Cataloging & Classification Quarterly
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/wccq20

Metadata Quality Control in Digital Repositories and Collections: Criteria, Semantics, and Mechanisms

Jung-Ran Park [a] & Yuji Tosaka [b]

[a] College of Information Science and Technology, Drexel University, Philadelphia, Pennsylvania, USA

[b] The College of New Jersey Library, Ewing, New Jersey, USA

Available online: 23 Sep 2010

To cite this article: Jung-Ran Park & Yuji Tosaka (2010): Metadata Quality Control in Digital Repositories and Collections: Criteria, Semantics, and Mechanisms, Cataloging & Classification Quarterly, 48:8, 696-715

To link to this article: http://dx.doi.org/10.1080/01639374.2010.508711

Cataloging & Classification Quarterly, 48:696–715, 2010

ISSN: 0163-9374 print / 1544-4554 online
DOI: 10.1080/01639374.2010.508711

Routledge
Taylor & Francis Group

Metadata Quality Control in Digital Repositories and Collections: Criteria, Semantics, and Mechanisms

JUNG-RAN PARK
College of Information Science and Technology, Drexel University, Philadelphia, Pennsylvania, USA
YUJI TOSAKA
The College of New Jersey Library, Ewing, New Jersey, USA

This article evaluates practices on metadata quality control in digital repositories and collections using an online survey of cataloging and metadata professionals in the United States. The study examines (1) the perceived importance of metadata quality, (2) metadata quality evaluation criteria and issues, and (3) mechanisms for building quality assurance into the metadata creation process. The survey finds wide recognition of the essential role of metadata quality assurance. Accuracy and consistency are prioritized as the main criteria for metadata quality evaluation. Metadata semantics greatly affects consistent and accurate metadata application. Strong awareness of metadata quality correlates with the widespread adoption of various quality control mechanisms, such as staff training, manual review, metadata guidelines, and metadata generation tools. And yet, metadata guidelines are used less frequently as a quality assurance mechanism in digital collections involving multiple institutions.

KEYWORDS *metadata quality control and evaluation, metadata semantics, metadata guidelines, semi-automatic metadata generation, digital repositories*

Received April 2010; revised June 2010; accepted June 2010.

This study is supported through an Early Career Development Award (2006–2010) from the Institute of Museum and Library Services. We thank the editor and reviewers for their invaluable comments and suggestions.

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

INTRODUCTION

As the rapid proliferation of digital library projects has raised various critical issues and concerns regarding metadata policies and implementations, one of the areas demanding more vigorous research is the quality of metadata in digital repositories. Initiating new digital collections from the ground up leads to a series of challenging metadata decisions, including selection of metadata schemes, semantics, content rules, controlled vocabularies, and metadata creation workflow. And yet, "once a metadata standard has been implemented within a system," Currier noted, "the specified fields must be filled out with real data about real resources, and this process brings its own problems."[1] Evaluation of metadata records created for specific digital projects might reveal that their quality is not sufficient to support successful end-user resource discovery and access. Metadata quality is an even more critical issue in the aggregated environment, as Shreeves, Riley, and Milewicz point out in their article on shareable metadata, because metadata interoperability based on accurate and consistent resource description is necessary to ensure that metadata will remain meaningful outside "its local context."[2]

The goal of this study is to assess practices that impact metadata quality in digital repositories through an online survey data mostly drawn from the community of cataloging and metadata professionals in the United States. We explore the following questions:

- What are prevailing perceptions of metadata quality in digital repositories?
- What are major criteria used to measure metadata quality?
- What are major issues encountered in ensuring metadata quality?
- What are major mechanisms used to improve metadata quality?

We will present an overview of recent studies relating to metadata quality, followed by the survey methodology and data employed to conduct this study and the general characteristics of the survey respondents. Then, we will examine how metadata quality issues are addressed across digital repositories. The final section will summarize the study findings and briefly present future research direction and practice in the area of metadata quality.

REVIEW OF RELATED STUDIES

While studies and best practices guides have been published on a wide spectrum of metadata-related questions, only recently has the importance of metadata quality become a major critical issue in the literature related to digital libraries. Accordingly, there is a lack of research on assuring metadata quality in digital repositories,[3] indicating the need for systematic examination

of metadata quality issues from a multiplicity of perspectives and methodologies. Even if a survey finds a strong awareness of metadata quality, for example, it could be that quality assurance has not been sufficiently built into existing best practices guides. Or, if quality control is addressed as a major issue in local guidelines, we might still find that poor quality metadata records diminish the value of digital collections for potential users. Multiple methodological approaches are integral to identify various layers of metadata quality issues and effective mechanisms for improving the quality of metadata across digital repositories.

The reviews that follow are not intended to be exhaustive; for a comprehensive review in the area of metadata quality, see the study by Park.[4] Our survey questions (see Appendix) were framed within the context of the literature on metadata quality, which has focused for the most part on (1) metadata quality evaluation criteria, (2) quality issues, and (3) quality assurance mechanisms. This section also reviews results from past surveys conducted on metadata quality policies and practices in digital repositories and collections.

