Analyzing the Goal Structured Requirements Engineering and Traceability Model

Vinay Kumar[1] and Reema Thareja[2]*

[1]*Professor, Department of IT, VIPS, GGSIPU, New Delhi 110 088, India*
[2]*Assistant Professor SPM Delhi University, Author of Oxford University Press, India*
[1]*vinay5861@gmail.com,* [2]*reema_thareja@yahoo.com*

Abstract

The principal role of Data warehouses in making strategy decisions makes its quality crucial for organizations. Quality of data models for data warehouse has significant effect on the quality of data warehouse. Therefore, we need methods, models and tools to have quantitative and objective measurement of quality of data models for data warehouses. Lack of a set of defined measures made us to define a set of metrics that can be used to measure the complexity, understandability, believability, usability, accessibility, data currency and modifiability which are in turn defined as quality characteristics for data warehouse systems.

The proposed metrics may act as objective indicators of the quality of data warehouse models and also help the developers to select quality data model among various semantically equivalent models. The other advantage of these metrics is that they are available during early phase of software development life cycle.

Keywords: metrics, data warehouse, complexity, quality

1. Introduction

Data warehouses form the core of most of the current decision support systems, providing organization with several years of historical information for making strategic decisions [13]. Poor quality of such an important system can result in technical and organizational losses like loss of clients, important financial losses or discontent amongst employees [1]. Therefore, it becomes very important for an organization to guarantee the quality of information stored in its data warehouse right from the early stages of data warehouse development project.

The information quality in data warehouse systems by and large depends on presentation quality and data warehouse quality (Figure 1). Data warehouse quality in turn depends on the DBMS quality, data quality and data model quality (which can be considered at conceptual, logical and physical levels) [2].
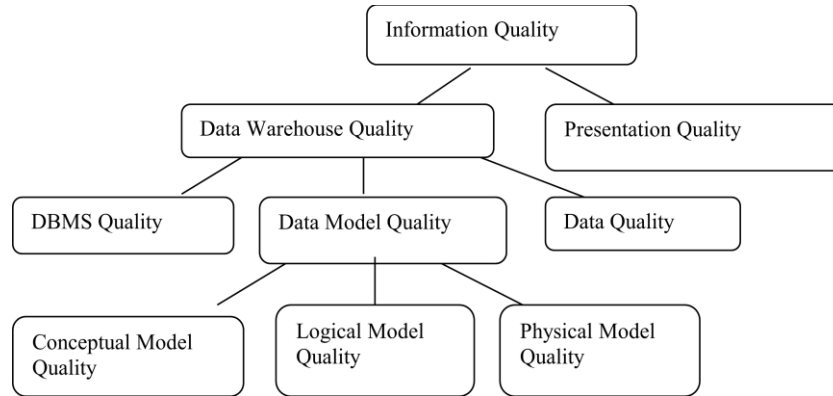
Figure 1. Data Warehouse Information Quality

In this paper, we will focus on the quality of conceptual model as the sooner the data warehouse development team concentrates on its quality, more is the likelihood of implementing a high quality data warehouse [3]. Moreover, conceptual data models lay the foundation of all later design work. So the conceptual model quality has a significant impact on the quality of the data warehouse.

In order to evaluate and if required to improve the quality of a conceptual data model first we must assess it in an objective way. This would help the data warehouse designers to make better decisions during design activities. Even when the designers have several conceptual schemas at hand, they can use this assessment to decide the best schema amongst them.

Till date, researchers all over the world have proposed many quality frameworks for conceptual data models but not much work has been done to specify valid metrics to evaluate the quality of conceptual data models in an objective way. The scarcity of such metrics made us to identify a set of measures that could analyze our conceptual model of Goal Dimension Table.

The paper is organized in section. Section 2 gives a brief introduction to the conceptual schema of the Goal Dimension Table. Section 3 details the process of metrics definition. Section 4 mentions the quality parameters and corresponding metrics which can be used to analyze the proposed model. Section gives a Quality Evaluation table and section 6 concludes the paer.

2. Conceptual Schema of Goal Dimension Table

In the conceptual model of Goal Strcutured Requirements Engineering and Taraceability Model, we have a Goal Table in the centre surrounded by Dimension Tables thereby forming a star type formation [4]. According to Figure 2, Goal table stores details about user's goal like a goal id, goal name, name of the stakeholder who owns that goal, metric related to the goal, the region where the goal is applicable, the product for which the goal is meant, time at which it entered the system and the remarks that explains whether the goal was accomplished and the technical details of the process carried out to satisfy the goal.

