DATA QUALITY THROUGH CONCEPTUAL MODEL QUALITY - RECONCILING RESEARCHERS AND PRACTITIONERS THROUGH A CUSTOMIZABLE QUALITY MODEL

(Completed Research Paper)

Kashif Mehmood
ESSEC Business School, Avenue Bernard Hirsch B.P. 50105, 95021 Cergy-Pontoise Cedex, FRANCE
CEDRIC-CNAM, 292 Rue Saint Martin, F-75141 Paris Cedex 03, France
Kashif.Mehmood@essec.fr

Samira Si-Said Cherfi
CEDRIC-CNAM, 292 Rue Saint Martin, F-75141 Paris Cedex 03, France
Samira.Cherfi@cnam.fr

Isabelle Comyn-Wattiau
CEDRIC-CNAM, 292 Rue Saint Martin, F-75141 Paris Cedex 03, France
ESSEC Business School, Avenue Bernard Hirsch B.P. 50105, 95021 Cergy-Pontoise Cedex, FRANCE
Isabelle.Wattiau@cnam.fr

Abstract: Data quality has emerged as an important and challenging topic in recent years. In this article we are addressing the conceptual model quality since it has been widely accepted that better conceptual models produce better information systems and thus implicitly improve the data quality. Unfortunately there is neither standard nor agreed framework for managing conceptual models quality. This article presents an overview of existing approaches with their advantages and limitations. It then proposes a comprehensive model for evaluating the quality of conceptual models. A survey involving practitioners has been used as an initial validation. This validation exercise aims to collect the responders' views on the holistic quality of the conceptual models in addition to their feedback over the newly proposed model. The received feedback has been evaluated and incorporated to the quality model. Furthermore, we propose a general approach for quality evaluation and improvement.

Key Words: Data Quality, Conceptual Modeling Quality, Quality Criteria, Quality Framework, Quality Metrics, Quality Evaluation, Quality Improvement

1. INTRODUCTION

Information Systems (IS) require high cost for their maintenance activities and therefore software quality is considered as an important issue in research laboratories and in IS firms. The relative cost for maintaining software and managing its evolution represents more than 90% of the total cost [4]. Many efforts are devoted towards the research and development of the methods to improve the software quality. Similarly, ensuring data quality is of utmost importance since data plays vital role in any information system and is central to any decision process. Consequently, data quality became an important and a challenging topic in recent years.

There are several means to improve data quality. Indeed, the cost relative to corrective actions aiming to improve data quality is rather high. In some situations, the data quality problems could be avoided by defining convenient constructs at the model level. A research stream addresses the problem of quality at the model level. The underlying hypothesis is that "good" models lead to "good" data. The approaches propose methods, metrics and tools for data model quality evaluation and improvement.

More precisely, the earlier we can measure the quality of future software, the more we can improve it by being able to correct errors at the specifications level and the less will be the cost of these corrections and data quality will be improved. We propose to measure software quality using conceptual representations of the information system for both static and dynamic aspects.

The subject of conceptual schema quality evaluation has occupied a substantial part of the effort devoted to conceptual modeling. The impact of conceptual schema quality is of central concern to computer scientists, as well as to end-users, and more generally to those who seek to evaluate software quality. The literature provides lists of desirable properties of conceptual schemas. The formalization of these properties is not yet sufficiently well understood and there is no general agreement on the list of desired properties and on the way they could be measured.

This article proposes a basis for constructing an agreed model for conceptual modeling quality management.

The structure of this paper is as follows: the next section reviews relevant literature in conceptual schema quality evaluation. Section 3 is devoted to the presentation of a quality aware conceptual modeling. This approach leans on a knowledge base constructed by the consolidation of several propositions from the literature. A validation of its content is done with both researchers and professionals for a larger agreement. The empirical validation is presented in Section 4. Finally, we conclude in Section 5 and discuss areas of future research.

2. LITERATURE REVIEW

Several research directions in the context of data quality exist in the literature. We concentrate our state-of-the-art on model quality.

Conceptual Models (CM) are the abstraction of the universe of discourse under consideration [2]. They are designed as part of the analysis phase and serve as a communicating mediator between the users and the development team. They provide abstract descriptions and hide the implementation details. Generally, the following three objectives are associated with the CMs [3]:

i. Meet the users' requirements,
ii. Provide a formal representation of the observed reality, and
iii. Be a basis for the implementation and evolution of the future information system.

Although a conceptual model may be consistent with the universe of discourse, it might not necessarily be correct [2]. This suggests that there is a strong urge for a quality-oriented approach that can help in ensuring the consistency and correctness of the conceptual models. As conceptual models precede the other development activities, therefore it will be more effective to catch requirements defects as soon as they occur [10]. Studies show that defect detection in the early stages of the application development can be 33 times more cost effective than testing done at the end of development [15]. Therefore, improvements in the quality of the conceptual models lead towards the improvements in the overall quality of the delivered systems [9]. Thus a higher quality conceptual model will yield a higher quality information system and will affect the efficiency (time, cost, effort) and effectiveness (quality of results) of IS development and maintenance.

