Assessing Skill Building in Metadata Instruction: Quality Evaluation of Dublin Core Metadata Records Created by Graduate Students

Oksana L. Zavalina and Mary Burke

The quality of metadata has a direct effect on the ability to find, identify, select, obtain, and explore information; therefore, training to create high-quality metadata is an important, yet challenging, task for LIS educators. To provide such training effectively, instructors need to develop a vision for where best to focus their efforts, informed by understanding of the common metadata quality problems in student-created records and the way in which observed patterns correlate with the content and methods of instruction. Our exploratory case study addresses this need through empirical data. The article provides an overview of the metadata-creation skill building in instructional design of an online graduate metadata course at a US university and includes results of a comparative content analysis of student-created Dublin Core metadata records in three semesters based on the criteria of accuracy, completeness, and consistency. Findings reveal that accuracy and completeness issues are prevalent and that metadata quality problems revolve mainly around representation of relations and subjects and the application of controlled vocabularies. Results also indicate student confusion regarding the purpose of several metadata elements. The article discusses the findings in relation to content and methods of instruction, considers how the observed quality issues in student-created metadata might be addressed in improving skill building through curriculum development, and proposes directions for further research.

Keywords: curriculum, Dublin Core, evaluation, metadata education, metadata quality, online learning, skill building

Introduction and relevant literature

Information discovery relies on metadata that represents information objects in databases. The library and information science (LIS) community has traditionally been involved in the development and application of standards for the organization of information that facilitate information discovery. This includes formulating requirements for early handwritten or printed metadata records in library catalogs, such as Antonio Panizzi's (1841) rules, Charles Cutter's (1904) objectives, and the development of the first machine-readable standard for library metadata records, MARC, widely used since the 1970s. This community also models the functional requirements for metadata records that represent information objects (bibliographic records) and their creators, subjects, and so forth (authority records) in the Functional Requirements for Bibliographic Records (FRBR), Functional Requirements for Authority Data (FRAD), Functional Requirements for Subject Authority Data (FRSAD), Library Reference Model (LRM), and Bibliographic Framework (BIBFRAME) conceptual models. This involves creation of detailed rules for representation of information objects in metadata based on these conceptual models (e.g., Resource Description and Access). The library community has also developed and implemented multiple knowledge organization systems enabling efficient discovery of information

KEY POINTS:

- To support information discovery and use, we need to evaluate the effectiveness of learning in metadata courses and reinforce metadata curriculum. These efforts should be informed by understanding of the common metadata quality problems in student-created records and the way in which observed patterns correlate with the content and methods of instruction, particularly in an online environment as more and more metadata courses are taught online.
- Information access and use are negatively affected by the observed persistent errors in student-created metadata: use of non-standardized forms of names and terms, misrepresenting relations between information objects, omitting fields that represent rights-related information.
- The efficiency of introductory metadata curriculum would benefit from further extension of experiential learning through student-led metadata evaluation based on defined quality criteria and learning about the specific ways in which metadata quality issues reduce the functionality of metadata in supporting user tasks.

by allowing databases to group metadata records that represent similar information objects: Dewey Decimal Classification, Universal Decimal Classification, Library of Congress Classification, Library of Congress Subject Headings, Faceted Application of Subject Terminology, and others.

Further, LIS education programs have traditionally provided training in theory, practice, and tools for organizing information. Information organization education has progressed through different stages, the "pitfalls and pendulum" (Hill, 2002), from its temporary de-emphasizing in the 1980s and 1990s (when many believed that improvements in information access afforded by emerging online catalogs and search engines would significantly reduce the need for information organization efforts) to the broad realization of its central role in LIS training (e.g., Gorman, 2002; Intner, 2002). Information organization is recognized as one of the key competencies for the graduates of LIS programs in Australia, Canada, New Zealand, and the United States[1] and is used as a criterion in the accreditation of such programs. For example, organization of recorded knowledge and information is one of the eight key competencies defined by the American Library Association and includes

- The principles involved in the organization and representation of recorded knowledge and information resources.
- The developmental, descriptive, and evaluative skills needed to organize recorded knowledge and information resources.
- The systems of cataloging, metadata, indexing, and classification standards, and methods used to organize recorded knowledge and information.

(ALA, 2009, pp. 2–3)

In the modern knowledge-based economy, the demand for highly qualified specialists in any area is rapidly growing (e.g., OECD, 2013). The nature of the work performed by these specialists and its knowledge base are ever changing in response to market and technological developments. This presents the need both for increased flexibility and adaptability and to keep learning to maintain qualifications (Bates, 2015). Arguably, all of these

characteristics apply to metadata specialists, as the landscape of metadata work has changed dramatically over the past two decades—for example, with the development of FRAD, FRSAD, LRM, BIBFRAME, the switch from Anglo-American Cataloging Rules (AACR) to Resource Description and Access (RDA), the proliferation of Linked Library Data—and continues to rapidly evolve. Analyses of job advertisements for metadata specialists (e.g., Han & Hswe, 2010) show that ability, willingness to learn, and flexibility are among the most frequently required traits, with an upward trend.

As a result of this shift to a knowledge-based economy, one of the two integral components of knowledge—skills—needs to receive more emphasis in designing the educational programs at any level than previously when education was more content focused (Bates, 2015). Research reveals that, unlike transmission of content, the effectiveness of skill transmission depends on significant amounts of practice. In addition, the appropriateness of teaching methods and technologies plays a crucial role in skill development (e.g., Fallows & Steven, 2000; Fischer, 1980).

Mitchell's (2009) framework defined "metadata literacy" as the "ability to conceptualize, create and work with metadata within information environments" (p. 63), consisting of three categories: skills, concepts, and contexts. Metadata skills are defined by Mitchell as "skills specific to metadata (such as recognize context, harvest, transform, archive) that are descriptive of the tasks that are required to use electronic metadata-rich documents" (p. 90). Professional associations such as the Association for Library Collections and Technical Services (ALCTS) publish information on the skill sets for metadata professionals.

Metadata-related skills have been studied by multiple researchers. The employer perspective has been examined by studies analyzing job descriptions (Dieckman, 2018; Hall-Ellis, 2006, 2015; Han & Hswe, 2010; Park & Lu, 2009; Turner, 2020). Skill-identifying surveys of metadata practitioners and educators have been conducted (e.g., Hider, 2006; Hsieh-Yee, 2004). Several studies have examined metadata-related course offerings in North America, Europe, and Asia at the snapshot level (e.g., Alajmi & ur Rehman, 2016; Bhakti, 2019; Davis, 2008; Joudrey, 2002, 2008; Joudrey & McGinnis, 2014). These evaluations included, for example, numbers and types of offered courses, course-offering frequency, whether metadata-related courses were required core courses or electives, and major topics mentioned in the syllabi. Due to their broader focus, these studies did not provide details on the content of metadata-related courses: specific metadata schemes and standards covered, assignments, or other evaluations. Several case studies focusing on how specific metadata skills are developed through curricula, activities, and assessments in metadata courses are also available (e.g., Glaviano, 2000, Hsieh-Yee, 2000; Or-Bach, 2005). One more recent study by Maron and Hill (2015) compared how metadata is being taught with the skills in job advertisements for metadata positions; however, detailed results have not yet been reported.

