Quality Metrics for Learning Object Metadata

Xavier Ochoa
Escuela Superior Politécnica del Litoral, Ecuador
xavier@cti.espol.edu.ec

Erik Duval
Katholieke Universiteit Leuven, Belgium
Erik.duval@cs.kuleuven.be

The quality of the learning objects metadata records stored in a repository is important its operation and interoperability. While several studies have tried to define and measure quality for metadata, a scalable and effective way to assess this quality is not currently available. This work converts the fuzzy quality definitions found in those studies into implementable measures (metrics). Several of these metrics are proposed. They are based on the same quality parameters used for human review of metadata: completeness, accuracy, provenance, conformance to expectations, logical consistency and coherence, timeliness, and accessibility. The information requirements to calculate the proposed metrics are also detailed. Some of these metrics are implemented and tested over to two collection of Learning Object Metadata, one of mainly human generated metadata, the other generated by automated means. Early results suggest that the metrics are indeed sensible to quality features in the metadata. Finally, this work recommends further work to validate and calibrate the proposed metrics.

1. Introduction and background

The quality of the metadata records stored in learning object repositories is perceived as an important issue for the operation and interoperability of those repositories (Barton, Currier, & Hey, 2003; Hughes, 2004; Guy, Powell, & Day, 2004). While there has been a lot of studies about how metadata quality could be measured (see section 2), the fuzziness of quality parameters, mainly designed to guide human reviewers, makes unfeasible to use them consistently and scaleably in real repositories. We need to convert these concepts into implementable measurements in order to be able to convert the metadata quality into something that can be measured and used to implement interesting functionality in Learning Object tools.

We will refer to quality as the measure of fitness for a task. When we refer to measuring quality of metadata we do not assess the quality of the metadata standard (these measurements try to be standard-agnostic) or the vocabularies used. Neither will it evaluate the quality of the learning resource itself. We refer to the quality of the information entered by indexers (human or automated) about a learning object and stored in a metadata record. This information is important, because it is used to recall the indexed learning object, to evaluate its fitness for a specific learning objective and to re-use it in some context.

There have been several studies that are related to quality of metadata in digital repositories:

1. Some of them (Barton et al., 2003; Dushay & Hillman, 2003; Guy et al., 2004) approach the problem from a library point of view: they typically manually review a statistical significant sample by comparing the values with those provided by metadata experts. Dushay and Hillman propose the use of visualization tools to help metadata experts in the task, which is still a manual activity. Human review could be useful for small-sized and slow-growing repositories, but it becomes impractical for large repositories or in the case of automatic indexers (Cardinels, Meire, & Duval, 2005) as humans "do not scale" (Weibel, 2005).
2. A different approach is followed by (Moen & McCluren, 2005; Najjar, Ternier, & Duval, 2003; Zeng, Subrahmanyam, & Shreve, 2004), who try to collect statistical information from the metadata instances in the repositories to evaluate the usage of the metadata standard. While this approach provides some information about the repository, it is not a real indicator of the quality of the metadata. Najjar et al (2004) tried to compare statistics on the metadata values in the repository and the values that are actually used in real searches. A similar approach is followed by the work of (Hughes, 2004). It tries to calculate simple metrics (completeness, vocabulary use) at the repository level for each of the repositories in the Open Language Archive. The Open Language Archive Report Card web page (2005) displays the results of this work.

A practical and useful measurement of quality for learning object metadata should comply with two requirements: to be possible to calculate automatically and to infer, to a certain degree of accuracy, the score that a hypothetical human expert would assign to the metadata record. None of the reviewed works comply with those requirements. The first group requires human intervention which is not scalable, while the second group uses measurements that can be automated, but that do not assess the quality as a whole, as metadata have can be used for other purposes other than just searching (e.g.: automatic assembly, DRM, impact measurement, etc). That is the reason why we propose the creation of more interesting metrics based on the same parameters used by human reviewers. This paper proposes several quality metrics for learning object metadata.

