Dublin Core metadata semantics: an analysis of the perspectives of information professionals

Jung-ran Park

College of Information Science and Technology, Drexel University, USA

Eric Childress

OCLC Research, OCLC Online Computer Library Center, USA

Abstract.

This study examines Dublin Core (DC) metadata semantics drawn from the perspectives and experiences of cataloguing and metadata professionals. The study ascertains the extent of difficulty in applying the DC metadata elements encountered by these professionals and examines factors engendering such difficulties during the metadata application process. Comments drawn from the survey participants (*n* = 141) show that conceptual ambiguities (41%) and semantic overlaps (45%) of the surveyed DC metadata elements are the most frequently cited factors causing difficulty and confusion, in turn leading to variant interpretations of DC metadata elements. This has the potential to bring forth inconsistent and inaccurate applications and implementation of the DC standard across institutions which can directly affect semantic interoperability across digital repositories. The high degree of difficulty (55.3%) engendered by the *Relation* field indicates that further examination of this element is needed.

Keywords: cataloguing and metadata professionals; conceptual ambiguities; Dublin Core metadata standard; metadata; metadata semantics; metadata creation practice; semantic overlaps

1. Introduction

Metadata, traditionally in the form of MARC (Machine-Readable Cataloguing) bibliographic records, has been the keystone of resource description and discovery. Standardized metadata facilitates semantic interoperability, which is a critical factor for resource sharing and discovery [1]. Dublin

Correspondence to: Jung-ran Park, College of Information Science and Technology, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA. Email: jung-ran.park@ischool.drexel.edu

Core (DC), designed for the description of electronic and web resources, is one of the most widely used metadata standards. Since its inception in 1995 in Dublin, Ohio, the DC metadata scheme has gained credibility and was approved as the NISO standard (Z39.85) in 2001 [2]. The standard was updated in 2009 as an ISO standard (15836) entitled *Information and documentation – the Dublin Core metadata element set* [3]. The Dublin Core Metadata Initiative (DCMI) serves as the ISO maintenance agency for ISO 15836.

The DC metadata scheme was developed primarily for one very important purpose – to facilitate cross-domain information resource description. Accordingly, DC semantics underscores the lowest common denominator for resource description. It is expected that document authors without an educational and professional background in cataloguing will be able to use the scheme. Park, for instance, shows how the Open Language Archives Community, mainly composed of linguists, utilizes DC for describing language-related resources worldwide [4].

However, there are issues inherent in the semantics of the DC scheme that hinder consistent and accurate application [5–7]. For instance, through an empirical analysis of usage of the DC metadata elements, Park [5] points out that conceptual ambiguities and semantic overlaps inherent among some DC metadata elements are partly responsible for inconsistency and inaccuracy in DC metadata application. Caplan [6, p. 78] also addresses an issue related to DC metadata semantics: 'Despite the simplicity of the Dublin Core Scheme, certain problems have arisen repeatedly in applications. One issue concerns the overlap in meaning in the definition of some elements'. In the meantime, Khoo and colleagues overview the problems drawn from lack of understanding of the basic concepts related to metadata and digital libraries by non-cataloguing professionals [8].

As will be discussed shortly, studies reporting issues in relation to DC metadata semantics are mostly derived from metadata analysis on the level of both data and service providers. In this sense, there has been a lack of research on DC metadata semantics drawn from the perspectives and experiences of cataloguing and metadata professionals who hold MLIS degrees or equivalent educational credentials and work experience in the area of information organization in general. With an eye toward addressing this research gap, this study aims at examining the extent of difficulty in applying DC metadata elements experienced by cataloguing and metadata professionals. It also aims to examine factors engendering such difficulties during the metadata creation process. The above goals are directly derived from the following research questions:

1. Which Dublin Core metadata elements pose the greatest degree of difficulty to cataloguing and metadata professionals during the metadata creation processes?

2. What are the primary factors engendering such difficulty?

2.   Literature review: Dublin Core metadata semantics

In this section, we will review studies related to DC metadata semantics. (The review is not meant to be exhaustive. For a state of the art review on metadata quality in digital repositories, see Park [9].)

The quality of metadata reflects the degree to which the metadata in question perform the core bibliographic functions of discovery and use. In this sense, the principal purpose of metadata is to a large degree related to that of the traditional online library catalogues and databases in finding, identifying, selecting and obtaining items. While examining metadata in e-print archives, Guy et al. [10] assessed metadata quality based on the functional requirements. Their definition of quality metadata concerns 'functional requirements of the system it is designed to support', which can be summarized as 'quality is about fitness for purpose'.

NISO [11] presents the following six principles of 'good' metadata which provide a framework of guidance for building good digital collections:

1. Conforms to community standards

2. Supports interoperability

3. Uses authority control and content standards

4. Includes a clear statement of the conditions and terms of use

5. Supports long-term curatorial functions and preservation

6. Should have the qualities of good objects, including authority, authenticity, archivability, persistence and unique identification.

Among the above-mentioned principles, support of interoperability and use of authority control and content standards to some degree relate to semantic factors for conditioning metadata quality. The importance of such semantic factors of quality metadata is also noted by Lei et al. [12].

