THE CONTINUUM OF METADATA QUALITY: DEFINING, EXPRESSING, EXPLOITING

Thomas R. Bruce, Legal Information Institute, Cornell Law School; Diane I. Hillmann, National Science Digital Library

Published in “Metadata in Practice,” ALA Editions, 2004.

Reader note: This document was created from the originally submitted version. The original appears to be a working draft and included numbered placeholders for footnotes that were never completed. These placeholders have been replaced with the text "(incomplete footnote)."

INTRODUCTION

Like pornography, metadata quality is difficult to define. We know it when we see it, but conveying the full bundle of assumptions and experience that allow us to identify it is a different matter. For this reason, among others, few outside the library community have written about defining metadata quality. Still less has been said about enforcing quality in ways that do not require unacceptable levels of human effort.

Some stage-setting work is in progress. In a 2002 study of element use by eighty-two OAI data providers, Jewel Ward reported that the providers used an average of eight Dublin Core (DC) elements per record. Five of the eight elements were used 71 percent of the time. Ward’s study indicates that most metadata providers use only a small part of the DC element set, but her study makes no attempt to determine the reliability or usefulness of the information in those few elements. In 2003, another paper published by Naomi Dushay and Diane Hillmann of the Digital Library Research Group at Cornell University described methods for evaluating metadata, and reported in detail some common errors and quality problems found in harvested metadata, as well as a technique for evaluating metadata using a commercially available visual graphical analysis tool (incomplete footnote). Both these efforts clearly have some definition of quality in mind, but neither states it explicitly.

Other recent papers by Barton, Currier and Hey and Moen, Stewart, and McClure have focused on the detection of defects in metadata records and the impact of defects on the utility of collections (incomplete footnote). Barton, et al., strongly believe that defect analysis has major implications for metadata generation practices (incomplete footnote). In preparing this paper, we too have found that it is difficult to talk about quality without also talking about things that betray its absence, but we believe that trying to comprehend quality by enumerating defects risks sacrificing an organized view of the forest to an overly-specific appreciation of the trees. Instead, we attempt a systematic, domain- and method-independent discussion of quality indicators.

PAST EXPERIENCE

The library community has repeatedly tried to define and enforce quality in its bibliographic and authority records, but until recently these attempts have been fairly inconsequential. The practice of bibliographic record sharing has been generally accepted for over a century, and such sharing has been the basis for most of the processing efficiencies realized by libraries during that time. Nevertheless, cataloging continues to be a labor intensive and costly function in libraries, requiring special knowledge and training, and the need for efficient cataloging is all the more keenly felt given the ever-increasing quantity of materials requiring cataloging attention. The resulting tension between "efficient" cataloging and "quality" cataloging has given rise to much conflict between cataloging staff and administrators, with record-selection techniques such as "white lists" giving way to criteria better suited to automation.

In the late 1990s, recognizing that these tensions were increasing even as automated selection of cataloging records from the bibliographic utilities was becoming the norm, the Program for Cooperative Cataloging (PCC) developed the BIBCO Core Record standards (incomplete footnote). BIBCO was an attempt to define a MARC record that could be trusted sufficiently to be re-used without human intervention. BIBCO took its cue from the success of the Name Authorities Cooperative Program (NACO), which had revitalized and diversified the production of name, subject and series authority records for reliable re-use. NACO and BIBCO were successful because they emphasized:

- acceptance of agreed upon standards for record quality;
- participation in a training program where each institution designated several catalogers as liaisons, training other staff members and later, other libraries, in a formalized "buddy" system; and
- Individual review of records by experienced catalogers (from the "buddy" institution) during the training period until an acceptable level of quality is reached.

The acceptance of these programs in the library community reinforces two important points: quality is quantifiable and measurable, and to be effective, enforcement of standards of quality must take place at the community level. Most metadata communities outside of libraries are not yet at the point where they have begun to define, much less measure, quality. However, other communities of practice, particularly those building digital library or e-print systems, are beginning to venture into discussions about metadata quality.

