International Conference on
Dublin Core and Metadata Applications
22 - 26 September 2008

A Conceptual Framework for Metadata Quality Assessment

Margaritopoulos Thomas (margatom@uom.gr)
Margaritopoulos Merkourios (mermar@uom.gr)
Mavridis Ioannis (mavridis@uom.gr)
Manitsaris Athanasios (manits@uom.gr)

Presentation Outline

- Related research on metadata quality
- Motivation
- The proposed framework
- A method for quality assessment
- Application of the method
- Conclusion – future work

Related research on metadata quality

- Statistical analysis on samples of metadata records – evaluation of the usage of the standard
- General principles and guidelines for the process of the creation of metadata
- Use of application profiles for quality assurance
- Identification of deficiencies that degrade quality
- Quality Assessment - Identification of quality parameters
- Metrics for quality evaluation

Metadata Quality Assessment

Metadata Quality

Quality parameters

- Accuracy
- Validity
- Cohesiveness
- Complexity
- Consistency
- Conformance to expectations
- Currency
- Provenance
- Completeness

…

…

…

Questions Arisen

- Could someone prove that he/she has found an exhaustive list of quality parameters?
- Could these parameters serve as a proof for quality?
- Are they sufficient to guarantee quality?

Metadata Quality Assessment

Metadata Quality

Quality parameters

- Accuracy
- Validity
- Cohesiveness
- Complexity
- Consistency
- Conformance to expectations
- Currency
- Provenance
- Completeness

...

...

...

The concept of metadata quality

What is metadata quality?

Or

What are the fundamental components of metadata quality?

Remember what metadata is?

The purpose of metadata is to provide adequate and correct information to their user so as to obtain a true picture of the content of a resource without having to access it.

- Metadata serve as the “mirror” of the resource
- True representation of the resource
- Absence of any distortion of its picture
- No previous knowledge of the resource

The Court Metaphor

Metadata world	Court of law
resources in a repository	facts of a case
metadata record (one record for each resource)	testimony for a single fact (one witness for each fact)
record fields	different aspects of the fact
metadata quality assessment	evaluation of the testimonies

