This is a preprint of an article published in Library & Information Science Research: (Stvilia, B., Gasser, L. (2008). Value based metadata quality assessment. Library & Information Science Research, *30*(1), 67-74. http://dx.doi.org/10.1016/j.lisr.2007.06.006)

Value-based metadata quality assessment

Abstract

In this article we propose a method that allows a value-based assessment of metadata quality, and construction of a baseline quality model. The method is illustrated on a large-scale, aggregated collection of Simple Dublin Core metadata records. An analysis of the collection suggests that metadata providers and end-users may have different value structures for the same metadata. To promote better use of the metadata collection, we propose making the value models for metadata in the collection transparent to the end-users and allowing end-users to exercise their own value models through participation in content creation and quality control processes.

Introduction

Large and small libraries, archives, and museums have now put their rapidly growing collections of digital content online, creating an immense wealth of scholarly and cultural information. However, problems with the quality of these metadata objects may impede their use and leave the collections underutilized (Anderson, 2006; Bruce & Hillman, 2004; Dushay & Hillman, 2003; Lagoze et al., 2006; Shreeves et al., 2005; Stvilia, Gasser, Twidale, Shreeves, & Cole, 2004). As a representational object, metadata may not accurately represent the actual information object because of inaccurate, incomplete, or inconsistent mapping, or because of changes in and the dynamism of the information object and the underlying reality. Ideally, one would apply a variety of information quality (IQ) assurance techniques to maintain the quality of the metadata collection as a whole at the highest possible level relative to the requirements of current strategies and tasks. However, in a world of limited resources, time, and attention, the uniform availability of such IQ assurance resources is problematic, especially for large information repositories whose sizes may actually increase with greater use and attention. A cycle of diminishing returns has been created, in which a greater need for high-quality metadata becomes coupled with greater difficulty in providing those metadata with limited resources as the metadata collection continues to grow and becomes increasingly diverse.

Until recently, libraries have created metadata for and provided access to a few genres of printed information resources. Proliferation of the Internet and networked digital information resources has changed this dramatically. Libraries now create and aggregate metadata for information resources with changing attributes and levels of quality, which may belong to various evolving genres (Huthwaite, 2001). As a result, they must continuously maintain the quality of their metadata collections by aligning them with the changing states of the information resources and the changing needs of their target communities.

Analyses of metadata collections and the activities they might be used in have shown that not all metadata are likely to be equally important for supporting a specific activity or set of activities (Greenberg, 2001; Stvilia et al., 2004). In addition, the importance of metadata and its quality levels is conditioned by the importance of current goals and the information-related activities driven by those goals, as traded off against alternative goals, tasks, and information uses (Eppler, 2003; Stvilia, 2006; Taylor, 1986). Not surprisingly, the level of metadata quality and the frequency of metadata use are not uniform and can vary even within the same collection (Lagoze et al., 2006; Shreeves et al., 2005). In fact, according to the guidelines of the Office of Management and Budget (2002), which implements Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001, also known as the “Data Quality Act,” federal agencies are advised to apply stricter quality control for important or “influential” information. Deciding what constitutes influential information, however, is left to the individual federal agency, based on the nature of its tasks and responsibilities.

Thus, relationships among goals, tasks, information, and the metadata are not uniform but probabilistic, and the utility of specific metadata in any given task will be probabilistic as well. Individual elements in a metadata base will be used over time in probabilistically varying ways, and with varying frequencies for varying tasks of varying strategic importance. It follows, then, that ideal quality levels for information or metadata elements in a large-scale information base need not be static, uniform, or even constant relative to the above aspects of information use.

A need exists to develop a theory with which to build mechanisms that can dynamically and differentially select and apply IQ assurance techniques. This theory should enable to condition IQ in general, and metadata quality in particular, for maximum effect based on the probability of increased *value*. Although a review of the *Library and Information Science* literature showed a growing interest in metadata quality, and several studies have proposed quality assessment criteria (i.e. Bruce & Hillman, 2004; Dushay & Hillman, 2003; Shreeves et al., 2005), those studies have stopped short of discussing the value of quality and of linking changes in metadata quality with changes in its value. Measuring the value of IQ change in sound and systematic ways is important, not only to optimize IQ assurance activities, but also to provide accountability and justification for the IQ assurance resources spent. Using concepts from reliability theory (Gertsbakh, 2000), decision theory (Radner, 1986), and information value models (Machlup, 1983; Marschak, 1971; Taylor, 1986), we propose a method for conducting a value-based assessment of metadata quality, and constructing a baseline quality model. The method is then illustrated on an aggregated collection of Simple Dublin Core (DC) records.

Literature Review

Information value and information value models: Linking quality and value changes

Definitions of information value and cost are elusive (Stvilia, 2006). Taylor (1986), in his value-added model of information systems, gave four interpretations of the concept of value: (1) the creation of wealth through production and distribution; (2) an increase in usefulness; (3) exchange-value; and (4) the impact of information on the user. He also proposed six categories of value added to information: (1) ease of use; (2) noise reduction; (3) quality; (4) adaptability; (5) time savings; and (6) cost savings. For reasons of space, we do not list the values here and refer the interested reader to this source (Taylor, 1986).