CHALLENGES IN APPROACHING QUESTIONS OF QUALITY

New metadata standards are being developed at a rapid pace, and their introduction into new communities has stimulated discussions of quality. This process has been swifter in some communities than in others, as the early adoption of metadata as a panacea for information overload is followed all too quickly by recognition that investments in quality are necessary for even modest gains. Furthermore, as communities of all kinds attempt to aggregate metadata (and ultimately services) via harvesting protocols like the Open Archives Initiative Protocol for

Metadata Harvesting (OAI PMH)—quality standards and measures are sorely missed (incomplete footnote).

Practitioners and implementers in these communities come from a variety of backgrounds and often have limited experience with information technology or library practices. The documentation and examples available to them may seem too "generalized," and research from ongoing projects that might assist them is buried in places foreign to their discipline. Specialist communities, who tend to see their data as unique, frequently resist the notion that there might be general strategies available to them that could inform their work and enable their metadata to interoperate. Isolation and a tendency to manufacture special solutions for what are really general problems create barriers for coordinated thinking about quality.

Resource constraints, particularly those that come into play as projects scale up, also militate against shared notions of quality. Specialists tend to consider only the attributes that matter to them, neglecting those that might make their data more useful to dimly-imagined, and hence easily-dismissed, groups of outsiders. Often the potential expense of creating interoperability is given as the reason for neglecting outside influences. Budgets for projects rarely contain sufficient funds to effectively plan and implement metadata components, and projects are quick to sacrifice investment that serve any but the most immediate target audiences. Quality that serves outsiders is seen as unaffordable altruism.

Even now, review panels for projects or grant proposals rarely include individuals versed in metadata standards. Planners' assessment of information technology needs is often limited to website or database design and construction, without considering how their information may function when exposed to aggregators or reused in other settings or by other services. To complicate matters, the rate of change in standards and technologies is so rapid that even careful planners and managers find it difficult to determine when they are compliant with current standards. Indeed, to the extent that a painstaking approach implies substantial time spent in consensus building and review, cautious efforts become even more likely to be superseded.

Legacy data presents special problems for many communities, as it rarely makes a clean transition into new metadata formats. Some data were heavily encoded during the days when expensive storage encouraged highly compressed encoding; most was designed specifically for niche rather than general use. Despite these challenges, a few areas of useful discussion are emerging. In 1997 the IFLA Study Group on the Functional Requirements for Bibliographic Records published a final report (incomplete footnote). As part of that effort, the group identified four generic user tasks to be accomplished using bibliographic records:

1. To find entities which correspond to the user's stated search criteria (i.e., to locate either a single entity or a set of entities in a file or database as the result of a search using an attribute or relationship of the entity)
2. To identify an entity (i.e., to confirm that the entity described corresponds to the entity sought, or to distinguish between two or more entities with similar characteristics)

3. To select an entity that is appropriate to the user's needs (i.e., to choose an entity that meets the user's requirements with respect to content, physical format, etc., or to reject an entity as being inappropriate to the user's needs)
4. To acquire or obtain access to the entity described (i.e., to acquire an entity through purchase, loan, etc., or to access an entity electronically through an online connection to a remote computer)

These tasks provide a useful, though not easily quantifiable, basis for testing the effects of metadata quality (or lack thereof) on potential users. In addition, they extend the conversation about metadata quality beyond simple support for resource discovery towards support of broader functionality more applicable to an expansive notion of digital libraries. It is important to note that the "user" in question is one who might be searching a website for materials, rather than the aggregator of metadata, who is also a user, though at a "wholesale" rather than "retail" level. Inevitably, quality is passed downstream from creator, to aggregator, to user. Most of the definitions of quality discussed in this paper affect the aggregator first, and only then the user at the website, who trusts that someone, somewhere, has been paying attention.

QUALITY MEASUREMENTS AND METRICS

In this section, we attempt to define general characteristics of metadata quality. Because we are interested in qualities that are domain-independent, they are necessarily abstract. One might think of these characteristics as places to look for quality in collection-specific schemas and implementations, rather than checklists or quantitative systems suitable for direct application. Recognizing that most metadata projects operate under serious resource constraints, our approach is pragmatic and managerial rather than idealistic. Too frequently, implementers fall into the trap of bipolar thinking, making the perfect the enemy of the good. Realistic approaches balance metadata functionality against applicable constraints to deliver maximum utility from valuable assets: the willingness of data providers, the data itself, and the effort and expense budgeted for metadata creation, organization, and review.