The paper is organized as follows: In section 2, a literature review is conducted to select quality indicators that have been used for manual assessment of metadata. In section 3,we develop several interesting quality metrics, relating them to the parameter that they try to infer. In section 4, two experiments are conducted to assess the proposed metrics. The result of those experiments, as well as some conclusions and possible uses of quality measurements are discussed on Section 5. Finally further work is presented.

2. Measuring the Quality of Metadata

It is difficult to assign an objective value to the quality of information. In order to reduce subjectivity in the assessment of information quality, several researchers have developed quality evaluation frameworks. These frameworks vary widely in their scope and goals. Some have been inspired by the Total Quality Management paradigm (Capezio & Morenhouse, 1995), such as (Strong, Lee, & Wang, 1997), (Dvir & Evans, 1996) and (Kanh, Strong, & Wang, 2002). Others, from the field of text document evaluation, especially of Web documents, are (Xu, 1996), (Segev, 1996) and (Zhu & Gauch, 2000). Particularly interesting for our work, because they are focused on metadata quality, are the frameworks that have evolved from the research on library catalogs, (Stuart, 1995), (Heery, 1996) and (Mangan, 1995).

Each framework differs from the others. Several studies have tried to summarize the recommendations made in previous frameworks to eliminated redundant or overly specific quality parameters. While no consensus has been reached on conceptual and operational definitions of metadata quality there are three main guides to make this evaluation: (Moen, Stewart, & McCluren, 1997), (Gasser & Stvilia, 2001) and (Bruce & Hillman, 2004). Moen et al, [1997] identifies 23 quality parameters. However, some of these parameters (ease of use, ease of creation, protocols, etc) are more focused on the metadata standard or metadata generation tools. Gasser & Stvilia [2001] uses most of Moen's parameters (excluding those not related with metadata quality), adds several more, and groups them in three dimensions of Information Quality (IQ): Intrinsic IQ, Relational/Contextual IQ and Reputational IQ. Some of the parameters (accuracy, naturalness, precision, etc) are present in more than one dimension. The Gasser & Stvilia framework describes 32 parameters in total. Bruce and Hillman [2004] condensed many of the parameters, in order to improve their applicability. They describe seven general characteristics of metadata quality: completeness, accuracy, provenance, conformance to expectations, logical consistency and coherence, timeliness, and accessibility. A comparison table between the frameworks of Bruce & Hillman and Gasser & Stvilia could be found in (Shreeves et al., 2005). In this paper we will use the Bruce & Hillman framework because its compactness will help us to easily operationalize the measurement of quality.

3. Quality Metrics for Learning Object Metadata

While Bruce and Hillman [2004] devised its framework to guide human reviewers, we will try to generate automated metrics to assess each one of the parameters describe in the framework. In this paper, the LOM standard (IEEE, 2002) is used, but other metadata schemas to represent information about learning objects or other digital resources could rely on the same methodology. The goal of these metrics is to be easily implementable in real environments and to produce a quality value that correlates well with the one produced by a human expert using the same framework.

3.1. Completeness Metrics

Completeness is the degree to which the metadata record contents all the information needed to have an ideal representation of the described object. This ideal representation varies according to the application and the community of use. While in LOM define all of its fields as optional, a working definition of the ideal representation could be considered as the mandatory and suggested fields defined by a community of use in the application-profiles (Najjar et al., 2003) of LOM. A first approach to assess the completeness of a metadata record will be to count the number of fields that content a no-null value. In the case of multi-valued fields, the field is considered complete if at least one instance exists.

$$Q_{completness} = \frac{\sum_{i=1}^{N} P(i)}{N}$$

Where P(i) is 1 if the ith field has a non-null value, 0 otherwise. N is the number of fields.
While straightforward, this metric does not reflect quality very well: not all data elements are relevant for all learning objects in all contexts. Moreover, not all data elements are equally relevant to all contexts. For example, a human expert may assign a higher degree of completeness to a metadata record that has a value for the title, but lacks one for coverage than to a metadata record that includes coverage, but lacks title. Human experts often assign a weight to each data element, representing its relative importance compared to other fields. This weighting factor can easily be included in the formula:

$$Q_{Wcompletness} = \frac{\sum_{i=1}^{N} \alpha_i * P(i)}{\sum_{i=1}^{N} \alpha_i}$$

Where α_I is the relative importance of the i^{th} field.
The α values should represent the importance (or relevance) of the data element for some context. This implies that a difference weighted completeness value could be calculated for different contexts. For example, the α_I value could represent the frequency with which searches have relied on that data element in a given period of time. Alternatively, the α_I value could represent the score that the ith field obtained in a survey that assesses the relevancy of data elements to evaluate learning objects.