Similar definitions of information value have been provided by Repo (1989) and Mowshowitz (1999), who have connected value with the amount of information used, or have mapped it onto equivalence exchange classes, which could be cash ("exchange value"). In addition, Mowshowitz observed that, although for the information producer, the critical factor of determining the value of an information product is the cost of production, for the user it is the impact of the information product on the making of other products or services for sale. This does not exclude a scenario in which the producer is a user of the product at the same time.

A survey of the information science and economics literature revealed two major groups or types of information value models. The first is the information theoretical approach, which uses the statistical structure of an information system. In this approach, information value equals information quantity or the gain in an information system, that is, how much information or how much unexpected information is contained in a given information object or item (Machlup, 1983). However, information value is a multidimensional concept; the "surprise" or "novelty" dimension is only one of those dimensions. Moreover, the information theoretical approach can be useful when one observes a sequence of informational events and compares the informational content of the new event with the information conveyed by the previous events. However, when the history or statistics of the past events are not available, one cannot assess the value of the new information event. Consequently, this approach may not be effective for discrete events.

In the second, the decision theoretical model, the value of information equals the size of the agent's welfare or net payoff increase achieved from the use of the information (Marschak, 1971; Radner, 1986). That is, the value of information equals the difference between the action payoffs obtained with and without the information. Hence, the value of information in the decision theoretical model is a function of the value or cost of the decision itself, or both. The success of a highly critical decision may lead to higher payoffs than the success of a less critical decision. Likewise, the cost of failure of a critical decision will be much higher than the negative impact of a less critical one. This is one of the major differences between the decision theoretical model and the information theoretical model, which assess the value of information based only on the probability of the event described by the information.

The decision theoretical model allows information items to be ranked by their efficacy to an organization's activity structure and information needs. This, in turn, allows a cost-benefit analysis to be applied to optimize the collection, retention, maintenance, and distribution of information. One must remember, however, that information needs are dynamic, and future needs and payoffs—and consequently, the value of a particular piece of information—can be difficult to predict. In addition, the causal relationship between information and decision making may not always be unidirectional. Sometimes the outcome may come before the decision, and information can be sought to justify the decision in retrospect or after the fact (Garfinkel, 1967).

Information cost and information cost models: Linking quality and cost changes

A term closely related to value is cost. In general, *cost* is defined as the units of resources one must spend to accomplish something, or "the enjoyment or utility that one anticipates having to forego as a result of selection among alternative courses of action" (Buchanan, 2000). Taylor (1986) defined the total cost of information as the sum of all costs incurred, from information generation through use. In addition, he categorized costs as *system costs* and *user costs*. That is, all the costs occurring up to the point when the user uses the system can be qualified as system costs, and the rest of the costs can be qualified as user costs. He cautioned, however, that even when this principle of cost apportioning holds in general, an exact ratio is context specific.

Models of IQ costs are scarce. At the same time, the cost of quality is one of the main variables used in reasoning about the effects of different quality levels. This helps organizations communicate the need for and importance of quality in planning or performance evaluations and in optimizing their processes. Recently, in an attempt to produce a comprehensive classification of IQ-related costs, Eppler and Helfert (2004) reviewed the IQ literature and defined cost as “a resource sacrificed or forgone to achieve a specific objective or as the monetary effects of certain actions or a lack thereof” (p. 313). In addition, they divided costs into two conceptually sound categories: (1) costs caused by low quality and (2) costs of improving or assuring quality. The low-quality cost category was divided into *direct* and *indirect* costs. Indirect costs do not usually provide immediate links to poor IQ. These are often difficult to identify and isolate from other costs, hence are difficult to quantify. Customer dissatisfaction costs and the costs of lost credibility are indirect costs. The category of quality improvement and assurance costs is divided into three subcategories: *improvement costs*, *detection costs*, and *repair costs*.

A substantial body of literature exists on the cost of quality in manufacturing and economics. Evans and Lindsay (2005) classified quality costs into four categories: (1) *prevention costs*—investing in a production improvement process to prevent nonconforming products from occurring and reaching the customer; (2) *appraisal costs*—costs associated with quality assessment and nonconformance detection; (3) *internal failure costs*—costs carried by a company because of product quality nonconformance before the product reaches the customer, which includes scrapping and reworking costs, correction costs, downgrading costs (when the product is sold at a lower price because of nonconformance), and process failure costs (when quality nonconformance causes machine downtime); and (4) *external failure costs*—costs incurred when a nonconforming product reaches the customer, which includes costs caused by customer complaints and returns, product warranty and quality claims, and product liability costs.