The quality of metadata directly affects users' ability to find, identify, select, and obtain information, as well as to explore relations between information objects of interest. Therefore, the training of metadata specialists to create high-quality metadata is an important task. Existing research reveals that perceived priorities in metadata education include metadata quality (Engelson, 2019; Park & Tosaka, 2010; Park, Tosaka, Maszaros, & Lu, 2010). Despite this acknowledgment of the importance of metadata quality coverage in metadata

education, this priority is not yet sufficiently reflected in the outline of core competencies by the ALCTS. This 10-page document presents numerous competencies under three categories: knowledge, skills and ability, and behavior. The metadata quality is currently included only under the "knowledge" category (as an example) but not under the "skills" or "ability" categories: "Knowledge of trends in cataloging and metadata profession: . . . ('Big Picture' Knowledge) Examples: . . . the impact of quality vs. lack of/poor metadata on user access to resources" (ALCTS, 2017, p. 4).

There are no reports about how existing courses and programs in LIS education address metadata quality, apart from two publications. Choi (2020) discussed metadata quality competencies in the context of the curriculum development of the Cultural Heritage Information Management specialization for librarians and museum workers. A brief report focused on the metadata quality skill building in LIS courses as the topic of growing importance (Zavalina, 2017). No studies so far have evaluated the effectiveness of metadata quality training.

As LIS departments and colleges (since 2005 organized in the iSchools movement) have been at the forefront of developing and offering online courses and programs since the late 1990s, some of them have accumulated significant experience in online graduate metadata education. This longstanding demand for online course offerings only intensified in 2020 and 2021 due to the adjustments necessary for the COVID-19 pandemic. Therefore, it is particularly timely to collect, analyze, and report empirical data to support online course development based on experience and lessons learned by long-term online instructors.

Our exploratory case study begins to address these needs through the examination of curriculum development to support efficient online learning of the important group of skills related to metadata quality. Moreover, the analysis of empirical data conducted as part of this study allows evaluating the effectiveness of learning and identifying the areas in need of reinforcement in metadata curricula. In this article, we provide an overview of the metadata-creation skill-building content of an online graduate metadata course at the University of North Texas, report results of the quality evaluation of student-created metadata records, and discuss possible solutions to improve metadata teaching and learning. We expect our study to support the transformation of LIS metadata education in the changing, interconnected world.

High-quality metadata-creation skill building in a graduate metadata course

The graduate online course "Introduction to Metadata" is a guided elective taken by students after completing the core course in information organization. Learning outcomes include understanding the purpose and building blocks of a metadata scheme, data content and data value standards (including controlled vocabularies), syntaxes for metadata encoding, and learning (both in theory and in practice) major standard item-level and collection-level metadata schemes. Students gain practical experience representing textual and non-textual information objects through two major assignments: the Portal to Texas History Exercise and the Creating Metadata Records Project.

The Portal to Texas History Exercise is a real-life project in which students create item-level metadata records using online submission forms in the University Metadata

Editor Tool. After obtaining this practical experience and learning about standard metadata schemes, students work on the Creating Metadata Records Project, which consists of four components. Students create item-level metadata records individually and collection-level metadata records in teams using XML and HTML templates. This allows students to practice creation of DCTERMS, MODS, and VRA Core 4.0 metadata records and obtain metadata-syntax–related skills, an opportunity that is not available when using online metadata submission forms in the Portal to Texas History Exercise, which relies on a locally developed metadata application profile.

The course is organized into eight learning modules, with each module building on the previous ones:

1. introduction to metadata, information organization, and retrieval
2. components of a metadata scheme
3. data content and data value standards
4. syntax for encoding metadata
5. Dublin Core item-level metadata
6. MODS item-level metadata
7. VRA Core 4.0 item-level metadata
8. collection-level metadata.

There is currently no designated learning module focusing on metadata quality. However, students' attention is brought to metadata quality regularly throughout the course. For example, as part of learning module 3, the instructor emphasizes how the metadata quality is improved by using authorized terms from controlled vocabularies. In learning module 4, the instructor highlights metadata quality problems that can be caused by incorrect application of a metadata syntax (e.g., issues with well-formedness and validity of XML documents). As part of the introduction to the first major standard metadata scheme, the instructor presents common quality problems found in available evaluations of Dublin Core metadata records in digital libraries and repositories (e.g., as summarized by Jackson, Han, Groetsch, Mustafoff, & Cole, 2008), demonstrates examples of such problems in DCTERMS records (along with suggested corrections), and encourages students to avoid these metadata quality problems. The instructor also explains the major quality criteria in the grading rubric for the three Creating Metadata Records Project submissions.

Learning modules 5–7, covering standard item-level metadata, include the following required readings:

1. instructor-developed lecture document containing
   a. the metadata scheme development history outline
   b. the metadata scheme design principles (e.g., Dublin Core principles)
   c. the metadata scheme structure (element set, cardinality, order of elements in metadata record, attributes and attribute values)
   d. definition and uses for each metadata element
   e. recommended controlled vocabularies (e.g., MARC Relator Code and Term list for MODS Role element, AAT for VRA Core 4.0 Cultural Context metadata

element, etc.) and data encoding standards (e.g., URI for DCTERMS Identifier metadata element, W3CDTF for MODS Date element, etc.)
   f. an illustrative example of metadata record creation with in-depth explanation for each field;
2. external readings, including the metadata standard itself, and the official usage guide for it (if separate from the metadata standard, as in the case of Dublin Core);
3. a two-hour live presentation delivered by the instructor in synchronous online class meeting(s), with PowerPoint slide set and Zoom archive (recording, transcription, and text chat log), including
   a. summary of lecture document and external readings
   b. detailed walkthrough of creating two or more example metadata records (different from the example in the lecture document)
   c. explanation of the Creating Metadata Records Project requirements
   d. question and answer session for clarification.

In the Creating Metadata Records Project, students implement understanding and skills developed in all eight learning modules. For example, in its first deliverable, which focuses on DCTERMS item-level metadata, students apply in practice what they learned in modules 1–5, as well as in the two preceding assignments: Portal to Texas History Exercise and the Metadata Research Presentation. Likewise, in the last component of the Creating Metadata Records Project—team creation and presentation of collection-level metadata—students implement what they learned in Modules 1–8, as well as in all previously completed practical exercises. Students are assigned two items each to represent with item-level metadata records in DCTERMS, MODS 3.7, and VRA Core 4.0 across three parts of the project. Each student creates records for both textual (academic writings and websites on metadata topics) and non-textual objects (paintings in gallery collections). The final component of the Creating Metadata Records Project is the team creation of two Dublin Core collection-level records to represent the team's collections of textual objects and paintings.

