Exploring Accuracy, Completeness, and Consistency of Student-Created VRA Core 4.0 Paintings Metadata

Vyacheslav Zavalin[1,*,†], and Oksana L. Zavalina[2,†]

[1] *Texas Woman's University, Denton, Texas, United States of America*
[2] *University of North Texas, Denton, Texas, United States of America*

Abstract

An important LAM professional competency is the ability to create metadata that effectively facilitates discovery of information resources in various settings. While some metadata courses for information professionals provide training in generating metadata that represents visual resources held by museums and galleries, reports assessing results of this learning are lacking. Our study is the first to evaluate student-created metadata that follows the specialized Visual Resources Association Core 4.0 metadata standard for artworks. This paper presents and discusses preliminary results of the exploration of accuracy, completeness, and consistency of metadata records that were created by students enrolled in the introductory graduate metadata course in an ALA-accredited LIS program. Our analysis identified metadata fields which pose challenges to learners in practicing creation of VRA Core 4.0 metadata records to represent paintings. Most common errors are presented and discussed in relation to findings of previous research and future research ideas.

Keywords

Metadata assessment; Visual Resources Association Core; Museum metadata

1. Introduction

Quality metadata is crucial for efficient discoverability of various information resources, including those held by museums and galleries. Metadata standards designed with the goal of improving access to these resources include Cataloging Cultural Objects, Categories for the Description of Works of Art, and Visual Resources Association (VRA) Core. In addition, the general-scope metadata schemes Dublin Core, Metadata Object Description Schema (MODS) etc. allow to represent different kinds of resources, including artworks.

Results of metadata quality evaluations inform improvements in metadata practice and metadata education. Such evaluations are guided by the frameworks that define metadata quality criteria and propose measures for assessing metadata against these criteria. The most influential Bruce and Hillman framework [5] consists of 7 criteria which have been adopted and adapted for analysis in many metadata studies and informed development of other frameworks: e.g., [20]. Three criteria – accuracy, completeness, and consistency – are considered the most important in quality assurance [18]. Developing knowledge and skills in creating quality metadata is a priority in the metadata specialists' preparation: e.g., [9], [10], [18]. Analyses of

* Corresponding author.
† These authors contributed equally.
vzavalin@twu.edu (V. Zavalin); oksana.zavalina@unt.edu (O. L. Zavalina)
0000-0001-9509-9135 (V. Zavalin); 0000-0002-3354-4923 (O. L. Zavalina)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

This article is licensed under a **Creative Commons Attribution 4.0 International License**, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and cite the source.

https://doi.org/10.23106/dcmi.952488807

syllabi – [1], [6], [11], [13] – found that Dublin Core is normally covered in practical assignments in metadata coursework, but it is unclear how often specialized metadata such as VRA Core is covered. Several authors reported how specific metadata skills are developed: e.g., [7], [10], [16], [23]. The quality of student-created Dublin Core records that represent text, paintings, audio, and video was recently evaluated in [3], [4], [22] and [24]. However, no studies so far examined learning of the specialized metadata schemes for representing artworks. Professionally created metadata for artworks has also not been evaluated in published studies but other visual resources (digital images and digital videos) metadata was included in assessments of metadata quality in digital repositories: [15] and [21]. Most digital repository metadata studies focused on Dublin Core or MODS (e.g., [12], [14], [17], [21]), and some on non-standard metadata developed locally: e.g., [2], [19], [23].

No research to date examined VRA Core 4.0 metadata accuracy, completeness, or consistency, and our exploratory study begins bridging this gap.

2. Methods

In this study, we quantitatively and qualitatively analyzed XML-encoded VRA Core 4.0 metadata records – created by students in the graduate metadata course at a US LIS Program – to represent paintings held by two US repositories. The following research questions guided our study:

1. Which VRA Core 4.0 metadata fields include more and less accuracy, completeness, and consistency mistakes?
2. What are the typical metadata mistakes observed? How are these mistakes distributed in the dataset?