The categorization of quality measures we use here was suggested in part by the Quality Assurance Framework (QAF) for statistical data developed by Statistics Canada (STC) and subsequently applied to metadata assessment by Paul Johanis (incomplete footnote). The original STC QAF described six dimensions of information quality: relevance, accuracy, timeliness, accessibility, interpretability, and coherence. We have reconceived these in a way that is better adapted to the growing number of large-scale projects in which metadata from multiple source providers is aggregated into a unified metadata resource. However, the considerations outlined below are relevant in any setting where metadata is shared. We will examine seven of the most commonly recognized characteristics of quality metadata: completeness, accuracy, provenance, conformance to expectations, logical consistency and coherence, timeliness, and accessibility.

Completeness

Metadata should be complete in two senses. First, the element set used should describe the target objects as completely as economically feasible. It is almost always possible to imagine describing things in more detail, but it is not always possible to afford the preparation and maintenance of more detailed information. Second, the element set should be applied to the target object population as completely as possible; it does little good to prescribe a particular element set if most of the elements are never used, or if their use cannot be relied upon across the entire collection.

Accuracy

Metadata should be accurate in the way it describes objects-- a uniquely non-controversial statement that houses platoons of devils. Minimally, the information provided in the values needs to be correct and factual. At the next level, accuracy is simply high-quality editing: the elimination of typographical errors, conforming expression of personal names and place names, use of standard abbreviations, and so on (incomplete footnote). In large or heterogeneous collections, accuracy may not be directly verifiable; sampling techniques, statistical profiles, or other alternatives to laborious inspection may be needed (incomplete footnote).

Provenance

The provenance of metadata often provides a useful basis for quality judgments. Sometimes this is a matter of knowing who prepared the metadata, how experienced he or she might be, how good his or her judgment is, or of having some sense of their expertise in the relevant domain and with metadata standards generally. We may also rely on well-understood or certified methodologies as proxies that ensure reliability and quality. Scientists and statisticians are quite at home making judgments about the quality of data based on the methods used to create and handle it. This is particularly true in situations where individual items cannot be directly verified. However, the use of creation and handling methodology as guarantor of quality is not limited to the sciences; all sorts of content standards and best-practices guides exist, the Anglo-American Cataloging Rules (AACR2) not least among them (incomplete footnote).

Information about creation is just the starting point of provenance. One should also know what transformations have been applied to the data since it was created. Metadata may come second- or third-hand, and beyond knowing who made it, how it was made, and where it has been, it is useful to know whether value has been added or subtracted since its creation.

Conformance to Expectations

Standard metadata element sets and application profiles that use them are promises from the metadata provider to the user. More, they are promises surrounded by the expectations of a particular community about what such promises mean, how realistic they are, and how they are to be carried out.

Element sets and application profiles should, in general, contain those elements that the community would reasonably expect to find. They should not contain false promises, i.e., elements that are not likely to be used because they are superfluous, irrelevant, or impossible to implement. Controlled vocabularies should be chosen with the needs of the intended audience in mind, and explicitly exposed to downstream users. Sometimes problems with conformance to expectations appear in disguise. Moen et al. correctly point out that problems with omitted metadata frequently occur because users see the particular element as irrelevant or unnecessary, so that what appears at first blush to be a completeness problem is in fact a problem with conformance to expectations (incomplete footnote).

Finally, metadata choices need to reflect community thinking and expectations about necessary compromises in implementation. It is seldom possible for a metadata project to implement everything that anyone would want; most often, the metadata provider cannot afford to make a project unimpeachable by making it comprehensive. It is therefore important that community expectations be solicited, considered, and managed realistically. Better an agreed-upon compromise that is well executed and documented than an approach that aspires to be all things to all people and ends up poorly and unevenly implemented.