In addition to this, the Goal table has a measure- frequency of use, which is a counter that is automatically incremented whenever a particular goal is accessed and refereed for making decisions.
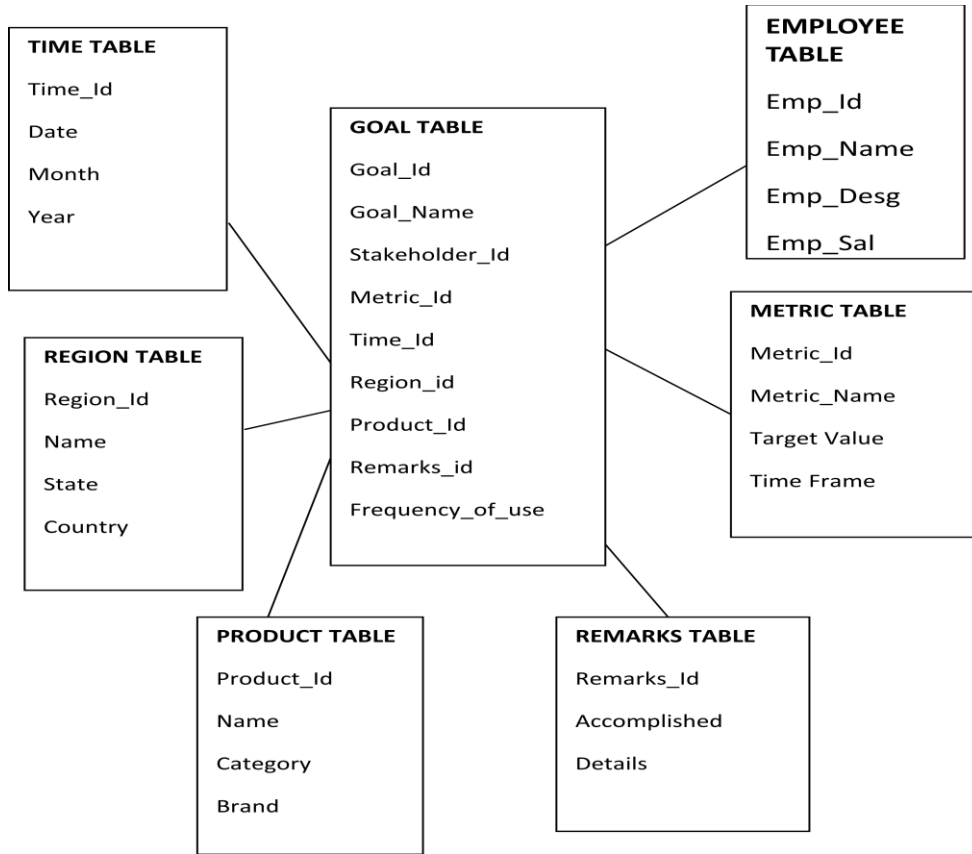
TIME TABLE

Time_Id

Date

Month

Year

REGION TABLE

Region_Id

Name

State

Country

GOAL TABLE

Goal_Id

Goal_Name

Stakeholder_Id

Metric_Id

Time_Id

Region_id

Product_Id

Remarks_id

Frequency_of_use

**EMPLOYEE
TABLE**

Emp_Id

Emp_Name

Emp_Desg

Emp_Sal

METRIC TABLE

Metric_Id

Metric_Name

Target Value

Time Frame

PRODUCT TABLE

Product_Id

Name

Category

Brand

REMARKS TABLE

Remarks_Id

Accomplished

Details

Figure 2. Goal and Dimension Tables

Identification of Goals

Formulation of Hypotheses

Metrics Definition

Theoretical Validation

Empirical Validation

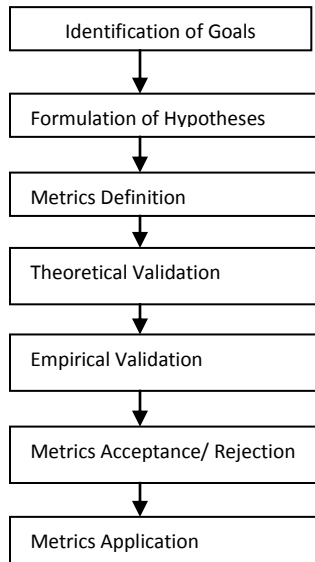Metrics Acceptance/ Rejection

Metrics Application

Figure 3. Method Definition Process

3. Metric Definition Process

The process of metric definition is based on the organization's measurement goals that are related to quality attributes of the conceptual schema. Fig 3 shows the steps involved in obtaining valid and useful metrics. The process has seven steps. First, the goals of the metrics that can assess and control the quality of conceptual model of data warehouse are identified. Then corresponding hypotheses are formulated. For example, if the goal is to assess the structural complexity of the conceptual model then the hypotheses could be- "lesser the structural complexity greater is the understandability (which is a quality attribute)."