As part of the Creating Metadata Records Project, each report containing student-created metadata records is submitted separately and graded by a Teaching Assistant (TA), an LIS PhD student with metadata expertise obtained through coursework and research projects. The instructor-provided rubric includes criteria for assessing records—accuracy, completeness, and consistency (based, e.g., on Bruce & Hillmann, 2004; Moen, Stewart, & McClure, 1998)—and the associated grade points. Using the rubric, the TA grades and annotates submissions with corrections and comments on metadata quality issues observed in the records. Individualized feedback—annotated submission with comments and corrections—is returned to each student through the course website. In addition, students receive general feedback: the instructor presents the common mistakes in student-created records during the following synchronous class meeting. Students are encouraged to ask questions on the individualized and general feedback and refer to it in future work.

Although the degree to which each student-created metadata record meets the expectations and demonstrates the student learning outcome development is routinely assessed through evaluation of submissions, the systematic comparative analysis of the quality of

student-created metadata across a class of students and across semesters has not been done before. We believe such an analysis will contribute to building a better understanding of the areas of strength and weakness in the students' skills, and that the results will be of interest to metadata educators at other institutions.

Study design

Our study explored the following research questions:

1. What metadata quality issues related to the major criteria of accuracy, completeness, and consistency are found in student-created metadata records, and how are these issues distributed in the data set?
2. How does the quality of student-created metadata records compare with the quality of metadata records reported by previous studies?
3. What is the relationship between the observed quality issues in student-created metadata records and the level and content of instruction on metadata quality?

Qualitative and quantitative content analysis was performed on the student-created records from three semesters—spring, summer, and fall of 2019—during which the Creating Metadata Records Project instructions and course materials for the relevant learning modules (instructor presentations in synchronous class meetings and lecture documents) remained consistent. In these semesters, the course was taught by the same instructor, which minimizes possible variation due to different teaching styles. Metadata records that represent textual objects (journal articles, book chapters, conference papers, standards, websites, etc.) were selected for analysis due to the potentially broader applicability of metadata quality findings, as opposed to the much more specialized metadata representing artwork. The focus on DCTERMS metadata makes results of our study more broadly useful for developers and instructors of metadata courses, as Dublin Core is most widely used in digital libraries and archives worldwide, its application was examined before, and it is commonly taught in metadata courses.

Three major criteria of metadata quality—accuracy, completeness, and consistency—were used in this study. Accuracy refers to the correctness of data values, including standardized spelling (Bruce & Hillman, 2004, p. 5). Accuracy problems were operationalized for this study as the use of data values and metadata elements not applicable to the item, the misuse of elements (e.g., Spatial used to represent the place of publication, or Date used when Issued is more appropriate), typographical errors, the combination of multiple values in a single element instance, and incorrect application of the xsi:type XML attribute to represent the controlled vocabulary. Based on Bruce and Hillman's (2004) definition of metadata completeness, our operational definition of completeness problems included the omission of elements that are applicable to the item, missing additional applicable instances of an element, or incomplete data values. Consistency, termed "logical consistency and coherence" by Bruce and Hillman (p. 7), is defined as the adherence to data value standards. Consistency issues were operationalized by us as failure to draw data values from controlled vocabularies (name authority files, subject headings, etc.).

The analysis focused on 38 DCTERMS metadata elements. Eleven additional DCTERMS metadata elements were excluded from analysis for the following reasons:

- Medium: not applicable to textual objects in general
- Source, Provenance, Conforms To, Requires, Is Required By: not applicable to any of the assigned textual information objects
- generic Date, Coverage, and Relation: should not be used when the more specific element applies (e.g., Issued for Date, Is Part Of for Relation, Spatial or Temporal for Coverage)
- References, and Is Referenced By: not feasible to include in records representing conference papers and journal articles, which tend to reference and be referenced by dozens of other works.

We collected for analysis all 73 DCTERMS metadata records created by 72 students enrolled in the University of North Texas metadata course in the spring, summer, and fall semesters of 2019. The records represented 37 unique textual information objects. One metadata record submitted by student in another metadata format (MODS) was excluded from analysis, so the remaining 72 records represented 36 information objects. Eleven objects had three student-created records, 14 objects had two student-created records, and 11 objects had one student-created record. This represented the entire population of DCTERMS metadata records for textual objects that were created by our students during 2019.

The binary coding approach was applied: a code 0 was used if a metadata field did not contain any quality problems; a code 1 was selected if there was one or more errors in a metadata field. To accommodate situations when a metadata field was applicable to some of the information objects but not to others, metadata fields not applicable to the object in question were marked with the code "n/a." Descriptive statistics indicators were measured for the overall level of metadata quality problems in student-created records, and separately for accuracy, completeness, and consistency errors, as well as individually for 38 DCTERMS metadata elements.

Study findings

Table 1 reports the overall statistics for metadata quality in student-created records. Accuracy issues were found in 3.76 metadata fields per record on average (*Mdn* = 3), completeness issues in 3.05 fields on average (*Mdn* = 3), and consistency issues in 1.64 fields on average (*Mdn* = 1). The highest maximum number of fields containing accuracy errors per student record was 26, followed by completeness (*n* = 19). Student-created records included errors in 7.84 metadata fields on average (*Mdn* = 8), and the lowest-quality record included errors in a total of 27 fields out of 38 analyzed. For the accuracy category, high variability was observed (*SD* = 4.098).

Figure 1 shows a typical example of a student-created metadata record. This record has eight metadata fields with quality issues, which is the median observed in the dataset. Five fields included in this record contained errors, while three additional applicable fields were missing. The Creator field exhibited both a completeness issue (missing name of the third creator) and a consistency issue (not using the authorized form of the second creator's name

Table 1: Number of fields with metadata quality issues per student-created record (*n* = 73)

	Range	*M*	*Mdn*	*SD*
Accuracy	0–26	3.76	3	4.10
Completeness	0–10	3.05	3	2.06
Consistency	0–19	1.64	1	2.48
Total	0–27	7.84	8	0.43

Figure 1: Example of a student-created metadata record showing metadata quality issues

from the Library of Congress Name Authority File controlled vocabulary). In this record example, five fields with quality problems contained completeness deficiencies.

