012).

Lagunas, J.S. and Basurto, L.F. (2005), "Model of metadata crosswalk on the Iberoamerican and Caribbean virtual library", *Proceedings of International Conference on Dublin Core and Metadata Applications 2005*, *September 12-15, Madrid*, available at: http://dcpapers.dublincore.org/ojs/pubs/article/view/814/810 (accessed March 20, 2009).

Lightle, K.S. and Ridgway, J.S. (2003), "Generation of XML records across multiple metadata standards", *D-Lib Magazine*, Vol. 9 No. 9, available at: http://dlib.ejournal.ascc.net/dlib/september03/lightle/09lightle.html (accessed March 19, 2009).

Lim, S., Seo, T., Lee, C. and Shin, S. (2012), "Study on the international standardization for the semantic metadata mapping procedure", *Lecture Notes in Artificial Intelligence*, Vol. 7239, pp. 243-249.

Machovec, G.S. (2002), "Challenges and issues with metadata crosswalks", *Online Libraries and Microcomputers*, Vol. 20 No. 4, pp. 1-3.

NISO (2004), "Understanding metadata", available at: www.niso.org/publications/press/UnderstandingMetadata.pdf (accessed March 19, 2009).

Nogueras-Iso, J., Zarazaga-Soria, F.J. and Muro-Medrano, P.R. (2005), *Geographic Information Metadata for Spatial Data Infrastructures: Resources, Interoperability and Information Retrieval*, Springer-Verlag, Berlin.

Park, J. (2002), "Hindrances in semantic mapping involving thesauri and metadata: a linguistic perspective", *Journal of Internet Cataloging*, Vol. 5 No. 3, pp. 59-77.

Rahm, E. (2011), "Towards large-scale schema and ontology matching", in Bellahsene, Z. *et al.* (Eds), *Schema Matching and Mapping: Data Centric Systems and Applications*, Springer-Verlag, Berlin, pp. 3-27.

Rahm, E. and Bernstein, P.A. (2001), "A survey of approaches to automatic schema matching", *The VLDB Journal*, Vol. 10 No. 4, pp. 334-550.

Shepherd, E. and West, V. (2003), "Are ISO 15489-1:2001 and ISAD(G) compatible: part 2", *Records Management Journal*, Vol. 13 No. 2, pp. 62-69.

St Pierre, M. and LaPlant, W.P. (1998), "Issues in crosswalking content metadata standards", available at: www.niso.org/publications/white_papers/crosswalk/ (accessed March 20, 2009).

Woodley, M.S. (2008), "Crosswalks, metadata harvesting, federated searching, metasearching: using metadata to connect users and information", in Baca, M. (Ed.), *Introduction to Metadata* (online edition, ver. 3.0), available at: www.getty.edu/research/conducting_research/standards/intrometadata/path.pdf (accessed March 20, 2009).

About the author

Dr Ya-Ning Chen is an Assistant Professor of Department of Information and Library Science of Tamkang University. His research interests are metadata, information organization, digital libraries, linked data, open access, research data management and social networks. Dr Ya-Ning Chen can be contacted at: arthur9861@gmail.com