Unlike the software engineering discipline where there is a proliferation of the methods and metrics for evaluating the quality of the product, there is significantly little literature devoted towards the quality of the conceptual models [2]. This literature includes a number of quality frameworks for evaluating the quality of the conceptual models. However, there are no generally accepted guidelines for evaluating the quality of the conceptual models and little agreement exists among the experts as to what makes a "good" conceptual model.

In [9], the author has listed the quality frameworks (complete and partial) existing in the literature. The findings are summarized as Table-1.

	Research[1]	Practice[1]	Collaboration[2]
Number of proposals	29	8	2
Percentage of Total	74 %	21 %	5 %
Empirically Validated	6	0	1
Percentage	20 %	0 %	50 %
Generalizable[3]	5	0	0
Percentage	17 %	0 %	0 %
Non Generalizable	24	8	2
Percentage	83 %	100 %	100 %

Table 1. Summary of the findings by Moody [9].

Following are some of the findings inferred from Table-1:

i. Researchers did not converge towards one quality framework (due to the proliferation of quality frameworks).
ii. Practitioners are not actively involved in evaluating the quality of the conceptual models (due to the scarcity of quality proposals originating from the practice).
iii. There is a lack of collaboration between researchers and practitioners (Just 5% of the quality frameworks were the result of mutual efforts from the researchers and the practitioners).
iv. There are few frameworks that have been empirically validated (Approximately 18% of the total).
v. There is a lack of generalization since there exist only 5 frameworks that can be generalizable while others are specific to some class of models (e.g. data models) or particular notation (e.g. ER models).

Moreover, the literature review on the quality of the conceptual models suggests that the researchers have a very little agreement on a "standardized" set of quality criteria for conceptual models [9]. Therefore, there is abundance of quality frameworks for conceptual models and only few of them inherit the ideas from the other frameworks. This has resulted in the existence of several definitions for the same concept. Nelson and Todd [11] have identified different definitions of the same quality concepts e.g. there exists nine different definitions for "completeness". Similarly, there exist numerous definitions for the same quality concept and identical names for some semantically different metrics [12]. Such issues have restricted the adoption of the existing quality frameworks in practice [9].

Furthermore, there does not exist any framework other than that of Lindland et al. [6] that has both a

[1] Research, Practice and Collaboration represent the source of the framework.
[2] Collaboration means that the researchers and practitioners formulated the framework in mutual agreement.
[3] Generalizable: This implies that the framework can be applied on to conceptual models in general and is not specific to a particular class of models (e.g. data models) or a particular notation (e.g. ER models).

theoretical basis and an empirical validation [7]. Similarly, most of the existing frameworks provide ways for quality evaluation but only a handful of them provide suggestions for defect correction [9].

3. QUALITY AWARE CONCEPTUAL MODELING

Our solution is based on a knowledge base containing quality dimensions, quality attributes, metrics etc. Conceptual models can be evaluated using a combination of quality dimensions or quality attributes from the knowledge base. The rest of this section describes the quality oriented modeling process.

3.1. The Knowledge base Structure

Figure-1 presents the structure of the knowledge base. The next sections detail the process of this knowledge base construction.

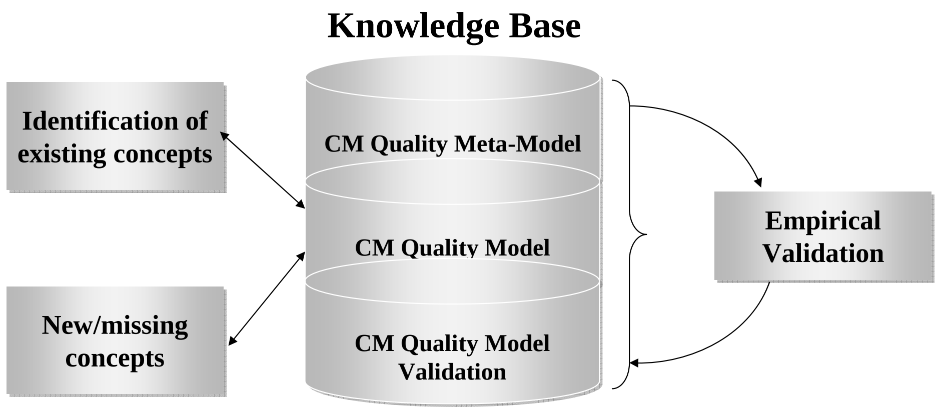

Fig. 1. Knowledge base structure

The knowledge base is composed of three abstraction levels namely: the quality meta-model, the quality model and the quality model validation material. It is interesting to note here that the same knowledge base structure can be used for data quality. However, in case of data quality the three abstraction levels could be the data quality meta-model, data quality model (containing the quality factors or quality metrics for evaluation) and data quality model validation.