The next step is to define metrics by considering the hypotheses, specific characteristics of the model that has to be asses and the experience of the designers of data warehouse systems. A goal-oriented approach as GQM (Goal-Question-Metric [5]) can also be very used for this step.

After defining the metrics, we need to validate them validation has to be done both theoretically as well as empirically. Theoretical validation specifies when and how to apply the metrics. It can be done either using formal frameworks or by measurement theory [2]. While formal framework merely specifies a set of formal properties defined for given software attributes for classifying the proposed metrics; the measurement theory, on the other hand, defines the scale to which a metric pertains. This helps us to know the statistics and transformations that can be applied to the metric.

Followed by theoretical validation, the metrics must be empirically validated to confirm and understand the implications of the measurement of the conceptual model. This can be done through experiments, case studies and surveys.

Once the metric has been validated, it can be simply accepted, redefined or discarded. This means that the final outcome of this step is a valid metric. The valid metric thus obtained is applied to be used in real world. Once it is applied, the metric must be monitored and if required adapted to application changing environment.

4. Parameters for Analyzing the Quality of Goal Structured Requirements Engineering and Traceability Model

In this section, we will define the parameters and the metrics that can be used to analyze the quality of our proposed conceptual schema of goal structured requirements engineering and traceability model. The quality parameters that we have identified are- understandability, modifiability, believability, accessibility, usability, timeliness and data currency. So, our goal would be to evaluate the proposed conceptual schema. The goal and its template has been shown in Figure 4.
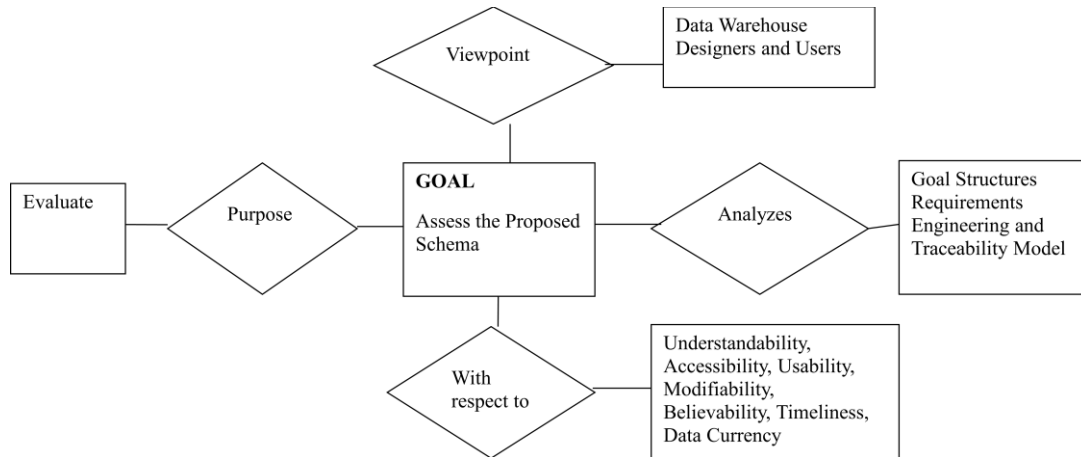
Figure 4. The Goal Template for Assessing the Conceptual Model

4.1 Understandability

Figure 5 depicts the relationship between structural complexity and quality of data warehouse schema. It has been found that the structural complexity of the conceptual schema affects the cognitive complexity which in turn affects its understandability which is an important external quality of any data warehouse system [2].
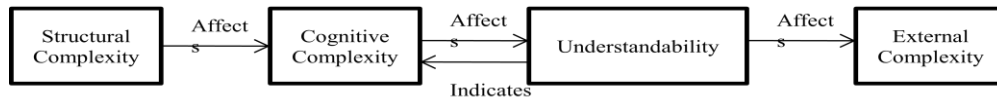

Figure 5. Effect of Structural Complexity and Quality of Data Warehouse Schema

So we see that in order to make the proposed conceptual model understandable, we need to analyze its structural complexity and cognitive complexity.