3.2. Accuracy Metrics

The accuracy is the degree to which the metadata values are "correct", i.e. how well they describe the object. The correctness could be a binary value, either "right" or "wrong", for objective information like file size or language, but, in the case of subjective information, it is a more complex spectrum with intermediate values (e.g.: A title of a picture, or the description of the content of a document). In general, the correctness and, thus the accuracy, could be considered as the semantic distance between the information that a user could extract from the metadata record and the information that the same user could obtain form the document itself. The shortest the distance, the higher the accuracy of the metadata.

Humans can asses with relative ease the accuracy of a metadata record. Computers, however, require complex artificial intelligence algorithms to simulate human understanding. Nevertheless, easy accuracy metrics are possible: one approach is to search for metadata already stored in the learning object or values that are easily obtainable. For example, if the learning object is a JPEG picture, we can use the information stored in EXIF (JEITA, 2002) to measure the accuracy of the LOM record. As an example, LOM contains the creation date. That date could be compared with the date the picture was taken, expressed in the EXIF header. After all such distances have been calculated, they can be added using a variation of the Euclidean distance.

$$Q_{accuracy} = 1 - \frac{\sqrt{\sum_{i=1}^{N} d(field_i)^2}}{\sum_{i=1}^{N} d(field_i)} \quad \text{(given that } \sum_{i=1}^{N} d(field_i) > 0\text{)}$$

The difficulty here is to define the distance measurement d(field). For different media types, a different function can be used. For numbers and dates it could be easily computed. The distance is equal to multi layered difference between values (e.g.: d(field)=0 if is exactly the same, 0.5 if it is in a 10% off, 0 otherwise). For categorical values (e.g.: human language of the object), it could be based in a pre-generated distance table (e.g.: the distance between Spanish and Italian could be set at 0.2, between Spanish and Japanese, at 1). Free text values require a more complex calculation. One approach for text will be to calculate the distance between the words present in the text values. First, we create a multi-dimensional space using the words used in both texts as dimensions. Each text can be represented as a vector, where the ith value represent the presence or absence of the word i in the text. Then we can use any method to calculate the distance between vectors, such as Euclidian distance or Cosine Similarity. The formula presents the Cosine similarity function:

$$d(textfield1, textfield2) = \frac{\sum_{i=1}^{N} tf1vector_i * tf2vector_i}{\sqrt{\sum_{i=1}^{N} tf1vector_i^{\,2} * \sum_{i=1}^{N} tf2vector_i^{\,2}}}$$

Where tf1vectori and tf2vectori, will be 1 if the text contains the word, 0 if not. N is the total number of different words in both texts.
This accuracy approach can be valid when the object has some extra metadata attached. It there are no additional metadata, the only remaining source of information is the content of the object itself. For example, a semantic distance can be calculated between the title or description metadata and the text present in a document. For this calculation, advanced text analysis algorithms can be used, such as Latent Semantic Analysis (LSA) (Landauer, Foltz, & Laham, 1998). For other types of media (pictures, videos, audio), there is still not a readily available computation to determine the semantic distance of the object with a text that describe it.

Is the perception of the authors, that while a simple accuracy metric for metadata (especially for metadata that describe text documents) could be implemented with existing algorithms, a more meaningful and general accuracy metric that could grade any metadata record is a complex task that could not be implemented with the current state of Information Extraction technologies. Nonetheless, the application of the simple accuracy metric could give at least some useful insight to spot and understand indexation errors.