Quality Evaluation

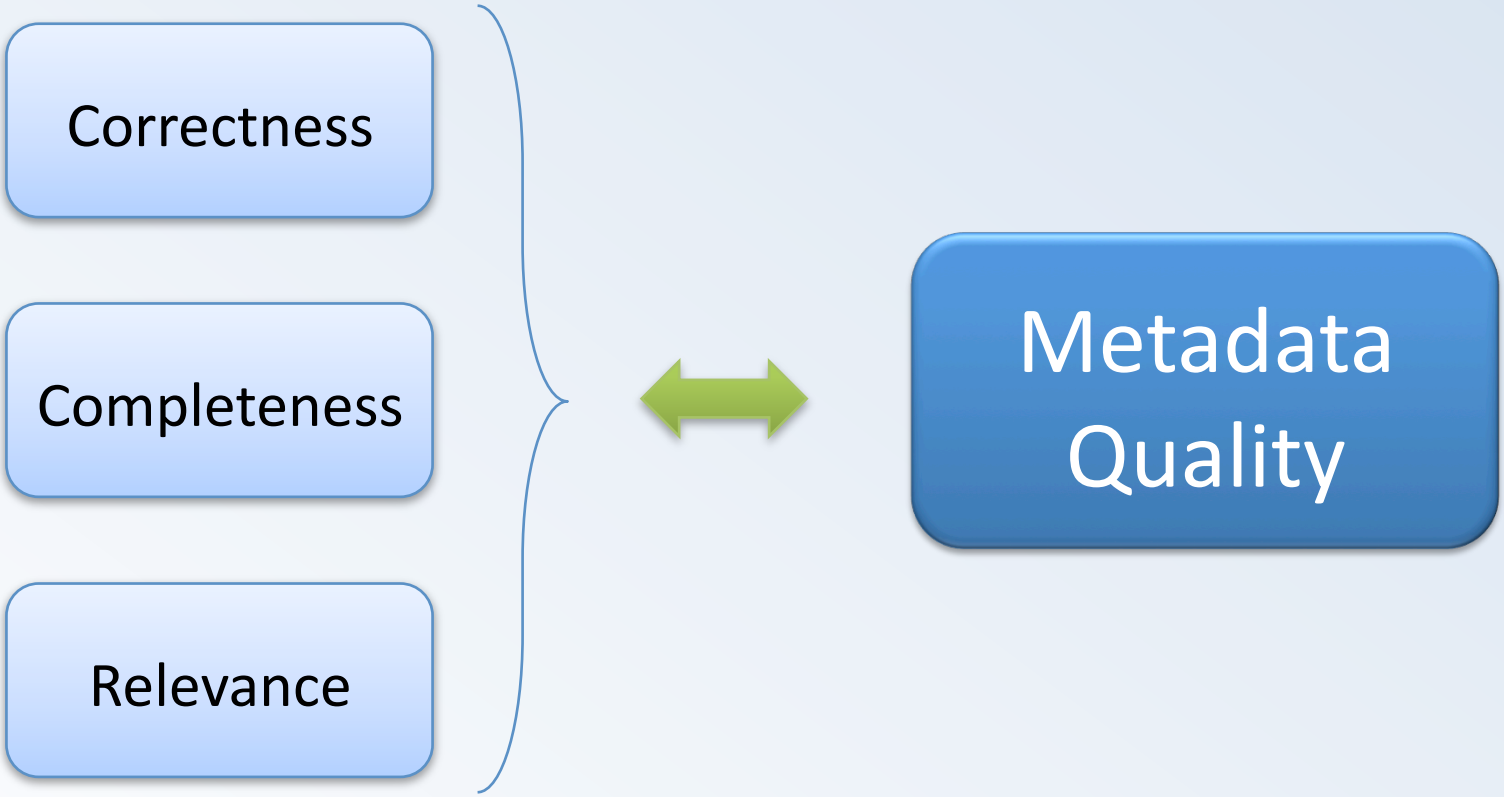
Testimony

- "…the truth,…"
- "…the whole truth,…"
- "…and nothing but the truth."

Metadata Record

- Correct
- Complete
- Nothing Irrelevant

Metadata Quality Components



Correctness (1/2)

Low Level

Values of the metadata fields must obey the grammatical and syntactical rules of the language and the metadata standard or the application profile used.

- Court metaphor: witness must be able to properly use the language so as to set the testimony fully understandable
- Objective

Correctness (2/2)

High Level

Semantical rightness of the values of the metadata fields (true representation of the reality and absence of any distortion).

- Court metaphor: the truthfulness of the testimony
- Subjective

Completeness

Sufficiency to fully describe a resource

- Court metaphor
  - choice of the questions (judge's task)
  - provision of answers to the questions (witness's task)
- Metadata world
  - choice of the metadata fields (application profile's task)
  - populating the fields with values (indexer's task)

Relevance

The compatibility with the context of use of the metadata in a given application

- Court metaphor: different perspectives between the question and the answer
- Subjective

Correctness or Relevance?

- Discriminating factor: the context
- An incorrect value is faulty regardless of the context, while an irrelevant value is associated with a particular context.
- A faulty value in a metadata field with a range of values out of a vocabulary will be, more possibly, attributed to incorrectness, rather than irrelevance.

A Method for Metadata Quality Assessment

Court Metaphor

- Check for contradicting descriptions...
  - regarding the aspects of a single fact.
  - regarding related facts.
- Any such contradictions violate implied logic rules and cause the testimonies to be considered unreliable.

A Method for Metadata Quality Assessment

Metadata World

- Check for dependencies among fields…
  - inside a single record (intra-record)
  - of records of related resources (inter-record)
- The dependencies of fields create implied logic rules.
- The method constitutes checking the validity of these rules.

Dependencies of metadata fields (intra-record)

- In some cases, the fields of a metadata record present some sort of correlation.
- The dependency is implicitly (if not explicitly) imposed by the specifications of the standard.

Example

LOM standard specifies that *«…a learning object with Structure="atomic" will typically have AggregationLevel=1…»* thus, setting fields "1.7 General.Structure" and "1.8 General.Aggregation Level" directly interdependent.

Dependencies of metadata fields (inter-record)

- Metadata fields of related resources might be influenced by each other.
- The influence is expressed as propositions in the form of logic rules.
- These logic rules are based on the semantics of the relations and the metadata.
- The rules constitute a set of validation principles that quality metadata must conform to.

Quality Assessment Logic Rules

Rules of Inclusion

The resource's metadata field values must include the values of the same metadata field of records of related resources

Example (using DC)

If resource a ***hasPart*** resource b, then all ***<DC:contributor>***s of b will be ***<DC:contributor>***s of a

Quality Assessment Logic Rules

Rules of Imposition

The resource's metadata field values must be equal to the result of a mathematical or logic expression of metadata field values of the records of related resources

Example (using LOM)

If learning object a ***IsRequiredBy*** learning object b, then "***5.11 Educational.Language***" of a must be equal to "***5.11 Educational.Language***" of b

Quality Assessment Logic Rules

Rules of Restriction

The range of a resource's metadata field values is not the complete value space defined by the standard, but a proper subset of it, computed from the values of the same metadata field of records of related resources

Example (using LOM)

If learning object a ***IsPartOf*** learning objects b, c, d,…
then "***5.5 Educational.Typical learning time***" of a must be < min("***5.5***" of b, "***5.5***" of c, "***5.5***" of d,…)

Application of Quality Assessment Logic Rules

The field ***<dc:language>*** of a resource must include the values of the same metadata field of its parts (***<dcterms:hasPart>***).

a ***hasPart*** b

a	b	Quality problem
en	en	-
	en	completeness(a)
en	en, fr	completeness(a), correctness(b)
en	fr	correctness(a), correctness(b), completeness(a)

Remarks

- Reduced set of resources with degraded quality
- No field is considered to be of high quality in advance
- The set of the related fields with problematic quality is the limit of the quality assessment's "granularity"
- Combination with other methods of metadata quality assessment in an integrated system

Presentation Review

- Related research
- Court metaphor
- Metadata quality components
- Method for quality assessment
- Application example
- Prospects for future work

Thank you!