4.1.1. Structural Complexity

The structural complexity of any conceptual schema including the proposed goal structured requirements engineering and traceability model can be evaluated based on the metrics given below:

1. Number of Tables
2. For each Dimension Table, we can calculate
    a. Number of attributes (NA)
    b. Relationship with other Dimension Tables
    c. Number of Dimension Hierarchies
    d. Depth of Chosen Hierarchy
3. Number of attributes repeating in different Dimension Tables
4. For Fact Table, we can calculate,
    a. Number of Attributes in Fact Table
    b. Number of Additive Measures
    c. Number of Semi- Additive Measures

    d. Number of Non- Additive Measures
5.   Number of Degenerated Dimension Tables

Let the Number of Tables be NT
Let the Complexity of Dimension Table$_i$ be calculated as DTC$_i$
Let the Complexity of Fact Table be calculated as FTC$_i$
Let the Number of Repeating Attributes be NRA
Let the Number of Degenerated Dimension Tables be NDD

Structural Complexity = NT * (Σ DTC$_i$) * (Σ FTC$_i$) * (1 + NRA) * (1 + NDD)

Here, we have added 1 to NRA and NDD because if we have a conceptual schema that has no attribute that repeats and no degenerated dimension, then its structural complexity must not evaluate to be zero.

4.1.2. Cognitive Complexity

Cognitive complexity is a characteristic that indicates how complex or simple is the concept, and structure and depends on the amount of knowledge required to perform a task using a specific application. It is actually a summation of several factors that make things hard to see, use, grasp, and understand, and contribute directly to our neural load [7].

In context of data warehouse conceptual schema, the main features that affect understandability are structural complexity and cognitive complexity. While structural complexity depends on parameters discussed in the above section, cognitive complexity, on the other hand, depends on people's (data warehouse users and designers) interests, capability and perceptions.
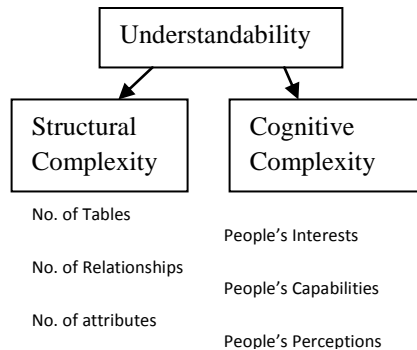

Figure 6. Understandability of Data Warehouse Conceptual Schema

Funke [8] has identified factors that affect the cognitive complexity. These factors include-

a.   Multiple goals. Different people have different goals and some of them may be contradictory so trade-offs are often required.

b.   Situation complexity. Data modeling often suffers when the data warehouse requirements analyst or designer fail to understand the requirements accurately due to complexity of domain.

c.   Connectivity. Connectivity under cognitive complexity does not imply cardinality as it used to in structural complexity. Here, connectivity means large number of interrelationships among a limited number of elements in complex problems.

d.   Novelty. Novelty refers to situations that are new and unfamiliar to the designer. A new situation may not seem to be complex if an analogue can be found. However, even then the design has to be made either from the first principles, or by a deft handling of mapping and manipulating from the analogue

4.2 Modifiability

Data warehouse collects its massive amount of data from production systems, external systems, internal systems and archived data. The source system's schema may change with time based on user's requirements. Such changes must also be incorporated in data warehouse systems. Such structural changes or schema changes in data warehouse systems may result in dimension updates, structural updates, instances updates, facts updates, attributes updates, hierarchy updates, quality updates and constraints updates.

We can assess our conceptual schema of goal structured requirement engineering and traceability model by making a checklist of following features and then analyzing whether the schema supports following features or not.

a. Add a table
b. Delete a table
c. Add a view
d. Delete a view
e. Alter view definition
f. Add an attribute
g. Delete an attribute
h. Change attribute domain
i. Add integrity constraint
j. Delete integrity constraint
k. Insert classification relationship
l. Delete classification relationship
m. Add a new level to the dimension
n. Delete an existing level from the dimension
o. Add a new attribute to given dimension or a given level
p. Delete an attribute from given dimension or a given level
q. Rename table, view, fact, dimension, level, or property

4.3 Accessibility

Data warehouse data elements must be well documented and easily accessible to the users so that if interested they can see the origin of data element and also validate their reports. The proposed model on goal structured requirements engineering and traceability model makes data sources clearly documented and traceable. For this, in the star schema, data elements have meaningful names so that their origin is clear to the users reading it. Moreover, the details of all these elements will be stored in the metadata according to a template specified by us in [4].