Figure 2 orders DCTERMS metadata elements based on the overall rate of observed metadata quality issues in our data set of metadata records created by 72 students. As seen in Figure 2, metadata elements with the highest rate of errors (between 41.1% and 61.6% of student-created records) included Subject, Is Part Of, Rights Holder, Creator, Rights, and Issued. Fewer than 3% of student-created records included metadata quality problems in Has Part, Has Version, Is Format Of, Has Format, Replaces, and Is Replaced By, all of which

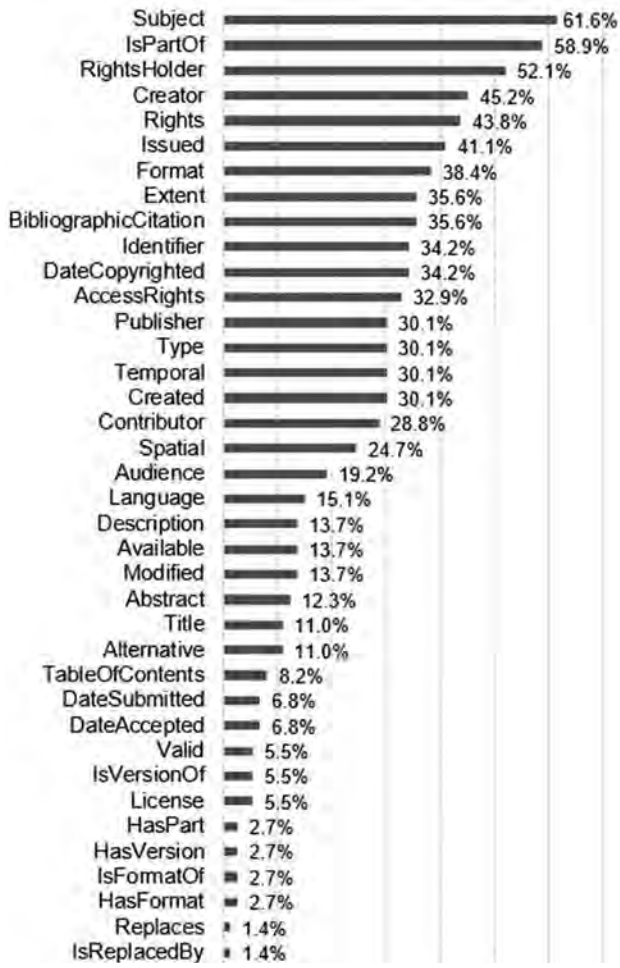

Figure 2: Percentage of student-created records with metadata quality issues by DCTERMS metadata element

belong to the Relation group of elements. Unlike Subject or Creator, the Relation elements are often not applicable to the information objects assigned to students. The quality issues in the metadata elements documented in Figure 2 represent situations where students included instances of one or more of these Relation elements when they were not applicable to an information object represented by the metadata record.

Figure 3 shows how accuracy, completeness, and consistency metadata quality problems were distributed in student-created records for each of the 38 DCTERMS metadata

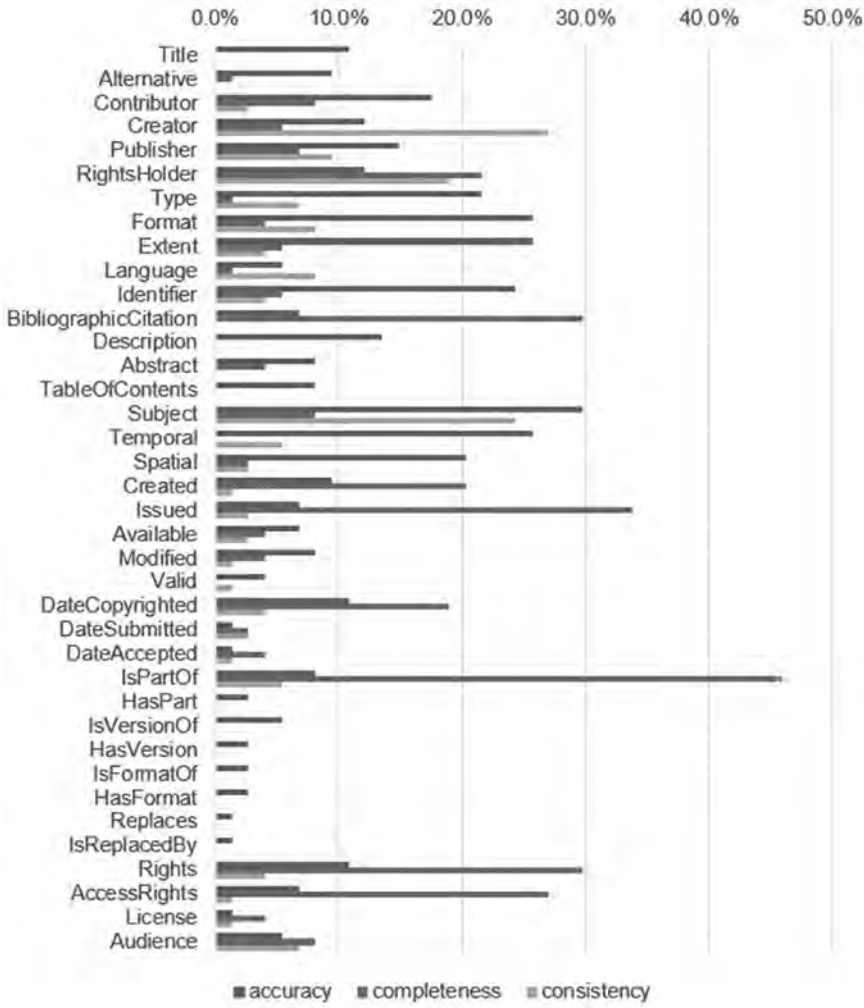

Figure 3: Accuracy, completeness, and consistency problems by DCTERMS metadata element in student-created records

elements examined in our study. In Figure 3, metadata elements are shown in the order in which they are presented in the official Dublin Core metadata creation guidelines.[2]

Accuracy issues were found to occur in each metadata element. The student-created record example in Figure 1 shows accuracy problems in the Alternative and Rights Holder fields. Six elements—Subject, Temporal, Format, Extent, Identifier, and Type—contained accuracy mistakes in the largest proportion of student-created records in our data set (between 21.9% for Type and 28.8% for Subject). Fewer accuracy issues were observed in the Date and Relation groups of elements. Still, students sometimes included these elements when they were not applicable, or with incorrect data values.

Completeness issues were found in a total of 26 fields out of 38 analyzed. They were prevalent in Is Part Of (46.6% of records), Issued (34.2%), Rights (30.1%), and Bibliographic Citation (28.8%). The student-created record example in Figure 1 illustrates this trend in Is Part Of and Bibliographic Citation fields. In our data set, 11 fields (Title, Description, Table of Contents, Temporal, Valid, Has Part, Is Version Of, Is Format Of, Has Format, Replaces, and Is Replaced By) did not contain completeness errors in any of the student records. In many cases, most of these metadata elements (except Title and Description) did not apply to objects being described. Is Part Of was the only one of the Relation group of elements that was applicable to most of the information objects represented by student records and was often missing; no completeness issues were observed for other Relation elements. In some cases, though, the element was included in the student record, but the data value was missing crucial information (e.g., only the tile of the journal in which the article appeared was included, but the volume and issue number were missing in Is Part Of).