Metadata Quality Evaluation Criteria and Quality Issues

Growing awareness of metadata quality control as a core element in building a good digital collection has been reflected in a functional perspective on metadata quality. "The utility of metadata," Hillmann, Dushay, and Phipps wrote in their study on metadata quality, "can be best evaluated in the context of services provided to end-users."[5] Likewise, Guy, Powell, and Day defined metadata quality in terms of "functional requirements," or "fitness for purpose." In other words, the quality of metadata is to a large degree related to the purpose of traditional bibliographic control, as specified in the International Federation of Library Associations and Institutions (IFLA)'s *Statement of International Cataloguing Principles*, in facilitating discovery, identification, selection, and use of the information resources needed by end users.[6]

Unlike traditional cataloging, however, a digital library project often involves decisions about handling new formats, various types of resources, and multiple options of using new and dynamic metadata standards for different communities with changing needs and expectations. In addition to the resulting interoperability issues in the aggregated environment, metadata creators must also pay attention to the newer functions of administration, provenance, rights management, and preservation. As a result, the quality of metadata in this increasingly complex environment is based on a much broader set of functional requirements beyond the application of established rules and standards in conventional descriptive and subject cataloging.[7]

The functional perspective on metadata quality is closely tied with the criteria that are used for metadata quality evaluation. While "good" metadata

reflects the degree to which it is fit for the intended functional purpose of supporting common user tasks and services, the types of required data elements can be defined only through specification of several specific, agreed-on dimensions for operationalizing the measurement of metadata quality. Toward that end, Park analyzed often overlapping criteria and matrices in her survey of research on metadata quality evaluation. Her study finds that completeness, accuracy, and consistency are the most commonly used criteria in measuring metadata quality.[8]

The completeness of metadata means that individual objects are described using all metadata elements that are relevant to their full access capacity in digital repositories.[9] This reflects the functional purpose of metadata in resource discovery and use. In other words, completeness is to a large degree defined by the characteristics of resource types or the nature of individual digital collections. A mandatory or conditionally mandatory (i.e., required if available) element for one resource type could be made optional for another resource type, while local policies and guidelines also may specify what data elements are needed to produce complete metadata records in local contexts for workflow and functional purposes (e.g., metadata creation, information access, and services).

Accuracy concerns the degree to which the data content of metadata elements corresponds to the individual objects being described and the way that it should be represented. It also concerns incorrect or missing data input, such as spelling and typographical errors. Inaccurate data values may also result from inaccuracies in metadata application.[10]

Consistency can be measured by looking at data value on the conceptual/semantic and structural levels, respectively. On the conceptual/semantic level, consistency is affected by the degree to which the same data values or elements are chosen for representing similar concepts in resource description. On the other hand, structural consistency concerns the extent to which the same data structure or format is used to represent information chosen for given metadata elements. For instance, different formats for recording a person's name (e.g., last name, first name, middle initial or last name, first and middle name initials) may result in inconsistent metadata use and authority control problems on the structural level.[11] Of the three major criteria, Caplan and others have shown that consistency especially seems to pose the greatest challenge in ensuring metadata quality in the heterogeneous, distributed context of digital repositories due to conceptual ambiguities and semantic overlaps of various metadata elements.[12]

Quality Assurance Mechanisms

Known issues in ensuring the quality of metadata have led to various efforts to build quality assurance into the metadata creation process. One of the essential mechanisms for improving metadata quality is metadata guidelines

and best practices documentation (also called application profiles). Metadata guidelines are typically composed of metadata element names, identifiers, definitions/descriptions, comments, content rules, and examples. They are designed to offer metadata creators instructions defining and specifying the use of selected metadata scheme(s) for specific digital collections. In this context, one promising approach to using metadata guidelines is to embed them within a metadata creation system so that it can assist metadata creators in creating quality metadata by referencing textual guidance on a Web form or using such features as drop-down menus and pop-up windows.[13]

It appears, however, that these metadata guidelines often contain varying specifications for given metadata schemes.[14] Metadata implementations in digital repositories have tended to emphasize simplicity and flexibility for managing the explosion of digital resources. Rather than create a common data model that can be readily referenced as a mediation mechanism during the metadata creation process, metadata creators have tended to "bend and fit metadata schemas for their own purposes," as Heery and Patel noted in their article on application profiles. Even when a common standard like the Dublin Core (DC) metadata scheme is used across digital repositories, Park and Lu have identified considerable divergence in what local metadata guidelines contain and how they are represented. Each guideline used different labels and included local additions and variants to describe local digital resources. This has the potential to hinder shareable metadata creation and diminishes the ability to obtain meaningful search results from diverse sets of metadata records created by different organizations. To make things worse, such locally created guidelines are often not made publicly available for more effective data sharing and interoperability.[15]

In addition to metadata guidelines, many researchers have explored a variety of tools for semi-automatic metadata generation as a mechanism to promote efficient and consistent metadata use. For instance, there are a number of systems that retrieve a Web page, automatically generate some metadata elements, and allow the generated metadata to be edited using the online form. Some promising studies also have developed a computer-aided system using automatic indexing and data extraction techniques to generate descriptive and other metadata values based on digital object content itself as well as its context, including object usage, user profile, metadata repositories, and domain ontologies. And yet, the feasibility and scalability of these research results have not been studied extensively in a realistic metadata creation environment.[16]

Metadata Policies and Practices Drawn from Survey Research

To supplement this research and evaluation, several recent surveys have been conducted to gain a broad, baseline understanding of current metadata

policies and practices in digital repositories. However, metadata quality issues have not been among their main concerns. Only two major surveys have addressed metadata quality control as part of the questionnaires distributed to metadata specialists and managers engaged in digital projects.[17]