VRA Core 4.0 includes 18 top-level metadata elements: *Agent*, *CulturalContext*, *Date*, *Description*, *Inscription*, *Location*, *Material*, *Measurements*, *Relation*, *Rights*, *Source*, *StateEdition*, *StylePeriod*, *Subject*, *Technique*, *TextRef*, *Title*, and *WorkType*. Nine of these have one or more subelements (e.g., *Agent*'s subelements *attribution*, *culture*, *name*, *role*, and *dates*). The *earliestDate* and *latestDate* are used as sub-subelements of both the *Agent*'s *dates* subelement (for representing dates of life or activity of the artist), and the *Date* top-level element (for representing the dates in the lifecycle of the artwork). Twenty-two VRA Core 4.0 elements have attributes: e.g., *unit* for *Measurements* element. The attribute *type* is used across the element set but with different meanings based on metadata element semantics. VRA Core's 9 global attributes include attributes intended for ensuring vocabulary control in *CulturalContext*, *Material*, *StylePeriod*, *Subject*, and *Technique* fields;; *name*, *culture*, and *role* subfields of *Agent*, *name* subfield of *Location*, etc. These attributes are *vocab* (i.e., the controlled vocabulary) and *refid* (i.e., the authority record number). For example, *name* subfield of the *Agent* field in the XML-encoded VRA Core 4.0 metadata record representing the painting "Roots" would look like this: <name vocab="ulan" refid="500030701">Kahlo, Frida</name>.

Our analysis focused on a total of 27 VRA Core 4.0 top-level metadata fields and subfields applicable to all paintings assigned to students in this course. The authors of this paper used the same evaluation rubric to manually assess accuracy, completeness, and consistency of metadata in each applicable field in each metadata record. The dataset, stripped of personally

OPEN ACCESS
This article is licensed under a **Creative Commons Attribution 4.0 International License**, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and cite the source.
https://doi.org/10.23106/dcmi.952488807

identifiable information, was collected after the end of semester to ensure truly unobtrusive data collection and avoid bias. To guide the analysis, we developed the following operational definitions of metadata mistakes based on 3 major criteria of metadata quality:

1. Metadata incompleteness: omission of an applicable metadata element or its applicable additional instance, omission of an element attribute, and incomplete data values (e.g., unfinished sentences and/or missing important pieces of information in the overly brief data value).
2. Metadata inaccuracy: misapplication of a metadata element (e.g., using it for the kind of data that this element is not intended for); misapplication of the metadata element attribute(s) (e.g., using the attribute value that is not allowed for the attribute or indicates the use of the controlled vocabulary not used in the data value); misrepresentation of the resource (e.g., using data values that do not apply, typographical errors).
3. Metadata inconsistency: failure to draw data values from controlled vocabularies where applicable.

3. Findings and discussion

The number of mistakes per VRA Core 4.0 metadata record ranged from 1 to 23, with the average of 6.94, the median of 5 and the standard deviation of 5.87. As seen in Table 1, metadata quality issues were found in 25 out of 27 metadata fields applicable to all paintings in our dataset. No metadata quality problems were observed in the *earliestDate* and *latestDate* subfields of the *Agent* top-level field's *dates* subfield. The total number of quality mistakes in other fields ranged between 1 (in the *dates* subfield of the *Agent* field and the *earliestDate* subfield of the *Date* top-level field) and 38 (in the *refid* subfield of the *Location* field). The probable source of confusion is that VRA Core 4.0 documentation uses the same label for this field and for one of the global attributes.

Table 1.
Distribution of metadata quality errors by 3 quality criteria (in % of records)

Top-level metadata fields and *subfields*	Accuracy	Completeness	Consistency
Work	15.63%	3.13%	0.00%
Agent	12.50%	0.00%	0.00%
name	9.38%	3.13%	3.13%
dates	3.13%	0.00%	0.00%
earliestDate	0.00%	0.00%	0.00%
latestDate	0.00%	0.00%	0.00%
culture	6.25%	0.00%	0.00%
role	12.50%	0.00%	9.38%
Cultural context	15.63%	0.00%	6.25%
Date	3.13%	6.25%	0.00%
earliestDate	0.00%	3.13%	0.00%
latestDate	6.25%	3.13%	0.00%
Description	6.25%	6.25%	0.00%
Location	6.25%	6.25%	9.38%
name	18.75%	18.75%	6.25%

OPEN ACCESS
This article is licensed under a **Creative Commons Attribution 4.0 International License**, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and cite the source.
https://doi.org/10.23106/dcmi.952488807

refID	15.63%	87.50%	3.13%
Material	15.63%	0.00%	15.63%
Measurements	3.13%	43.75%	0.00%
Relation	34.38%	21.88%	3.13%
Rights	12.50%	0.00%	0.00%
rightsHolder	6.25%	0.00%	0.00%
StylePeriod	28.13%	6.25%	0.00%
Subject	9.38%	12.50%	6.25%
term	53.13%	12.50%	9.38%
Technique	12.50%	6.25%	9.38%