Logical Consistency and Coherence

Consistency and coherence are usually seen as problems only for heterogeneous, federated collections, or perhaps for single collections that are presented in successive "releases" over time. But in fact, very few collections exist in isolation, even at their inception. There is almost always a need to ensure that elements are conceived in a way that is consistent with standard definitions and concepts used in the subject or related domains and presented to the user in consistent ways.

The use of standard mechanisms like application profiles and common crosswalks enhance the ability of downstream users to assess the intended level of coherence. Standard mechanisms create a track record of intent over time, thus enabling metadata implementers to easily make comparisons between instantiations.

The quality of "searchability" nicely illustrates the value of consistency. Users expect to be able to search collections of similar objects using similar criteria, and increasingly they expect search results and indicative indexes to have similar structures and appearance (incomplete footnote). This common, reliable user experience depends crucially on metadata being coherent and consistently presented across collections.

Barton et al. describe an interesting and dysfunctional variation on a notion we might term "over-coherence" or "false coherence" (incomplete footnote). It is really a problem with accuracy. In this scenario, the same metadata records are applied inappropriately to multiple components of an object or objects, as if by rote. Similarly, the same study cited by Barton found problems with over reliance on software-supplied default values for some elements.

Timeliness

We use two different terms to refer to two different aspects of metadata timeliness: "currency", and "lag." "Currency" problems occur when the target object changes but the metadata does not. "Lag" problems occur when the target object is disseminated before some or all metadata is knowable or available.

Currency

Stale Uniform Resource Identifiers (URIs) are poster children for problems with metadata currency, but almost any element or value can, in time, become detached from its original target or purpose. Information objects move around, whether on shelves, websites, or conceptual maps of an intellectual discipline. Metadata loses quality over time if it loses synchronization with those movements. Beyond knowing that the metadata is in synchronization with its target object, it is important that synchronization itself has been recently reviewed and verified. This underscores a recurring theme in our analysis: high-quality practices are those that not only accurately describe target objects but also enhance user confidence in the description.

Lag

The dissemination of metadata is not necessarily synchronized with the dissemination of the object to which it applies. New objects take time to describe, categorize, and catalog, particularly if human judgment is involved. The problems thus created become particularly acute if the item being described must be disseminated quickly, leaving metadata lagging behind.

The official citation of judicial opinions provides an instructive if horrifying example. In many jurisdictions, official citation is derived from the page numbers of a printed volume, and hence must wait for the appearance of the volume. In the case of the federal appellate courts in the US, this can take as long as eighteen months from the time the opinion is originally handed down. Many public archives on the Internet never revisit the issue, leaving opinions accessible but without official citation.

TIMELINESS AS AN ARENA FOR CULTURAL DIFFERENCES

The aging of metadata presents obvious problems in the form of potentially broken URIs, drifting controlled vocabularies, and evolving, sometimes divergent, conceptual maps of the underlying corpus. These are problems that are easy to grasp, though not necessarily easy to solve given cultural differences among collection developers and maintainers. There are subtle difficulties rooted in the deeply embedded and divergent expectations that library and computing communities have about audiences, the permanence or persistence of metadata design decisions, and the stability of technology. The cataloging of printed materials is generally done with the expectation that metadata creation is a one-time proposition. Library catalog records are seldom revisited, and new views of metadata are seldom created unless there is great economic or political incentive to do so. As an example, one need only look at the

techniques developed by libraries to avoid costly updates to authority records, i.e. personal name headings, when established authors die. Current practice is to add the death date and source of information to the body of the name authority record, rather than the heading (even when a birth date is already present) so as to avoid the necessity of updating existing catalog records with a new heading with death date.

By contrast, computer technologists come from a world in which techniques are continually changing and often improving. This encourages a more experimental, iterative approach to metadata extraction and other machine-processing efforts. In addition, the more fluid view of metadata offered by searchable and dynamic databases makes audience-customization seemingly more attainable. Unfortunately, this group has been slow to recognize and accommodate practical and efficient updating techniques. As an example, the OAI Protocol for Metadata Harvesting originally proposed a definition of a metadata record "update" that would have required harvesters to unnecessarily replace records in their systems for changes that affected only the administrative portions of the metadata record.