Ma's 2007 survey of ARL (Association of Research Libraries) member libraries ($N = 67$) suggested a widespread awareness of the essential role of metadata quality assurance. She asked the respondents what metadata quality assurance mechanisms were used in their digital collections. Manual evaluation was used as a quality control method in 56 libraries (83.6%). Reflecting the fact that many item records were created by users or content creators themselves, 41 libraries (61.2%) reported that such metadata was checked and approved by metadata librarians, catalogers, or other library staff. Twenty-one libraries (31.3%) used various tools to check metadata consistency and accuracy, such as authority control, XML and other software validation, compliance with application profiles, and locally developed scripts.[18]

In 2007, Zeng, Lee, and Hayes conducted a survey on metadata implementations in preparing a chapter on metadata in the *IFLA Guidelines for Digital Libraries*. Their survey, which received over 400 responses from 49 countries, focused on major issues and concerns regarding metadata standards and controlled vocabularies used for various projects. In regard to metadata quality issues, nearly half the respondents (49.4%) reported that "to learn how to measure and control metadata quality" was a major concern in designing and planning digital projects. Although four other choices, such as workflow design and metadata reuse, were ranked slightly higher, the survey data indicated that metadata quality had a high perceived value as an essential building block for developing good digital collections. In addition, the survey showed that providing guides for consistent metadata creation was one of the top priorities for digital library planners (68.5%). Accuracy (i.e., "to learn how to provide correct information in a record") also seemed to be of major importance among the respondents (51.0%). And yet, among the seven choices offered about metadata standards, the creation of application profiles was ranked of least value (33.0%), a result that seems to contradict the perceived importance of providing guides for consistent metadata creation as indicated in the other part of the survey.[19]

These two surveys point to the importance of metadata quality control among practitioners in the field. However, more research needs to be done to redress the lack of a cumulative body of research on metadata quality issues from the perspectives and experiences of catalogers and other metadata specialists. To address this research gap, we examine how information professionals approach, define, and attempt to improve metadata quality in their digital repositories.

RESEARCH METHOD AND DATA

For this study, we conducted a Web survey using the WebSurveyor system (http://www.vovici.com/). The survey included both multiple-choice and open-ended questions. The survey was extensively reviewed by members of the advisory board, a group of three experts in the field, and was pilot-tested prior to being officially launched. The survey included many multiple-response questions that asked respondents to check all answers that applied.

Participants were recruited through survey invitation messages and subsequent reminders to electronic mailing lists of interest to metadata and cataloging professionals. Table 1 shows the 10 mailing lists employed for the study. They were selected based on their representative characteristics in the field. Individual invitations (approximately 600) were sent to department heads in cataloging and technical services in academic libraries. We also sent out individual invitations and distributed flyers to selected metadata/cataloguing sessions during the 2008 annual ALA midwinter conference held in Philadelphia.

During the 62-day period from August 6 through October 6, 2008, a total of 303 completed responses were received by the online survey system. The survey attracted a large number of initial participants ($N = 1{,}371$). Among the participants who started the survey, a total of 303 (22.1%) completed it. We suspect that many participants failed to complete the survey when they decided that the subject matter was outside the scope of their regular job responsibilities. The length of the survey also may have been a factor in the number of incomplete responses.[20] The survey had a total of 49 questions, many of which consisted of several multiple-choice questions in a matrix format intended to capture the complexity of the current metadata environment. Other issues such as metadata creation practices, semi-automatic metadata application, locally added metadata elements, and the continuing education of cataloging and metadata professionals are reported on in-depth in separate studies.[21]

TABLE 1 Electronic Mailing Lists for the Survey

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

TABLE 2 Job Titles of Participants (Multiple Responses Allowed)

Job Titles	Count/Percentage
Other	135(44.6)
Cataloger/cataloging librarian/catalog librarian	99(32.7)
Metadata librarian	29(9.6)
Catalog & metadata librarian	26(8.6)
Head, cataloging	26(8.6)
Electronic resources cataloger	17(5.6)
Cataloging coordinator	15(5.0)
Head, cataloging & metadata services	15(5.0)

The overall validity of our collected survey data is illustrated clearly by the respondents' profiles regarding job titles (see Table 2) and job responsibilities (see Table 3). Most of the individuals who completed the survey questionnaire are engaged professionally in activities directly relevant to the research objectives, such as descriptive and subject cataloging, metadata creation and management, authority control, non-print and special material cataloging, electronic resource/digital project management, and integrated library system management. Thus, our respondents were in an appropriate position to offer first-hand information about the current state of metadata quality control practice in their institutions.

Although the largest proportion of participants (44.6%) chose the "Other" category on the question of job title (Table 2), their answers show that the vast majority can be categorized as cataloging and metadata professionals. When they were further asked to specify their professional positions, most job titles given in their responses were associated with one of the cataloging and metadata/digital library-related professional activities as listed in Table 4.

TABLE 3 Participants' Job Responsibilities (Multiple Responses Allowed)

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

TABLE 4 Professional Activities Specified in "Other" Category

Professional Activities	Count/Percentage
Cataloging & metadata creation	31(10.2)
Digital projects management	23(7.6)
Technical services	17(5.6)
Archiving	16(5.3)
Electronic resources & serials management	6(2.0)
Library system administration/Other (e.g., LIS education)	6(2.0)

Less than half (121 or 39.9%) of the entire sample of survey participants provided institutional information. This question was designed to be optional, following a suggestion from the Institutional Review Board at Drexel University. The majority of participants who did provide institutional information were from academic libraries (75.2%), followed by public libraries (17.4%), and from other institutions (7.4%). Nearly half (49.8%) reported that their institutions were doctorate-granting universities. These data suggest that survey results tended to represent the experiences and perspectives of cataloging and metadata professionals in academic, particularly research, libraries.