Thus, quality-related expenses occur not only in the production stage (prevention and appraisal costs), but also in the use stage (internal and external failure), and these costs are inversely related. This fundamental tradeoff was observed by Taguchi, Elsayed, and Hsiang (1989) in their concept of the quality loss function (see Figure 1). Taylor (1986) referred to the same tradeoff when he divided the total quality cost into system and user cost categories. The total loss function of Taguchi et al. for an attribute *g* sums all losses for *n* products manufactured and is defined as follows:

This is a preprint of an article published in Library & Information Science Research: (Stvilia, B., Gasser, L. (2008). Value based metadata quality assessment. Library & Information Science Research, *30*(1), 67-74. http://dx.doi.org/10.1016/j.lisr.2007.06.006)

$$\bar{L}(\mu,\sigma) = L(\mu) + \frac{\partial^2 L}{2\partial g^2}\sigma^2 ,$$

where the variance is

$$\sigma^2 = (1/n)\sum_{j=1}^{n}(g_j - \mu)^2 , \quad L(\mu) = L(g_T) + \frac{\partial^2 L}{2\partial g^2}(\mu - g_T)^2$$

and the mean is

$$\mu = (1/n)\sum_{j} g_j$$

(Cook, 1997).

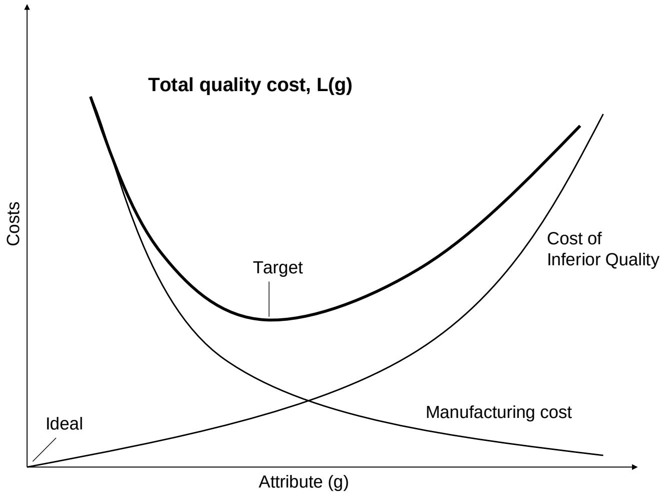

Figure 1: Total quality cost (source: Cook, 1997).

According to this equation, one can improve quality by reducing the variance σ, by reducing the distance between the mean μ and target value g_T, or by reducing the slope or curvature of the loss function $\frac{\partial^2 L}{2\partial g^2}$. This means that quality can be improved either by *improving the production process* (reducing the variance and moving the mean toward the target value through a better process), which will consequently translate into higher production and preventive maintenance costs, or by *adopting stricter quality control* of the ready products (reducing the variance and moving the mean toward the target value through quality control), which increases costs related to scrapping and reworking, or by *increasing robustness* to parameter deviations, which may too result in a process cost increase. Search engines, like Google, often use the later approach when suggesting the end-user a more frequently found spelling for a search keyword and reducing this way the chance of search failure.

Balancing quality-related costs against expected benefits and finding an optimal quality level are emphasized in four elegant principles, which are known as the return on quality approach: (1) quality is an investment; (2) quality efforts must be made financially accountable; (3) it is possible to spend too much on quality; (4) not all quality expenditures are equally valid (Rust & Keiningham, 1999).

Although a rather sophisticated and comprehensive methodology of quality cost assessment and optimization has already been developed in manufacturing, it is not yet clear how much of it is transferable to the realm of IQ. The scarcity of scholarly works and data on the costs of IQ clearly point to the need for more research in this area.

Value-based evaluation of metadata quality

In this section, we propose a method for value-based metadata quality assessment, which consists of techniques for identifying baseline metadata requirements and contextualizing IQ metrics. The techniques are then illustrated on an aggregated collection of DC records.

It is well understood that IQ in general and metadata quality in particular are multidimensional concepts (Bruce & Hillman, 2004; Eppler, 2003; Moen, Stewart, & McClure, 1998; Stvilia et al., 2004; Wang & Strong, 1996). Furthermore, to justify expenditures for quality assurance and to make effective investments in quality improvement, one may often need to quantify the effects of a quality change in those different dimensions. In general, one would like to say that the value of quality is the value of an activity outcome with and without the quality. Likewise, the effectiveness of a change in metadata quality can be calculated as the normalized change in an activity outcome value:

$$E[\Delta MQ(t+1, t)] = V[\Delta AO|\Delta MQ(t+1, t)]/C[\Delta MQ(t+1, t)],$$

where *AO* is an activity outcome, *MQ* is metadata quality, *V* is value, *E* is effectiveness, *C* is cost, and *t* is time.