Through the use of focus groups Howarth [13, 14] presents the challenges of matching between a common namespace set of 17 labeled categories, based on an analysis of the mapping of nine metadata schemes (i.e. DC, MARC, EAD, GILS, TEI, VRA, CIMI, CSDGM, ONIX) and their corresponding definitions. According to the findings of the study, only three category names (i.e. language, physical format, and date and time period) are most readily understood; others ($n = 14$) require clarification of definition and additional definitions. Howarth [8, p. 183] underscores the critical role played by context in which a term and the corresponding definition occur: 'the provision of brief, simple definitions devoid of topical or other perspective was insufficient to clearly and unequivocally identify all elements'.

Informed by Howarth, Park [15] conducted a mapping between unqualified DC metadata elements and their corresponding definitions at the start of a metadata course delivered in early 2006. The course has a prerequisite of basic cataloguing and classification; accordingly the participating students ($n = 18$) have background knowledge in information organization and some vocabulary schemes.

For the mapping task, DC metadata element names and corresponding definitions are arbitrarily rearranged from an unqualified DC metadata scheme. Considering the number ($n = 15$) of DC metadata element names and definitions, students were allocated 25 minutes to complete the task. Instructions were given for the mapping task: one-to-one mapping, one-to-many mapping and null mapping were allowed. After completion of the mapping task, students were encouraged to provide comments on any problems and difficulties they faced during the mapping between DC metadata element names and their corresponding definitions.

The most frequently cited unclear and ambiguous element names were the following: the relationship among *Format*, *Type* and *Description*; the relationship between *Source* and *Relation*; the relationship between *Type* and *Subject*; the relationship between *Description* and *Subject*; and the relationship among *Creator*, *Publisher* and *Contributor*. The most frequently occurring element names vis-à-vis the accuracy of mapping were the following: *Language*, *Date* and *Title*. During an open discussion forum (ca. 10 minutes), students commented on the lack of semantic clarity in most element names and definitions ($n = 11$) which in turn engendered difficulty in the mapping task. The semantic link that operates most clearly between DC metadata element names and corresponding definitions concerns *Language*, *Date*, *Title* and *Right*. The results of the mapping task and the open discussion forum indicate that unclear semantic links between DC metadata element names and corresponding definitions may lead to confusion and inaccuracy in the mapping task involving the above-mentioned DC metadata elements.

Park [5] examined DC metadata semantics through two mechanisms:

1. Linguistic semantic analysis of DC metadata element names and corresponding definitions

2. Analysis of 659 DC metadata item records drawn from three digital repositories.

The findings of the study suggest that conceptual ambiguities and semantic overlaps inherent among some DC metadata elements hinder semantic interoperability. Vital issues, in terms of resource sharing and discovery across digital collections owing to inaccurate and inconsistent metadata usage, were also brought to the fore. The most frequently occurring inaccurate metadata element uses are partially owing to confusion between *Format* and *Description* and between *Type* and *Format* (they are interchangeably used); the DC elements *Source* and *Relation* are also inconsistently used.

The confusion of application of DC *Type* and *Format* is also reported in the study by Zeng [16] on metadata quality analysis of the open source Metadata Repository at the National Science Digital Library (NSDL): 'Particular areas where confusion occurs are between Type and Format'. The same confusion is also observed by Godby, et al. [7]. By examining 400 Dublin Core records, the study reported on incorrect and inconsistent metadata uses: 'Format and Type both contain names of media types such as photograph'.

As stated earlier, *Source* and *Relation* are also interchangeably used in Park's [5] study. Zeng's [16] analysis of the NSDL metadata repository also presents this: 'Particular areas where confusion occurs are between … Relation and Source …'. Park [5] reported that *Source* (54.9%) and *Relation* (33.8%) are the most infrequently employed elements from the surveyed three digital image collections.

Infrequent and inaccurate usage of *Source* and *Relation* is also the case in the NSDL metadata repository [17]. The NSDL comprises over 100 collection sets submitted from various data providers. The lack of consistency in metadata usage in the NSDL is partially due to the fact that metadata in the repository derives from many different data providers. As well, these data providers utilize a variety of schemes other than the DC metadata scheme. However, for data harvesting purposes all metadata schemes in the NSDL are mapped onto the DC metadata scheme. It is in this mapping process that inaccurate and inconsistent mappings occur. According to Bui and Park [17], the usage of *Source* in 111 collections is less than 15% and that of *Relation* is less than 7%.

Sokvitne [18] presents inaccuracies in DC metadata application drawn from 20 Australian government and educational organizations. He found that the high incidence (58%) of duplication of the same data value occurs in metadata elements such as *Creator*, *Contributor* and *Publisher*. Regarding these elements, Caplan [6] points out that at one point a proposal to combine these three elements into a single element called 'agent' was considered and rejected due to impracticability. Thus, at this point there are no refining qualifiers to specify the meaning of these elements. This semantic overlap engenders confusion and inaccuracy in the usage of the DC metadata elements *Creator*, *Contributor* and *Publisher*.