3.3. Provenance Metrics

Provenance covers the reputation that a metadata record has in a community. For example, a user may trust more metadata generated by a metadata expert that he knows, than metadata generated by a software tool. While the automated generated metadata may be of a better quality (according to the other metrics), provenance is more related to the subjective perception that the user has about the origin of the metadata.

In order to be able to capture this subjective perception, a comprehensive system to log actual usage and ratings is needed. This system should include components that actively register the approval of the user (for example, asking the user if the metadata shown has been useful) or that infer that approval from the actions of the user (for example, which metadata records have lead to more downloads or actual use of objects). This information should be collected from the different tools and stored in a repository (or a group of interoperable repositories). The information that we recommend these systems should log is described in Table 1 in order of relevance for this metric. The table also shows how to convert the log information into a provenance metric.

Table 1. Information needed to compute Provenance Metric

Information to be logged:	The provenance metric is equal to:
The rating that the user gives to the metadata-producer	The normalized rate of its metadata-producer
The rating that the user gives to a metadata record	The average of the rate of the other objects belonging to the same metadata-producer
Network of trusted friends of a user (in XFN (GMPG, 2003) for example)	The normalized inverse distance between the user and the metadata-producer
The objects downloaded by a user (reviewing of the metadata record)	The normalized value of the number of downloaded objects that belongs to a metadata-producer
The objects integrated (inside an LMS for example) by a user	The normalized value of the number of integrated objects that belongs to a metadata-producer

While such usage repositories are not widespread at this time, ongoing work on Attention.XML (Technorati, 2004), (Attention Trust, 2005) and (Najjar, Meire, & Duval, 2005) and logging repositories (Broisin, Vidal, & Duval, 2005) suggest that this approach is feasible.

3.4. Conformance to Expectations Metrics

The conformance to expectations measures the degree in which the metadata record fulfills the requirements of a given community of users. There are several parameters that affect this quality: The vocabularies words used in the metadata record should be meaningful for the user, the metadata fields filled the ones needed to perform the task that the user intended (search, evaluation, integration, etc), the amount of information is enough to describe the learning object. As LOM solves the problem of vocabulary terms, and the weighted completeness metrics asses the second problem, we will develop a metric to assess the third issue.

One of the main requirements of any community towards the LOM record is that it contains enough information to describe uniquely its referred learning object. A human would consider that the metadata is of high quality if after reading it, he/she knows (or he/she thinks he/she knows) what the learning object is about and what it contents. While there is no computational algorithm that could claim to be able to measure the grade in which the metadata describes the object, the amount of useful (unique) information present in the metadata relative to the repository where it belongs could be estimated using Information Theory. A proposed conformance-to-expectation metric could be calculated measuring the Information Content of the metadata fields.

$$Q_{conformance} = \frac{\sum_{i=1}^{N} Icontent(field_i)}{N}$$

Where N is the number of fields and Icontent(fieldi) is the estimation of the amount of unique information contained by the field ith.

Each type of field (free text, numerical values or vocabulary values) has a different way in which to calculate the amount of information. For categorical fields, the Information Content is equal to 1 minus the entropy of the value (the entropy is the negative log of the probability of the value) (Resnik, 1995). For numerical values it can be calculated as 1 minus the entropy (in this case the negative probability of the value assuming a normal distribution). For free text, on the other hand it is more difficult to handle because the calculation of the "importance" of a word is directly proportional to how much that word appear in the document and inversely proportional to how much documents contain that word. This relation is handled by the Term Frequency-Inverse Document Frequency (Aizawa, 2003) calculation. The number of times that the word appear in the document is multiplied by the negative log of the amount of documents that contain that word (could be considered as a weighted entropy measurement).

$$Icontent(categorical\_field) = -\log(P(value))$$

$$Icontente(numerical\_field) = -\log(P(value)) \text{ (in a normal distribution)}$$

$$Icontent(freetext\_field) = \frac{\sum_{i=1}^{N} term\_frequency(word_i) * \log\left(\frac{1}{document\_frequency(word_i)}\right)}{N}$$

