International Journal for Quality research
UDK- 004.738.5:005.6
Short Scientific Paper (1.03)

MEASURING THE DATA MODEL QUALITY IN THE E-SUPPLY CHAINS

Zora Arsovski[1], Dragana Rejman Petrovic[1], Igor Milanovic[1], Vladimir Rankovic[1], Zoran Kalinic[1]

1) Faculty of Economics, University of Kragujevac, Djure Pucara 3, 34000 Kragujevac, Serbia
{zora, rejman, djidji, vladar, zkalinic}@kg.ac.rs

Abstract: *The implementation of Internet technology in business has enabled the development of e-business supply chains with large-scale information integration among all partners.The development of information systems (IS) is based on the established business objectives whose achievement, among other things, directly depends on the quality of development and design of IS. In the process of analysis of the key elements of company operations in the supply chain, process model and corresponding data model are designed which should enable selection of appropriate information system architecture. The quality of the implemented information system, which supports e-supply chain, directly depends on the level of data model quality.One of the serious limitations of the data model is its complexity. With a large number of entities, data model is difficult to analyse, monitor and maintain. The problem gets bigger when looking at an integrated data model at the level of participating partners in the supply chain, where the data model usually consists of hundreds or even thousands of entities.The paper will analyse the key elements affecting the quality of data models and show their interactions and factors of significance. In addition, the paper presents various measures for assessing the quality of the data model on which it is possible to easily locate the problems and focus efforts in specific parts of a complex data model where it is not economically feasible to review every detail of the model.*
Keywords: *e-supply chain, data model, quality, measurement*

1. FOREWORD

Regardless the fact that modern business operations always needs more efficiently approaches, managers all across the world are trying to find new ways to improve business. One of these, e-supply chain management (e-SCM), emerges as integration of Internet based e-business concept and supply chain management. In this paper we have presented the research related to data model quality metrics in e-SCM environment. Design of data model presents basis for database implementation that will serve for business improvement for all partners in the e-supply chain. On this way, quality of data model directly affecting on information integration and sharing among all partners in the e-supply chain. Information visibility and velocity of response throughout chain becomes timely and that directly reflects on companies' success in global open market competition. For measuring the quality of data model, we have used metrics presented by Moody [1] and Genero [2]. By using these metrics, quality of data model presented in our previous research paper [3] was analysed. Results of this case study are given bellow.

2. E-SUPPLY CHAIN

The supply chain is a concept that is rapidly developing in recent years. One of definitions could be – supply chain is a system of organizations, people, technology, activities, information and resources involved in moving a product or service from supplier to customer.

The supply chain is made up of all the activities that are required to deliver products to the customer - from designing product to receiving orders, procuring materials, marketing, manufacturing, logistics, customer service, receiving payment and so on. Anyone, anything, anywhere that influences a product's time-to-market, price, quality, information exchange, delivery, among other activities is part of the supply chain.

The fastest growing and highest volume by far is business-to-business transactions in supply chains. Internet capabilities already have, and will continue, to fundamentally change business-to-business supply chain models.

Even more intriguing is the rapid evolution of the digital marketplace which allows buyers and sellers to transact in a single intelligent, multidimensional marketplace that connects multiple trading exchanges. This allows buyers to consolidate orders from multiple vendors and subsequently provide for the effective integration of the final logistical activities. Putting intelligence into super portals so customers can get their information their way is essential.

E-SCM is defined as the impact that the Internet has on the integration of key business processes from end user through original suppliers that provides products, services, and information that add value for customers and other stakeholders. It combines the concepts of electronic business (e-business) and supply chain management (SCM), and depicts how trade channel members are working together to optimize resources and opportunities. E-SCM helps trade channel members collaborate to save money and deliver products with better quality. The Internet and advanced software solutions greatly enhance the benefits of SCM collaborating. Through SCM collaboration, suppliers can gain access to inventory and logistics data of retailers in order to meet just-in-time inventory demands and to provide efficient responses to customized orders.