Consistency issues were also observed in 26 DCTERMS metadata fields: the most often in Creator (27.4% of records), Subject (24.7%), and Rights Holder (19.2%) and significantly less frequently in other elements. The student-created record example in Figure 1 illustrates this trend in Creator and Subject fields. Overall, the metadata elements with consistency issues observed in student-created records were those that represent names of agents, as well as those that commonly draw data values from controlled lists (Subject, Language, Format, Type). Major quality issues in student-created metadata records are detailed in the Discussion section.

Discussion

Items represented by student-created metadata included electronic journal articles, conference papers, technical reports, and standards. Certain metadata fields are relevant only to some of these. For example, while journal articles, conference papers, and some technical reports have formal abstracts that can be copied into the Abstract field, the Description element is typically more applicable for representing standards, which lack a formal abstract. Similarly, standards more commonly contain a table of contents, while journal articles do not. More than one type of date is often included on the landing page of an electronic journal article: beyond the dates of publication/copyright, these include the dates of submission, revision, and acceptance. Likewise, dates of modification are often included in standards. Thus, some of the metadata elements in the Date group that do not apply for representing other objects should be included in records representing journal articles or standards:

Issued, Date Submitted, and Date Accepted, or Modified. Leaving these differences aside, our analysis revealed common problems with metadata quality observed in 72 student-created records across three semesters. These common problems are categorized below.

Semantic accuracy

The DCTERMS Coverage elements, Temporal and Spatial, were found to be commonly used by students to describe the time period and geographical location of publication rather than the subject matter. For instance, the term "North America" was applied by students to several items published in a US journal, though not focused on North America more than any other geographical location. The purpose of some other elements was also misinterpreted. Students used the Description element for a formal abstract, for example, either repeating the same value in the Abstract element or omitting the Abstract element completely. Similarly, the Table of Contents element was commonly applied to objects with no formal table of contents (e.g., journal articles). These issues in student-created metadata records in the Description, Abstract, and Table of Contents metadata fields are similar to those observed in the Dublin-Core–based metadata in the Digital Public Library of America by Tarver, Zavalina, and Phillips (2016).

Subject representation and authority control

The DCTERMS Subject field was found to contain errors in more than 60% of the student-created records overall, with a substantial proportion of errors in all three areas: accuracy, completeness, and consistency. Some subject terms were included in wrong fields (e.g., genre terms such as "Technical manual" that belong to the Type metadata element were included in the Subject element instead) or were too broad for the metadata-focused information objects that students were representing (e.g., "Federal government").

Subject terms applicable to the object were also often not taken from controlled vocabularies. For example, the data value "digital repositories" was used instead of "Digital libraries," which is the authorized Library of Congress Subject Headings (LCSH) term for digital repositories. This trend was also observed in the Type element. Students also struggled to use the authorized name forms in all fields with names of people and organizations: Creator, Rights Holder, and Publisher. When authorized name forms were used, students tended to omit the xsi:type XML attribute pointing to the controlled vocabulary referenced (as instructed in the course materials).

Subject representation and authority control have been known in the LIS field as crucial, yet very complex, topics (e.g., Aluri, Kemp, & Boll, 1991; Miksa, 1983; Pettee, 1946). That complexity presents challenges to teaching and learning relevant skills in these areas (e.g., Cabonero & Dolendo, 2013; Mugridge & Furniss, 2002; Snow & Hoffmann, 2015; Taylor & Joudrey, 2002). With these challenges in mind, and with the goal of developing the student learning outcomes in creating high-quality metadata records, special attention is given to subject analysis and representation in the University of North Texas introductory metadata course design. Two general learning modules (modules 2 and 3) provide substantial coverage of these topics through readings and in-class mini-exercises. In learning modules that focus on specific metadata schemes, including Dublin Core, required readings, instructor

presentations, examples of completed metadata records, and grading rubrics place emphasis on subject analysis and representation using controlled vocabularies. Course materials convey the importance of authority control, not only for subjects (topical, geographical) but also for agent names, as well as for data values in other metadata elements that play a significant role in information discovery. It is worth noting that all students enrolled in the University of North Texas introductory metadata course either completed in the previous semesters, or are taking concurrently, a core course in information organization, which provides practical training on subject analysis and representation as well as on authority control for names.

Representing relations

Metadata fields intended for representing relations were used differently depending on the type of textual object described by a student-created metadata record. A metadata record describing a standard is more likely to require the use of Replaces and/or Is Replaced By DCTERMS metadata elements as older versions of the standard are replaced by the newer versions and tend to be archived on the same website. Representation of a book chapter, journal article, or conference paper, on the other hand, requires the use of a metadata element that conveys a different kind of relation—Is Part Of—as these information objects are published in another larger information object: a book, a journal issue, or a volume of conference proceedings. Some other Relation metadata elements (e.g., Replaces, Is Replaced By, Has Part) do not apply to book chapters, journal papers, or conference papers. Student-created metadata records representing journal articles most frequently contained the Is Part Of metadata element. However, the data values in Is Part Of were often incomplete (e.g., provided the journal title but omitted the volume and issue number).

Students also often used the metadata elements in the Relation group of elements redundantly (e.g., included instances of both the generic Relation and the specific Is Part Of metadata elements with identical data values). Other Relation elements—Has Part, Is Version Of, Has Version, Is Format Of, Has Format, Replaces, and Is Replaced By—were frequently included in student-created metadata records when these elements were not applicable, all with the same data value. This reflected a lack of understanding of the purpose of the metadata elements in the Relation group.

Overall findings in comparison to existing studies of Dublin Core metadata records

Persistent errors that students tend to make negatively affect information access. For example, non-standardized forms of names and terms and misrepresenting relations between information objects prevent metadata from supporting the exploration user task. Omitting fields that represent rights-related information results in metadata that fails to provide necessary context for information use.

One might assume that, because students are relatively new to metadata creation and Dublin Core is the first standard metadata scheme they encounter in the Creating Metadata Records Project, growing familiarity with the assigned information objects and metadata quality feedback obtained from instructors would result in most students avoiding some of the typical metadata quality issues in the future. In our course, after creating DCTERMS records in module 5, students create Metadata Object Description Scheme (MODS) metadata

records for the same objects in module 6. Although systematic analysis of student-created MODS records was not part of this study, anecdotal evidence obtained through years of evaluating student submissions suggests these common types of metadata quality issues with representation of subjects, agents, and relations are also observed in the later stages in the course, when students develop more expertise. These observations align with findings of studies that examined digital repository metadata, much of which is generated by metadata creators who are more experienced than students in our course.