3.2. The Quality Meta-Model

We base our quality meta-model on a Goal Question Metric (GQM) approach [1] to start managing quality from the user's needs. Inspired from the ISO/IEC 9126 Software Product Quality Model, the meta-model in Figure-2 is generic and simple. It is based on the notion of quality goal. Our knowledge base proposes a set of predefined dimensions, attributes, metrics etc. However new dimensions, attributes, metrics etc. can be defined or added by the user. Now, any user can define his quality goal and then chooses the dimensions that he thinks are important for achieving his quality goals. Similarly he can choose the quality attributes and the corresponding metrics for its evaluation. Once the metrics are calculated, he can use the corresponding predefined transformations for improving the quality of his conceptual model based on his quality goal.

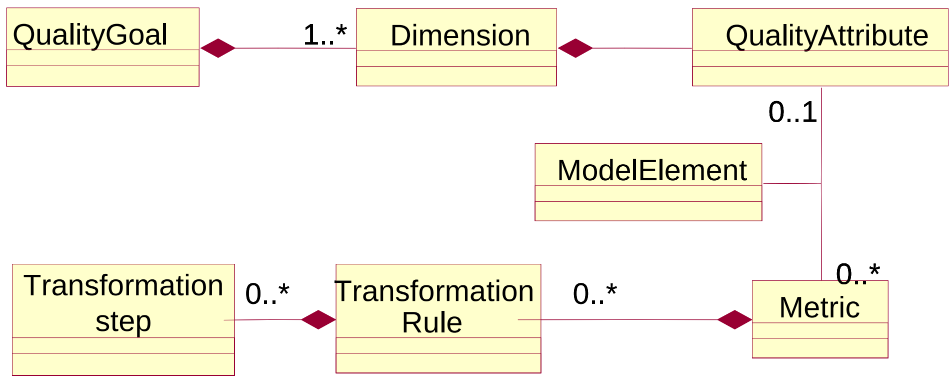

Fig. 2. Proposed Meta-Model

3.3. The quality model: an instantiation of the quality meta-model

Consider a user who is interested in evaluating the quality of his conceptual model with respect to the easiness with which it could be changed. Therefore, the quality goal is to evaluate the "ease of change". The instantiation of the meta-model according to this goal is illustrated in Figure 3. It is to be noted here that this model is the result of our enrichment methodology and uses our knowledge base.

The first level of the tree is the Ease of change Goal. The second level depicts two dimensions namely Complexity and Maintainability related to this goal. The other levels are devoted to quality attributes, metrics and transformation rules.

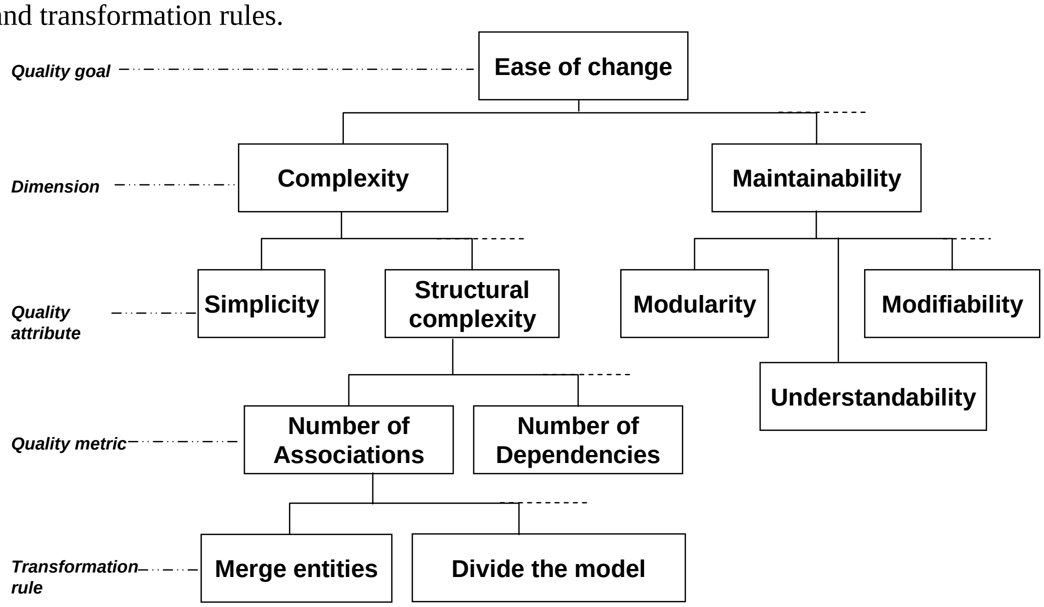

Fig. 3. Instantiation of the proposed meta-model

3.3.1. Quality attributes

Our solution uses the knowledge base that was the result of the enrichment methodology described in Section-4.3. Our proposed solution suggests two dimensions with multiple attributes that can help in evaluating the user's desired goal. Moreover, each of these attributes will employ numerous metrics (existing or newly defined) to evaluate its value of quality. Furthermore, based on the measured value of this metric, different actions/transformations can be proposed that will try to optimize this value.