Inaccurate and inconsistent metadata uses bring forth the importance of metadata quality assurance and essential needs for metadata education (see Park [9]). For instance, through research on cross-collection searches for visual resources, Hegg and Knab [19] point out that the solution for optimal cross-collection searching depends on 'the curator's ability to accurately map the MDID file onto DC elements and refinements'. Godby, et al. [7] also address the importance of metadata education through analysis of 400 Dublin Core records: 'Without extensive human-mediated correction, or training that promotes more consistent application of the Dublin Core element semantics when the records are created, even the goal of limited interoperability is compromised'.

In line with this, while reporting on hindrances in integrating DTDs (Document Type Definitions), Heflin and Hendler [20] stress the indispensability of cataloguing professionals and human indexers in the metadata creation process: 'it is difficult for machines to make determinations of this nature, even if they have access to a complete automated dictionary and thesaurus'. Through examining learning objects and e-prints of communities of practice, Barton et al. [21] point out the problems inherent in the metadata creation process, such as inaccurate data entry and the lack of formal investigation into the process. They also argue for the essential need for metadata education: 'not all problems of metadata quality can be addressed effectively by machine solutions'.

As shown, several studies report semantic issues with the DC metadata standard. However, there is a lack of studies on DC metadata semantics drawn from the perspectives and experiences of cataloguing and metadata professionals. To address this research gap, in the following sections we will examine the extent and degree of difficulty in applying certain DC metadata elements and the factors impinging on this.

3.   Research procedures

This study is a part of a three-year (2006–2009) project entitled 'Metadata Creation and Metadata Quality Control across Digital Repositories: Evaluation of Current Practices' funded through the

Table 1
Electronic mailing lists for the survey

1. AUTOCAT: AUTOCAT@LISTSERV.SYR.EDU
2. Dublin Core listserv: DC-LIBRARIES@JISCMAIL.AC.UK
3. Metadata librarians listserv: metadatalibrarians@lists.monarchos.com
4. Library and Information Technology Association listserv: lita-l@ala.org
5. OnLine Audiovisual Catalogers electronic discussion list: OLAC-LIST@LISTSERV.ACSU.BUFFALO.EDU
6. Subject Authority Cooperative Program listserv: SACOLIST@LISTSERV.LOC.GOV
7. SERIALST: SERIALST@LIST.UVM.EDU
8. Text Encoding Initiative listserv: TEI-L@LISTSERV.BROWN.EDU
9. Electronic Resources in Libraries listserv: ERIL-L@LISTSERV.BINGHAMTON.EDU
10. Encoded Archival Description listserv: EAD@LISTSERV.LOC.GOV

Institute of Museum and Library Services. The aim of the project is to gain a better understanding of issues surrounding the metadata creation process, the employment of controlled vocabulary schemes, metadata quality control measures and new competencies and skill sets demanded of cataloguing and metadata professionals in the digital era, with associated consequences to library and information science education.

The objective of the research reported in this paper is to investigate the current state of application of the DC metadata standard in libraries across the USA. Toward this end, we selected some of the DC metadata elements that show the most frequent issues demonstrated in the literature in terms of consistency and accuracy in application of the DC standard (see Section 2). These elements include: *Creator*, *Contributor*, *Publisher*, *Type*, *Format*, *Source* and *Relation*. The selection of these elements is particularly informed by the results of the study by Park [5], also discussed in Section 2.

Park [5] extensively reviewed DC metadata use and related semantics in relation to semantic interoperability across digital repositories. To reiterate, the results of her study based on an analysis of 659 DC metadata item records and linguistic semantic analysis of DC metadata element names and corresponding definitions show evidence of frequent inaccurate and inconsistent usage of the following metadata elements: *Format*, *Type*, *Source* and *Relation*.

As stated earlier, there is a lack of studies on DC metadata semantics drawn from the perspectives and experiences of cataloguing and metadata professionals. To ascertain the extent of difficulty in applying DC metadata elements encountered by these professionals and to examine factors engendering such difficulties during the metadata creation process, we conducted a web survey. The survey was designed using *websurveyor* (vovvich.com) and included structured and open-ended questions. It was extensively reviewed by members of the advisory board, a group of three experts in the field, and pilot-tested prior to being officially launched.

Participants were recruited through survey invitation messages and subsequent reminders to the electronic mailing list of communities of metadata and cataloguing professionals. Table 1 indicates the mailing lists employed for the study.

We also sent out individual invitations and distributed flyers to selected metadata and cataloguing sessions during the 2008 annual ALA Midwinter Conference held in Philadelphia.

4. Results

During the 62-day period from 6 August 2008 to 6 October 2008, a total of 303 completed responses were received by the survey management system. The survey attracted a large number of initial participants ($n = 1371$). Among the initial participants who started the survey, a total of 303 (22.1%) people completed it. We suspect that incompletion of the survey may stem from the fact that the subject matter may have been outside the scope of the participants' job responsibilities. The length of the survey may also have been a factor in the incompletion rate.