Furthermore, the survey data indicated that metadata creation and management were relatively newer professional responsibilities for many respondents. In terms of educational backgrounds, the vast majority of the respondents (93.6%) had obtained MLS/MLIS or more advanced degrees. In terms of work experiences (see Table 5), more than half (57.4%) reported over 5 years of professional experience: 6 to 15 years (31.1%) and 16 years and more (26.4%). Approximately one-third of the respondents (34.5%) reported 1 to 5 years of experience, while the rest (8.1%) reported less than a year of professional experience. When asked how many years they had been creating metadata for digital library materials, about one quarter of the respondents (26.0%) reported over five years of experience, 31.4% reported

TABLE 5 Years of Professional Experience vs. Metadata Creation Experience

Year	Professional Experience	Metadata Creation Experience
<1 year	8.1%	16.7%
1–5 years	34.5%	57.4%
1–2 years		26.0%
3–5 years		31.4%
5 years >	57.4%	26.0%
6–15 years	31.1%	
15 years <	26.4%	

Note. Percentages do not total 100% due to rounding.

three to five years of experience, 26.0% reported one to three years of metadata creation experience, and 16.7% reported that they had begun to create metadata within the previous year.

RESULTS

The findings of this survey can be grouped into the following three main areas: (1) the perceived importance of metadata quality among cataloging and metadata professionals; (2) criteria used for metadata quality evaluation and their relative importance, and factors causing difficulty in ensuring metadata quality; and (3) mechanisms used to build quality assurance into the metadata creation process. (Percentages in the following tables may not add to 100% because responses in the "no opinion" category are not included in the survey data presented.)

Perceived Importance of Metadata Quality in Digital Repositories

The survey respondents believe strongly that the quality of metadata is a critical component for building successful digital repositories. As shown in Table 6, the vast majority ("strongly agree"—50.8%, "agree"—30.4%) considers quality control to be essential for resource discovery and sharing. This emphasis on quality metadata is also seen clearly in the fact that "metadata control mechanism" ranks highest on the list of future continuing education topics of interest for the survey participants.[22]

As defined by Caplan, the key components of metadata standards are semantics of metadata elements, content standards used for supplying appropriate data values for each metadata element and syntax used for encoding metadata elements and values.[23] Metadata semantics concerns the semantics associated with metadata such as concepts, conceptual relations, and definitions of metadata elements. Metadata syntax is typically defined in metadata standards (e.g., XML in MODS) or system-supplied; thus, the survey only asked the participants about their opinion regarding the importance of

TABLE 6 Perceived Importance of Metadata Quality Control

Response	Percent
Strongly agree	50.8
Agree	30.4
Neither agree nor disagree	3.6
Disagree	0.3
Strongly disagree	8.9

TABLE 7 Perceived Importance of Content Standards for Metadata Quality

Response	Percent
Strongly agree	47.9
Agree	34.7
Neither agree nor disagree	3.0
Disagree	0.7
Strongly disagree	6.6

content standards and metadata semantics relevant to the creation of quality metadata.

On one hand, the survey responses reveal a broad consensus on the importance of content standards for the quality of metadata. As shown in Table 7, nearly half of the respondents (47.9%) "strongly agree" that "content standards for metadata creation are critical to the maintaining of quality metadata," while over one-third (34.7%) "agree" to that statement. Only 7.3% ("disagree"—0.7%, "strongly disagree"—6.6%) do not see the correlation between content standards and metadata quality.

On the other hand, the survey reveals an interesting contrast in regard to the perceived importance of metadata semantics to the quality of metadata. As shown in Table 8, the majority of the respondents (68.3%) still agree that "the semantics of metadata elements is critical to the maintaining of quality metadata," although only about a quarter (27.7%) choose the "strongly agree" category, as opposed to the 47.9% "strongly agree" responses for the importance on content standards.

Metadata semantics appears to be considered less important for metadata quality assurance; 16.2% of the respondents take a neutral position ("neither agree nor disagree") on the importance of semantics to the creation of quality metadata. This may be derived from the fact that the term "metadata semantics" was not clearly defined in the survey questionnaire. However, as will be discussed in the following sub-section, conceptual ambiguity and semantic overlaps appear to cause much difficulty in ensuring metadata quality.

TABLE 8 Perceived Importance of the Semantics of Metadata Elements for Metadata Quality

Response	Percent
Strongly agree	27.7
Agree	40.6
Neither agree nor disagree	16.2
Disagree	2.0
Strongly disagree	6.6

TABLE 9 Criteria Used for Measuring Metadata Quality

Response	Percent
Accuracy	76.9
Consistency	74.3
Completeness	65.0
Currency	25.1
Other	3.0

Major Metadata Quality Evaluation Criteria and Quality Issues

When asked what criteria are used for measuring metadata quality in their digital repositories, the survey responses report that the accuracy and consistency of metadata description are the most important quality measurement metrics (Table 9), and are used by about three-fourths of the respondents in their digital collections (76.9% and 74.3%, respectively). On the other hand, the importance of complete metadata description is ranked lower than accuracy and consistency, although it is still widely used as a quality measurement metric by nearly two-thirds of the respondents (65.0%). That accuracy and consistency are prioritized over completeness appears to be consistent with the fact that unlike the other two criteria, usage of metadata elements can vary, as discussed earlier, depending on such characteristics as resource types and the functional purpose of individual collections. Finally, the survey responses show that currency is used as a metadata quality measurement metric only by about a quarter of the respondents (25.1%). While the currency of data values is essential for maintaining reliable metadata records utilized for resource discovery, many digital repositories seem to place much less emphasis on developing a system for performing regular metadata maintenance as the described objects and/or given metadata standards change.