Completeness and accuracy problems were much more widespread than consistency problems in student-created metadata records. While we are not aware of any studies evaluating student-created metadata, several existing studies on the level of application of various metadata fields in Dublin Core metadata in digital libraries offer a point of comparison. Our analysis demonstrates that students often omit instances of widely applicable metadata fields representing dates in their records: DCTERMS metadata elements, such as Issued (missing in 34.2% of student-created records), Created (20.5%), and Date Copyrighted (19.2%). Comparable data from existing studies of Dublin Core metadata in digital libraries and repositories published in 2008 and 2010 shows substantially higher metadata completeness in relation to Date fields. For example, Jackson et al. (2008) observed systematic inclusion of the Dublin Core *Date* element (86%). Studies by Weagley, Gelches, and Park (2010) and Kurtz (2010) revealed an even higher level of application of the Dublin Core *Date* metadata element (96% and 100% of metadata records, respectively).

On the other hand, our analysis demonstrates a relatively high level of completeness of subject representation in student-created metadata when compared with findings of previous studies. A moderate proportion of student-created metadata records analyzed in our study missed the DCTERMS Subject field or included only one instance of it when more than one was applicable (8.2% of records). This finding is similar to Jackson et al.'s (2008) finding that the Subject field was missing in 6% of metadata records. By contrast, both Kurtz (2010) and Weagley et al. (2010) found that the Dublin Core Subject field was missing in 35% of records. Another applicable DCTERMS field used for subject representation, Spatial, was missing in our study in only 2.7% of student-created metadata records to which it was applicable. This metadata field was found to be included in 7%, 21%, and 51% of metadata records in Kurtz, Weagley et al., and Jackson et al., respectively.

Conclusion

The findings of our study empirically support the observations made by practitioners and educators in library cataloging and digital library metadata management. These existing works state that determining the aboutness of an information object is one of the most complex and intellectually challenging tasks in the process of creating metadata records in traditional or digital libraries (e.g., Aluri et al., 1991; Cabonero & Dolendo, 2013; Miksa, 1983; Pettee, 1946; Taylor & Joudrey, 2002). We also found that skills in applying controlled vocabularies for representing subjects and agents need improvement. Likewise, our findings support with empirical data the anecdotal evidence that representation of relationships among various information objects and other entities is conceptually difficult (especially the logical pairs of reciprocal relations between entities).

This emphasizes the need for focused attention on teaching how to analyze and represent aboutness, agents, and relationships, and using controlled vocabularies. We believe these topics should be introduced to students repeatedly at various levels, using a number of examples, and extensive practical exercises to reinforce the knowledge through learning-by-doing. We examined the metadata records created by students who have received this type of instruction in at least two graduate courses—the core course on information organization and the introductory metadata course—prior to creating metadata records analyzed in this study. Some of the student record creators also received this training in one or more library cataloging courses. However, as our analysis reveals, certain metadata fields in student-created records demonstrate prevailing problems: Subject, Creator, and Relation group of elements. Information in Subject and Creator metadata fields has traditionally been among the main ways (along with Title field data values) for users to find information. Quality in representing relationships gains more and more importance as it enables users to explore information, especially in the Linked Data environment.

We believe the following recommendations for curriculum development may address the identified gaps in skills of metadata students:

- Providing additional practice in assigning terms from controlled vocabularies for representation of:
  - aboutness (e.g., for DCTERMS Subject, Spatial, and Temporal metadata elements)
  - agents: personal names and organization names (e.g., for DCTERMS Creator and Contributor metadata elements)
- Providing additional practice using metadata element attributes and attribute values to indicate what controlled vocabulary the term or name form is from (e.g., xsi:type XML attribute for DCTERMS metadata, authority and authorityURI attributes for MODS metadata, refid attribute for VRA Core 4.0 metadata).

Another solution to these persistent metadata quality errors that impede information retrieval would be further raising the student awareness of metadata quality. Based on our study results, more emphasis needs to be placed on teaching the metadata quality criteria, evaluation of metadata quality against these criteria, and the specific ways in which metadata quality issues negatively affect the functionality of metadata in supporting user tasks of finding, identifying, selecting, obtaining, and exploring information, as defined by the Library Reference Model (IFLA, 2017). This could be achieved through a learning module focusing entirely on metadata quality, and an associated practical exercise in which students evaluate the quality of metadata records created by themselves and/or their peers. At present, in our program, this is done in the next course in the sequence. However, the advanced metadata course is taken by a much smaller number of students than the introductory one: 8–12 per academic year, as opposed to 70–90. This ratio is consistent with the comparative frequency of offering introductory and advanced cataloging and metadata courses across various universities (e.g., Davis, 2008; Hsieh-Yee, 2004; Joudrey, 2002; Velucci, 1997). To ensure that beginning metadata librarians are adequately prepared for creating high-quality metadata that fully supports the functions of metadata at providing access to information in our interconnected world, metadata quality education is needed in the introductory metadata courses taken by most students.

Limitations and future research

There is a need for studies that continue exploration of the quality of student-created metadata and how metadata courses' curricula can address the prevalent quality issues. While providing rich qualitative data, results of this small-scale case study may not be statistically generalizable beyond graduate metadata instruction at the University of North Texas. Future studies that analyze student-created metadata across metadata courses offered by different universities is needed. Such a study might also compare online synchronous courses with online asynchronous courses to ascertain their efficacy in the development of critical skills through metadata instruction.

A comparative analysis of multiple metadata records representing the same information object created by different students who received the same instruction would provide additional insight. While our data set included metadata records representing the same information object by different students (normally in two different semesters), the sample size did not allow for meaningful comparisons.

Also, future research might investigate how the similarity of examples used in instruction affects the quality of student-created metadata records. For example, if the instructor provides walk-through examples of DCTERMS record creation for a conference paper and a website, are the student-created records representing websites and conference papers of a better quality overall than those representing other types of information objects, such as journal articles, book chapters, and so on? This would allow us to assess more precisely the extent to which students benefit from lecture examples representing their assigned object type, and which method of delivery of record examples is more effective: the learning module lecture document that students read on their own (ours currently includes a walk-through example of a record creation for a website) or the live and recorded presentation (ours currently includes a walk-through record creation example for a conference paper).

In recent years, the Dublin Core Metadata Initiative community had been finalizing the new Dublin Core User Guide (Rühle, Baker, & Johnston, n.d.), which was officially released on the DCMI website in the summer of 2019. The new user guide (first the draft version available on GitHub, then the final official version) informed the guidelines for student-created DCTERMS records in the course instructor presentations for all three semesters covered by this study. For the period when neither the draft nor the final version of the new user guide was available online (in spring and summer 2019 semesters), students in our metadata class were instructed to consult the older official user guide available on DCMI website (Hillmann, 2005). In the fall 2019 semester, students were instructed to consult the new official user guide. Our study did not examine similarities and differences between semesters in student metadata quality based on the official user guide version used. Future research will address this limitation.