For realization of e-SCM concept, information infrastructure is necessary as well as information integration of all partners through adequate database derived from data model.

3. DATA MODEL

Information Engineering is an integrated and evolutionary set of tasks and techniques that enhance business communication throughout an enterprise enabling it to develop people, procedures and systems to achieve its vision. The first step in any Information Engineering project is the Information Strategy Plan or ISP. The ISP would look at the data, process, organization, technology and interactions of an enterprise. Three key deliverables of an ISP are functional decomposition, an interaction matrix and a **data model**. [3]

Many different definitions of data model could be found in the literature:

"The activity of discovering and documenting information requirements." [4]

"Data model is used for describing entities and their relationships within a core domain." [5]

A data model describes the data items of a certain part of the perceived reality (business domain) relevant for a specific application or a specific user in a structured way. This model includes the relationships between the data.

On the other side, the term data modelling has a different meaning in nearly every organization and often the meaning changes even between different units within the same organization. [6]

"Data modelling is the art and science of arranging the structure and relationship of data." [7]

"Data modelling is a design discipline." [8]

Data modelling is a method used to define and analyse data requirements needed to support the business processes of an organization. It is process which realization is always supported by both, users and developers.

Entity relationship modelling is a relational schema database modelling method, used in software engineering to produce a type of conceptual data model (or semantic data model) of a system, often a relational database, and its requirements in a top-down fashion.

One of the most serious limitations of the entity relationship model in practice is its inability to cope with complexity. With large numbers of entities, data models become difficult to understand and maintain. This is the major reason why data modelling techniques have not realized their full potential in practice. The problem is multiplied many times over at the enterprise level, where models typically consist of hundreds or even thousands of entities.

The entity relationship diagram is the standard data technique for creating data models. The entity relationship diagram enables an analyst to create a graphical view of the data concepts of an organization and their relationships. Traditional system development dictates creation of an entity relationship diagram that is converted to a database design of a relational database.

The main objective when a SCM data model is developing is to supply it with purchasing and supplier information as much as it possible does. Important consideration when designing the model is to include large number of future users in the discussions. Knowledge of people familiar with data, which they use in every day operations, can provide meaningful input to the data model design.

It is important to include as many fields as possible in data model, but problem could be if too many fields are included. If data model is too complex, it will be difficult to provide the necessary data. If the vast majority of the users cannot provide data for a specific field, it may be better to not include that field at all.

Each of the entities must be able to be joined using key fields.In addition to specify what entities are going to be needed for data model, the data model should also document what data columns or attributes are going to be included in each entity. Adding or changing attributes can be tricky after begin populating the data model. Additionally, changes requested by one user that may adversely affect another user should be avoided.It may also be necessary to freeze the data model at some point in time during development because the stability of the data model can be jeopardized by constant change. It should not be forgotten that the data model is designed with an implementation in mind.By nature, a normalized entity relationship diagram tends to separate the data concepts into separate entities. A traditional approach to entity relationship modelling is concerned with three concepts: entities, relationships and attributes

4. QUALITY METRICS

The quality of data models can be understand as the match between the data model and the requirements of its users with regard to the defined application areas of data models

Quality of data modelling is not the same as quality of data models. Data modelling, as it is used in most organizations, is a method, which supports the process of adaptation, standardization and integration in the development of application systems. During this process, the individual perceptions and understandings of the members of the organization involved are brought together, discussed and integrated. [6]Data modelling quality consist of process quality and product quality. Process quality is the quality of the development and application of data models and product quality is the quality of the results of the modelling process.

Evaluating the quality of data models is difficult task because quantitative measurement of quality is almost non-existent. The quality factors and the primary stakeholders involved in evaluating them (given in brackets) are shown below [1]:

- Completeness (Business User),
- Integrity (Business User),
- Flexibility (Business User),
- Understandability (Business User and Application Developer),
- Correctness (Data Analyst),
- Simplicity (Data Analyst),
- Integration (Data Administrator),
- Implementability (Application Developer).