Neither library nor computer technology approach is necessarily better. The point is that different members of a project team will approach data-aging issues with different biases. Some will be inclined toward a "do-it-once-right-and-forget-about-it" approach, dismissing iterative approaches as impractical. On the other hand, there are others who will take an exclusively iterative approach, dismissing front-loaded strategies as unnecessarily expensive and time-consuming. It may then be useful to ask not only what yields the most utility for the user, but what yields the most utility for the user the soonest, and what yields the most robust utility over the long term. Balance between these competing concerns is needed.

Accessibility

Metadata that cannot be read or understood by users has no value. The obstacles may be physical or they may be intellectual. Barriers to physical access come in several forms. Metadata may not be readily associated with the target objects, perhaps because it is physically separated, comes from a different source, or is not properly keyed or linked to the object being described. Or it may be unreadable for a wide variety of technical reasons, including the use of obsolete, unusual or proprietary file formats that can only be read with special equipment or software. In some cases, metadata is considered "premium" information that is accessible only at extra cost to the user, or proprietary information that is not released publicly at all, often because it represents a competitive advantage that the creator or publisher wishes to retain. In other words, the barriers may be economic or trade-related rather than technical or organizational.

It is hard to reduce or eliminate barriers to intellectual access in a world where both objects and metadata are used by multiple audiences and the extent of dissemination is unpredictable. Controlled vocabularies are particularly difficult in this respect. While systems such as the West key-number system for classifying legal materials provide excellent, fine-grained organization for experts, they are of little value to those whose perspective is different – for example,

hospital administrators interested in public-benefits law. There is a need to offer different views or arrangements of metadata to meet the expectations and needs of diverse audiences.

Although metadata providers are powerless to force understanding of any particular element or set of elements on the user, some intellectual barriers can be lowered by careful consideration of potentially diverse audiences when designing and documenting metadata implementations. Above all, one needs to avoid the notion that concise and formal expressions of metadata structure are sufficient documentation in and of themselves. Extensible Markup Language (XML) schemas do not convey thinking or intentions. For proper intellectual access, there needs to be more, in the form of practice guides and other similarly rich forms of documentation.

It is natural to ask which of the seven dimensions previously described is most important, or which most urgently needs to be present for a particular project. Where and how should we begin to foster quality? We believe that the way in which one might prioritize these various criteria is far from uniform, and are dictated by the nature of the objects to be described; whether the implementer is a source provider or an aggregator; and perhaps most importantly how the metadata is to be constructed or derived. Three familiar scenarios illustrate the diversity of options for metadata creation: a collection using author self-submission as the principle means of collecting both data and metadata; a project relying heavily on human judgment to create classificatory metadata; and a project using automated text-extraction techniques to pull metadata from a text corpus.

Each of these methods will have different ways of achieving high-quality results. Assuming a relatively stable corpus (a sometimes-dangerous but reasonable assumption), a computer program that extracts metadata will produce absolutely consistent results over an indefinite period of time, where a churning pool of student employees assigned to a markup project will not. A project where one person classifies information can make some assumptions about coherence and accuracy that a project relying on voluntary submissions cannot. A project that makes use of topical classification only as a means of creating rough boundaries for full-text search or a current-awareness service will not be as concerned about accurate classification as a project that is intended to produce a fine-grained taxonomic survey of a large body of literature.

DEFINING LEVELS OF QUALITY FOR METADATA

Any definition of quality must address attributes of the metadata at several levels: the semantic structure (sometimes called the “format” or “element set”), the syntactic structure (including the administrative wrapper, generally expressed via a “schema”), and the data values themselves. All of these can be validated to some extent by automated means.

For the purposes of this paper, the term “element set” will be used for metadata semantics instead of “format” or “schema,” while “schema” will be used for syntax and syntactic binding. We recognize that “schema” is often used as a more general term referring to both areas, but such use in this context would create unhelpful confusion.