Another promising extension to the current study would be a comparative analysis of metadata quality in student-created metadata using different metadata schemes. For example, a future study may compare records that students create for the same information objects in three metadata schemes: DCTERMS, MODS, and VRA Core 4.0. This would allow us to evaluate the extent to which the teaching team's feedback affects the metadata quality in subsequently created records, and the extent to which the varying level of detail

in official user guides for different metadata standards—concise for DCTERMS and much more detailed for MODS and VRA Core 4.0—affects metadata quality.

Quality evaluation of student-created metadata for non-textual information objects and comparison of findings to those obtained for metadata representing textual information objects is also needed. This will allow for identifying specific areas of attention for metadata educators in revising curricula to ensure that our graduates are adequately prepared to organize information, not only in libraries but also in other cultural heritage institutions.

Oksana L. Zavalina, *Department of Information Science, University of North Texas*, is an Associate Professor at the University of North Texas. She develops and teaches introductory and advanced courses on metadata for digital libraries and archives, library cataloging and classification. She is also a developer and instructor of the American Library Association eLearning professional development course on metadata and Linked Data offered since 2016. Her research interests include descriptive and authority metadata evaluation (including quality and change of metadata) in brick-and-mortar and digital libraries and archives, user needs for information organization, and how these needs are currently met.

Mary Burke, *Department of Information Science, University of North Texas*, is a student at the University of North Texas Information Science Ph.D. Program with a concentration in Linguistics. Her research interests include metadata for information retrieval, language data repositories, and the information behavior of underrepresented groups. Currently, she is focusing on developing infrastructure for archiving data from endangered languages.

References

Alajmi, B., & ur Rehman, S. (2016). Knowledge organization trends in library and information education: Assessment and analysis. *Education for Information*, *32*(4), 411–420. https://doi.org/10.3233/efi-160084

Aluri, R., Kemp, D. A., & Boll, J. J. (1991). *Subject analysis in online catalogs*. Englewood, CO: Libraries Unlimited.

American Library Association (ALA). (2009). *ALA's core competencies of librarianship*. Retrieved from http://www.ala.org/educationcareers/careers/corecomp/corecompetences

Association for Library Collections and Technical Services (ALCTS). (2017). *Core competencies for cataloging and metadata professional librarians*. Retrieved from https://alair.ala.org/bitstream/handle/11213/7853/Core%20Competencies%20Cataloging%20Metadata%20Professional.pdf

Bates, A.W. (2015). *Teaching in a digital age: Guidelines for designing teaching and learning in a digital age*. Retrieved from https://opentextbc.ca/teachinginadigitalage/

Bhakti, G. (2019). Cataloging skills of the 21st century: A mapping of cataloging education in LIS departments across India. *Pearl: A Journal of Library and Information Science*, *13*(4), 325–333. https://doi.org/10.5958/0975-6922.2019.00039.1

Bruce, T. R., & Hillmann, D. I. (2004). The continuum of metadata quality: Defining, expressing, exploiting. In D. Hillman & L. Westbrook (Eds.), *Metadata in practice* (pp. 238–256). Chicago, IL: American Library Association.

Cabonero, D. A., & Dolendo, R. B. (2013). Cataloging and classification skills of library and information science practitioners in their workplaces: A case analysis. *Library Philosophy and Practice* (e-journal), *960*. Retrieved from https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=2337&context=libphilprac

Choi, Y. (2020). Developing a specialization for LAM convergence using a competency-based approach in an LIS graduate curriculum. *Journal of Education for Library and Information Science*, *61*(2), 212–228. https://doi.org/10.3138/jelis.2019-0044

Cutter, C. (1904). *Rules for a dictionary catalog* (4th ed.). Washington, DC: Govt. Printing Office.

Davis, J. M. (2008). A survey of cataloging education: Are library schools listening? *Cataloging & Classification Quarterly*, *46*(2), 182–200. https://doi.org/10.1080/01639370802177604

Dieckman, C. S. (2018). Qualifications for serials catalogers in the 21st century: A content analysis of job advertisements. *Cataloging & Classification Quarterly*, *56*(5-6), 487–506. https://doi.org/10.1080/01639374.2018.1493011

Engelson, L. A. (2019). Sufficiency of cataloging education: School librarians respond. *Journal of Education for Library and Information Science*, *60*(4), 285–311. https://doi.org/10.3138/jelis.2018-0072

Fallows, S., & Steven, C. (Eds.) (2000). *Integrating key skills in higher education: Employability, transferable skills and learning for life*. Sterling, VA: Stylus.