In manufacturing, improving the quality of a product means finding new optimal target specifications and requirements for each quality dimension at the product level and then transferring those quality requirements down into component-level quality requirements and specifications. Meeting quality requirements at the component level should guarantee, with some confidence, that the product (system)-level quality targets will be met too. The other way of improving quality is to reduce variance around the existing targets. In both instances, whether by setting a new quality target or by reducing variance around the old quality targets, the value of new quality specifications is calculated from the market's value structure and cost for the product (Cook, 1997; Montgomery, 1985). That is, the cost of improving the quality of a given system attribute must be met with an increase in product value, as exemplified by an increase in cash flow. In library settings, monetary metrics may not always be directly applicable to metadata objects. However, a cost-benefit analysis is part of the traditional information-retrieval system evaluation (Lancaster, 1979). One can still treat metadata objects as products and measure the value of a quality change based on the change in an information activity outcome.

Similarly, a quality change may affect not only the value of the activity outcome, but also the cost of the activity. Although an activity outcome may remain the same (success or failure), the cost of completing the activity may go down or up as a result of a quality change. For instance, the user may spend less time on locating a desired item in a catalog after an additional element has been added to the item's metadata. However, if the user does not use that particular element, the user may incur an increased cost by actually spending more time on browsing or reading longer records, or by obtaining a slower response from the catalog because of the increase in index size. Thus, the utility of an IQ change is determined by the value of the activity outcome change adjusted by the costs of this change, and there are a number of ways to estimate the value of a quality change indirectly (see Table 1).

Ways of estimating the value of quality	**Explanation**
A function of the activity success or failure	The quality of metadata may ultimately determine the success or failure of an activity. For instance, inaccurate or missing metadata may lead to a failed search.
A function of the cost and rework	There is a direct relationship between cost and value. Metadata are often an important organizational asset, especially in organizations such as libraries, which use metadata as the cornerstone of their services. One can calculate its dollar cost, and consequently its value change (loss) due to a quality change (degradation), based on the average time a cataloguer spends on creating or reworking a record or an element of the record.
A function of the amount of use	The value of a metadata quality change can be a function of the change in the amount of metadata use. For instance, adding a new element or improving the accuracy of an existing element(s) may increase the amount of metadata and information resource use.
A function of the activity cost	The value of a metadata quality change can be assessed based on a change in the cost of an activity in which the metadata are used. For instance, adding a metadata element with high content entropy (see Table 3) may reduce the time required to find a desired information resource.
A combination of the above factors	The impact of a quality change on metadata value can be some combination of the above factors, conditioned by the criticality of the metadata to the activity outcome.

Table 1: Indirect ways of evaluating the value of a metadata quality change.

Criticality conditions the value of quality

To evaluate the value of metadata quality meaningfully, one needs to be able to determine a baseline quality model for a particular activity context. There are at least two distinct approaches to do that. The first, an *analytical* approach, may involve an analysis of the activity system in which the metadata are used. By modeling and analyzing different scenarios for metadata use, one may identify the quality requirements and determine the baseline levels for metadata quality. Alternatively, if a representative collection for that provider is available, one may use an *empirical* approach and construct the baseline quality representations based on the statistical profile of the collection. Measures of centrality tendency, such as a mean, median, or mode, or a graphical representation of relative cumulative frequencies can be used to determine the baseline levels of quality.

In some cases, a community may already have developed a conceptual model for an activity, which can be used in the analytical approach to contextualize quality metrics.

The International Federation of Library Association's (IFLA's) Functional Requirements for Bibliographic Records (FRBR) model for discovery activity is one such model (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998). The model consists of four actions—find, identify, select, and obtain—and can help to identify some of the metadata requirements for completing the activity successfully. For instance, the model can suggest a set of relevant elements for each of the FRBR actions (see Table 2), and these sets can then be used to deduce a baseline quality model.

Alternatively, when a representative metadata collection is available, one can use the relative cumulative frequencies of quality levels to infer an active baseline quality model for a community. The relative cumulative frequencies and relative value (*RV*) of a quality level change against the baseline value can be calculated as follows:

$$RV = \frac{\sum_{min}^{max} q \times p_q}{\sum_{min}^{baseline} q \times p_q},$$

where q is a quality level measured by some quality metric function; p_q is the portion of the collection having that level of quality; *max* is the highest level of q encountered in the collection; *min* is the lowest level of q encountered in the collection; and *baseline* stands for a baseline level of q (Cook, 1997).

Dublin Core element name	Find	Identify	Select	Obtain
Title	x	x		x
Creator	x	x		
Subject	x	x		
Description	x	x	x	x
Publisher		x	x	x
Contributor[1]	x	x		
Date		x	x	x
Type	x	x		
Format				x
Identifier		x		x
Source		x		x
Language		x	x	x
Relation		x	x	x
Coverage	x	x		
Rights		x	x	x

Table 2: Dublin Core element—Find activity mapping (adapted from: Delsey, 2002, using the Library of Congress MARC to DC Crosswalk[2]).