Complexity. Complexity has been regarded as an important quality criterion [7,8,1]. In our proposed model, it serves as one of the two quality dimensions. It has been well argued and empirically proven that complexity in models leads to degradation in the quality and affects the maintainability of the models. Furthermore it has been shown that structurally complex models tend to be less understandable and less modifiable. Following are some of the attributes that can represent the complexity dimension:

Simplicity. This attribute is based on the hypothesis that if a model employs simple concepts for representation then it will decrease the model complexity. Modeling elements in a conceptual schema have structural interconnections and thus imply interdependencies that in return affect schema complexity. This attribute is dependent on other attributes such as size, structural complexities etc.

Structural Complexity. This attribute represents the model complexity due to the existence of different relational elements within the model. These elements can include associations, aggregations, generalizations, dependencies, transitions, relationships etc. This attribute contains several metrics that will measure the different aspects of the structural complexity of the model.

Maintainability. This dimension is based on the attributes that evaluate the ease with which a model can be maintained. This dimension is based on the hypothesis that the quality of the model can be represented by its degree of maintainability. Research [8] has showed that complexity in models affects their maintainability. Similarly, the ability to understand a model contributes towards its maintainability. However, understandability can be ameliorated by the use of standard notations or guidelines. Some of the attributes that can be part of this dimension include the following:

Modularity. This attribute is based on the object-oriented practices and hypothesize that high modularity leads to ease in maintainability. However, this attribute extends its domain and evaluates the modularity by verifying the cohesion and coupling of the modules. Cohesion can be defined as a measure to calculate the grouping of the common responsibilities or functionalities within one module. It is deemed to have high cohesion in the module to have the common responsibilities grouped in the same class. Similarly, coupling can be defined as the dependency of the module to rely on other existing modules. Coupling is deemed to be as low as possible to have less reliance on other modules and thus to have ease in maintainability.

Modifiability. This attribute is based on the notion of the ease with which a model can be modified. Modification is a common and an important activity. The authors in [5] have showed that structural complexity in models leads to low modifiability and thus directly affects the maintainability of the model. Moreover, this attribute will try to evaluate the quality of the model by accessing its degree of modifiability.

Extendibility. Software extension is an important and recurring activity as new requirements continue to emerge within the organization. Therefore, if conceptual models are designed in such a way that they can be extended to accommodate the new requirements then it will ease the software extension process. This attribute is based on the ease with which a model can be extended to include new concepts or functionalities. However, we hypothesize that if conceptual models have high degree of modularity then they can accommodate new concepts easily and thus are highly extendible.

Understandability. This attribute is based on the hypothesis that the ease in model understanding leads to ease in its maintenance. This attribute will try to evaluate the ease with which the model can be understood. However, understandability is dependent on other attributes such as those in Complexity dimension. Indeed structural complexity in models leads to low understandability and thus affects the maintainability of the model.

Usage of Standard Notation. This attribute will evaluate the model on its usage of standard language or notations. By 'standard language or notations' we mean the use of widely acceptable notations to represent the models such as the use of Unified Modeling Language or Entity Relationship notations. Here we hypothesize that the use of standard notation for the formulation of models will lead to higher understandability and analyzability and thus will contribute to the ease in change.

3.3.2. Quality Metrics.

The above mentioned quality model (Figure-3) lists some of the metrics that can be used to quantify our quality attributes with respect to ease of change. Due to space constraints, we are listing only some of the metrics that are available in our knowledge base.

Metrics	Definition
Number of Associations	Total number of associations in a model.
Number of Dependencies	This metrics is used to calculate the total number of dependency relationships within the class diagram.
Number of Aggregations	It calculates the number of aggregation relationships within a class diagram.
Depth Inheritance Tree	It calculates the longest path from the class to the root of the hierarchy in a generalization hierarchy

Table 2. Metrics for Structural Complexity [8]

Metrics	Definition
Cohesion	This metric calculates the cohesion of different modules.
Coupling	It calculates the coupling between different modules.

Table 3. Metrics for Modularity

3.3.3. Transformation rules for improvement.

Corrective actions or transformation rules are the essence of our proposed solution. Once the quality metrics are calculated, corresponding corrective actions or transformations can be proposed to optimize the model. Due to space constraints, we are including only two corrective actions for the above mentioned two quality metrics.