Here, the Goal is accessibility and metric is percentage of data elements with their data sources properly documented.
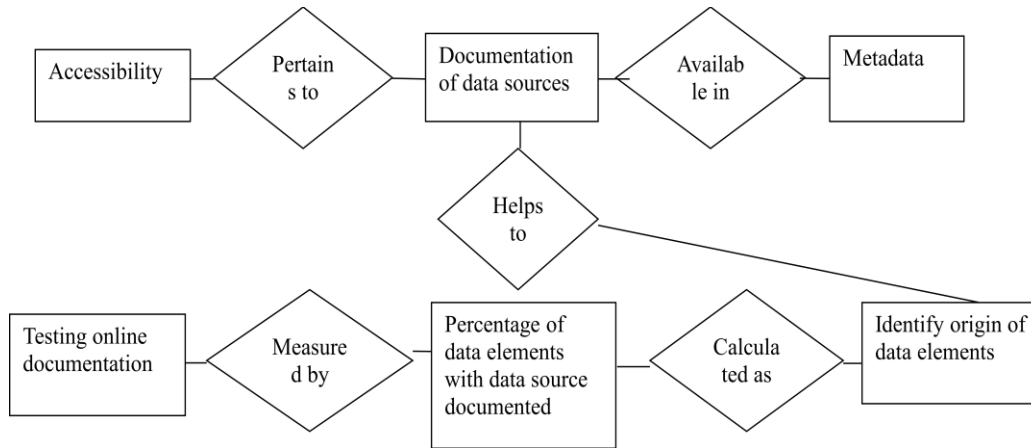
Figure 7. Data Warehouse Conceptual Model Quality Template- Accessibility

4.4 Data Currency

While collecting requirements for the upcoming data warehouse project, the team must also ask the users how frequently they want the data to be refreshed. As a normal practice data warehouse data must be refreshed at least once in a week for example, during the weekends when data warehouse may not be available for strategic decision making. To measure this, the data warehouse development team must measure the frequency of refreshing the physical tables that are based on our proposed conceptual schema. The team must also measure the time it takes to refresh these tables. Here, the Goal is data currency and metric is frequency of data refreshes and the time taken to complete the refresh process.
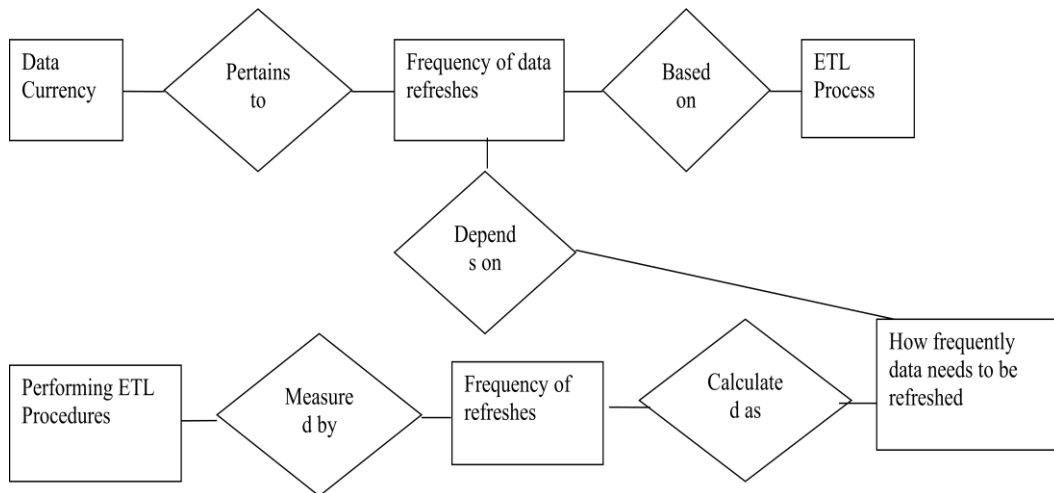