The survey data show that DC is the most frequently used non-MARC metadata scheme.[24] To gain insight into metadata quality issues, survey participants using DC metadata scheme are specifically asked about the difficulties they have experienced in applying the DC scheme during metadata creation. Because DC metadata application is reported in a separate study, we only briefly sketch the topic here in the context of metadata quality control.[25]

Regarding this question, over 20% (Creator element) to 50% (Relation element) of survey participants using DC scheme report that they have found it "very difficult" or "somewhat difficult" to apply various DC metadata elements in their digital repositories (e.g., Source element, 42.2%; Contributor element, 29.6%; Publisher element, 26.0%; Type element, 25.0%; Format element, 23.1%). Furthermore, the respondents report that semantic overlaps (45%) and ambiguities (41%) are by far the two most critical factors

causing difficulty in the correct application of the DC metadata scheme. Another issue is that DC metadata semantics is overly broad; this may engender inconsistency in the application of the standard across digital repositories (11%). These survey data raise concern in that metadata semantics, as discussed above, is considered to be less important to quality metadata creation and maintenance while it appears to pose a particularly serious challenge in creating quality metadata.[26]

Mechanisms for Improving Metadata Quality

Strong interest in metadata quality (see Table 6) is evidenced by more than 90% of the survey respondents who utilize at least some metadata quality control mechanisms in their digital repositories. As shown in Table 10, staff training is the most commonly used quality control mechanism (69.9%). To a large degree, this seems to reflect the fact that metadata creation in a single digital library project often involves different metadata creators within an institution as well as across multiple institutions and includes many paraprofessionals who may be unfamiliar with metadata application in digital repositories. This situation demands that quality assurance needs to be built into the metadata creation process to ensure accurate, complete, and consistent interpretations and application of given metadata standards.

The high usage of metadata creation guidelines (63.0%), as shown in Table 10, provides strong evidence of the perceived need to develop a clear, documented mechanism for consistent metadata application. Some participants using the DC scheme report that its simplicity creates a pressing need for the development of local, project-specific metadata guidelines to ensure internal consistency in metadata creation.

Such guidelines are not without problems, however. As Park and Lu have demonstrated, these tend to diverge from metadata standards. To make

TABLE 10 Types of Metadata Quality Control Mechanisms Used by the Respondents' Organizations

Response	Percent
Staff training	69.9
Metadata creation guidelines	63.0
Metadata creation tools (e.g., self-checking metadata entry templates, drop-down selections for certain metadata fields)	43.6
Periodic sampling of metadata records for quality review	37.6
Embedding metadata creation guidelines into system	21.5
None	9.6
Other	8.6

things more difficult, the survey shows that the use of locally added home-grown metadata elements is allowed in nearly 70% of them. Only about one-fifth of local application profiles (19.6%) are made available online to the public. This means that not only is it difficult to create shareable metadata but also it is very difficult to have a quality assurance mechanism that is shareable beyond the local environment.[27]

As shown in Table 10, other types of commonly used metadata quality control mechanisms include various tools for metadata creation (43.6%). More than one-fifth of the respondents (21.5%) reported that their institutions have embedded metadata creation guidelines within a digital collection management system so that metadata creators would be guided to create metadata with a higher level of consistency in resource description within a collection.[28] Nearly 40% of the respondents (37.6%) also report that they rely on periodic sampling or peer review of original metadata records as a manual mechanism for quality assurance.

One of the most critical problems in metadata quality control may be found in repositories involving multiple institutions. A mixed metadata environment based on multiple workflows adds challenges to metadata creation, management, and access in a digital library project. The survey finds that nearly 40% of the respondents (37.6%) are involved with such multi-institution projects. Given the importance of ensuring conformance to repository-wide metadata standards, our questionnaire was designed to discover mechanisms that are used to ensure metadata consistency across multiple institutions. Survey responses show that the vast majority use at least some type of mechanisms to overcome this metadata issue (Table 11). About two-thirds report that selecting a unified metadata standard is key to establishing a common approach to metadata creation. Staff training (59.6%) is also employed as a way to promote consistent metadata application.

TABLE 11 Mechanisms Used to Ensure Metadata Consistency across Multiple Institutions

Response	Percent
Selecting unified metadata standard	66.7
Staff training	59.6
Using the same software	44.7
Using one unified metadata creation guidelines	39.5
Periodic quality review	36.0
Metadata creation tools (e.g., self-checking metadata entry templates, drop-down menus for certain metadata fields)	33.3
Using crosswalks for different metadata standards	28.1
Embedding metadata creation guidelines into system	18.4
None	15.8
Other	8.8

Furthermore, the same software (44.7%) and various metadata creation tools (33.3%) are seen as important semi-automatic mechanisms for ensuring consistent metadata application across institutions, as are crosswalks for converting different metadata standards into a common repository metadata standard (28.1%). Periodic quality review of metadata records (36.0%) and embedding metadata creation guidelines within a digital collection (18.4%) are used in multi-institution digital projects at rates similar to single-institution collections (Table 10).