Reduce the Number of Associations

If the "Number of Associations" metric reveals that the model has too many associations then the corresponding corrective actions could include the following:

i. Merge the entities (for example, merge the entities having multiplicity at-least and at-most one and that are semantically close).
ii. Use the inheritance mechanism to factorize associations.
iii. Examine cycles and remove the redundant associations.

Increase cohesion

If the "Cohesion" metric shows a very low cohesion value for a model then the corrective actions could include:

i. Decompose the model into functionally independent modules.
ii. If an entity has bad values for cohesion then divide the entity and redefine the relationships.

The next section is devoted to the empirical validation of our proposal.

4. EMPIRICAL SUPPORT

Validation serves as the backbone of our methodology. We have identified different quality concepts and enriched the knowledge base. These quality concepts must go through some validation methodology for verification.

4.1. Research methodology for validation

We combine quantitative and qualitative research methods for validation. However, other suitable methodologies can also be used or employed for validation.

4.1.1. Quantitative Methods.
As depicted in Figure-1, our proposed methodology includes an important step of empirical validation of our quality concepts from the professionals including practitioners. We have used survey-based techniques to gather data from the respondents over the contents of our knowledge base. We have first used these data to analyze the conformance of the proposed quality criteria to practitioners' viewpoint. Similarly, we can verify the correctness and completeness of our knowledge base through this survey. We also used the survey feedbacks to enrich the knowledge base.

4.1.2. Qualitative Methods.
We plan to use of qualitative methods such as ethnography, action research, use case study etc. as a validation exercise for our knowledge base. For example, action research can be very productive in understanding the real world setting and in identifying the practices employed by practitioners for evaluating the conceptual models.

4.2. The process for constructing and enriching the knowledge base

A central contribution of our work includes the identification and creation of certain artifacts (quality dimensions, quality attributes, quality metrics etc.) in addition to their validations.

In this article, we have created the following artifacts:

i. Conceptual model quality meta-model
ii. Goal oriented quality model
iii. Different set of quality attributes for evaluation
iv. Numerous metrics for quantifying the quality attributes
v. Transformation rules for improving the quality value obtained by the metrics

Our goal is to propose a multi-faceted quality approach for conceptual modeling that should be generic, flexible and remains valid for different types of conceptual models (ER models, UML diagrams etc.). Our approach has tried to aggregate the existing quality frameworks in addition to providing the missing elements or concepts and has a theoretical, practical and epistemological foundation.

4.2.1. Theoretical Foundations.

The proposed approach uses existing literature on conceptual modeling as its theoretical foundation to formulate a multi faceted quality knowledge base. We have extracted and filtered different concepts, from the previously existing quality frameworks or literature to enrich the knowledge base. This enrichment activity provides a more comprehensive and flexible quality approach for conceptual models. The enriched knowledge base classifies the existing concepts into different dimensions and thus helps in identifying the uncovered areas of conceptual modeling quality. Moreover, this knowledge base is evolutionary and open to future additions and modifications.

4.2.2. Practical Foundations.

Proposed approach involves practitioners' viewpoint as its practical foundation. As mentioned above, our approach requires a validation methodology. In this paper, we have proposed a first validation involving practitioners. We also plan to repeat this experience to actively involve professionals by using surveys, interviews etc. The objective of this validation exercise is to collect the responders' views on the holistic quality of the conceptual models in addition to their feedback over the enriched knowledge base. This activity can help in extracting the general practices of the practitioners and in identifying their view on the quality of conceptual models. The received feedbacks from the first validation have been evaluated and incorporated in the knowledge base. We will proceed similarly with all the future validations.

4.2.3. Epistemological Foundations.

Many studies in the domain of conceptual models evaluation address the problem of miscommunication between business and IT actors. One of the research directions aiming to understand this problem considers an epistemological point of view based on well-founded assumptions [13, 14]. We studied and used some of these approaches for enriching our knowledge base.

4.3. Survey for Validation

Our proposed approach includes practitioners' viewpoint as its practical foundation. Therefore, a web-based survey, built on our proposed framework (above), was formulated. The purpose of this survey was twofold:

i. To serve as a validation exercise in providing feedback from professionals including practitioners over the efficacy of our knowledge base.
ii. To study the general practices and views of the professionals over the quality of conceptual models. This includes the identification of attributes or factors important to professionals for evaluating the quality of conceptual models.

As mentioned above, a web-based survey was formulated to conduct this study. This was a closed survey and was accessible through a special link, provided to the invited participants only to avoid unintended participants. This was a comprehensive survey containing 42 general questions and our knowledge base specific questions. However, all the questions were directly related to the quality of conceptual models. These questions include the two feedback questions where the participants were required to mention:

i. Attributes/factors that in their view are crucial to the quality of conceptual models (They can identify up to seven attributes/factor).
ii. How do they compare two conceptual models representing the same reality or modeling the same problem? They were required to identify and mention up to seven attributes/properties that they think they will employ in choosing the best model with respect to their perception of quality.