This metric could be refined assigning a weighting factor to account for different importance of the fields (similar to has been done in the Weighted Completeness metric). In this way, for example, the Information Content of 4.7 Duration will be less important than the Information Content of 1.2 Title. These importance values could be determined by an expert or be based on the preference of the users

3.5. Logical consistency and coherence Metrics

The logical consistency and coherence is the degree to which a metadata record matches a standard definition and the values used in the fields, correlate positively among them. For the particular case of Learning Objects, the first part of this metric should measure how well the metadata record adjusts to the LOM standard. This calculation is trivial using any validation parser (for example Xerces (Apache Foundation, 2005)), when the metadata record is in the XML binding of LOM. The second measurement is more subjective, trying to asses the internal consistent and coherence of the information stored in the metadata record.

While not being mandatory, LOM suggest certain combination of values to maintain the internal consistency of the record. For example, it is recommended that if the value of 1.7 Structure is "atomic", the 1.8 Aggregation Level should be marked as 1, other structure value could be pared with 2, 3 o 4 in the aggregation field. Table 2 present other internal checks that could be followed to asses the consistency of the record.

Table 2. Fields to compare to access Consistency

Field 1	Field 2	Example
1. 7 Structure	1.8 Aggregation Level	Structure=atomic => an aggregation level=1
5.1 Interactivity Type	5.3 Interactivity Level	Interactivity type=active => high values of

		Interactivity level.
5.2 Learning resource type	5.1 Iteractivity Level	Learning resource type=narrative text => interactivity level=expositive
5.4 Semantic Density	5.8 Difficulty	A high semantic density => a high difficulty
5.6 Context	5.7Typical Age Range	If context=higher education=>age range should start at least at 17 years

To obtain a numerical value of this analysis we could create a function that return 1 when the restrictions are met, and reduce its value based on the distance of the observed value from the desired one.

$$Q_{consistency} = \frac{\sum_{i=1}^{N} level\_of\_compliance(field_i, value_i)}{N}$$

Where N is the number of fields to be analyzed.

The coherence could be measuring analyzing the free text fields. A coherent metadata record describe to the same topic in title, description and keywords. To assess this coherence, the semantic distance is calculated between the free text fields. To obtain that measure, we will use the LSA technique (Foltz, Kintsch, & Landauer, 1998).

3.6. Timeliness Metrics

The timeliness relates specially to the degree to which a metadata record remains current among certain community. The currency of LOM record could be measured as the how useful the metadata remains with the pass of time. For example, if a the LOM describing a course was created 5 years ago, and the users could still find and correctly evaluate the content of the course and download them, then the metadata could be considered current. On the other hand, if the metadata record misleads the users, because the referred object has change to the point where the contents differed from the original, the metadata registry is obsolete and must be replaced.

An automated metric to assess the currency of the object is necessarily composed by several calculations, because it needs to weight several factors. First the accuracy of the metadata record; second the age of the metadata (older metadata have a higher probability of being obsolete) and how often the metadata is used (appear in a search).

$$age = present\_year - publication\_year$$

$$frequency\_of\_use = \frac{times\_retrived}{total\_records\_retrieved} \text{ (Over a period of 1 year)}$$

$$Q_{currency} = Q_{acurracy} * age(\ ) * frequency\_of\_use(\ )$$

To calculate this metric, the repository of usage information, described in the provenance metric should exist.

3.7. Accessibility Metrics

Accessibility measure the degree to which a LOM metadata is accessible, both in terms of cognitive accessibility as well as physical/logical accessibility. The cognitive accessibility measure how easy is for a user to understand the information contained in the metadata record. In librarian review of this characteristic (Guy et al., 2004) several simple metrics are used: measuring spelling errors, conformance with the vocabulary, but they always include a human evaluation of the difficulty of the text. The difficulty assessment could be automated using one of the available readability indexes, for example the Flesch Index (Foltz et al., 1998), principally to analyze the description of the learning object. Readability indexes access the difficulty of a text based in the characteristics of the sentences and the words used.