Table 2
Job titles of participants (n = 227)

Job titles	Number of participants (%)
Other	135 (44.6)
Cataloguer/cataloguing librarian/catalogue librarian	99 (32.7)
Metadata librarian	29 (9.6)
Catalogue and metadata librarian	26 (8.6)
Head, cataloguing	26 (8.6)
Electronic resources cataloguer	17 (5.6)
Cataloguing coordinator	15 (5.0)
Head, cataloguing and metadata services	15 (5.0)

Table 3
Professional activities specified in 'Other' category (n = 99)

Professional activities	Number of participants (%)
Cataloguing and metadata creation	31 (10.2)
Digital projects management	23 (7.6)
Technical services	17 (5.6)
Archiving	16 (5.3)
Electronic resources and serials management	6 (2.0)
Library system administration	6 (2.0)

4.1. Participant profiles

The job titles of the participants are presented in Table 2. The largest proportion of participants (135, 44.6%) chose the 'Other' category instead of choosing one of the given job titles. Following the survey request, participants who chose this category further specified their job titles. We classified these job titles based on common characteristics of professional activities. This, in part, was informed by the matrices of job responsibilities developed by Park, Lu and Marion [22] and Park and Lu [23] for job description analyses of cataloguing and metadata professionals. Most of the job titles given as 'Other' are associated with one of the professional activities listed in Table 3.

As shown in Tables 2 and 3, job activities of survey participants mostly centre on cataloguing and metadata creation and management. The question related to job responsibilities further evidences this characteristic of the participant profile (see Table 4).

The job responsibilities shown in Table 4 show that the survey participants engage with the core activities of cataloguing such as descriptive and subject cataloguing, metadata creation and management, authority control, non-print and special material cataloguing, electronic resource/digital project management and integrated library system management (see [22, 23] for details on job responsibilities and competencies of cataloguing and metadata professionals).

The 'Other' category encompasses activities such as collection development, department/personnel management and supervision, public services, acquisitions, preservation and conservation, digital library projects design and development, metadata schema and system development and archival processing. As discussed in the studies [22, 23], there are a wide array of responsibilities including collection development and public services expected of cataloguing and metadata professionals.

In terms of work experience, more than half of the respondents (n = 170, 58%) reported over five years of experience in cataloguing and metadata creation: 6–15 years (n = 92, 31.6%); over 16 years

Table 4
Participants' job responsibilities (*n* = 303)

Job responsibilities	Number of participants (%)
General cataloguing (e.g. descriptive and subject cataloguing)	171 (56.4)
Metadata creation and management	153 (50.5)
Authority control	147 (48.5)
Non-print cataloguing (e.g. microform, music scores, photographs, video-recordings)	133 (43.9)
Special material cataloguing (e.g. rare books, foreign language materials, government documents)	126 (41.6)
Digital project management	101 (33.3)
Electronic resource management	62 (20.5)
Integrated library system management	59 (19.5)
Other	51 (16.8)

Table 5
Use of Dublin Core metadata standard (*n* = 283)

Response category	Response rating (%)
Yes	141 (49.8)
No	142 (50.2)

(*n* = 78, 26.4%). Approximately one-third of the respondents (*n* = 102, 34.5%) reported that they had been a cataloguing/metadata librarian for 1–5 years. The rest (*n* = 24, 8.1%) reported experience of less than a year.

Concerning the participants' institutional background, of the survey participants (*n* = 303) fewer than half (*n* = 121) provided institutional information. We believe that this is mostly owing to the fact that the question was designed to be optional, following a suggestion from the Institutional Review Board of the study. We analysed the institutional background from 121 responses. The majority of participants (*n* = 91) were from academic libraries followed by public libraries (*n* = 21) and from other institutions (*n* = 9).

4.2. DC metadata semantics: usage and difficulty

In order to understand the use of DC metadata among participants, we first questioned participants on whether they and their fellow cataloguers/metadata librarians use the DC metadata standard. Table 5 shows the percentage of participants (*n* = 283) answering 'Yes' and 'No' to this question.

As shown in Table 5, there was almost an exact split: half of the participants (*n* = 141, 49.8%) confirmed that they and their fellow cataloguers and metadata librarians utilize the DC metadata while the other half do not (*n* = 142, 50.2%).

To examine the extent of difficulty experienced by cataloguing and metadata professionals in applying DC metadata elements, we further asked participants to present their experience with the following metadata elements: *Creator*, *Contributor*, *Publisher*, *Type*, *Format*, *Source* and *Relation*. Table 6 illustrates the experiences of survey participants regarding the relative difficulty in applying these metadata elements.