Accuracy problems were observed in 24 metadata fields. The lowest non-zero accuracy error level (3.13% records) was observed for two top-level fields (*Date* and *Measurements*), and *Agent*'s *dates* subfield. The highest accuracy error level (53.13% records) was exhibited by the *Subject*'s subfield *term*. Most often, students used non-aboutness terms: the painting's genre or style the name of the artist, the geographical location not shown in the painting, etc. These mistakes had been commonly observed in the previous studies of Dublin Core metadata: [22], [24]. One VRA-Core-4.0-specific kind of accuracy error – related to application of the AAT thesaurus in *Subject's* subfield *term* – was misrepresenting the kind of topic: using the "descriptiveTopic" *type* attribute value with the AAT conceptual subject term (e.g., "Military exercises"), or the "conceptTopic" for a descriptive AAT term representing a specific object or animal depicted in the painting (e.g., "Osprey").

Three more fields exhibited accuracy errors in over 20% records each. Almost 22% of records had accuracy errors in the *WorkType* field, most often using the data value that another field is intended for: the technique term (e.g., "oil painting (technique)"), or the process term ("painting (image-making)"). The *StylePeriod* field – intended for representing artistic style or period (e.g., impressionism) – included accuracy mistakes in 28.13% of records. Students included data that do not belong to this field: for example, terms representing a category of artists (e.g., "vedutisti" – artists working in the "vedute" genre). This field also included typographical errors and inappropriate attribute values (for example, referring to the AAT as a controlled vocabulary when the data value is a name from the TGN thesaurus). The *Relation* field exhibited accuracy errors in 34.38% of records. Although instructed to use *Relation* to represent the downloadable digital image of the painting, students sometimes included the landing page containing information about the painting, or inaccurately represented the kind of relation in the *type* attribute value.

Inaccurate attribute values were often observed for attributes of the *Work*, field intended to describe the metadata record itself (16.75%). These were mainly formatting errors: e.g., the *id* attribute value formatting prescribed by the documentation for the *Work* field – including "w_" that indicates "work" at the beginning – was ignored, or the number was entered without the punctuation "_". Some students misinterpreted guidelines and entered their initials instead of "w" in *id* attribute or misused the *source* attribute (intended for recording the name of the repository and /or collection) for bibliographic-citation-like information.

Completeness problems were observed in 17 fields. No completeness issues were observed in *CulturalContext*, *Material*, 5 subfields of *Agent*, and *rightsholder* subfield of *Rights*. The highest levels of completeness errors occurred in the *Location*'s *refid* subfield intended for accession numbers and other gallery-specific artwork identifiers. Like in the equivalent *Identifier* field in Dublin Core records representing paintings [22], most students omitted one of the two

OPEN ACCESS
This article is licensed under a **Creative Commons Attribution 4.0 International License**, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and cite the source.
https://doi.org/10.23106/dcmi.952488807

identifiers that each painting in the dataset has. Students also often included the *Measurements* field for only one of several sets of dimensions: e.g., recorded overall dimensions but omitted framed ones, recorded measurements in inches but omitted those in centimeters. The *Relation* field was the 3rd most prone to completeness errors: skipped the field instance to represent the digital image.

Consistency errors were observed in the lower number of fields (n=12) than accuracy (n=24) or completeness (n=17). The highest level of consistency errors was found in the *Material* field (15.63% of the records). Students tended to use the non-preferred form of the AAT controlled vocabulary term (e.g., "canvas" instead of "canvas (textile material)"). Two top-level fields (*Location* and *Technique)* and two subfields (*Subject's* subfield *term*, and *Agent's* subfield *role)* exhibited the same level of consistency mistakes: 9.38% of records. Both the *Subject's* subfield *term* and the *Technique* field were found to include non-preferred forms of terms as their data values. The *Agent's* subfield *role* contained the outright non-authorized form of the term in 1 record, and punctuation or capitalization errors resulting in mismatch with the authorized form of the term in others. In *Location* field, students also included the *type* attribute value that is not authorized for use with this attribute (e.g., "art museum" instead of "repository").