Some quality dimensions can be more critical than others. Consequently, improving the quality on that dimension can be more valuable than improving the quality on others for a specific activity and action (Office of Management and Budget, 2002). Furthermore, the effects of a quality change may not necessarily be linear, or monotonic. For instance, although more than one DC element can be used to search for an information object, certain elements, such as a Creator, can more effectively reduce a search space because of its high level of information or entropy compared with a Type element (see Table 3). A Language element, on the other hand, although useful for a Select action, may have little effect on the outcomes of Identify or Find actions. Understanding activity structures and their relationships with information objects, creating activity–component mappings similar to the one shown in Table 2, and ranking activities by their criticality or value to a given organization or community can be helpful for contextualizing IQ metrics.

[1] MARC to DC Crosswalk does not use the Contributor element. It uses a Creator element instead.

[2] MARC to DC Crosswalk is available from http://www.loc.gov/marc/marc2dc.html.

Element	Total no.	Unique no.	Entropy
Identifier	205,719	184,769	0.98
Title	133,108	87,689	0.88
Subject	304,661	80,702	0.71
Source	29,537	11,008	0.68
Description	153,088	59,523	0.67
Creator	84,829	18,385	0.65
Date	189,661	11,068	0.62
Coverage	12,103	1,738	0.59
Contributor	16,813	2,882	0.54
Relation	80,629	3,115	0.35
Publisher	114,305	3,347	0.35
Rights	68,228	341	0.33
Type	124,853	191	0.15
Format	111,647	2,308	0.13
Language	85,397	95	0.10

Table 3: Dublin Core elements ordered by their average information content or entropy (Stvilia, Gasser, Twidale, Shreeves, & Cole, 2004).

To develop a criticality-sensitive model of metadata quality measurement, one may use concepts from reliability theory (Gertsbakh, 2000). Indeed, the importance to the activity of a particular quality dimension can be evaluated based on causal relationships among the level of the quality dimension, the probability of the failure of an action, and the cost of failure. If the probability of taking action a_a for the collection is $P(a_a)$, and the probability that a_a will fail when the quality of a metadata element e_i on a y dimension $q_y(e_i) < q_{yi}(required)$ is $p[q_y(e_i) < q_{yi}(required)|Fail(a_a) = 1]$, then the relative criticality of quality level q on the y dimension for metadata element e_i can be evaluated as follows:

$$C(q_y(e_i)) = \left(P(a_a) \times P\left(q_y(e_i) < q_{yi}^{required} \mid Fail(a_a) = 1\right)\right) / \left(\sum_{j=1; j \neq i}^{n} P\left(q_y(e_j) < q_{yj}^{required} \mid Fail(a_a) = 1\right) \right)$$

One can expect that the use of metadata elements in a collection will not be uniform either. Clearly, this needs to be reflected in the criticality function:

$$C(q_y(e_i)) = \left(P(e_i) \times P(a_a) \times P\left(q_y(e_i) < q_{yi}^{required} \mid Fail(a_a) = 1\right)\right) / \left(\sum_{j=1; j \neq i}^{n} P\left(q_y(e_j) < q_{yj}^{required} \mid Fail(a_a) = 1\right) \right)$$

where $p(e_i)$ is the probability of the use of e_i in the collection.

Next, the criticality function can be transformed into a value function by connecting it to the cost of failure. The value of the e_i for an activity A can be evaluated as a weighted sum of its criticalities for each action within the activity multiplied by the cost of the activity failure:

$$v(e_i)_{Activity} = \left(\sum_{a \in Activity} (c_i \times w_a) \right) \times CostOfFailure$$

where weights w_a can be assigned, based on the relative impact of the failure of action a on the success or failure of the activity as a whole.

Finally, the value of a metadata object can be evaluated as follows:

$$ObjectValue = \sum_{i=1}^{n} v(e_i)$$

The function could be used as a value-sensitive completeness quality metric for the object.

Application of the method

In this section, we illustrate how a combination of the analytical and empirical approaches could be used to estimate baseline levels of metadata quality for a particular community. For this purpose, we use the aggregated collection of OAI Simple DC metadata objects harvested by the IMLS Digital Collections and Content (IMLS DCC) Project. The size of the collection at the time of analysis was more than 150,000 objects, collected from more than 20 different metadata providers.

The metadata objects from the IMLS DCC collection use the DC schema and are intended mostly for resource discovery. Hence, it would be appropriate to use the FRBR conceptual model for the discovery activity and develop a community-specific metric function. Metadata schema-specific best practices and implementations of the FRBR model may allow a further contextualization of those metrics. For instance, by using the mapping of DC elements to the FRBR actions from Table 2, we could evaluate the probabilities of DC element use in individual actions, or could evaluate the FRBR activity as a whole. The table suggests that the probability of using the Title element in the activity could be set to ¾, and the probability of using the Description element could be set to 1. Furthermore, according to the table almost every DC element could be used in the Identification action.