The survey consists of the following four sections to target the different types of information:

i. General Information about the respondents
ii. Respondents' knowledge about CM.
iii. Respondents' knowledge about CM quality and their feedback over our knowledge base.
iv. Respondents' general practices about conceptual modeling quality

The survey provides the dictionary and instant help about the definitions and details of all the terms and concepts that were present in the survey including the definitions of all the attributes in our knowledge base. Respondents were asked to provide their feedback over each of our quality attribute by selecting any one of the following four scenarios:

i. Attribute is not related to quality
ii. I am not sure if it is related to quality
iii. Attribute is directly related to quality
iv. Attribute is indirectly related to quality.

4.3.1. Sample.
In total 179 professionals (including IS managers, IS developers, researchers etc.) were contacted to complete the survey. However, 57 professionals completed the survey that resulted in the response rate of 31.8%. Among the received 57 responses, three were discarded due to errors in the provided data or incomplete information.

Respondents were required to select their occupation from a list of fifteen pre-defined occupations. The distribution of respondents and detailed information about each of the occupation is summarized in the following Table-4:

Occupation	**Analysts**	**IS Project Manager**	**Lecturer/ Professor**	**Researcher**	**Software Developer**
Number of Respondents	7	12	7	9	19
Average Age	31.4	30.42	33.2	30.3	28.9
Min Age	27.5	27.5	27.5	27.5	27.5
Max Age	50	40	50	40	40
Average Experience	9.7	7.9	8.7	3.6	4.8
Min Experience	4	4	2	1	1
Max Experience	26	15	26	7.5	15
Average Modeling Exp.	6.8	4.1	8.8	2.1	2.7
Min Modeling Exp.	2	1	1	1	1
Max Modeling Exp.	15	7.5	15	4	7.5

Table 4. Classification of respondents with respect to their occupation

Respondents belong to different organizations ranging from small organizations having less than 50 employees to as big as having more than 1000 employees. Distribution of respondents with respect to their organizational strength is shown in Table-5:

Company Size:	Count	Percent
Less than 50 employees	15	27.8
Between 50 & 100 employees	7	13
Between 101 & 200 employees	8	14.8
Between 201 & 500 employees	11	20.4
Between 501 & 1000 employees	2	3.7
More than 1000 employees	9	16.7
Not answered	2	3.7
TOTAL	54	

Table 5. Respondents with respect to their organization size.

4.3.2. Data Analysis.

The collected data shows that 85% of the respondents consider the imposition of quality approach on the conceptual models to directly influence the quality of the final product. However, it is interesting to note that 87% of the respondents have never used any method or framework to evaluate the quality of conceptual models. This shows that despite the appreciation of importance of implementing quality approach, professionals do not employ any methods to improve the quality. As mentioned above, such a behavior has resulted due to the gap between research and practice. To date there does not exist any quality framework that is standardized and comprehensive enough to accommodate the requirements of the practitioners. However, our proposed methodology is unique in a way that it is generic, simple, customizable, and easy to implement. Moreover, our proposed knowledge base follows a hierarchy of different quality levels starting from quality goals and ending at the corrective suggestions.

As mentioned above, we asked respondents to provide feedback over the efficacy of the attributes in the knowledge base. They were required to mark these attributes into either 'not related to quality', 'I am not sure', 'directly related to quality' or 'indirectly related to quality'. However since the last two options affirm that the attribute is related to quality therefore we have merged these two options as one to have a clear distinction between the attributes that are related to and not related to quality. Table-6 summarizes the responses.

Attributes	NOT Related to Quality	Related to Quality	Not answered	I am not sure
Simplicity	11.1	75.9	3.7	9.3
Structural complexity	11.1	75.9	1.9	11.1
Modularity	5.6	72.2	1.9	20.4
Modifiability	11.1	70.4	1.9	16.7
Understandability	7.4	87	3.7	1.9
Usage of standard notations	3.7	79.6	3.7	13
Extendibility	3.7	83.3	3.7	9.3

Table 6. Respondents' feedback on some of the proposed quality attributes.

In Table-6, all the values are in percentages of the responses and are rounded off to the nearest tenth digit. Each row should be read as, for example, 75.9 % of the respondents think that 'Structural Complexity' is related to quality against 11.1% that think 'Structural Complexity' is not related to quality. Similarly, 11.1% of the respondents declare their inability to categorize 'Structural Complexity' in any of four classes.