4. Conclusions and future research

This study is the first one to examine VRA Core 4.0 metadata quality created by learners. As such, there are no previous studies that produced directly comparable data. However, certain themes emerging from our analysis seem to be common with findings from previous research focusing on other metadata schemes. One of these commonalities is the difficulty with subject analysis and representation that non-experienced metadata creators often exhibit. This includes but is not limited to confusing aboutness and isness. Another finding of this study of VRA Core 4.0 metadata that resonates with reports from previous research is the difficulty metadata learners evidently experience in clearly identifying and representing different kinds of relationships between information resources. Also, importantly, as in multiple other metadata quality studies done in the past, this study found accuracy errors to be much more widespread than completeness and consistency errors. This indicates the need for both stronger emphasis on accuracy in metadata education and for in-depth metadata research focusing on metadata accuracy.

Our study's dataset was collected from the introductory graduate metadata course, which most students enter with little preexisting metadata skills. Learners gradually develop these skills through regular practice of increasing complexity, using the same two information objects (including a painting) that each student is assigned to first create Dublin Core DCMI Metadata Terms records, then MODS, and then VRA Core 4.0 metadata records. The teaching team provides multiple demonstrations showcasing how to create metadata using these standards for similar resources to those assigned to students. Detailed feedback is also provided on the quality of student-created records in each submission. Thus, one may assume that by the time students create VRA Core 4.0 metadata, they accumulate metadata experience through learning-by-doing and studying examples and feedback and develop better understanding of the process and resulting quality of metadata.

The preliminary findings of this exploratory study indicate that the overall quality of student-created VRA Core 4.0 metadata is higher than the quality of the Dublin Core records

OPEN ACCESS
This article is licensed under a **Creative Commons Attribution 4.0 International License**, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and cite the source.
https://doi.org/10.23106/dcmi.952488807

created earlier in the semester (e.g., [22], [24]). However, some persistent errors were observed (with entering identifiers and measurements of the artwork). Also, recent research demonstrated that quality of student-created metadata later in the semester (in individual projects) was lower than in the team projects that students pursue before that [4]. Thus, the amount of learner's experience level alone is not a reliable predictor of improved quality. To be able to assess the influence of this and other factors on the quality of student-created VRA Core 4.0 metadata (e.g., more detailed metadata creation guidelines, availability of record examples in official documentation of the metadata standard, possibility of consulting with experts or peers, etc.), future in-depth studies are needed. In addition to comparing the Dublin Core, MODS, VRA Core 4.0, etc. records originating from the same metadata creator, these studies will collect and analyze other relevant data (e.g., surveys of metadata creators).

References

[1] B. Alajmi, S. Rehman, Knowledge organization trends in library and information education: Assessment and analysis, Education for Information 32(4) (2016) 411–420. doi: 10.3233/EFI-160084

[2] S. Aljalahmah, O.L Zavalina, A case study of information representation in a Kuwaiti archive, Diversity, Divergence, Dialog: 16th International Conference, iConference 2021, Beijing, China, March 17–31, 2021, Poster Descriptions, pp.1-6. IDEALS, Urbana, Illinois (2021). http://hdl.handle.net/2142/109683

[3] S. Aljalahmah, O.L Zavalina, Audiovisual resources metadata: Analysis of records originating from novice metadata creators in Kuwait, Journal of Library Metadata, 24(3) (2024) 189-214. doi: 10.1080/19386389.2024.2343577

[4] S. Aljalahmah, O.L Zavalina, Student-created Dublin Core metadata representing Arabic language eBooks: Comparison of individual and group work outcomes, Journal of Education for Library and Information Science 65(3) (2024) 325-344. doi: 10.3138/jelis-2023-0016

[5] T.R. Bruce, D.I. Hillman, The continuum of metadata quality: defining, expressing, exploiting, Metadata in Practice, American Library Association, Chicago (2004) 238–256.