Figure 8. Data Warehouse Conceptual Model Quality Template- Data Currency

4.5. Believability

While designing he conceptual schema of the data warehouse, the designers must give due consideration to the hierarchies that exist in the data and represent them in the star schema designs. Data warehouse users must be able to view summary data as well as detail data, helps them to deduce how summaries have been calculated and also increases their trust in

accuracy and completeness of data. In the proposed goal structured requirements engineering and traceability model, users can get a complete information at the lowest level of detail. Here, the Goal is believability and metric is the level of detail at which the data is stored.
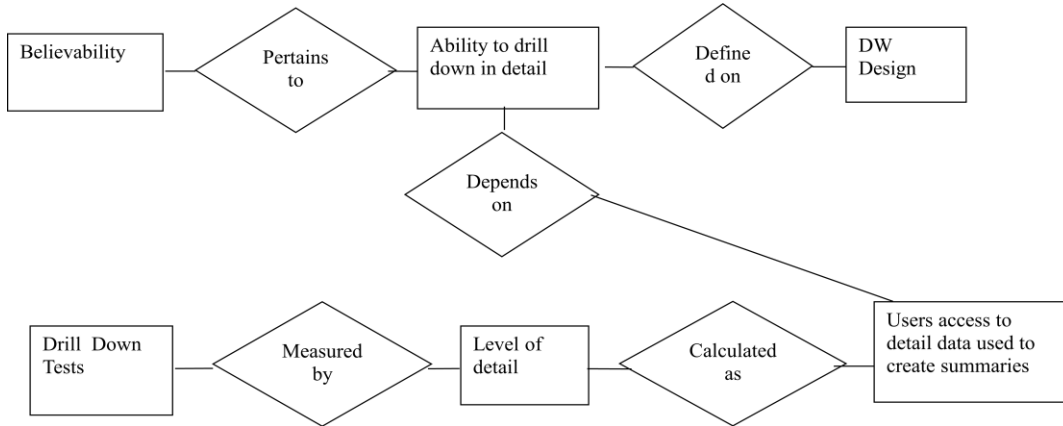

Figure 9. Data Warehouse Conceptual Model Quality Template- Believability

4.6 Timely

Quality of a data warehouse is also affected by processing time or the time required to run reports. To evaluate the response time, each available report must be processed and the time taken to process that report must be measured and optimized.

Here, the Goal is data timeliness and metric is time required (in seconds/minutes) to generate requested reports.
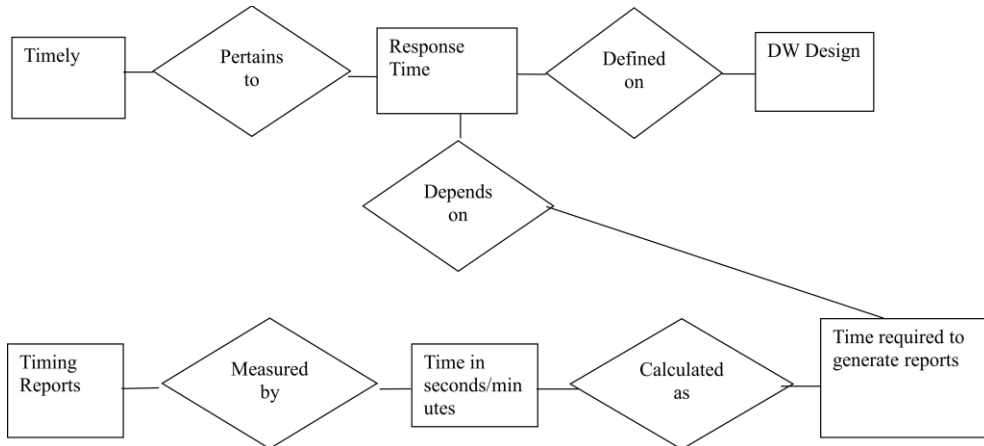