However, the statistical profile of the aggregated collection suggests a different value structure for the metadata. The Title element appeared almost twice as often as the Description element (80 versus 47%). Furthermore, the statistical analysis of the aggregated collection showed a significant dependency of DC element use on the *provider, provider type*, and *object type* variables. For instance, academic libraries on average used a higher number of distinct elements per record than did public libraries, 11 versus 9, respectively. Likewise, the mean total number of elements used was much higher for academic libraries than for public libraries and museums, 21 versus 14 and 17, respectively. The metadata records generated by academic libraries on average were larger in size than the records generated by other types of institutions. Indeed, clustering a 2,000-record sample from the aggregated collection by the number of distinct DC elements, using the K-means clustering technique with two clusters and 10-fold cross-validation, almost perfectly discriminated the public library records from the academic library records. Most of the public library records went into cluster 1, with a center of 10 distinct elements (title, subject, description, publisher, date, type, identifier, language, relation, rights), whereas the academic library records were placed in cluster 2, with a center of 12 elements (title, creator, subject, description, publisher, date, format, identifier, source, language, relation, rights). The museum records in our sample were split almost equally between these two clusters. Thus, implicitly or explicitly, public and academic libraries might use different baseline requirements for quality when generating metadata records, based on the needs of their marginal users and cost structures, or on professional norms and rules of information organization. It could also be caused simply by different types of providers supplying records for different types of objects, although these two variables—*provider type* and *object type*—did not show significant correlations.

Thus, a community's active model for quality can be different from its conceptual models, including formal models. Empirical data (representative metadata collections and their use logs) can help one to infer the active model and use it to better align the formal model of the community with its actual metadata needs and practices. For instance, an analysis of the number of distinct elements used (a Completeness metric) in the DC objects of the aggregated collection, relative to the suggested best practice of 8 distinct elements, showed that the metadata provider community may consider setting up the number of required elements to 11, the point at which the effect of diminishing returns begins.

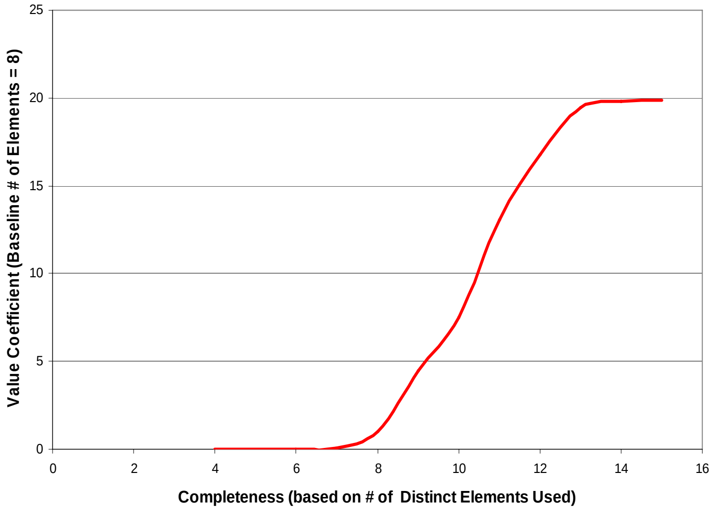

Figure 2: Relative value of schema completeness based on the number of distinct elements used (the aggregated collection).

Application of the same approach to the search log statistics of another metadata aggregator, OAIster suggests that the end-user's value model could be significantly different from that of the data providers. The majority of searches used only a single metadata element (see Figure 3), suggesting that enabling a search by individual metadata elements on top of a full-text search option may not result in a substantial increase in value to the end-user.

This is a preprint of an article published in Library & Information Science Research: (Stvilia, B., Gasser, L. (2008). Value based metadata quality assessment. Library & Information Science Research, *30*(1), 67-74. http://dx.doi.org/10.1016/j.lisr.2007.06.006)

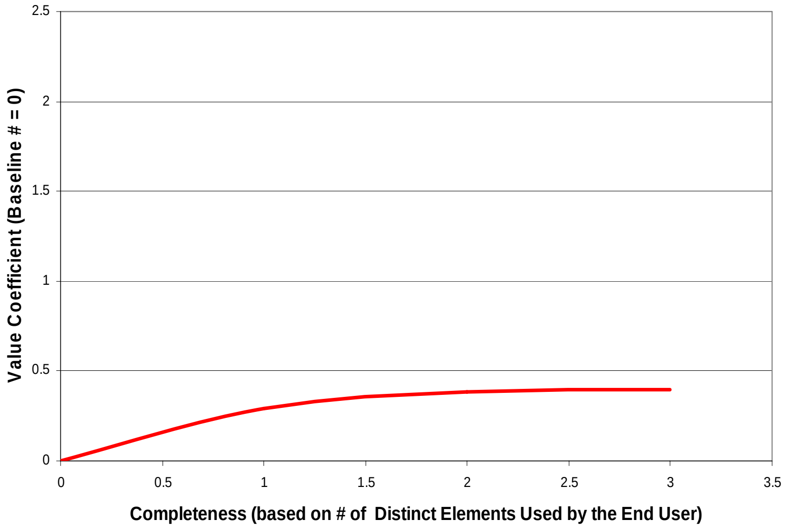

Figure 3: Relative value of completeness based on the number of distinct elements used by the end-user.