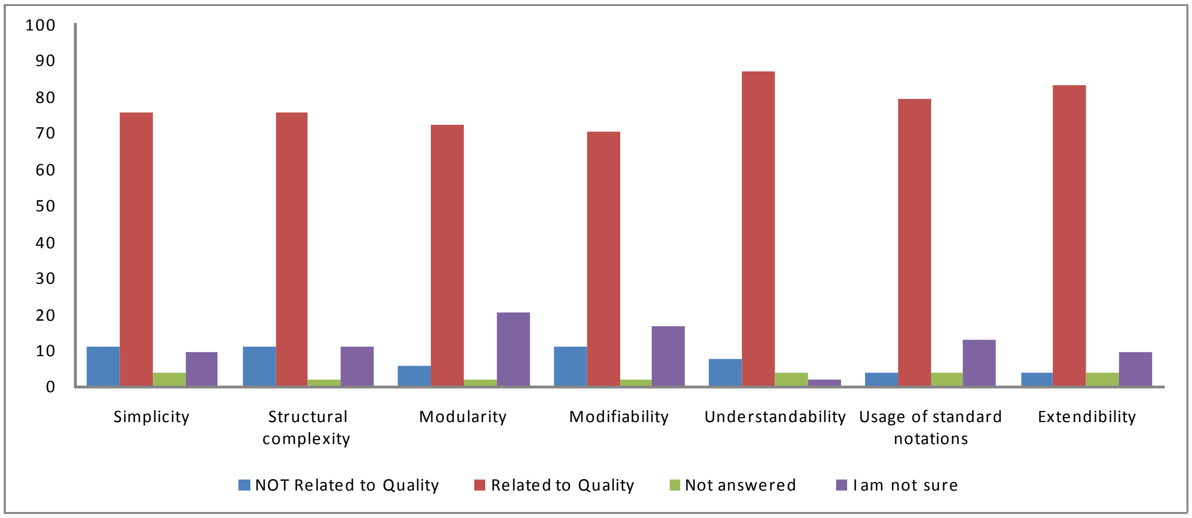

Fig. 4. Bar chart depicting respondents' feedback on some of the proposed quality attributes.

After viewing the above feedback, we can say that the attributes in our knowledge base are well identified and represent the same attributes and factors that are required by the professionals. However, respondents have also identified some attributes that they think are important to quality. These attributes were evaluated and incorporated to our knowledge base. Due to space constraints, these attributes along with several other quality attributes are not mentioned in this article. We have included only some of the quality dimensions and attributes from the knowledge base in this article.

Next step of our work involves the quantification of these attributes. This involves the identification and formulation of the metrics to quantify these attributes. It must be noted here that not all of these metrics could be generic. Similarly, there will be metrics specific to particular type of conceptual models such as class diagrams, state diagrams etc.

Similarly, the pie charts (figure-5) show the respondents' views about our dimensions. We can see that the respondents agree that the dimensions in our quality model are related to quality as well.

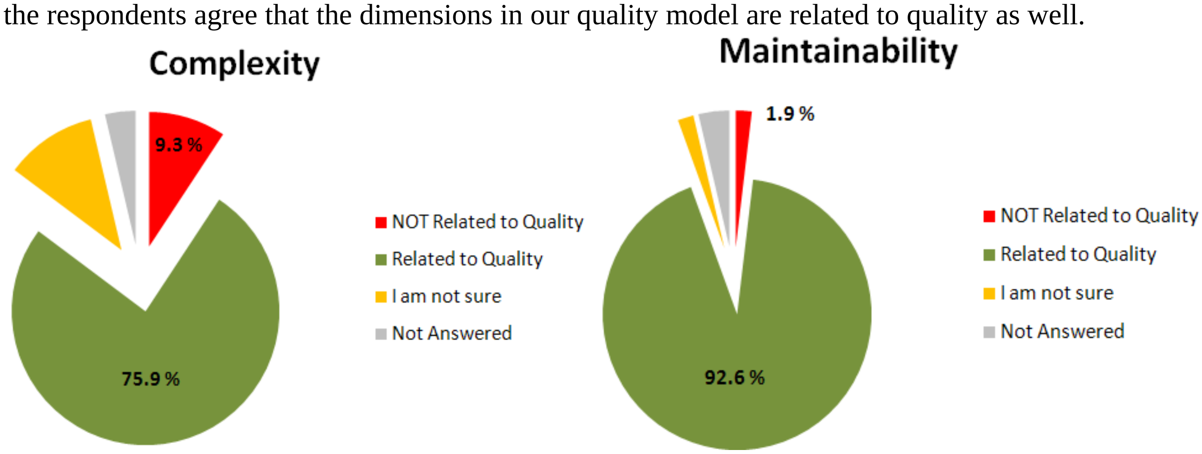

Fig. 5. Pie chart depicting respondents' feedback on Complexity and Maintainability dimensions.

5. CONCLUSION AND IMPLICATIONS FOR FURTHER RESEARCH

Two research directions, data quality assessment and models quality assessment have emerged as an important and challenging topic in recent years. However, there is no concrete contribution on the precise relationship between the two qualities. We claim that data quality can be achieved through model quality assessment and improvement. In this article we have addressed the conceptual model quality. We started from Moody's statement on the lack of agreement on a quality framework and a lack of collaboration between researchers and practitioners. This paper proposed a comprehensive quality meta-model for evaluating the quality of conceptual models.