As shown, more than half the participants (55.3%) expressed the greatest difficulty in applying the *Relation* field. In the case of *Source*, participants also perceived a great degree of difficulty (42.4%). Almost one-third of the survey participants (29.6%) expressed difficulty in using *Contributor*. In the case of *Publisher*, *Type* and *Format*, approximately one-fourth of the survey

Table 6
Degree of difficulty of DC metadata elements. Numbers in brackets

Element names	% Very difficult	% Somewhat difficult	% Somewhat easy	% Very easy
Relation	6.9 (9)	45.4 (59)	32.3 (42)	15.4 (20)
Source	6.2 (8)	36.2 (47)	31.5 (41)	26.2 (34)
Contributor	2.3 (3)	27.3 (36)	42.4 (56)	28.0 (37)
Publisher	2.3 (3)	23.7 (31)	38.9 (51)	35.1 (46)
Type	5.3 (7)	19.7 (26)	31.8 (42)	43.2 (57)
Format	2.3 (3)	20.8 (27)	33.1 (43)	43.8 (57)
Creator	3.0 (4)	18.2 (24)	38.6 (51)	40.2 (53)

Table 7
Factors engendering difficulty in DC metadata application

Factors	Response rating (%)
Ambiguity	29 (41)
Semantic overlaps	32 (45)
Broadness/simplicity	8 (11)
One-to-one principle	4 (6)
Other	7 (10)

participants (26%) expressed difficulty in using these elements. Approximately one-fifth of the survey participants (21.2%) expressed difficulty in using the *Creator* element.

To examine the factors causing difficulties in applying the above-mentioned DC metadata elements, we further asked participants to explain the perceived cause of the difficulty. One-fourth of the participants ($n = 72$) responded to this question. For the purposes of the study, we classified the descriptions into five categories based on common characteristics of the factor engendering difficulty. Table 7 illustrates the five categories and the response rating.

In the following subsections, we will discuss each category separately.

4.2.1. Ambiguity

As shown, ambiguity in DC metadata elements and their definitions is the second most frequently cited factor ($n = 29$, 41%) causing difficulty in application. Survey participants specifically pointed out the ambiguous nature of DC metadata. For instance, some of the participants commented on the vagueness of the definitions and difficulty in understanding definitions of the surveyed elements, as in: 'Dublin Core does not provide clear enough descriptions of what is contained in each element'. Such conceptual ambiguities seem to create confusion and difficulty in applying DC metadata.

Survey participants raised difficulties in using all the surveyed DC metadata elements owing to ambiguity, as indicated below:

• Type is a subjective value.

• Source is a confusing field. It is difficult to apply it consistently.

• Creator can be very varied and it can be tricky determining exactly who the creator is.

• The information from the publisher is vague.

• Can't define different role of contributor.

• There is often great ambiguity in terms of Type and Relation.

Some participants commented on unclear and vague semantic links by noting the possibility of variant interpretations resulting in inconsistent and inaccurate applications and implementation across institutions. This in turn may directly affect semantic interoperability across digital repositories resulting in hindrances in resource discovery. The following comments foreground this issue:

- The issue seems to be that everyone uses them differently.

- Definition of the 'source' is sufficiently ambiguous to result in various interpretations in implementation.

- Interpreting the standard and attempting to remain consistent across collections.

- There have been discussions nearly every time we set up a metadata template deciding how to interpret guidelines in conjunction with other statewide and consortial practices.

- The elements are not used consistently by institutions applying them.

This indicates that in the current DC metadata scheme instead of one-to-one mapping between DC metadata element names and corresponding definitions there may be inherent mapping possibilities of either many-to-one or one-to-many mappings [24]. One participant's comment directly points to this issue: 'Many options could satisfy fields'. This mapping issue is partially derived from semantic overlaps. Further analysis of the descriptions of participants shows considerable semantic overlaps of the surveyed elements as discussed below.

4.2.2. Semantic overlaps

The survey participants extensively addressed semantic overlaps as a factor ($n = 32$, 45%) causing difficulty in applying the surveyed DC elements. Their comments evidence semantic overlap between *Format* and *Type*, resulting in confusion and a difficulty in discerning a difference between the two. One participant reported that it was necessary to provide extensive documentation to distinguish between them: 'I have had to write extensive documentation to differentiate between type and format'. These semantic overlaps may bring forth inconsistency of DC metadata application as commented on by a participant: 'Format and Type are not self-evident labels, I find, and can be used differently by different metadata creators within an institution as well as across institutions'.

In the same way as with *Type* and *Format*, survey participants presented confusion and difficulty in distinguishing the DC metadata elements *Creator*, *Publisher*, and *Contributor*. One participant explained in detail challenges faced in using names and organizations denoting intellectual responsibility:

> Because we're a special collections library and museum, some of the added entries we want in our records are for people like former owners or recipients of correspondence. MARC provides relator terms that accurately describe their relationship to the items cataloged, but Dublin Core has no good place for them. They're neither creators of nor contributors to the items; their relationship began after the creation but is an important one to reflect in the catalog. Yes, the information can be included in provenance notes or, in the case of letters, in the title, but the names will not be searched or displayed along with other personal name fields.

There have been persistent complaints about not being able to capture the nature of the relationship of agents to the object described. DCMI has attempted to provide a guideline entitled *Using Agent Roles in Dublin Core* [25]; however, it is unclear how and to what extent DCMI's guideline has been adopted as a successful strategy.

Semantic overlaps are observed in the definitions of DC elements *Creator*, *Contributor* and *Publisher*. Definitions from the unqualified DC metadata scheme [26] illustrate that *Creator* is 'An entity primarily responsible for making the resource', *Contributor* is 'an entity responsible for making contributions to the resource', and *Publisher* is 'an entity responsible for making the resource available'. In these definitions, *Creator* can be seen as a particular type of *Contributor* and *Publisher*. As indicated in comments of the survey participants, this semantic overlap engenders confusion and difficulty in usage. The study by Sokvitne [18] also demonstrates this issue through an examination of DC metadata application drawn from 20 Australian government and educational organizations.