Discussion

Although conceptual models for metadata can be intuitive and generalizable, they may not represent the actual metadata practices of individual data providers and end-users. The proposed quality value function assumes that increasing the quality above the required level does not provide additional value. An examination of the aggregated collection showed that this could often be the case. The individual providers may optimize the quality of their metadata records to meet their baseline quality requirements and may not be motivated to go beyond that (to meet the needs of an aggregator or of other communities who may use their metadata). This result echoes the findings of an earlier study by Moen et al. (1998) on the use of the now-defunct Government Information Locator Service (GILS) metadata schema, in which they found “a set of agency GILS rather than a uniform and coherent government-wide locator service” (p. 254).

Furthermore, the analysis showed that end-users’ model for metadata quality could be significantly different from the model used by data providers. It could be that the end-users are mostly interested in descriptive metadata, whereas providers may also use other kinds of metadata, such as administrative metadata, to maintain the collection. It could also be that the users were not aware of the available metadata and the levels of quality, or were limited to fewer options by the interface of the system.

One would expect, however, that by opening the “black box” and exposing information about the metadata and quality-level distributions of its collection, an aggregator or a provider may enable end-users to form or revise their models for the collection and make more informed selections and decisions. Like providers, different end-users may have different value functions for particular metadata, and consequently different value functions for its quality. A metadata quality value function for a collection needs to be an aggregate of these individual user value functions. Hence, the function itself, and the subsequent collection and system decisions about quality or user interface based on that function, tend to be aligned with the preferences of the common users and uses. These might not closely match the quality preferences of a specific user, especially if the user is of a marginal type. Allowing individual users to have direct access to the collection objects and assemble their own subcollections, based on their information needs and quality preferences for the metadata, may lead to higher user satisfaction and better utilization of the collection as a whole.

Furthermore, research shows that, in certain cases, opening a content creation process to end-user contributions could help improve the quality of the collection as a whole. Users may be willing to contribute content and metadata to the areas that are valuable to them (Anderson, 2006; Stvilia, 2006; Twidale & Marty, 1999).

Conclusion

Libraries and universities creating, aggregating and putting online metadata of various kinds of digital content create fertile ground for studying dynamic IQ problems. It also leads to growing understanding of the importance of applying systematic and standardized models of IQ control to enable effective use and reuse of these metadata.

Although it is desirable to maintain all metadata objects in a collection in a uniform, high-quality state, achieving this is often an unrealistic goal because of the scale and cost of quality assurance. In a world of limited resources, there is a need to prioritize IQ assurance based on the baseline models of quality for a particular activity system and the value of a quality change.

In this article we proposed a method using a combination of analytical and empirical approaches to estimate the value of a metadata quality change and to construct a baseline quality model. The analytical approach allows to construct a conceptually sound and more complete set of metadata requirements for a particular activity system, whereas the empirical approach helps to infer the actual or active model for quality of a particular data provider or end-user.

Future work includes testing and refining this method by applying it to different collections. We also plan to develop a registry of general IQ metrics, which could be reused when developing context-specific models of quality assessment.

Acknowledgements

This work was partially supported under grant no. 04135 from the University of Illinois Campus Research Board and an IMLS National Leadership Grant. We thank Michael B. Twidale, Linda C. Smith, Tim Cole, and Sarah Shreeves for helpful comments and suggestions.

References

Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*. New York, NY: Hyperion.

Bruce, T., & Hillman, D. (2004). The continuum of metadata quality: Defining, expressing, exploiting. In D. Hillman & E. Westbrooks (Eds.), *Metadata in practice* (pp. 238–256). Chicago, IL: ALA Editions.

Buchanan, J. (2000). *Collected works of James M Buchanan. Vol. 6: Cost and choice: An inquiry in economic theory.* Indianapolis, IN: Liberty Fund.

Cook, H. (1997). *Product management: Value, quality, cost, price, profits, and organization*. London, UK: Chapman & Hall.

Delsey, T. (2002). *Functional analysis of the MARC 21 bibliographic and holdings formats*. Retrieved July 8, 2004, from http://www.loc.gov/marc/marc-functional-analysis/source/analysis.pdf

Dushay, N., & Hillmann, D. (2003). Analyzing metadata for effective use and re-use. In *DC-2003, proceedings of the international DCMI metadata conference and workshop, September 28-October 2, 2003, Seattle, Washington USA.* [United States]: DCMI.

Eppler, M. (2003). *Managing information quality: Increasing the value of information in knowledge-intensive products and processes*. Berlin, Germany: Springer-Verlag.

