Skill-Building in Online Metadata Instruction: Quality Evaluation of Student-Created Metadata

Oksana L. Zavalina and Mary Burke
Department of Information Science, University of North Texas, USA
oksana.zavalina@unt.edu, mary.burke@unt.edu

ABSTRACT

As metadata quality directly affects access to information, training LIS students to create high-quality metadata is an important task. To provide an effective training, a vision is needed for where best to focus the efforts. That vision should be informed by empirical data on the common quality problems in student-created metadata records in relation to the content and methods of instruction. We attempt to address this need through an overview of the metadata creation skill-building content of the online introductory graduate metadata course, results of the analysis of quality in student-created metadata records, and discussion of how the observed common metadata quality issues might inform curriculum development.

ALISE RESEARCH TAXONOMY TOPICS AND AUTHOR KEYWORDS

metadata; education; curriculum; online learning; students

AUTHOR KEYWORDS

skill-building instruction; metadata education; online pedagogy; quality evaluation

INTRODUCTION

In the knowledge-based economy, the demand for highly qualified specialists rapidly grows, with the nature of the work performed by knowledge workers ever-changing in response to market and technological developments. As a result, one of the two integral components of knowledge – skills – must receive greater emphasis in instructional design. Importantly, effectiveness of skill transmission depends on amounts of practice and appropriateness of teaching methods and technologies for skills development (Bates, 2015).

The dramatically changed landscape of metadata work continues to rapidly evolve. Analyses of employer-posted job ads, surveys of metadata practitioners and educators conducted to identify important skills and traits, place metadata quality and its evaluation among the perceived priorities in metadata education – along with willingness to learn and flexibility – as metadata quality has a direct effect on the ability of users worldwide to access information (e.g., Hall-Ellis, 2015; Park & Lu, 2009, etc.). To evaluate how skills are developed in the metadata-related coursework, most studies conduct the snapshot-level analyses for measures like number and type of courses, offering frequency, required vs. elective course status, and lists of topics taught (e.g., Joudrey & McGinnis, 2014). Several more in-depth reports focused on how the specific metadata skills are developed through curricula, assignments and other activities in metadata courses; however, with exception of Zavalina (2017) brief report, none of them focused on the metadata quality skill-building as a topic of growing importance.

Training of future metadata specialists to create high-quality metadata is an important, yet challenging, task for LIS educators. To be able to provide such training in an effective way, the field needs understanding of the common patterns in metadata quality of the student-created records and the way these patterns correlate with the content and methods of instruction. The LIS educators have been at the forefront of developing and offering online education, and some iSchools have accumulated significant experience in teaching graduate metadata courses online. With the long-term demand for online course offering, that has significantly intensified this year due to the need for adjustment to the circumstances dictated by the pandemic, the new normal of LIS education (including metadata education) is online delivery mode. In this situation, it is particularly timely to provide empirical support for online course development based on the lessons learned by long-term online metadata instructors. Such reports will help transform LIS metadata education – with the focus on improving its effectiveness – in the rapidly changing, interconnected world.

We address this need for empirical data to evaluate the effectiveness of curriculum that supports online learning of the metadata-quality-related group of skills and identify the areas in need of reinforcement. In this paper, we provide a brief overview of the quality metadata creation skill-building in the graduate metadata course offered in synchronous online mode at the University of North Texas (UNT), report results of the analysis of quality in student-created metadata records, and discuss possible solutions to improve metadata teaching and learning.

OVERVIEW OF HIGH-QUALITY METADATA CREATION SKILL-BUILDING IN THE ONLINE GRADUATE METADATA COURSE

The INFO 5223 introductory graduate level metadata course has been offered at UNT continuously since 2000, mostly online. Since 2014, it is offered 3 times a year (to a total of 70-90 students annually), with weekly synchronous class meetings. Learning outcomes include understanding the purpose and various components of metadata schemes (e.g., structure, elements, syntax, semantics), data content and data value standards (including controlled vocabularies), and XML and HTML syntaxes for metadata encoding.

The course is organized into eight Learning Modules, where each module builds on the previous ones:

1. Introduction and metadata role in 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.

Students develop practical experience with standard metadata schemes (Dublin Core, MODS, and VRA Core 4.0), and the UNT Libraries metadata application profile as they gain skills in representing textual and non-textual information objects through two main assignments. The Portal to Texas History Metadata Exercise provides the first opportunity to create item-level metadata records using online metadata submission forms. After obtaining this real-life metadata

experience and learning about standard metadata schemes, students work on the Creating Metadata Records Project. In that major project, which consists of 4 components, students practice creating item-level metadata records individually and collection-level metadata records in teams, using XML and HTML syntax templates.