He found that a high incidence (58%) of duplicative data value occurs in metadata elements such as *Creator*, *Contributor* and *Publisher*.

The DC metadata elements *Source* and *Relation* seem to be challenging to survey participants. In the case of the former, participants perceived a greater degree of difficulty (42.4%) vis-à-vis other surveyed metadata elements. Among the surveyed DC elements, the *Relation* field shows the highest percentage of difficulty in the experience of participants. More than half of the participants (55.3%) expressed difficulty in applying this field.

In the same way as with DC elements discussed earlier, semantic overlap between *Source* and *Relation* is also observed in the comments of survey participants. To differentiate between these two elements, a participant reported that extensive documentation was necessary. Semantic overlaps between the two elements are observed in the DC metadata standard. In other words, the unqualified DC scheme [26] defined *Source* as 'a related resource from which the described resource is derived'; on the other hand, according to the qualified DC scheme [27], the refinement for the element *Relation* (is Version of), is 'the described resource is a version, edition, or adaptation of the referenced resource'. These definitions render *Source* as a particular type of *Relation*.

As stated, the *Relation* field seems to pose a particularly high degree of difficulty (55.3%) to survey participants. In addition to the semantic overlaps, difficulty in applying the DC element *Relation* also seems to be derived from the complex nature of the element itself which entails relationships between information objects (documents) and between an information object and its context. In other words, discernment of the dynamic and interrelated nature of information objects presents challenges in using the *Relation* element.

Some comments demonstrate this issue:

> Relation is relatively difficult, because it requires identification of formal and informal relationships to other resources. Considering the diversity of metadata creators and of course, the size of our digital resources, we may not be able to identify all possible relationships accurately.

Discernment of the connection between the individual objects and their parent collection(s) also brings forth particular challenges in applying *Relation*. Survey participants' comments evidence this issue as well: 'Determining the formal identification system to use for this field has been difficult, especially since we have varied collections and no single numbering or classification system'.

4.2.3. Broadness and simplicity

One of the most salient characteristics of the DC metadata scheme concerns its simplicity: 15 elements constitute the simple and unqualified DC element set. The DC metadata scheme was developed primarily for one important purpose – to facilitate cross-domain resource description and discovery. This inherent simplicity of the DC scheme has strengths as well as weaknesses. Simplicity facilitates adoption across communities; as stated earlier, the DC standard is one of the most widely used metadata standards. Since its inception in 1995, the DC metadata scheme has gained formidable credibility and was approved as the NISO standard [2] in 2001. However, compared to other metadata schemas, the DC lacks semantic richness because of simplicity geared toward cross-domain resource description.

Although there are advantages in such simplicity, other issues may cause difficulty in using DC metadata standards. Some survey participants ($n = 8$, 11%) reported difficulty in applying DC metadata elements owing to simplicity and broadness: 'Not rich enough relationships to reflect real depth of material'. One participant raised an issue of semantic loss during crosswalking from other schemas to DC: 'difficulty arises more from the simplicity of Dublin Core. How does one cram specific metadata from other schemas into the broad/simplified elements of Dublin Core'.

It is noteworthy that the simplicity of the DC metadata scheme in terms of semantics may also engender inconsistency of the application of the standards across institutions. Survey participants raise this issue directly, as in the following comments: 'The categories are generic and need project-specific interpretation';

> The fields are very broad and rigorous, local, content standards must be created to ensure any kind of consistency. In addition, although the content standards we create make us internally consistent we may not be consistent with other institutions.

4.2.4. One-to-one principle

The Dublin Core metadata schema is intended to be used in accordance with the one-to-one principle; that is, one metadata description for one resource type (e.g. physical painting: digital image of the original painting). However, in practice, there seems to be a great deal of confusion in applying this principle: 'I often wrestle with whether to indicate the digitizer or the original creator; many people mix contributor with institution who contributed the digital collection which should be digital publisher'.

Some survey participants ($n = 4$, 6%) report that difficulty in applying the DC elements stems from the lack of provision of how to apply the one-to-one principle. The following comment is illustrative:

> The problem with DC aside from its ambiguity is the failure of the 1:1 principle. Inevitably implementors use a single record to describe both the original (what users are really interested in) and the digital. DC is still really only for digital manifestations. VRA has addressed this with a more complex hierarchical schema. But that's not necessary. What's needed in DC is a simple subset of elements clearly designated for the digital manifestation with the primary element set reserved for the original object.

As discussed in the literature review section, there is a growing body of literature dealing with DC metadata semantics. However, there is a lack of studies focusing on how the one-to-one principle is reflected in the metadata creation process among cataloguing and metadata professionals as well as in actual DC metadata item records.

Lastly, the 'Other' category ($n = 7$, 10%) encompasses various other issues such as choosing appropriate controlled vocabulary schemes, lack of information for creating metadata, familiarity with DC (e.g. 'just not familiar enough with Dublin Core') and local policies.