One noticeable difference between multi-institution and single-institution collections is found in the use of metadata creation guidelines. The same metadata guidelines are used in no more than 40% of cross-institutional collections, in contrast to the widespread usage of metadata guidelines (63.0%) as a whole. While the same metadata standards are used in two-thirds of multi-institution digital library projects, it appears problematic that the same metadata guidelines are not used in the majority of such projects, since given metadata standards may remain open to local interpretation and metadata application guidelines may vary institution by institution. This has the potential to hinder consistency in metadata creation. Finally, survey data indicate that 15.8% of multi-institution digital collections use no mechanism to ensure metadata consistency across institutions. The lack of quality assurance mechanism is a cause for concern in ensuring a minimum level of consistency in resource description within such collections.

CONCLUSION

Metadata quality plays an essential role in building good digital collections. The core functions of bibliographic control in facilitating discovery, identification, selection, and use of digital resources needed by end users, not to mention the newer functions of administration, provenance, rights management, and preservation, depend on it. The rapid proliferation of distributed digital repositories creates a pressing need for systematic evaluation of the current status of metadata quality control practices given the critical importance of quality metadata for optimal resource sharing and reuse in the rapidly evolving networked environment.

While the importance of metadata quality appears to be widely recognized by practitioners in the field, these survey results reveal several areas that may cause difficulty in creating quality metadata that is interoperable across digital repositories. Accuracy and consistency are prioritized over completeness as the most commonly used criteria in measuring metadata quality. That currency ranks much lower as a metadata quality evaluation metric indicates that many repositories currently may not have systems and/or resources for regular metadata quality maintenance. The semantics of metadata elements is perceived to be less important overall than content standards for

metadata quality assurance. However, the survey reveals a gap between the low ranking for metadata semantics and the fact that it appears to be a source of great difficulty in consistent metadata creation.

Strong awareness of the importance of metadata quality is clearly seen in the widespread adoption of various metadata quality control mechanisms in most digital repositories. Training complements manual quality review as the most commonly used mechanism for implementing quality assurance procedures and assisting technical staff through the metadata creation process. Other mechanisms such as metadata creation guidelines (sometimes embedded into the metadata creation system) and metadata generation tools also have been adopted to promote consistent metadata application in many digital repositories.

It appears, however, that well-coordinated mechanisms for quality assurance are sometimes lacking in digital collections encompassing multiple institutions. The most frequently utilized mechanisms are the selection of common metadata standards and software as well as staff training. On the other hand, it is disconcerting that common metadata creation guidelines are used in less than 40% of libraries involved in multi-institutional digital collections. Metadata guidelines seem to be fundamental in ensuring a minimum level of consistency in resource description within a collection and across distributed digital repositories.[29] We face critical issues in relation to the creation of quality, interoperable metadata if such guidelines are not shared widely across multiple institutions creating metadata for the same digital collections.

There are certain limitations to this study. Although an online survey instrument allows us to ask many complex questions and collect data from a large sample quickly, it may not be the best method available for capturing the fuller context of metadata quality control decisions made in individual digital repositories. For future studies, the incorporation of other research methods, such as follow-up telephone surveys and focus groups, are necessary to gain a fuller understanding of the current status of metadata quality control practices. The survey reveals that current metadata practices still fall short of creating quality, shareable metadata and that there is a need to build a common data model that is interoperable across libraries. Development of such a common model demands in-depth studies in relation to semantic factors of metadata schemes and development of a common framework for measuring and improving metadata quality.

NOTES

1. Sarah Currier, "Metadata Quality in E-Learning: Garbage in—Garbage out?" *Centre for Educational Technology Interoperability Standards*, April 2, 2004, http://assessment.cetis.ac.uk/content2/20040402013222 (accessed March 26, 2010).

2. Sarah L. Shreeves, Jenn Riley, and Liz Milewicz, "Moving towards Shareable Metadata," *First Monday* 11, no. 8 (2006), http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1386/1304/ (accessed February 9, 2010).

3. Jung-ran Park, "Metadata Quality in Digital Repositories: A Survey of the Current State of the Art," *Cataloging & Classification Quarterly* 47 (2009): 213–228.

4. Ibid.

5. Diane Hillmann, Naomi Dushay, and Jon Phipps, "Improving Metadata Quality: Augmentation and Recombination," *International Conference on Dublin Core and Metadata Applications*, October 11–14, 2004, 7, http://dcpapers.dublincore.org/ojs/pubs/article/view/770/766 (accessed March 26, 2010).

6. Marieke Guy, Andy Powell, and Michael Day, "Improving the Quality of Metadata in Eprint Archives," *Ariadne* 38 (2004), http://www.ariadne.ac.uk/issue38/guy/ (accessed February 11, 2010); International Federation of Library Associations and Institutions, *Statement of International Cataloguing Principles* (The Hague: IFLA, 2009), http://www.ifla.org/files/cataloguing/icp/icp_2009-en.pdf (accessed March 26, 2010).

7. NISO Framework Working Group, *A Framework of Guidance for Building Good Digital Collections*, 3rd ed. (Baltimore: National Information Standards Organization, 2007), http://www.niso.org/publications/rp/framework3.pdf (accessed March 26, 2010).

8. Park, "Metadata Quality in Digital Repositories," 217–221.

9. Thomas R. Bruce and Diane Hillmann. "The Continuum of Metadata Quality: Defining, Expressing, Exploiting," in *Metadata in Practice*, ed. Diane Hillmann and E. L. Westbrooks (Chicago: American Library Association, 2004), 238–256.