These quality factors may be used as criteria for evaluating the quality of individual data models or comparing alternative representations of requirements. Together they incorporate the needs of all stakeholders, and represent a complete picture of data model quality. For each quality factor, it is possible to define quality measures given in addition.

Completeness means that user requirements must be embeeded into the data model as set of attributes. It is the most important quality factor since if user requirement are not corresponding to data model, then all other quality factors are not important. If the requirements as expressed in the data model are inaccurate or incomplete, the system which results will not satisfy users, no matter how well designed or implemented it is. Quantitative measures are:

1. Number of items in the data model that do not correspond to user requirements
2. Number of user requirements that are not represented in the data model
3. Number of items in the data model that correspond to user requirements, but are inaccurately defined
4. Number of inconsistencies with process model - a critical task in verifying the completeness of the data model is to map it against the business processes that the system needs to support. This ensures that the model can meet all functional requirements. The result of this analysis can be presented in the form of a CRUD (Create, Read, Update, and Delete) matrix. Analysis of the CRUD matrix can be used to identify gaps in the data model as well as to cancel unnecessary data from the model.All of these numbers should tend to zero for maximum completeness.**Integrity** means that integrity constraints or business rules applied on data are built in data model. These rules are necessary to maintain the consistency and integrity of data stored, as well as to enforce business policies (e.g., it should be forbidden to input any other date of contract except present date).

Quantitative measures are:

5. Number of business rules which are not enforced by the data model
6. Number of integrity constraints included in the data model that do not accurately correspond to business policies - too weak (the rule allows invalid data to be stored) or too strong (the rule does not allow valid data to be stored)

All of these numbers should tend to zero for maximum integrity.**Flexibility** means that data model could be easily changed according to business changes in the company's environment. Environmental changes could not affect to data model changes in greater volume. Quantitative measures are:

7. Number of elements in the model which are subject to change in the future
8. Estimated cost of changes
9. Strategic importance of changes - expressed as a rating by business users of the need to respond quickly to the change

Above group of factors cannot be easily measured since it is based on prediction.

Understandability means if data model could be understood easily from the aspect of both, business users and application developers. If business user cannot understand the data model, then it will be harder for him to use it. Quantitative measures are:

10. User rating of understandability of model
11. Ability of users to interpret the model correctly - can be measured by the number of errors in populating the model, which is much more important from the point of view of verifying the accuracy of the model
12. Application developer rating of understandability

Above group of factors cannot be easily measured since it is based on personal judgement.

Correctness means that data model has been developed by using the adequate data modelling techniques without redundancies. Quantitative measures are:

13. Number of violations to data modelling conventions
14. Number of normal form violations
15. Number of instances of redundancy between entities

All of these numbers should tend to zero for maximum correctness.

Simplicity means that the data model contains the minimum possible constructs. Quantitative measures are:

16. Number of entities (NE)
17. Number of entities and relationships (NE+NR)
18. Number of constructs (NE+NR+NEA) - where NE is the number of entities, NR is the number of relationships and NEA is the total number of entities attributes

Genero, Jiménez and Piattini also defined metrics for ER Diagram Complexity (what is opposite from Simplicity) [2] through following formulas:

RvsE Metric measures the relation that exists between the number of relationships and the number of entities in an ER diagram.

$$RvsE = \left(\frac{N^R}{N^R + N^E} \right)^2 \quad (1)$$

More the N^R is greater than N^E, RvsE tends to be 1 and more the N^E is greater than N^R, RvsE tends to be 0. Higher RvsE corresponds to higher complexity and lower simplicity.

EAvsE Metric measures the relation that exists between the number of entity attributes and the number of entities in an ER diagram.

$$EAvsE = \left(\frac{N^{EA}}{N^{EA} + N^E} \right)^2 \quad (2)$$

More the N^{EA} is greater than N^E, EAvsE tends to be 1 and more the N^E is greater than N^{EA}, EAvsE tends to be 0. Higher EAvsE corresponds to higher complexity and lower simplicity.