Figure 10. Data Warehouse Conceptual Model Quality Template- Timely

4.7 Usability

Data warehouse users always appreciate the availability of pre-defined reports that are frequently used by them this eases their work and also reduces the time they have to put in to generate that report. Here, the Goal is usability and metric is percentage of percentage of reports that are already available for the data warehouse users.
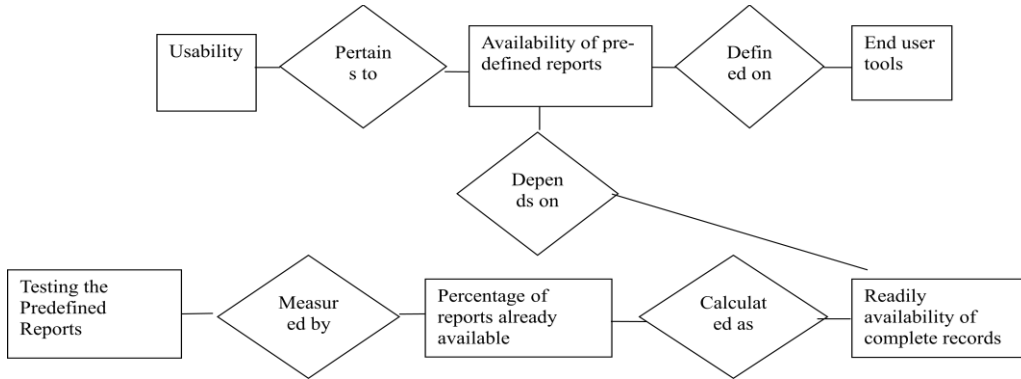
Figure 11. Data Warehouse Conceptual Model Quality Template- Usability

4.7. User Friendliness

User friendliness of a data warehouse in other words means the ease with which it can be accessed by its users. User friendliness is usually measured by number of steps the users perform to produce a particular report. Generally, users don't want to perform more than five steps to make a report.

Here, the Goal is user friendliness and metric is number of steps the users have to perform to make a report.
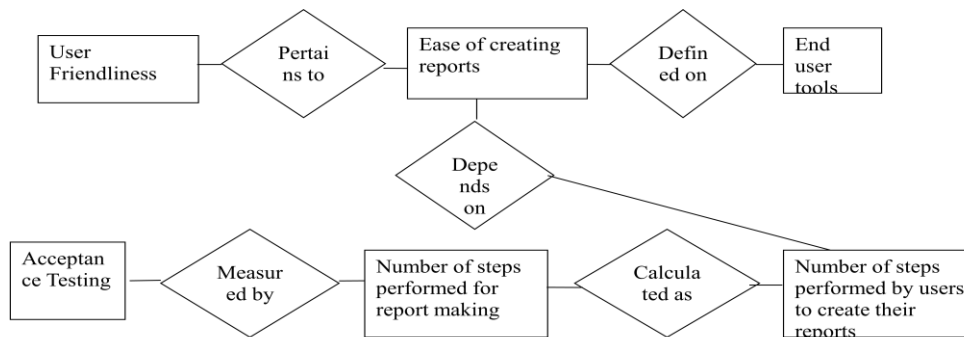

Figure 12. Data Warehouse Conceptual Model Quality Template- User Friendliness

5. Quality Evaluation Table

In order to assess the proposed conceptual schema of goal structured requirements engineering and traceability model, we can use the table given below and get it filled by the data warehouse designers. While finalizing the schema, the designers may have more than one schema, so they can use this table to identify the best schema so that it can be used further with development process.

In this paper, we have given metrics for eight goals. To use this table, the designers must follow the steps given below.

Step 1: Prioritize the quality goals

Step 2: Assess the schemas based on the proposed metrics. For this, give rating from 0-100 for every individual metric such that 0 is the least and 100 is the maximum rating

Step 3: Normalize the assessment result of top four quality goals so that its value lies in range of 0-15.

Step 4: Normalize the assessment result of next four quality goals so that its value lies in range of 0-10.

Step 5: Add the assessment of all the quality goals to get a final value in the range of 0-100.

Quality Goal	Metric	Value	Normalized Value
Understandability	Structural complexity		
Modifiability	Provision for enhancement		
Accessibility	Percentage of data elements with documented data sources		
Data Currency	Frequency of data refreshes		
Believability	Level of detail		
Timely	Time taken to make a report		
Usability	Percentage of reports already available		
User friendliness	Number of steps performed to make a report		

6. Conclusion and Scope for Future Work

Data modeling plays a vital role in data warehouse development. The quality of these systems can suffer because of poor data modeling practices. So, in order to assure the quality of data warehouse systems, we need to guarantee the quality of the models used in their design. As of now, many research studies have focused on comparisons of data modeling formalisms, but little has been said on developing techniques to manage complexity and enhance their quality.

In this paper, we have proposed a set of quality goals and their corresponding metrics to assure the quality of the proposed conceptual schema (goal structured requirements engineering and traceability model) used in the early stages of data warehouse design. These metrics will help us measure the understandability of users and designers, modifiability, timeliness, data currency, believability, usability, user friendliness and accessibility.

After giving this set of metrics to measure aforementioned aspects, we will now work to define the correct metrics thresholds under which several design options can be taken.