As a practical matter, we begin with the notion that automated metadata validation or evaluation is usually cheaper than human validation. Automated techniques potentially enable humans to use their time to make more sophisticated assessments. Cost-effective machine-based techniques represent "the least we can do" to ensure metadata quality, possibly with more expensive human techniques following on.

We might define a "first tier" of quality indicators as:

- the ability to validate the schema—implying that there is some defined schema, whether XML or some other syntax, that can be checked for validation by programmatic means;
- the use of appropriate namespace declarations—each data element present must be defined within a specified namespace, which may or may not be machine-readable; and
- the presence of an administrative "wrapper" containing basic provenance (metadata identifier, source, and date)— each metadata record so described should carry an identifier that serves to specify it uniquely. In addition, information about the source of the metadata and the date it was created or modified should be present.

Beyond those basics, we might assert that quality of metadata is improved by the presence of the following (noting that "presence" is something that can often be confirmed by automated means):

- controlled vocabularies, expressed by means of unique tokens linked to publicly available sources of terms, such as Internet Mime Types
- elements defined by a specific community as important to discovery of that community's resources, as defined by a publicly available application profile
- a full complement of general elements relevant to general discovery, independent of any particular community, free of assumptions about who will be using the metadata. The five ubiquitous elements identified by Ward probably define an effective minimum for primarily textual objects (incomplete footnote).
- provenance information at a more detailed level including (in addition to source, date, and identifier) information about the methodology used in the creation of the metadata.

Beyond this point, it is less likely that quality determinations can be made by automated means. But the following are nevertheless useful quality indicators:

- an expression of metadata intentions based on an explicit, documented application profile, endorsed by a specialized community, and registered in conformance to a general metadata standard
- source of trusted data with a known history of regularly updating metadata, including controlled vocabularies. This includes explicit conformance with current standards and schemas.

- full provenance information, including nested information, as original metadata is harvested, augmented, and re-exposed. This may not record changes at the element level, but should reference practice documentation that describes augmentation and upgrade routines of particular aggregators.

Applying this system of tiered quality indicators to the seven criteria explained earlier yields the chart below in table 15-1. It is not meant to be a comprehensive procedural checklist or an all-embracing list of indicators. Rather, it is a series of suggested questions that the project manager seeking to create (or assess) quality practices might ask, as well as some indicators he or she might use to answer them. We emphasize approaches and tools that we have found useful, particularly use of "visual view" (graphical analysis software) described in Dushay and Hillmann (incomplete footnote). We expect that those with different experiences will undoubtedly suggest other approaches.

TABLE 15-1 Quality: What to ask for and where to look

Quality Measure	**Quality Criteria**	**Compliance Indicators**
Completeness	Does the element set completely describe the objects?	Application profile; documentation
Completeness	Are all relevant elements used for each object?	Visual view;*; sample
Provenance	Who is responsible for creating, extracting, or transforming the metadata?	OAI server info,† File info, TEI Header‡
Provenance	How was the metadata created or extracted?	OAI Provenance; colophon or file description
Provenance	What transformations have been done on the data since its creation?	OAI About
Accuracy	Have accepted methods been used for creation or extraction?	OAI About; documentation
Accuracy	What has been done to ensure valid values and structure?	OAI About; visual view; sample; knowledge of source provider practices; documentation for creator-

Quality Measure	Quality Criteria	Compliance Indicators
		provided metadata; known-item search tests
Accuracy	Are default values appropriate, and have they been appropriately used?	Known-item search tests; visual view
Conformance to expectations	Does metadata describe what it claims to?	Visual view; external documentation; high ratio of populated elements per record
Conformance to expectations	Are controlled vocabularies aligned with audience characteristics and understanding of the objects?	Visual view, sample, documentation; expert review
Conformance to expectations	Are compromises documented and in line with community expectations?	Documentation; user assessment studies
Logical consistency and coherence	Is data in elements consistent throughout?	Visual view
Logical consistency and coherence	How does it compare with other data within the community?	Research or knowledge of other community data; documentation
Timeliness	Is metadata regularly updated as the resources change?	Sample or date sort of administrative information
Timeliness	Are controlled vocabularies updated when relevant?	Test against known changes in relevant vocabularies
Accessibility	Is an appropriate element set for audience and community being used?	Research or knowledge of other community data; documentation

Quality Measure	Quality Criteria	Compliance Indicators
Accessibility	Is it affordable to use and maintain?	Experience of other implementers; evidence of licensing or other costs.
Accessibility	Does it permit further value-adds?	Standard format; extensible schema

*By "visual view" we mean the process of evaluating metadata using visual graphical analysis tools, as described in the Dushay and Hillmann paper cited earlier.