RAvsR Metric measures the relation that exists between the number of relationship attributes and the number of relationships in an ER diagram. Number of relationship attributes N^{RA} is total number of attributes included in all relationships (primary and foreign keys).

$$RAvsR = \left(\frac{N^{RA}}{N^{RA} + N^R} \right)^2 \quad (3)$$

More the N^{RA} is greater than N^R, RAvsR tends to be 1 and more the N^R is greater than N^{RA}, RAvsR tends to be 0. Higher RAvsR corresponds to higher complexity and lower simplicity.We suggest one more formula that can describe ratio between entity and relationship attributes:

RAvsEA Metric measures the relation that exists between the number of relationship attributes and the number of entity attributes in an ER diagram.

$$RAvsEA = \left(\frac{N^{RA}}{N^{EA}} \right) \quad (4)$$

More the N^{RA} tends to N^{EA}, RAvsEA tends to be 1 and more the N^{RA} is less than N^{EA}, RAvsEA tends to be 0. Higher RAvsEA corresponds to higher complexity and lower simplicity.

Integration means that developed data model corresponds to other organisation's data if it was built additionally. It shows if data model fits to previously developed models. Quantitative measures are:

19. Number of data conflicts with the Corporate Data Model - entity conflicts (number of entities whose definitions are inconsistent with the definition entities in the corporate data model), data element conflicts (number of attributes with different definitions or domains to corresponding attributes defined in the corporate data model), naming conflicts (number of entities or attributes with the same business meaning but different names to concepts in the corporate data model).
20. Number of data conflicts with existing systems - number of data elements whose definitions conflict with those in existing systems, number of key conflicts with existing systems or other projects and number of naming conflicts with other systems.
21. Number of data elements which duplicate data elements stored in existing systems or other projects
22. Rating by representatives of other business areas as to whether the data has been defined in a way that meets corporate needs rather than the requirements of the application being developed

All of these numbers should tend to zero for maximum integration factor.**Implementability** means that data model should be easily implemented from both, time and budget aspect. Quantitative measures are:

23. Technical risk rating
24. Schedule risk rating
25. Development cost estimation

Above group of factors cannot be easily measured since they are based on estimation.

5. CASE STUDY

Metrics defined in previous section have been further analysed and decision was made that some of measures could be used for quality evaluation of data model – completeness, integrity, correctness and simplicity.

Other metrics (flexibility, understandability, integration and implementability) were ignored in this analysis because they cannot be easily quantified.

Previously realized integral data model for e-SCM in automotive industry (Figure 1) developed in our research [3] is subject for quality measurement.