$$Q_{readability} = \frac{Flesch(description\_text)}{100}$$

The physical/logical accessibility could be understood as how easy is to find the LOM record in a repository. While it could be possible to automatically access this characteristic using the frequency of retrieval (explained in 3.6), it just represent the present accessibility with the current searching tool. A more interesting (and accurate) metric should access the potential accessibility regardless of the accessing tool. We propose the use of the linkage of an object as its accessibility measure. The linkage value of a record is equal to the number of other records that reference to it. The reference information is stored in the section 7 of the LOM standard.

$$Q_{linkage} = \frac{number\_of\_ongoing\_links}{average\_links\_in\_the\_repository}$$

Unfortunately, the Section 7 of LOM is hardly used in the repositories (ISO/IEC JTC1 SC36, 2004). That is because linking information is difficult to maintain manually. Automated Indexers promise to be able to generate and maintain this linkage information (Ochoa, Cardinels, Meire, & Duval, 2005).

4. Implementation and evaluation of the Metrics

The proposed metrics were tested over two different samples of data, one is a collection of 4000 LOM records of mostly manually generated metadata from ARIADNE repository (Ariadne Foundation, 2005b), the other is a batch of 1000 LOM records automatically generated from the Open CourseWare (Ariadne Foundation, 2005b) documents by the AMG framework (Cardinels et al., 2005). The metrics have been normalized to a 0 to 10 scale. Details of the implementation can be found in (Ariadne Foundation, 2005a). Table 3 presents a summary of the results. Accuracy and Coherence were not implemented because the lack of availability of a tool to perform LSA in the test implementation. In these early evaluation experiments, the lack of usage information made it impossible to calculate the provenance metric and imposed some restrictions (the frequency of use was not available) on the calculation of the currency of the metadata record.

Table 3. Results of the quality metrics applied to two sets of metadata records

Metric # instances	ARIADNE 4406	OCW 992
Completeness	6.43	4.03
Weighted Completeness	9.99	8.05
Consistency	10.00	10.00
Conformance	2.08	4.76
Currency (only age)	5.53	10.00
Readability	0.99	5.42
Linkage	0.00	0.00
Average (no weighting)	5.00/10	6.01/10

5. Result Analysis and Conclusions

The only purpose of the experiment was to test the feasibility of the implementation of the metrics. The metrics need validation in order to be considered viable to assess the quality of a repository data. In order to be able to validate the metrics, a complete comparative study should be carried out, correlating the quality values assigned by human experts in metadata and the quality values generated by the metrics (in each quality parameters and combined). We will undertake this analysis in further work. Nonetheless, a preliminary analysis of the results can be performed:

- ARIADNE obtained a higher completeness than OCW. That is easily explained as ARIADNE metadata has filled all the mandatory fields in the ARIADNE application-profile. In the case of OCW, some of those fields were missing, because the AMG extractor did not generate them.
- The weighted completeness was calculated using the times that the fields were used in search as reported in (Najjar, Ternier, & Duval, 2004). As those fields were mainly the mandatory fields, ARIADNE obtained an almost perfect score, while OCW is staying behind.
- Consistency looked for inconsistencies in the fields, but because the indexing software used of both cases prevent those errors, no inconsistencies were found.
- Conformance metric took in consideration the uniqueness of categorical and textual information in the metadata records. The advantage of OCW could be explained as all the OCW records have a lengthy description, while only ~500 objects have any description at all. The low numbers obtained in this metric could be a signal that the normalization factors used should be calibrated.
- The lack of descriptions is also a reason for the extremely low Readability value of ARIADNE (absence of text result in a low grade). The 5.42 obtained for OCW obey to the difficult terms used.
- The currency metrics gives almost no information as it is only calculated with the age of the metadata record. In the case of ARIADNE, it seems that the metadata records have an average age of four and a half years, while the OCW records were generated just 1 week before the experiment.
- The linkage shows that neither collection has reference information between objects.

The higher final result of OCW was expected as the metadata records seems (no formal proof could be given at these moment) of better quality (especially because of a better description) that the metadata stored at ARIADNE.