5. Conclusion

In this study, we examined DC metadata semantics drawn from the perspectives and experiences of cataloguing and metadata professionals. Through a web-survey method, the study ascertained the extent of difficulty in applying DC metadata elements encountered by cataloguing and metadata professionals and examined factors engendering such difficulties during the metadata creation process. Approximately half ($n = 141$, 49.8%) of the survey participants specified that the DC metadata standard has been utilized for metadata creation and management in their libraries.

Comments from the survey participants show that semantic overlaps (45%) engender confusion and difficulty in applying DC metadata elements accurately and consistently. Comments show that that extensive documentation is necessary to disambiguate such semantic overlaps among the surveyed DC metadata elements in metadata creation practice. The comments of survey participants also show that conceptual ambiguities of the surveyed DC metadata elements is the second most frequently cited factor (41%) causing difficulty and confusion in applying the DC metadata standard. Moreover, such conceptual ambiguities lead to variant interpretations of DC metadata element use. This, in turn, has the potential to bring forth inconsistent and inaccurate applications and implementation of the DC standard across institutions. This may directly affect semantic interoperability across digital repositories resulting in hindrances in resource discovery.

It is noteworthy that the experiences of survey participants on conceptual ambiguity and semantic overlaps of the surveyed DC metadata elements are consonant with the findings of studies reporting issues in relation to DC metadata semantics (see Section 2). These studies point out that semantic overlaps and conceptual ambiguity among certain DC metadata element names and corresponding definitions hinder accurate and consistent application of the DC metadata scheme [5–7].

The major limitation of this study derives from the participant population. Although the 10 electronic mailing lists used for recruiting survey participants constitute the major communication channels for the communities of metadata and cataloguing professionals (see Table 1), there exist other sources for recruiting the survey participants in the field. Thus, the findings of the study need to be further validated by utilizing a larger pool of data sources and participants.

Analysis of the comments of survey participants foregrounds the major issues in DC metadata semantics. The conceptual ambiguities and semantic overlaps inherent in some DC metadata elements need to be re-examined in order to disambiguate semantic relations of some DC metadata elements that engender confusion leading to variant interpretations. The high degree of difficulty (55.3%) caused by the *Relation* field indicates that further study of this element is needed in terms of a mechanism for discerning and representing the dynamic inter-relationships between information objects and between an information object and its context.

As shown by conceptual ambiguities and semantic overlaps, difficulties in application of the *Relation* element also seem to be derived from the poorly defined element itself. Accordingly, for its full application extensive local documentation and a rather steep learning curve for gaining familiarity with the element seem to be necessary. This highlights the importance of examining locally developed documentation encompassing metadata application profiles, best practices and guidelines [28]. The comparison of the usage of DC elements across locally developed metadata best practices and guidelines may further reveal current practice of DC metadata application.

As stated, there is a lack of studies examining the way the one-to-one principle is reflected in the metadata creation process as well as in DC metadata item records. The confusion in applying the one-to-one principle of the DC metadata standard indicates that there is a need for in-depth investigation on a mechanism for applying the one-to-one principle in cataloguing and metadata practice. The results of the study also foreground the importance of training and educating cataloguing and metadata professionals as well as LIS students in the areas of metadata semantics and quality metadata generation.

Acknowledgements

References

[1] R. Heery, Metadata futures: steps toward semantic interoperability. In: D.I. Hillman and E.L. Westbrooks (eds), *Metadata in Practice* (American Library Association, Chicago, 2004) 257–71.

[2] National Information Standards Organization, *ANSI/NISO Z39.85 – The Dublin Core Metadata Element Set* (2007). Available at: www.niso.org/kst/reports/standards/kfile_download?id%3Austring%3Aiso-8859-1=Z39-85-2007.pdf&pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsJDs4R-jLz0t90_d5_ymGsj_IKVa86hjP37r_hFEijh12LhLqJw52B-5udAaMy22WJJl0y5GhhtjwcI3V (accessed 20 February 2009).

[3] International Organization for Standardization, *Information and Documentation – The Dublin Core Metadata Element set ISO 15836* (2009). Available at: www.iso.org/iso/search.htm?qt=15836&searchSubmit=Search&sort=rel&type=simple&published=on (accessed 15 April 2009).

[4] J.R. Park, Language-related open archives: impact on scholarly communities and academic librarianship, *E-JASL: The Electronic Journal of Academic and Special Librarianship* 5(2–3) http://southernlibrarianship.icaap.org/content/v05n02/park_j01.html (accessed 15 April 2009).

[5] J. R. Park, Semantic interoperability and metadata quality: an analysis of metadata item records of digital image collections, *Knowledge Organization* 33(1) (2006) 20–34.

[6] P. Caplan, *Metadata Fundamentals for All Librarians* (American Library Association, Chicago, 2003).

[7] C.J. Godby, D. Smith and E. Childress, Two paths to interoperable metadata. In: *DC-2003: Supporting Communities of Discourse and Practice – Metadata Research and Applications* (28 September–2 October 2003, Seattle, WA). Available at: www.oclc.org/research/publications/archive/2003/godby-dc2003.pdf (accessed 20 February 2009).