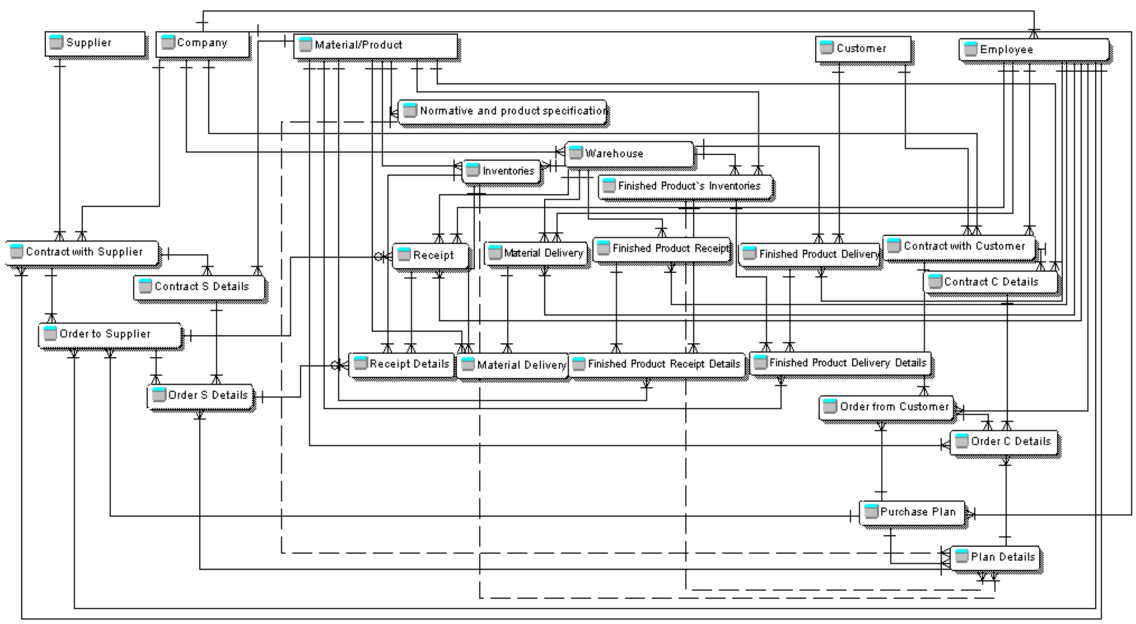

Figure 1.

Completeness means that user requirements must be embedded into the data model as set of attributes. Business users (stakeholders) define user requirements. We identified following stakeholders: owners, customers, suppliers and government and special interest groups.

Owners articulate their demands globally and express them through percentage of sales growth, percentage reduction of transportation costs, optimization of the supply chain and purchase plan realization. In analysis bellow, each owner requirement will be linked to adequate attributes given in entities of data model.

Percentage of sales growth could be derived from entities Contract with Customer and Contract C Details in a form of query designed from attributes Contract Date and Quantity. *Percentage reduction of transportation costs* cannot be derived from presented data model, but it can be easily added in a form of Payment entity with attributes Payment Date, Payment and Transportation Costs. *Optimization of the supply chain* could be derived from entities Material/Product in a form of query designed from attributes Transportation Days from Supplier to Company, Transportation Days from Company to Customer and Safety Days. *Purchase plan realization* could be derived from entities Purchase Plan and Plan Details in a form of query designed from attributes Quantity, From Date and To Date.

The requirements of *customers* in the supply chain include procurement of specified materials/products at previously contracted prices, technical specifications and specified quality, on time. One of the requirements that customers can request is related to the for the purpose of the supply chain.

Procurement of specified materials/products could be derived from entities Contract with Customer, Contract C Details, Order from Customer, Order C Details, Finished Product Delivery and Finished Product Delivery Details in a form of query designed from attributes Ordered Quantity, Price, Order Date, Estimated Delivery Date, Delivery Date, Shipment Date and Shipment Quantity. *Optimal finished product`s inventories* are present in Inventories entity in a form of attribute.

The dominant request of the *supplier* is related to accurate predictions inside of fluctuations of ± 20%, which is business standard.

Accurate predictions could be derived from entities Order to Supplier, Order S Details, Inventories and Finished Product`s Inventories in a form of query designed from attributes Ordered Quantity-fixed for n week, Ordered Quantity-prediction for n+1 week, Ordered Quantity-prediction for n+2 week, Ordered Quantity-prediction for n+3 week and Current Inventory Quantity.

The requirements of *governments and special interest groups* are articulated through the sets of laws and regulations that must be strictly respected in the business. Standard specifies requirements for quality management system developed specifically for the global automotive industry. Some requirements are: the obligation that all purchased products meet the requirements of regulations, the obligation to determine the quality of products purchased by a predefined method and the tracking of suppliers in terms of its history of delivery.

The obligation that all purchased products meet the requirements of regulations is present in Normative and Product Specification entities in a form of attribute. *Next two requests* could be derived from entity Receipt Details in a form of query designed from attributes Receipt Quantity, Rejected Quantity, Inadequate Packing, Rest Quantity and More Quantity.

Based on the above-performed analysis, it can be concluded that data model meets 9 of 10 requests that is 90% of completeness. Quantitative measures are:

1. Number of items in the data model that do not correspond to user requirements – 0,
2. Number of user requirements that are not represented in the data model – 1,
3. Number of items in the data model that correspond to user requirements, but are inaccurately defined – 0,
4. Number of inconsistencies with process model – 0.