The proposed quality meta-model is generic, flexible, customizable, and remains valid for different types of conceptual models (ER models, UML diagrams etc.).

The meta-model is simple and could be easily instantiated. The instantiation process produces a tree structure that could be used to incrementally guide an IS designer or a quality engineer to achieve a quality goal. This goal is refined through quality dimensions that are further measured by quality metrics. The last level of the tree is composed of transformation rules that propose actions leading to the quality improvement according to the desired quality goal. These trees, representing both researcher's and practitioner's quality practices, are capitalized in an evolutionary knowledge base.

Our approach uses existing literature on conceptual modeling as its theoretical foundation to formulate a multi faceted quality knowledge base. Different concepts, from the previously existing quality frameworks or literature, are extracted and filtered to construct the proposed knowledge base. An extract from the knowledge base is presented in the paper as an instantiation of the quality meta-model.

Finally, the paper presented a first validation on the basis of the feedback received from different populations (researchers, professionals, students etc.) who have either used the proposed knowledge base for evaluating and improving the quality of their models or have been interviewed or surveyed over the efficacy of the proposed quality model.

To enhance the validation process, the knowledge base, as well as the validation results, will be available on a web site to be used by researchers and professionals to go a step forward in the construction of this knowledge base.

Acknowledgements. This research was partially supported by the ANR ARA Masses de Données Program (2005-2008), "Grant ARA-05MMSA-0015" (QUADRIS project).

REFERENCES

[1] Basili, V. R., Gianluigi, C., and Rombach, H. D., The Goal Question Metric Approach, In *Encyclopedia of Software Engineering*, Wiley, 1994.

[2] Cherfi, S.S., Akoka, J., and Comyn-Wattiau, I., Measuring UML Conceptual Modeling Quality-Method and Implementation, In *the Proceedings of the BDA Conference*, Ed. P. Pucheral, Collection INT, France, 2002.

[3] Cherfi, S.S., Akoka, J., and Comyn-Wattiau, I., A Framework for Conceptual Modeling Quality Evaluation, In *the Proceedings of ICSQ'03*, October 6-9, 2003.

[4] Erlikh, L., Leveraging Legacy System Dollars for E-Business, *(IEEE) IT Pro, May/June* 2000, pp. 17-23.

[5] Genero, M., Piatini, M., and Manso, E., Finding "Early" Indicators of UML Class Diagrams Understandability and Modifiability, In *the proceedings of the 2004 International Symposium on Empirical Software Engineering (ISESE'04) IEEE Computer Society*, 2004.

[6] Lindland, O.I., Sindre, G., and Sølvberg, A., Understanding Quality in Conceptual Modeling, *IEEE Software* 11(2) 1994, pp. 42-49.

[7] Maes, A., and Poels, G., Evaluating Quality of Conceptual Modeling Scripts Based on User Perceptions, *Data*

& Knowledge Engineering 63(3) 2007, pp. 701-724.

[8] Manso, E., Genero, M., and Piattini, M., No-Redundant Metrics for UML Class Diagram Structural Complexity, *Lecture Notes on Computer Science* (2681) 2003, pp. 127-142.

[9] Moody, D.L., Theoretical and Practical Issues in Evaluating the Quality of Conceptual Models: Current State and Future Directions, *Data & Knowledge Engineering* (55) 2005, pp. 243-276.

[10] Moody, D.L., Sindre, G., Brasethvik, T., and Sølvberg, A., Evaluating the Quality of Information Models: Empirical Testing of a Conceptual Model Quality Framework, In *the proceedings of the 25th international Conference on Software Engineering,* May 03–10 2003, pp. 295-305.

[11] Nelson, R.R., and Todd, P.A., Antecedents of Information and System Quality: An Empirical Examination within the Context of Data Warehousing, *Journal of Management Information Systems* 21(4) 2005, pp. 199-235.

[12] Purao, S., and Vaishnavi, V., Product Metrics for Object-Oriented Systems, *ACM Computing Surveys* 35(2) 2003, pp. 191-221.

[13] Ribbert, M., Niehaves, B., Dreiling, A., and Holten, R., An Epistemological Foundation of Conceptual Modeling, In *the Proceedings of the 12th European Conference on Information Systems,* Turku, Finland, 2004.

[14] Schütte, R., Architectures for Evaluating the Quality of Information Models - A Meta and an Object Level Comparison, In *the proceedings of the 18th International Conference on Conceptual Modeling*. Lecture Notes in Computer Science (Springer-Verlag), 1999, pp. 490-505.

[15] Walrad, C., and Moss, E., Measurement: The Key to Application Development Quality, *IBM Systems Journal* 32(3) 1993, pp. 445-460.