The quality metrics proposed in this work demonstrate that it is possible to operationalize quality parameters. The results of the experiments performed suggest that the proposed metrics are sensitive to quality aspects of the metadata records, because their values reflect identifiable quality flaws in the measured metadata. The quality score produce could be attached to the metadata records and used to implement interesting functions as:

- **Selective publishing:** When a metadata instance is submitted, a quality analysis could determine whether the record complies with a minimum standard to be accepted. This could be especially important for automatic generators of metadata.
- **Filtering of Federated search results:** A federated search across multiple repositories could establish a minimum level of quality for records to be included in search results.
- **Help in the ranking of search results:** The quality of the metadata record could be a component of the ranking algorithm used in search tools.
- **Selection of the most appropriate metadata instance:** A learning object can have several metadata instances that describe it. The quality could be taken in account when selecting the most appropriate record for each situation (search, evaluation, assembly, etc).

6. Further Work

The fitness of the proposed metrics needs to be validated in much more detail. This evaluation is left for a future work. Once the metrics are successfully validated, they can be used to make comparative qualitative analyses of metadata. For example, such metrics will allow us, in a scientifically sound way, to answer the question: "how good is automated generated metadata?"(Ochoa et al., 2005)

7. References

- Barton, J., Currier, S., & Hey, J. (2003). Building Quality Assurance into Metadata Creation: an Analysis based on the Learning Objects and e-Prints Communities of Practice. In Proceedings 2003 Dublin Core Conference: Supporting Communities of Discourse and Practice - Metadata Research and Applications (pp. 39-48). Seatle, Washington.
- Hughes, B. (2004). Metadata quality evaluation: Experience from the Open Language Archives Community. Digital Libraries: International Collaboration and Cross-Fertilization, Proceedings, 3334, 320-329.
- Guy, M., Powell, A., & Day, M. (2004). Improving the quality of metadata in Eprint archives. Ariadne.
- Dushay, N. & Hillman, D. (2003). Analyzing Metadata for Effective Use and Re-Use. In International DCMI Metadata Conference and Workshop.
- Cardinels, K., Meire, M., & Duval, E. (2005). Automating metadata generation: the simple indexing interface. In Proceedings of the 14th international conference on World Wide Web (pp. 548-556). New York, NY, USA: ACM Press.
- Weibel, S. (2005). Border Crossings: Reflections on a Decade of Metadata Consensus Building. D-Lib Magazine, 11.
- Moen, C. & McCluren, C. (2005). An Evaluation of the Federal Government's Implementation of the Government Information Locator Service (GILS) (Rep. No. General Services Administration Office of Information Technology Integration Washington, DC).
- Najjar, J., Ternier, S., & Duval, E. (2003). The actual use of Metadata in Ariadne. In Proceedings of the 3rd Annual ARIADNE Conference (pp. 1-6).
- Zeng, M. L., Subrahmanyam, L., & Shreve, G. M. (2004). Metadata quality study for the National Science Digital Library (NSDL) Metadata Repository. Digital Libraries: International Collaboration and Cross-Fertilization, Proceedings, 3334, 339-340.
- The Open Language Archive Report Card (2005). OLAC: Open Language Archives Community [On-line]. Available: http://www.language-archives.org/tools/reports/archiveReportCard.php?archive=1
- Capezio, P. & Morenhouse, D. (1995). Taking the Mystery out of TQM. Career Press.
- Strong, D., Lee, Y., & Wang, R. (1997). Data quality in context. Communications of the ACM, 40, 103-110.
- Dvir, R. & Evans, S. (1996). A TQM Approach to the Improvement of Information Quality. In Proceedings of the 1996 Conference on Information Quality (pp. 207-220).
- Kanh, B., Strong, D., & Wang, Y. (2002). Information quality benchmarks: Product and service performance. Communications of the ACM, 45, 184-192.
- Xu, A. (1996). Accessing Information on the Internet: Feasibility Study of USMARC Format and AACR2. In Proceedings of the OCLC Internet Cataloging Colloquium.
- Segev, A. (1996). On Information Quality and the WWW Impact: A Position Paper. In Proceedings of the Conference on Information Quality (IQ'96) Cambridge, MA, USA: Massachusetts Institute of Technology, Sloan School of Management.