10. Besiki Stvilia et al., "A Framework for Information Quality Assessment," *Journal of the American Society for Information Science and Technology* 58 (2007): 1720–1733; Sarah J. Currier et al., "Quality Assurance for Digital Learning Object Repositories: Issues for the Metadata Creation Process," *ALT-J Research in Learning Technology* 12 (2004): 5–20; Jeffrey Beall, "Metadata and Data Quality Problems in the Digital Library," *Journal of Digital Information* 6, no. 3 (2005), http://journals.tdl.org/jodi/article/view/65/68 (accessed March 26, 2010); Lloyd Sokvitne, "An Evaluation of the Effectiveness of Current Dublin Core Metadata for Retrieval" (paper presented at VALA [Victorian Association for Library Automation] 2000, Melbourne, Australia, February 16–18, 2000), http://www.vala.org.au/vala2000/2000pdf/Sokvitne.PDF (accessed March 26, 2010).

11. Park, "Metadata Quality in Digital Repositories," 221.

12. Priscilla Caplan, *Metadata Fundamentals for All Librarians* (Chicago: American Library Association, 2003), 78–79; Jung-ran Park, "Semantic Interoperability across Digital Image Collections: A Pilot Study on Metadata Mapping," in *Data, Information, and Knowledge in a Networked World*, ed. Liwen Vaughan (proceedings of the 2005 annual conference of the Canadian Association for Information Science, London, Ontario, Canada, June 2–4, 2005), http://www.cais-acsi.ca/proceedings/2005/park_J_2005.pdf (accessed March 26, 2010); Jung-ran Park, "Semantic Interoperability and Metadata Quality: An Analysis of Metadata Item Records of Digital Image Collections," *Knowledge Organization* 33 (2006): 20–34.

13. Jane Greenberg et al., "Author-Generated Dublin Core Metadata for Web Resources: A Baseline Study in an Organization," *Journal of Digital Information* 2 (2001): 1–10.

14. Rachel Heery, "Metadata Future: Steps toward Semantic Interoperability," in *Metadata in Practice,* ed. Diane Hillmann and E. L. Westbrooks (Chicago: American Library Association, 2004), 257–271.

15. Rachel Heery and Manjula Patel, "Application Profiles: Mixing and Matching Metadata Schemas," *Ariadne* 25 (2000), http://www.ariadne.ac.uk/issue25/app-profiles/ (accessed February 3, 2010); Diane I. Hillmann and Jon Phipps, "Application Profiles: Exposing and Enforcing Metadata Quality," in *Proceedings of the International Conference on Dublin Core and Metadata Applications*, *August 27–31, 2007, Singapore* (Singapore: Dublin Core Metadata Initiative and National Library Board Singapore, 2007): 53–62, http://www.dcmipubs.org/ojs/index.php/pubs/article/viewFile/41/20 (February 3, 2010); Jung-ran Park and Caimei Lu, "An Analysis of Seven Metadata Creation Guidelines: Issues and Implications" (paper presented at 2008 Annual ER&L [Electronic Resources & Libraries] Conference, Atlanta, Georgia, March 18–21, 2008); Park and Tosaka, "Metadata Creation Practices."

16. Jane Greenberg, Kristina Spurgin, and Abe Crystal, *Final Report for the AMeGA (Automatic Metadata Generation Applications) Project* (2005), http://www.loc.gov/catdir/bibcontrol/lc_amega_final_report.pdf (accessed March 26, 2010); Dublin Core Metadata Initiative, "DCMI Tools and Software," http://dublincore.org/tools/ (accessed March 26, 2010); Marek Hatala and Steven Forth, "A Comprehensive System for Computer-Aided Metadata Generation" (paper presented at the WWW 2003 Conference, Budapest, Hungary, May 20–24, 2003); Kris Cardinaels, Michael Meire, and Erik Duval,

"Automating Metadata Generation: The Simple Indexing Interface" (paper presented at the 14th International World Wide Web Conference, Chiba, Japan, May 10–14, 2005); Y. Li, C. Dorai, and R. Farrell, "Creating Magic: System for Generating Learning Object Metadata for Instructional Content" (paper presented at the 13th ACM International Conference on Multimedia, Singapore, November 6–11, 2005); G. W. Paynter, "Developing Practical Automatic Metadata Assignment and Evaluation Tools for Internet Resources" (paper presented at the 5th ACM/IEEE-CS Joint Conference on Digital Libraries, Denver, CO, June 7–11, 2005); Michael Meire, Xavier Ochoa, and Erik Duval, "SAMgI: Automatic Metadata Generation v2.0" (paper presented at the World Conference on Educational Multimedia, Hypermedia and Telecommunications, Vancouver, Canada, June 25, 2007).

17. University of Houston Libraries Institutional Repository Task Force, *Institutional Repositories*, SPEC Kit 292 (Washington, DC: Association of Research Libraries, 2006); Karen Smith-Yoshimura, *RLG Programs Descriptive Metadata Practices Survey Results* (Dublin, OH: OCLC Programs and Research, 2007), http://www.oclc.org/programs/publications/reports/2007–03.pdf (accessed February 20, 2010); Karen Smith-Yoshimura and Diane Cellentani, *RLG Programs Descriptive Metadata Practices Survey Results: Data Supplement* (Dublin, OH: OCLC Programs and Research, 2007), http://www.oclc.org/programs/publications/reports/2007–04.pdf (accessed February 20, 2010); Karen Markey et al., *Census of Institutional Repositories in the United States: MIRACLE Project Research Findings* (Washington, DC: Council on Library and Information Resources, 2007), http://www.clir.org/pubs/reports/pub140/pub140.pdf (February 20, 2010); Jin Ma, *Metadata*, SPEC Kit 298 (Washington, DC: Association of Research Libraries, 2007); Marcia Lei Zeng, Jaesun Lee, and Allene F. Hayes, "Metadata Decisions for Digital Libraries: A Survey Report," *Journal of Library Metadata* 9 (2009): 173–193.