The following readings are required for learning modules 5-7:

1. the learning module lecture document that includes:
    a. history of the metadata scheme development
    b. structure of the metadata scheme (element set, cardinality, order of elements in metadata record, element attributes and attribute values)
    c. definition and uses for each metadata element
    d. recommended controlled vocabularies and data encoding standards
    e. an illustrative example metadata record with in-depth explanation on the application of each element in the record
2. external readings, including the metadata standard itself, and the official usage guide for it
3. a two-hour live interactive presentation delivered by instructor in synchronous online class meeting, with the slide set, Zoom recording, transcription, and text chat log posted for students, including:
    a. introduction of the learning module lecture document and external readings
    b. detailed walkthrough of the process of creating example metadata record(s)
    c. in class mini exercises
    d. explanation of the Creating Metadata Records Project requirements related to the topic of the learning module
    e. question and answer session.

In the Creating Metadata Records Project, students implement the knowledge and skills obtained in all eight Learning Modules. Each student creates DCTERMS, MODS, and VRA Core 4.0 metadata records for two unique items: a textual object (academic writing or a website on a metadata topic) and a non-textual object (painting in the National Gallery of Art or Art Institute of Chicago collection). The final component of the Creating Metadata Records Project is the team creation of Dublin Core Collection Application Profile records for two team's collections: textual objects and paintings.

As part of the Creating Metadata Records Project, students submit three individual reports that contain student-created metadata records in XML syntax. Each submission is graded by a Teaching Assistant, a Ph.D. student in Information Science, who must have metadata experience obtained through coursework and research projects. Course developer and instructor provides the TA with a rubric for evaluation of records, including criteria of accuracy, completeness, and consistency, and the associated grade points. For example, for DCTERMS student-created metadata records, the completeness evaluation includes submission of both metadata records and the presence of all applicable elements (with repeated elements instances whenever applicable). TAs markup student submissions with corrections and comments on metadata quality issues observed in the records, and assign the grade based on the rubric. Individualized feedback – annotated submission with TA comments and corrections – is returned to each student through the course website, and students are encouraged to examine the feedback and ask questions.

Metadata quality is currently covered in the course at a somewhat fragmented level, without a designated learning module focusing on this topic. As part of introduction to the first major standard metadata scheme to the class, instructor presents the common quality problems found by available evaluations of Dublin Core metadata records in digital libraries and repositories (as summarized by Jackson et al., 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 briefly explains to students the major quality criteria in the grading rubric for the three Creating Metadata Records submissions – accuracy, completeness, and consistency – when introducing the project. Finally, during three additional class meetings, instructor presents to students the generalized summary of common metadata quality problems in student-created records after each of the three reports is graded.

STUDY PURPOSE AND METHODS

The goal of this exploratory study is to develop understanding of the overall quality of student-created metadata records, identify the metadata fields where student errors commonly occur, the typical metadata quality issues, and how this relates to the level and content of instruction received by students on the creation of high-quality metadata records. We explored the following research questions:

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

To address the research questions, content analysis of student-created metadata records annotated by Teaching Assistants was used. Data from three semesters – Spring, Summer, and Fall of 2019 – was selected for the following reasons:

1. the course was taught by the same instructor (who is also the developer of the course); this minimized possible instruction variation due to different teaching styles
2. assignment descriptions and course materials for the relevant learning modules remained consistent throughout these three semesters, after the substantial update in early January of 2019.

Metadata records that represent textual objects were selected for analysis due to the potentially broader applicability of findings regarding metadata quality in records representing textual works (journal articles, book chapters, conference papers, standards, websites etc.) as opposed to the much more specialized metadata representing artwork. Analysis focused on the DCTERMS metadata records because Dublin Core is the most common metadata scheme in digital libraries and archives, its application was examined before, and it is commonly taught in metadata courses. This approach allows for comparisons and makes results more broadly useful for developers and instructors of metadata courses.

The binary coding approach was utilized in evaluation of a total of 74 student-created metadata records representing 37 unique information objects. A code '0' was used if a metadata field did not contain any quality problems. If there were one or more quality errors in a metadata field, a code '1' was selected, and the comment describing the error(s) was added. Metadata fields not applicable to the information object in question were marked with the code 'n/a'. Descriptive statistics indicators such as median, range, standard deviation, and percentages were measured for the overall number of fields with metadata quality problems per metadata record, as well as for the overall number of metadata quality problems per metadata field across a set of student-created records. The same indicators were also measured for each of the three specific categories of metadata quality based on the major metadata quality criteria: accuracy, completeness, and consistency (Bruce & Hillmann, 2004).

PRELIMINARY FINDINGS

Our analysis revealed common problems with metadata quality that are briefly summarized here. Metadata quality problems mainly revolved around authority control: assigning subject terms from controlled vocabularies and using name authority files. Also, XML attribute-value pairs that should represent a controlled vocabulary that served as a source of a data value were not applied by students consistently or accurately. The varied use of *Coverage*, *Description*, and *Relation* groups of elements also point to an insufficient understanding of these metadata elements' semantics. Overall, completeness and accuracy problems were found to be much more widespread in student-created metadata records compared to consistency problems.