These metrics may not be unique and complete, but we strongly believe that they can help to enhance the quality of data models for data warehouse. Some of the metrics proposed in this work are available during early stages (design phase) of data warehouse development and hence, may be used to develop quality data warehouse systems. In future, we also intend to validate these metrics both theoretically using Measurement theory as well as empirically, so that there practical utility may be proven and then they can be used for proposing effective data modeling techniques, develop corresponding tool prototypes, test the prototypes in a

controlled laboratory environment, improve the prototypes, and validate the tools in realistic settings [9].

Acknowledgment

We are grateful to all those who have been constantly encouraging us to go for such application oriented study work besides the regular work which we are doing at our respective departments.

References

[1]   L. English, Information Quality Improvement: Principles, Methods and Management, Brentwood, Information Impact International,Inc., 1996.

[2]   Serrano, M., Trujillo, J., Calero, C., and Piattini, M. 2007. Metrics for data warehouse conceptual models understandability. Journal of Information and Software Technology. 49, 8 (Aug. 2007), 851-870.

[3]   M. Serrano, Definition of a Set of Metrics for Assuring Data Warehouse Quality, Univeristy of Castilla, La Mancha (Spain),2004.

[4]   V. Kumar, R. Thareja, Goal Structured Requirements Engineering and Traceability Model, International Journal of Information Technology and Computer Science, October 2013.

[5]   V. Basili, D. Weiss, A Methodology for Collecting Valid Software Engineering Data, IEEE Transactions on Software Engineering 10 (1984) 728–738.

[6]   Kieras D, Polson P (1985) An approach to the formal analysis of user complexity. Int J Man Mach Stud 22:365–394.

[7]   D. Batra, Cognitive complexity in data modeling: causes and recommendations, Springer-Verlag, London, 2006.

[8]   Funke J (1991) Solving complex problems: exploration and control of complex social problems. In: Sternberg RJ, Frensch PA (eds) Complex problem solving: principles and mechanisms. Erlbaum, Hillsdale, pp 185–222.

[9]   Hevner A, March S, Ram S, Park J (2004) Design science research in information systems. MIS Quart 28:75–105

[10] Piattini, M., Genero, M., and Jimenez, L. 2001. A metric based approach for predicting conceptual data models maintanability. International Journal of Software Engineering and Knowledege Engineering. 11, 6 (Dec. 2001),703-729.

[11] Berenguer, G., Romero, R., Trujillo, J., Serrano, M., and Piattini M. 2005. A Set of Quality Indicators and Their Corresponding Metrics for Conceptual Models of Data Warehouses. In proceedings of the Data Warehousing and knowledge Discovery (Copenhagen, Denmark, August 22-26, 2005). DaWaK' 05. Lecture Notes in Computer Science 3589, Springer. ISSN: 0302-9743. 95-104.

[12] Calero, C., Piattini, M, Pascual, C., and Serrano, M. A. 2001. Towards Data warehouse quality metrics. In Proceedings of the 3rd International workshop on design and Management of Data warehouses (Interlaken, Switzerland, June 4, 2001). DMDW'2001.

[13] Kumar V., Thareja R., "A review on best practices to improve data warehouse performance", International Conference on Computing for Sustainable Development, 5th-7th March 2014, BVICAM, New Delhi.

[14] Kumar V., Thareja R., "A simplified Approach For Quality Management In Data Warehouse", International Journal of Data Mining & Knowledge Management Process (IJDKP), vol. 3, no. 5, September 2013., pg 61-69.

[15] Thareja R., Data Warehousing, Oxford University Press, India 2009.

Authors

Vinay Kumar is a Professor in Vivekananda Institute of Professional Studies, Delhi. Earlier he worked as Scientist in National Informatics Centre, MoCIT, Government of India. He completed his Ph.D. in Computer Science from University of Delhi and MCA from JNU, Delhi. He has authored a book on Discrete Mathematics and has contributed many research papers to refereed journals and conferences. His area of interest is graph algorithm, steganography, data security, data mining and e-governance. He is member of CSI and ACM.

Reema Thareja is working as an Assistant Professor in Department of Computer Science, Shyama Prasad Mukherjee College for Women, University of Delhi. She is author of Programming in C, Data Structures, Data Warehousing, Data and File Structures (GTU), Computer Fundamentals and Programming in C, Introduction to C Programming and co-author Computer Programming and Data Structures (JNTU) all published by Oxford University Press.