† Open Archives Initiative (home page)

‡ Text Encoding Initiative (home page), http://www.tei-c.org/ (accessed 28 July 2003)

IMPROVING METADATA QUALITY IN THE SHORT- AND LONG-TERM

Better documentation at several levels has long been at the top of metadata practitioners' wish list. First and most general is the application of standards. Basic standards documents should be accompanied by best practice guidelines and examples. Though such documentation has been prescribed and described many times over, volunteer documentarians remain few. Most projects do not budget for documentation for internal purposes, much less donate time for their staff to create such services for the community at large. Furthermore, many do not expose their internal practices and materials for the use of others. Until this support is forthcoming, and seen as necessary and rewarding (perhaps remunerated) work, it will never be as readily available as all agree it should be.

Better documentation and exposure of local vocabularies used by specialist communities would greatly enhance the willingness of implementers to use them in metadata. Support should include making a vocabulary available in a number of ways, perhaps as harvestable files or via a web interface. In addition, a community process for updating and maintenance must be supported. Admittedly the withering away of the market for printed products and the lack of effective business models for web-accessible vocabularies has led to a fear of revenue loss; this in turn has limited the availability of machine readable vocabulary files for purposes of quality assurance.

Application profiles are just beginning to emerge as the preferred method for specialized communities to interact with the general metadata world. Application profiles, by their nature models created by community consensus, demand a level of documentation of practice that is rarely attempted by individual projects or implementers. Because they mix general and specific metadata elements, they also provide an alternative to the proliferation of metadata standards that re-use the same general concepts with definitions different enough to undermine interoperability.

As critical as documentation is to improving overall quality, cultural change may be even more critical. We must encourage the growth of an implementer and aggregator culture that not only supports better documentation practices, but also sees dissemination of training, tools, methodologies and research results as essential. In the library world, the Library of Congress has spurred efforts in this area, but there is no single organization that can take on this role for the galaxy of specialist communities that could benefit from such leadership. Like the data itself, leadership in the non-library metadata communities is likely to be distributed. Marshalling these organizations and their members to contribute time, server space, technical expertise and training materials for the general good rather than the good of a particular group is a significant challenge. There are already some recognized ways to channel this sort of effort—the Dublin Core Metadata Initiative, OAI and other metadata and technical standards communities offer many opportunities for interested and experienced volunteers, as well as the technical infrastructure for the distribution of tools, ideas and research.

Another essential component to consider is more focused research on practical metadata use and the influence of quality on that use. One current project with a great deal of potential for supplying answers is the National Science Digital Library (NSDL) MetaTest Project headed by Elizabeth Liddy at the Syracuse University School of Information Studies (incomplete footnote). This project seeks to determine how and whether metadata can assist users in locating resources and to understand whether automatically generated metadata is as effective as manually generated metadata in assisting users. This research has the potential of moving the digital library community past the untested, sometimes dogmatic, assumptions underlying current metadata discussions and towards more solid understanding of the role of metadata in future implementations.

CONCLUSION

Another way in which metadata quality resembles pornography is that –as Susan Sontag once remarked -- pornography is a theater of types, and not of individuals. It is difficult to come up with checklists of specific quality-assurance techniques that will apply across a wide range of domains, media types, and funding levels. This may be why we have not tried to do so here. But a playbill (or in our case, a chart) that describes the types is useful to project managers. It gives aggregators hints about where they might look for trouble in legacy and multiple-source data. It can serve as a point of departure for communities and for implementers as they develop standards and documentation, the two indicators of quality most often missing in action. The ubiquity of quality concerns is one sign that the metadata community is growing up. The ability to use generalized thinking across community boundaries as a touchstone for practical solutions will be the next.