Existing studies of Dublin Core metadata in digital libraries and repositories offer a point of comparison to our data. We observed that instances of applicable metadata fields representing dates are missing in almost 34% of student-created records. This indicates a substantially lower level of application than in Jackson et. al. (2008), Kurtz (2010), and Weagley, Gelches, and Park (2010) studies which found that 86%-100% of records included *Date* field. The most often omitted fields in student-created metadata records were those representing relations between information objects: applicable but missing in almost 50%. This is a significantly higher level of omission than that observed by Jackson et al., and Weagley et al. (33%-34% of records).

Our analysis demonstrates a relatively high level of completeness in subject metadata (only 8% and 2.7% of student-created records omitted *Subject* and *Spatial Coverage* fields respectively) compared with findings of Kurtz (2010) and Weagley et al. (2010) that *Subject* and *Coverage* fields were missing in 35% and 49% or more of Dublin Core records respectively. On the other hand, crucial for intellectual access subject metadata fields contained a high number of accuracy and consistency errors in student-created records. The same was true about other fields under authority control: *Creator*, *Contributor*, etc.

LIMITATIONS, DISCUSSION AND CONCLUSION

Results of this small-scale case study may not be statistically generalizable beyond UNT graduate metadata instruction. However, they empirically support observations made by practitioners and educators in cataloging and digital library metadata management (based on the studies and anecdotal evidence) and allow to draw meaningful conclusions. For example, our

study confirms earlier findings that subject analysis and subject representation, as well as authority control, are the most complex and intellectually challenging tasks in the process of creating metadata records (e.g., Cabonero & Dolendo, 2013; Snow & Hoffmann, 2015). Likewise, our findings support existing anecdotal evidence that representation of relationships among various information objects and other entities is conceptually difficult. This emphasizes the need for improving instruction on how to analyze and represent aboutness and relationships (especially the logical pairs of reciprocal relations between entities, including information objects, concepts, etc.). To ensure knowledge retention, we believe these topics should be taught repeatedly at various levels (e.g., in general core courses and specialization courses), using a variety of examples, and with extensive practical exercises to reinforce the knowledge through learning-by-doing.

Another possibly viable solution to persistent metadata quality errors that impede information retrieval would be further raising the students' awareness of metadata quality. This study results suggest that in teaching introductory metadata courses, more emphasis needs to be placed on the metadata quality criteria, evaluation of metadata against these criteria, and the specific ways certain metadata quality issues negatively impact 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 entirely focusing on metadata quality, and an associated practical exercise, in which students would evaluate the quality of metadata records created by themselves and/or their peers. At UNT, this is currently implemented in the advanced metadata course INFO 5224. However, it appears that reliance on advanced metadata courses for in-depth metadata quality coverage should be reconsidered. Advanced metadata courses are taken by significantly smaller number of students compared to introductory courses (e.g., 8-12 a year as opposed to 70-90 at UNT), and it is not a widespread practice in LIS schools to regularly offer advanced metadata courses (e.g., Davis, 2008, etc.). To ensure metadata education fulfills its mission and maintains its value in the changing environment, LIS programs need to adequately prepare future librarians for creating high-quality metadata that would fully support the functions of metadata at providing access to information in a connected world. Therefore, focused metadata quality training would be best placed in the introductory metadata courses that are taken by high proportion of LIS students worldwide.

REFERENCES

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/

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

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.

Davis, J.M. (2008). A survey of cataloging education: are library schools listening? *Cataloging & Classification Quarterly, 46*(2), 182-200.

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.

IFLA FRBR Review Group Consolidation Editorial Group. (2017). *IFLA Library Reference Model: A conceptual model for bibliographic information.* International Federation of Library Associations and Institutions.

Jackson, A.S., Han, M., Groetsch, K., Mustafoff, M., & Cole, T.W. (2008). Dublin Core metadata harvested through OAI-PMH. *Journal of Library Metadata, 8*(1), 5-21.

Joudrey, D.N., & McGinnis, R. (2014). Graduate education for information organization, cataloging and metadata. *Cataloging & Classification Quarterly, 52*(5), 506-550.

Kurtz, M. (2010). Dublin Core, DSpace, and a brief analysis of three university repositories. *Information Technology and Libraries, 29*(1), 40-46.

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.

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.

Weagley, J., Gelches E., & Park, J. (2010). Interoperability and metadata quality in digital video repositories: a study of Dublin Core. *Journal of Library Metadata, 10*(1), 37-57.

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 and Workshop on Dublin Core and Metadata Applications* (pp.93-96). Retrieved from https://dcpapers.dublincore.org/pubs/article/view/3856/2041