Integrity means that integrity constraints or business rules applied on data are built in data model. Quantitative measures are:

5. Number of business rules which are not enforced by the data model – 0,
6. Number of integrity constraints included in the data model that do not accurately correspond to business policies - too weak (the rule allows invalid data to be stored) or too strong (the rule does not allow valid data to be stored) – 0.

Correctness means that data model has been developed by using the adequate data modelling techniques without redundancies. Quantitative measures are:

7. Number of violations to data modelling conventions – 0,
8. Number of normal form violations – 0,
9. Number of instances of redundancy between entities – 0.

Zeroes for metrics from 4 to 9 are consequence of using methodology described in [3].

Simplicity means that the data model contains the minimum possible constructs. Quantitative measures are:

10. Number of entities (NE) – 27,
11. Number of entities and relationships (NE+NR) – 92,
12. Number of constructs (NE+NR+NEA) – 331.

- **RvsE Metric**

$$RvsE = \left(\frac{65}{92}\right)^2 = 0,499 \quad (5)$$

- **EAvsE Metric**

$$EAvsE = \left(\frac{239}{266}\right)^2 = 0,807 \quad (6)$$

- **RAvsR Metric**

$$RAvsR = \left(\frac{108}{173}\right)^2 = 0,390 \quad (7)$$

- **RAvsEA Metric**

$$RAvsEA = \left(\frac{108}{239}\right)^2 = 0,204 \quad (8)$$

Results derived from the above equations (5-8) give us picture that analysed data model meet the quality requests. Only one of these numbers is higher (6), which means that number of entity attributes could be smaller. On the other side, less entity attributes could lead to inaccurate data model that cannot meet the user requirements.

6. CONCLUSION

Although it is hard to measure quality of data model, metrics exists and could be performed which is shown on the data model for e-SCM. Results of quality metrics contribute to timely detect and eliminate errors in the data model. This is prerequisite for time and money savings, which is the main goal of every successful company in the e-SCM.

REFERENCES

[1] Moody L., *Metrics for Evaluating the Quality of Entity Relationship Models*, Proceedings of the 17th International Conference on Conceptual Modelling (ER ´98), Singapore, pp. 213-225, November 1998

[2] Genero M., Jiménez L. and Piattini M., *Measuring the Quality of Entity Relationship Diagrams*, Proceedings of the 19th International Conference on Conceptual Modeling (ER 2000), Salt Lake City, USA, pp. 513-522, October 2000

[3] Arsovski Z., Rejman Petrović D., Milanović I., Ranković V., Kalinić Z., *Data Modeling of Automotive Industry Supply Chain*, Proceedings of the International Scientific Conference EBM 2010, Kragujevac,

Serbia, pp. 529-542, December 2010
[4] DeAngelis M. C., *Data Modeling with Erwin*, Indiana, Sans Publishing, p.7, 2000
[5] Topi H. , Ramesh V., *Human Factors Research on Data Modeling: A Review of Prior Research, An Extended Framework and Future Research Directions*, Journal of Database Management, Vol. 13 (2): pp. 3-19 , 2002
[6] Maier R., *Evaluation of Data Modeling*, Proceedings of the 3rd East European Conference on ADBIS, Maribor, Slovenia, Septembar, pp. 232-246, 1999
[7] McComb D., *Semantics in Business Systems*, California, Elsevier, p. 293, 2004
[8] Simsion G. C., Witt G. C., *Data Modeling Essentials*, San Francisco, Morgan Kaufmann publishers, p. 7, 2005
[9] Lysons K., Farrington B., *Purchasing and Supply Chain Management,* Prentice Hall, New York, 2006
[10] Gattorna J., *Dynamic Supply Chains*, FT Prentice Hall, Dorchester, 2010
[11] Arshinder, Kanda., A., Deshmukh, S.G., *Supply Chain Coordination: Perspectives, Empirical Studies and Research Directions,* International Journal of Production Economics, Vol. 115, pp. 316-335, 2008

Received: 17.09.2011 Accepted: 10.01.2012 Open for discussion: 1 Year