- Zhu, X. & Gauch, S. (2000). Incorporating quality metrics in centralized/distributed information retrieval on the world wide web. In Proceedings of the 23rd Annual International ACM/SIGIR Conference (pp. 288-295).
- Stuart, E. (1995). Fitness for purpose: The Future Evolution of Bibliographic Records and Their Delivery. Catalogue & Index, 116.
- Heery, R. (1996). Review of Metadata Formats. Program, 30.
- Mangan, E. (1995). The Making of a Standard. Information Technology and Libraries, 14, 99-110.
- Moen, W., Stewart, E., & McCluren, C. (1997). The Role of Content Analysis in Evaluating Metadata for the US Government Information Locator Service (GILS): results from an exploratory study. http://www.unt.edu/wmoen/publications/GILSMDContentAnalysis.htm [On-line].
- Gasser, L. & Stvilia, B. (2001). A new framework for information quality (Rep. No. Technical report ISRN UIUCLIS--2001/1+AMAS.). Champaign, Ill, USA: University of Illinois at Urbana Champaign.
- Bruce, T. & Hillman, D. (2004). The continuum of metadata quality: defining, expressing, exploiting. In D.Hillman & L. Westbrooks (Eds.), Metadata in Practice (Chicago: American Library Association.
- 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 Proceedings of the Twelfth National Conference of the Association of College and Research Libraries Chicago, IL: Association of College and Research Libraries.
- IEEE (2002). IEEE Standard for Learning Object Metadata. http://ltsc.ieee.org/doc/wg12/ [On-line].
- JEITA (2002). EXIF standard. http://www.exif.org/ [On-line].
- Landauer, T., Foltz, P., & Laham, D. (1998). Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284.
- GMPG (2003). XFN: XHTML Friends Network. http://gmpg.org/xfn/ [On-line].
- Technorati (2004). Attention.XML specification. http://developers.technorati.com/wiki/attentionxml [On-line].
- Attention Trust (2005). Attention Trust. http://www.attentiontrust.org/ [On-line].
- Najjar, J., Meire, M., & Duval, E. (2005). Attention Metadata Management: Tracking the Use of Learning Objects through Attention.XML. In Proceedings of ED-MEDIA 2005 World Conference on EducationalMultimedia, Hypermedia and Telecommunications.
- Broisin, J., Vidal, P., & Duval, E. (2005). A Model driven Approach for the Management of Learning Object Life Cycle. In Proceedings of Intelligent, Interactive, Learning Object Repositories Network (I2LOR).
- Resnik, P. (1995). Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In Proceedings of International Joint Conferences on Artificial Intelligence (pp. 448-453).
- Aizawa, A. (2003). An information-theoretic perspective of tf–idf measures. Information Processing and Management, 39, 45-65.
- Apache Foundation (2005). Xerces Parser. http://xerces.apache.org/ [On-line].
- Foltz, P., Kintsch, W., & Landauer, T. (1998). The measurement of textual coherence with Latent Semantic Analysis. Discourse Processes, 25, 285-307.
- ISO/IEC JTC1 SC36 (2004). International LOM Survey: Report.
- Ochoa, X., Cardinels, K., Meire, M., & Duval, E. (2005). Frameworks for the Automatic Indexation of Learning Management Systems Content into Learning Object Repositories. In Proceedings of ED-MEDIA 2005 World Conference on EducationalMultimedia, Hypermedia and Telecommunications (pp. 1407-1414).
- Ariadne Foundation (2005). Ariadne Foundation. http://www.ariadne-eu.org [On-line].
- Ariadne Foundation (2005). Quality Metrics for ARIADNE. http://ariadne.cti.espol.edu.ec/QMetrics [On-line].
- Najjar, J., Ternier, S., & Duval, E. (2004). User Behavior in Learning Objects Repositories: An Empirical Analysis. In Proceedings of ED-MEDIA 2004 World Conference on EducationalMultimedia, Hypermedia and Telecommunications.