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, *87*(6), 477–531. https://doi.org/10.1037/0033-295x.87.6.477
Glaviano, C. (2000). Teaching an information organization course with Nordic DC metadata creator. *OCLC Systems & Services: International Digital Library Perspectives*, *16*(1), 33–40. https://doi.org/10.1108/10650750010371400
Gorman, M. (2002). Why teach cataloging and classification? In J. S. Hill. (Ed.), *Education for cataloging and the organization of information: Pitfalls and the pendulum* (pp. 1–13). New York, NY: Haworth.
Hall-Ellis, S. D. (2006). Cataloging electronic resources and metadata: Employers' expectations as reflected in *American Libraries* and *AutoCAT*, 2000–2005. *Journal of Education for Library and Information Sciences*, *47*(1), 38–51. https://doi.org/10.2307/40324336
Hall-Ellis, S. D. (2015). Metadata competencies for entry-level positions: What employers expect as reflected in position descriptions, 2000–2013. *Journal of Library Metadata*, *15*(2), 102–134. https://doi.org/10.1080/19386389.2015.1050317
Han, M.-J., & Hswe, P. (2010). The evolving role of the metadata librarian. *Library Resources and Technical Services*, *54*(3), 129–141. https://doi.org/10.5860/lrts.54n3.129
Hider, P. (2008). A survey of continuing professional development activities and attitudes amongst catalogers. *Cataloging & Classification Quarterly*, *42*(2), 35–58. https://doi.org/10.1300/j104v42n02_04
Hill, J. S. (2002). Pitfalls and pendulum: Reconsidering education for cataloging and the organization of information: Preface. In J. S. Hill. (Ed.), *Education for cataloging and the organization of information: Pitfalls and the pendulum* (pp. xv–xix). New York, NY: Haworth.
Hillmann, D. I. (2005). *Using Dublin Core*. Dublin Core Metadata Initiative. Retrieved from https://www.dublincore.org/specifications/dublin-core/usageguide/
Hsieh-Yee, I. (2000). Organizing Internet resources: Teaching cataloging standards and beyond. *OCLC Systems & Services: International Digital Library Perspectives*, *16*(3), 130–143. https://doi.org/10.1108/10650750010345256
Hsieh-Yee, I. (2004). Cataloging and metadata education in North American LIS programs. *Library Resources and Technical Services*, *48*(1), 59–68. https://doi.org/10.5860/lrts.48n1.59.
IFLA (International Federation of Library Associations and Institutions). (2017). IFLA Library Reference Model: A conceptual model for bibliographic information. Den Haag, Netherlands: Author. Retrieved from https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf
Intner, S. (2002). Persistent issues in cataloging education: Considering the past and looking toward the future. In J. S. Hill. (Ed.), *Education for cataloging and the organization of information: Pitfalls and the pendulum* (pp. 15–29). New York, NY: Haworth.
Jackson, A. S., Han, M.-J., Groetsch, K., Mustafoff, M., & Cole, T. W. (2008). Dublin Core metadata harvested through OAI-PMH. *Journal of Library Metadata*, *8*(1), 5–21. https://doi.org/10.1300/j517v08n01_02
Joudrey, D. N. (2002). A new look at US graduate courses in bibliographic control. *Cataloging & Classification Quarterly*, *34*(1–2), 57–99. https://doi.org/10.1300/j104v34n01_06
Joudrey, D. N. (2008). Another look at graduate education for cataloging and the organization of information. *Cataloging & Classification Quarterly*, *46*(2), 137–181. https://doi.org/10.1080/01639370802177588
Joudrey, D. N., & McGinnis, R. (2014). Graduate education for information organization, cataloging, and metadata. *Cataloging & Classification Quarterly*, *52*(5), 506–550. https://doi.org/10.1080/01639374.2014.911236
Kurtz, M. (2010). Dublin Core, DSpace, and a brief analysis of three university repositories. *Information Technology and Libraries*, *29*(1), 40–46. https://doi.org/10.6017/ital.v29i1.3157
Maron, D., & Hill, J. (2015). How should we teach metadata? What comparisons between job ad and classroom trends can tell us about preparing LIS students. In Proceedings of the International Conference on Dublin Core and Metadata Applications (pp. 203–205). Retrieved from https://dcpapers.dublincore.org/pubs/article/view/3781/1971
Miksa, F. L. (1983). *The subject in the dictionary catalog from Cutter to the present*. Chicago, IL: American Library Association.
Mitchell, E. (2009). Metadata literacy: an analysis of metadata awareness in college students (Doctoral dissertation). University of North Carolina at Chapel Hill, Chapel Hill, NC. Retrieved from https://cdr.lib.unc.edu/concern/dissertations/rf55z846z
Moen, W. E., Stewart, E. L, & McClure, C. R. (1998). *The role of content analysis in evaluating metadata for the U.S. Government Information Locator Service (GILS): Results from an exploratory study. UNT Scholarly Works.* Retrieved from https://digital.library.unt.edu/ark:/67531/metadc36312/
Mugridge, R. L., & Furniss, K. L. (2002). Education for authority control: Whose responsibility is it? *Cataloging & Classification Quarterly*, *34*(1–2), 231–241. https://doi.org/10.1300/j104v34n01_14
Organisation for Economic Cooperation and Development (OECD). (2013). *Competition policy and knowledge-based capital: Key findings*. Paris, France: Author. Retrieved from http://www.oecd.org/daf/competition/Knowledge-based-capital-%20KeyFindings2013.pdf

Or-Bach, R. (2005). Educational benefits of metadata creation by students. *ACM SIGCSE Bulletin*, *37*(4), 93–97. https://doi.org/10.1145/1113847.1113885
Panizzi, A. (1841). *Rules for the compilation of the catalogue*. London, England: Nichols and Son.
Park, J.-R., & Lu, C. (2009). Metadata professionals: Roles and competencies as reflected in job announcements, 2003–2006. *Cataloging & Classification Quarterly*, *47*(2), 145–160. https://doi.org/10.1080/01639370802575575
Park, J.-R., & Tosaka, Y. (2010). Metadata quality control in digital repositories and collections: Criteria, semantics, and mechanisms. *Cataloging & Classification Quarterly*, *48*(8), 696–715. https://doi.org/10.1080/01639374.2010.508711
Park, J.-R., Tosaka, Y., Maszaros, S., & Lu, C. (2010). From metadata creation to metadata quality control: Continuing education needs among cataloging and metadata professionals. *Journal of Education for Library and Information Science*, *51*(3), 158–176.
Pettee, J. (1946). *Subject headings: The history and theory of the alphabetic subject approach to books*. New York, NY: H. W. Wilson.
Rühle, S., Baker, T., & Johnston, P. (n.d.). Creating metadata. Dublin Core Metadata Initiative. Retrieved from https://www.dublincore.org/resources/userguide/creating_metadata/
Snow, K., & Hoffmann, G. L. (2015). What makes an effective cataloging course?: A study of the factors that promote learning. *Library Resources and Technical Services*, *59*(4), 187–199. https://doi.org/10.5860/lrts.59n4.187
Tarver, H., Zavalina, O. L., & Phillips, M. (2016). An exploratory study of a Description field in the Digital Public Library of America. In Proceedings of the International Conference and Workshop on Dublin Core and Metadata Applications (pp. 34–44). Retrieved from https://dcpapers.dublincore.org/pubs/article/view/3821
Taylor, A. G., & Joudrey, D. N. (2002). On teaching subject cataloging. *Cataloging & Classification Quarterly*, *34*(1–2), 221–230. https://doi.org/10.1300/j104v34n01_13
Turner, R. (2020). Analyzing cataloging job descriptions: Are cataloging jobs disappearing, changing, or merging? *Cataloging & Classification Quarterly*, *58*(6), 591–602. https://doi.org/10.1080/01639374.2020.1795768
Velucci, S. (1997). Cataloging across the curriculum: A syndetic structure for teaching cataloging. *Cataloging & Classification Quarterly*, *24*(1–2), 35–59. https://doi.org/10.1300/j104v24n01_04
Weagley, J., Gelches, E., & Park, J.-R. (2010). Interoperability and metadata quality in digital video repositories: A study of Dublin Core. *Journal of Library Metadata*, *10*(1), 37–57. https://doi.org/10.1080/19386380903546984
Zavalina, O. L. (2017). Integrated learning of metadata quality evaluation and metadata application profile development in a graduate metadata course. In Proceedings of the International Conference on Dublin Core and Metadata Applications (pp. 93–96). Retrieved from https://dcpapers.dublincore.org/pubs/article/view/3856

Notes

1. See https://librarianship.ca/resources/competencies/
2. See https://www.dublincore.org/resources/userguide/creating_metadata/