18. Ma, *Metadata*, 13, 28–30.

19. Zeng, Lee, and Hayes, "Metadata Decisions," 173–187.

20. Matthias Schonlau, Robert D. Fricker, and Marc N. Elliott, *Conducting Research Surveys via E-Mail and the Web* (Santa Monica, CA: Rand Corporation, 2002).

21. Jung-ran Park and Yuji Tosaka, "Metadata Creation Practices in Digital Repositories and Collections: Schemata, Selection Criteria, and Interoperability," *Information Technology and Libraries*, 29, no. 3 (2010); Jung-ran Park and Caimei Lu, "Application of Semi-Automatic Metadata Generation in Libraries: Types, Tools, and Techniques," *Library & Information Science Research* 31 (2009): 225–231; Jung-ran Park, Yuji Tosaka, and Caimei Lu, "Locally Added Homegrown Metadata Semantics: Issues and Implications," in *Paradigms and Conceptual Systems in Knowledge Organization: Proceedings of the Eleventh International ISKO Conference, 23–26 February 2010, Rome, Italy*. Advances in Knowledge Organization, vol. 12, ed. Claudio Gnoli and Fulvio Mazzocchi (Würzburg [Germany]: Ergon, 2010), 283–290; Jung-ran Park et al., "From Metadata Creation to Metadata Quality Control: Continuing Education Needs among Cataloging and Metadata Professionals," *Journal of Education in Library and Information Science* 51, no. 3 (2010): 158–176.

22. Park et al., "From Metadata Creation to Metadata Quality Control," 169–170.

23. Caplan, *Metadata Fundamentals,* 5–7.

24. Park and Tosaka, "Metadata Creation Practices in Digital Repositories and Collections."

25. Jung-ran Park and Eric Childress, "Dublin Core Metadata Semantics: An Analysis of the Perspectives of Information Professionals," *Journal of Information Science* 35, no. 6 (2009): 727–739.

26. Ibid., 734.

27. Park and Lu, "An Analysis of Seven Metadata Creation Guidelines"; Park, Tosaka, and Lu, "Locally Added Homegrown Metadata Semantics," 283–290.

28. Park and Lu, "Application of Semi-Automatic Metadata Generation in Libraries," 225–231.

29. Jung-ran Park, ed., "Metadata Best Practices: Current Issues and Future Trends," *Journal of Library Metadata* 9, nos. 3/4 (2009).

APPENDIX: EXCERPT FROM THE SURVEY INSTRUMENT

24. Please indicate the degree to which you agree with the following statement: Metadata quality control is critical to resource discovery and sharing.

___Strongly agree
___Agree
___Neither agree nor disagree
___Disagree
___Strongly disagree

25. Please indicate the degree to which you agree with the following statement: Content standards for metadata creation are critical to the maintaining of quality metadata.

___Strongly agree
___Agree
___Neither agree nor disagree
___Disagree
___Strongly disagree

26. Please indicate the degree to which you agree with the following statement: The semantics of metadata elements is critical to the maintaining of quality metadata.

___Strongly agree
___Agree
___Neither agree nor disagree
___Disagree
___Strongly disagree

27. Do you use the Dublin Core metadata standard?

___Yes
___No (skip to question #30)

28. Which Dublin Core metadata field(s) do you experience difficulty in applying during metadata creation? (VD—Very difficult, SD—Somewhat difficult, SE—Somewhat easy, VE—Very easy)

- Creator
  ___VD ___SD ___SE ___VE
- Publisher
  ___VD ___SD ___SE ___VE
- Contributor
  ___VD ___SD ___SE ___VE
- Type
  ___VD ___SD ___SE ___VE
- Format
  ___VD ___SD ___SE ___VE
- Source
  ___VD ___SD ___SE ___VE
- Relation
  ___VD ___SD ___SE ___VE

29. If you experience difficulty in applying any of the above Dublin Core element(s), please explain the cause of the difficulty.

30. What types of metadata quality control mechanisms does your organization utilize? (Please specify all that apply)

___Metadata creation guidelines
___Staff training
___Embedding metadata creation guidelines into system
___Metadata creation tools (e.g., self-checking metadata entry templates, drop-down selections for certain metadata fields)
___Periodic sampling of metadata records for quality review
___None (skip to question #32)
___Other (please specify)

31. What criteria do you use for measuring metadata quality? (Please specify all that apply)

___Completeness
___Consistency
___Accuracy
___Currency
___Other (please specify)

32. Is your organization involved with digital library related projects that include multiple institutions?

___Yes
___No (skip to question #34)

33. What mechanisms do you use to ensure metadata consistency across multiple institutions? (Please specify all that apply)

___Selecting unified metadata standard
___Using crosswalks for different metadata standards
___Using one unified metadata creation guidelines
___Using the same software
___Staff training
___Embedding metadata creation guidelines into system
___Metadata creation tools (e.g., self-checking metadata entry templates, drop-down menus for certain metadata fields)
___Periodic quality review
___None
___Other (please specify)