[8]   M. Khoo, J.R. Park and X. Lin, The user-centered design of a non-specialist metadata tool and interface for the Internet Public Library, *Proceedings of the American Society for Information Science and Technology* (2009).

[9]   J.R. Park, Metadata quality in digital repositories: a survey of the current state of the art, *Cataloging and Classification Quarterly* 47(3) (2009) 213–28.

[10] M. Guy, A. Powell and M. Day, Improving the quality of metadata in e-print archives, *Ariadne* 38 (2004), www.ariadne.ac.uk/issue38/guy/ (accessed 15 April 2009).

[11] National Information Standards Organization, *A Framework of Guidance for Building Good Digital Collections* (NISO Press, Bethesda, MD, 2007) 61–2.

[12] Y.G. Lei, M. Sabou, V. Lopez, J.H. Zhu, V. Uren and E. Motta, An infrastructure for acquiring high quality semantic metadata. In: J. Dominque (ed.). *Semantic Web: Research and Applications* (Springer, Berlin, 2006) 230–44.

[13] L.C. Howarth, Designing a metadata-enabled namespace for enhancing resource discovery in knowledge bases. In: S. Guerrini and L. Sardo (eds), *Proceedings International Conference Electronic Resources: Definition, Selection and Cataloguing* (2001). Available at: http://w3.uniroma1.it/ssab/er/relazioni/howarth_eng.pdf (accessed 20 February 2009).

[14] L.C. Howarth, Designing a common namespace for searching metadata-enabled knowledge repositories: an international perspective, *Cataloging and Classification Quarterly* 37(1–2) (2003) 173–85.

[15] J.R. Park, Semantic mapping of Dublin Core metadata element names and their definitions (Unpublished manuscript, 2006).

[16] M. Zeng, Metadata quality study for the National Science Digital Library (NSDL) metadata repository, Presented paper at the *Research and Teaching Talk Series* (Information Science and Technology, Drexel University, 2006)

[17] Y. Bui and J.R. Park, An assessment of metadata quality: a case study of the National Science Digital Library metadata repository. In: H. Moukdad (ed.), *Information Science Revisited: Approaches to Innovation*, CAIS/ACSI Proceedings of the 2006 annual conference of the Canadian Association for Information Science (Toronto, Ontario, 2006). Available at: www.cais-acsi.ca/proceedings/2006/bui_2006.pdf (accessed 20 February 2009).

[18] L. Sokvitne, An evaluation of the effectiveness of current Dublin Core metadata for retrieval (Victorian Association for Library Automation, 2000). Available at: www.vala.org.au/vala2000/2000pdf/Sokvitne.PDF (accessed 20 February 2009).

[19] K.J. Hegg and A.R. Knab, Using Dublin Core to facilitate cross-collection searches in an enterprise image repository. In: *Dublin Core Conference: Supporting Communities of Discourse and Practice–Metadata Research and Applications* (Information Institute of Syracuse, Syracuse, NY, 2003). Available at: http://dc2003.ischool.washington.edu/Archive-03/03hegg.pdf (accessed 20 February 2009).

[20] J. Heflin and J. Hendler, Semantic interoperability on the Web. In: *Proceedings of Extreme Markup Languages* (Graphic Communications Association, 2000). Available at: www.cs.umd.edu/projects/plus/SHOE/pubs/extreme2000.pdf (accessed 20 February 2009).

[21] J. Barton, S. Currier and J.M.N. Hey, Building quality assurance into metadata creation: an analysis based on the learning objects and e-prints communities of practice. In: *DC-2003: Supporting Communities of Discourse and Practice–Metadata Research and Applications* (28 September–2 October 2003, Seattle, WA). Available at: http://purl.oclc.org/dc2003/03barton.pdf (accessed 20 February 2009).

[22] J.R. Park, C. Lu and L. Marion, Cataloging professionals in digital environment: a content analysis of job descriptions, *Journal of American Society of Information Science and Technology* (Published online, 8 January 2009).

[23] J.R. Park and C. Lu, Metadata professionals: roles and competencies as reflected in job announcements, 2003–2006, *Cataloging and Classification Quarterly* 47(2) (2009) 145–60.

[24] J.R. Park, Hindrances in semantic mapping among metadata schemes: a linguistic perspective, *Journal of Internet Cataloging*, 5(3) (2002) 59–79.

[25] Dublin Core Metadata Initiative, *Using Agent Roles in Dublin Core* (2005). Available at: http://dublincore.org/documents/usageguide/appendix_roles.shtml (accessed 20 February 2009).

[26] Dublin Core Metadata Initiative, *DCMI Metadata Terms, Section 3: Properties in the legacy/elements/1.1/namespace* (2008). Available at: http://dublincore.org/documents/dcmi-terms/ (accessed 20 February 2009).

[27] Dublin Core Metadata Initiative, *Using Dublin Core – Dublin Core Qualifiers*, (2005). Available at: http://dublincore.org/documents/usageguide/qualifiers.shtml (accessed 20 February 2009).

[28] J.R. Park (ed.), Special issue on metadata best practices, *Journal of Library Metadata* (forthcoming).