[6] J.M. Davis, A survey of cataloging education: are library schools listening? Cataloging & Classification Quarterly, 46(2) (2008) 182–200. doi: 10.1080/01639370802177604

[7] C. Glaviano, Teaching an information organization course with Nordic DC metadata creator. OCLC Systems & Services: International Digital Library Perspectives, 16(1) (2000) 33–40. doi: 10.1108/10650750010371400

[8] M.F.A. Hady, A.K. Shaker, Cataloging and classification education in Egypt: Stressing the fundamentals while moving toward automated applications, Cataloging & Classification Quarterly, 43(3/4) (2006) 407–429. doi: 10.1300/J104v41n03_11

[9] P. Hider, A survey of continuing professional development activities and attitudes amongst catalogers. Cataloging & Classification Quarterly, 42(2) (2008) 35–58. doi: 10.1300/J104v42n02_04

[10] I. Hsieh-Yee, Organizing Internet resources: Teaching cataloging standards and beyond, OCLC Systems & Services: International Digital Library Perspectives, 16(3) (2000) 130–143. doi: 10.1108/10650750010345256

OPEN ACCESS

This article is licensed under a **Creative Commons Attribution 4.0 International License**, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and cite the source.

https://doi.org/10.23106/dcmi.952488807

[11] M. Hudon, The status of knowledge organization in library and information science master's programs. Cataloging & Classification Quarterly, 52(5) (2021) 506–550. doi: 10.1080/01639374.2021.1934766
[12] A.S. Jackson, M. Han, K. Groetsch, M. Mustafoff, T.W. Cole, Dublin Core metadata harvested through OAI-PMH, Journal of Library Metadata, 8(1) (2008) 5–21. doi: 10.1300/J517v08n01_02
[13] D. Joudrey, R. McGinnis, Graduate education for information organization, cataloging, and metadata, Cataloging & Classification Quarterly, 59(6) (2014) 576–596. doi: 10.1080/01639374.2014.911236
[14] M. Kurtz, Dublin Core, DSpace, and a brief analysis of three university repositories, Information Technology and Libraries, 29(1) (2010) 40–46. doi: 10.6017/ital.v29i1.3157
[15] S. Lim, C. Li Liew, Metadata quality and interoperability of GLAM digital images. Aslib Proceedings, 63(5) (2011) 484–498. doi: 10.1108/00012531111164978
[16] R. Or-Bach, Educational benefits of metadata creation by students, ACM SIGCSE Bulletin, 37(4) (2005) 93–97. doi: 10.1145/1113847.1113885
[17] J.R. Park, S. Maszaros, Metadata Object Description Schema (MODS) in digital repositories: An exploratory study of metadata use and quality, Knowledge Organization, 36(1) (2009) 46–59. doi: 10.5771/0943-7444-2009-1-46
[18] J.R. Park, Y. Tosaka, Metadata quality control in digital repositories and collections: criteria, semantics, and mechanisms, Cataloging & Classification Quarterly, 48(8) (2010) 696–715. doi: 10.1080/01639374.2010.508711
[19] M. Phillips, O.L. Zavalina, H. Tarver, Exploring the utility of metadata record graphs and network analysis for metadata quality evaluation and augmentation, International Journal of Metadata, Semantics, and Ontologies, 14(2) (2020) 112–124. doi: 10.1504/IJMSO.2020.10030296
[20] B. Stvilia, L. Gasser, M.B. Twidale, L.C. Smith, A framework for information quality assessment, Journal of the American Society of Information Science, 58 (2007) 1720–1733. doi: 10.1002/asi.20652
[21] J. Weagley, E. Gelches, J.R. Park, Interoperability and metadata quality in digital video repositories: a study of Dublin Core, Journal of Library Metadata, 10(1) (2010) 37–57. doi: 10.1080/19386380903546984
[22] V.I. Zavalin, O.L. Zavalina, Exploration of accuracy, completeness, and consistency in metadata for physical objects in museum collections, Normality, Virtuality, Physicality, Inclusivity: 18th International Conference, iConference 2023, Proceedings (2023). 83-90. doi:10.1007/978-3-031-28032-0_7
[23] O.L. Zavalina, Integrated learning of metadata quality evaluation and metadata application profile development in a graduate metadata course, DCMI'17: Proceedings of the 2017 International Conference on Dublin Core and Metadata Applications, Dublin Core Metadata Initiative, Washington, DC (2017) 93–96. doi: 10.23106/dcmi.952137905
[24] O.L. Zavalina, M. Burke, Assessing skill building in metadata instruction: Quality evaluation of Dublin Core metadata records created by graduate students. Journal of Education for Library and Information Science, 62(4) (2021) 423–442. doi: 10.3138/jelis.62-4-2020-0083

OPEN ACCESS
This article is licensed under a **Creative Commons Attribution 4.0 International License**, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and cite the source.
https://doi.org/10.23106/dcmi.952488807