NOTES

1. Ward, Jewel Hope. “A Quantitative Analysis of Dublin Core Metadata Element Set (DCMES) Usage in Data Providers Registered with the Open Archives Initiative ,” paper presented at the JCDL 2003. http://foar.net/research/mp/Jewel_Ward-MPaper-November2002.pdf (accessed July 28, 2003).

2. Naomi Dushay, and Diane I. Hillmann. “Analyzing Metadata for Effective Use and Re-Use,” paper accepted for the DC 2003 Conference, http://www.cs.cornell.edu/naomi/DC2003/dushay_hillmann_draft.pdf (accessed July 28, 2003). Dushay & Hillmann characterize their visual graphical analysis techniques as “being able to view the forest as well as the trees”, an important component of effective evaluation of quality. The techniques rely on a combination of automated sorting and display combined with human analysis of graphical and textual output.
3. Jane Barton, Sarah Currier, and Jessie M.N. Hey. “Building Quality Assurance into Metadata Creation: an Analysis based on the Learning Objects and e-prints Communities of Practice,” --Paper accepted for the DC 2003 Conference; William E. Moen, Erin L. Stewart, and Charles R. McClure, “The Role of Content Analysis in Evaluating Metadata for the U.S. Government Information Locator Service: Results From an Exploratory Study,” http://www.unt.edu/wmoen/publications/GILSMDContentAnalysis.htm. (accessed September 19, 2003).
4. “In this paper we seek to challenge four of the assumptions which underlie both the absence of inquiry into how metadata should best be created, and the trend for authors of learning objects and e-Prints to create the metadata for their own resources” (Barton, Currier and Hey, “Building Quality Assurance”). The authors go on to list those four assumptions, among them the idea “that given a standard metadata structure, metadata content can be generated or resolved by machine”.
5. The Program for Cooperative Cataloging developed the BIBCO Core Record standards. Information on the standard is available at http://lcweb.loc.gov/catdir/pcc/bibco/coreintro.html (accessed July 28, 2003).
6. Open Archives Initiative (home page), http://www.openarchives.org (accessed July 28, 2003).
7. IFLA Study Group on the Functional Requirements for Bibliographic Records, “Functional Requirements for Bibliographic Records,” 1997, http://www.ifla.org/VII/s13/frbr/frbr.htm (accessed July 28, 2003).
8. Statistics Canada’s Quality Assurance Framework. (Ottawa: Statistics Canada, 2002), also available at http://www.statcan.ca/english/freepub/12-586-XIE/12-586-XIE02001.pdf (accessed July 28, 2003); Paul Johanis, “Assessing the Quality of Metadata: the Next Challenge,” presentation for a Work Session on METIS, Luxembourg, 6-8 March 2002, http://www.unece.org/stats/documents/2002/03/metis/19.add.2.e.pdf (accessed July 28, 2003).
9. Barton, reporting experience at the Higher Level Skills for Industry (HLSI) Repository, collects a rogue’s gallery of defects in this category, including spelling errors, inconsistent use of terminology, and confusion of content with structure.
10. Dushay and Hillmann, “Analyzing Metadata”.
11. American Library Association, Anglo-American Cataloging Rules, 2d ed., 1988 revision. (Chicago: American Library Association, 1988).
12. Moen, Stewart, and McClure, “Role of Content Analysis.”
13. Moen, Stewart, and McClure discusses many of the same criteria we describe here under the heading of “serviceability,” though the assessment methods used were very much from an end-user perspective. For example, the “serviceability” assessment also

comprehended accuracy problems like misspellings insofar as they interfered with users' ability to retrieve information.
14. Barton, at section 4.1, describing the findings of the HLSI project.
15. Ward, "Quantitative Analysis."
16. Dushay and Hillmann, "Analyzing Metadata."
17. Center for Natural Language Processing, "NSDL MetaTest," http://www.cnlp.org/research/project.asp?recid=21 (accessed 28 July 2003).