Eppler, M., & Helfert, M. (2004). A classification and analysis of data quality costs. In S. Chengulur-Smith, L. Raschid, J. Long, & C. Seko (Eds.), *Proceedings of the 9th International Conference on Information Quality (ICIQ-04)* (pp. 311–325). Cambridge, MA: MIT

Evans, J., & Lindsay, W. (2005). *The management and control of quality* (6th ed.). Mason, OH: Thomson.

Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Gertsbakh, I. (2000). *Reliability theory: With applications to preventive maintenance*. Berlin, Germany: Springer-Verlag.

Greenberg, J. (2001). Quantitative categorical analysis of metadata elements in image applicable metadata schemas. *Journal of the American Society for Information Science and Technology*, *52*, 917–914.

Huthwaite, A. (2001). AACR2 and its place in the digital world: Near-term goals and long-term direction. In A. Sandberg-Fox (Ed.), *Proceedings of the Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the challenges of networked resources and the Web*. Washington, DC: Library of Congress, Cataloging Distribution Service. Retrieved June 3, 2007, from http://www.loc.gov/catdir/bibcontrol/huthwaite_paper.html

IFLA Study Group on the Functional Requirements for Bibliographic Records (1998). *Functional requirements for bibliographic records: Final report*. Retrieved March 23, 2004, from http://www.nlc-bnc.ca/ifla/VII/s13/projects.htm

IMLS Digital Collections and Content (IMLS DCC) Project (http://imlsdcc.grainger.uiuc.edu/)

Lancaster, F. W. (1979). *Information retrieval systems: Characteristics, testing, and evaluation* (2nd ed.). New York: Wiley.

Lagoze, C., Krafft, D., Cornwell, T., Dushay, N., Eckstrom, D., & Saylor, J. (2006). Metadata aggregation and "automated digital libraries": A retrospective on the NSDL experience. In *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 230–239). Chapel Hill, NC: ACM Press.

Machlup, F. (1983). Semantic quirks in studies of information. In F. Machlup & U. Mansfield (Eds.), *The study of information: Interdisciplinary messages* (pp. 641–671). New York: Wiley.

Marschak, J. (1971). Economics of information systems. *Journal of the American Statistical Association*, *66*(333), 192–219.

Moen, W., Stewart, E. & McClure, C. (1998). Assessing metadata quality: Findings and methodological considerations from an evaluation of the U.S. Government Information Locator Service (GILS). In *Proceedings of the IEEE International Forum on Research and Technology Advances in Digital Libraries: ADL'98* (pp. 246–255). Los Alamitos, CA: IEEE Computer Society Press.

Montgomery, D. (1985). *Introduction to statistical quality control*. New York: Wiley.

Mowshowitz, A. (1999). On the market value of information commodities. III. Demand price. *Journal of the American Society for Information Science*, *43*, 242–248.

Office of Management and Budget. (2002). Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of information disseminated by federal agencies. *Federal Register*, *67*(36), 8451–8460.

OAIster (http://www.oaister.org/).

Repo, A. (1989). The value of information: Approaches in economics, accounting, and management science. *Journal of the American Society for Information Science*, *40*(2), 68–69.

Rust, R., & Keiningham, T. (1999). Return on quality at Chase Manhattan Bank. *Interfaces*, *29*(2), 62–73.

Shreeves, S., Knutson, E., Stvilia, B., Palmer, C., Twidale, M., & Cole, T. (2005). Is 'quality' metadata 'shareable' metadata? The implications of local metadata practices for federated collections. In H. A. Thompson (Ed.), *Proceedings of the Twelfth National Conference of the Association of College and Research*

Libraries (pp. 223–237). Chicago, IL: Association of College and Research Libraries.

Stvilia, B., Gasser, L., Twidale, M., Shreeves, S., & Cole, T. (2004). Metadata quality for federated collections. In S. Chengulur-Smith, L. Raschid, J. Long, & C. Seko (Eds.), *Proceedings of the International Conference on Information Quality—ICIQ 2004* (pp. 111–125). Cambridge, MA: MIT.

Stvilia, B. (2006). *Measuring information quality.* Unpublished doctoral thesis, University of Illinois at Urbana-Champaign.

Taguchi, G., Elsayed, E., & Hsiang, T. (1989). *Quality engineering in production systems.* New York: McGraw-Hill.

Taylor, R. (1986). *Value-added processes in information systems*. Norwood, NJ: Ablex.

Treasury and General Government Appropriations Act for Fiscal Year 2001. Pub. L. 106–554. 21 Dec. 2000. *Stat.* 114.2763.

Twidale, M., & Marty, P. (1999). *An investigation of data quality and collaboration*. Technical Report ISRN UIUCLIS--1999/9+CSCW. Champaign, IL: University of Illinois at Urbana–Champaign, Graduate School of Library and Information Science. Retrieved December 26, 2007, from http://people.lis.uiuc.edu/~twidale/pubs/dq.html.

Wang, R., & Strong, D. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5–35.