Integrating Customer Requirements into Data Product Design

Ec., Ph.D. candidate Otilia PÎRLOG

Successful data products are ones that meet customer needs, are innovative and that offer value. Quality Function Deployment offers a continuous process for ensuring that all the different quality issues are considered during the whole design and production chain. It provides a plane view of designed process and is typically applied to small systems. Its extension for large systems consists of few tree-dimensional cascading structures. This is an important tool for understanding the customer and deploying a structured manner throughout the data life cycle. A set of matrixes links customer desires, data quality characteristics and system functions with subsystems / procedures dedicated to generate data product desired. The formalization of the methodology given here demonstrates the capacity of using it like a key component to assist in creative thinking and problem solving. The subsystems / procedures proposed to be used are related with the corresponded costs.

Data Quality Function Deployment for Large Systems

The complexity of data product administration involves many systems. The main goal of data quality is to translate the customer's demands (Voice of Customer) into designed targets. *Quality Function Deployment (QFD)* invests time in planning to reap a profit in a shorter overall development cycle. The requirements are decomposed in a logical fashion, from the level of customer wants to the detailed process, using a set of correlated matrixes. The central matrix mediates the customer wants with supplier plans to provide them. This basic matrix can be expanded to provide additional insight to the supplier and cascaded to identify process parameters that must be controlled. There are many varieties of QFD and many variations of the charts used.

Data Quality Function Deployment (DQFD) is a method of translating customer desires throughout the entire data product cycle. It relates ideas to ideas, ideas to data, data to data process. This is one of the most powerful approaches for delivering a valuable data product to customers. It helps to achieve high quality data products by improving data design and process quality. It can also be viewed as the allocation of requirements to subsystems in that a subsystem must meet a set of requirements.

DQFD is a pointed way of listening to customers to understand exactly what they want and then, using a logical system, to determine how to best fulfill those requirements with available recourses. By relating these customer demands to the technical requirements (Voice of Developer), a set of target values for each technical requirement is logically derived. It addresses a set of dimensions: customer desire, data quality characteristics, functions, subsystems, procedures. Since requirements and technology largely dictate the complexity of the system, analysis is focused on the complexity of data characteristics.

Because the performance increases, the first unit cost of large systems has been escalating. The primary cost parameters are the size and the complexity of systems. In order to estimate and optimize the cost of data product, we have to analyze a lot of variables. The solution is *Data Quality Function Deployment for Large Systems (DQFDLS)*. It links the concurrent engineering process, the robust design process and the costing process to the data quality features. The effect is to generate a tightly project management process of high dimensionality which flushes out issues early to provide a high quality, low cost, and, hence, competitive data product.

The Three-dimensional Model

The first understanding of the basic structure for DQFD is giving a plane view with multiple correlations in all directions. This technique demonstrates the meaning of its name, ‘a matrix of matrixes’. With regard to DQFDLS, the picture is changed in a set of three-dimensional derived each-another shapes. This offers the possibility of a better look inside the system and to decide witch way is the best one for our design. Every level of analyze consists of a tree-dimensional representation (figure 1) in which are presented those categories of information that are correlated at that moment.

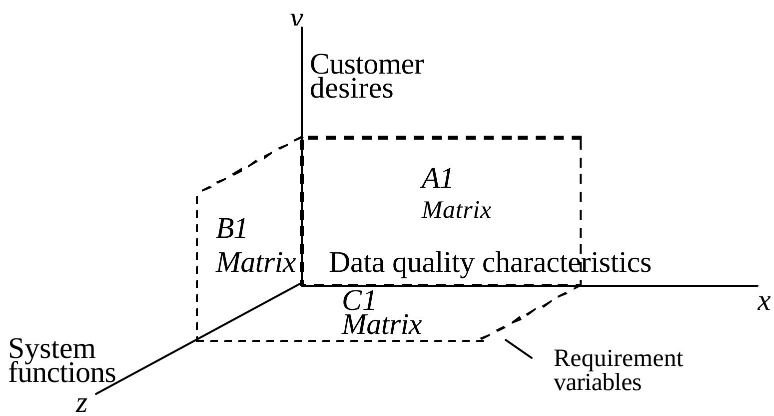

Fig. 1. Requirement Definition

The start is represented by the customer desires, the quality demanded by him. The customer desires are placed on the axis of a matrix. A data quality characteristic is a measurable attribute by which one can measure whether a customer is getting the demanded quality. For each customer desire are settled the data quality characteristics. Quality characteristics are defined through brainstorming to generate an affinity diagram. After forming a tree diagram of the chosen data quality characteristics, those on the lowest level are placed on the other axis of the analyzing matrix.

Each data quality characteristics are compared with each customer desire to determine the type of correlation: there is no correlation (value=0), a weak correlation (value=1), a moderate correlation (value=3) or a strong correlation (value=5). By summarizing the customer desires for the specific data quality characteristic it provided a value for that characteristic. This may be interpreted as the value of a data quality characteristic for a specific customer desire valuation. Mathematically speaking, the vector of values of customer desires is transformed into vector values for data quality characteristics, using the customer desire / data quality characteristics correlation matrix *(A1 matrix).*

The process is continued with identifying functions. A function is something the system must do to ensure the demanded quality. Usually it is defined in the form <verb, noun>. Quality characteristics and system functions intersect and define a requirement variable of the form <function, attribute> *(C1 matrix)*. Each requirement variable is a measurable attribute for which a correlation exists between the function and the data quality characteristic. Requirement variables can be fixed to create requirements or they can be used as design guidelines for improving the system. If these variables can be related through equations, they provide a parametric behavioral description of the product.

Data quality characteristics and functions can be ranked in terms of transformed customer value to determine which are the most important. If resources are constrained, then the most priority can be given to those with the highest customer value. The same way is followed to correlate the functions with customer desires using the customer desire/function correlation matrix *(B1 matrix).*

The determined axis for the first stage of analyze are:

- *xx’* – characteristics of data product for satisfying customer desires;
- *yy’* – customer desires;

- *zz'* – functions used to achieve data characteristics.

The process is continued by earmarking subsystems / procedures to the determined functions. This forms the new spatial structure of the model (figure 2). It can also be viewed as the allocation of requirements to subsystems in that a subsystem must meet a set of requirement variables.

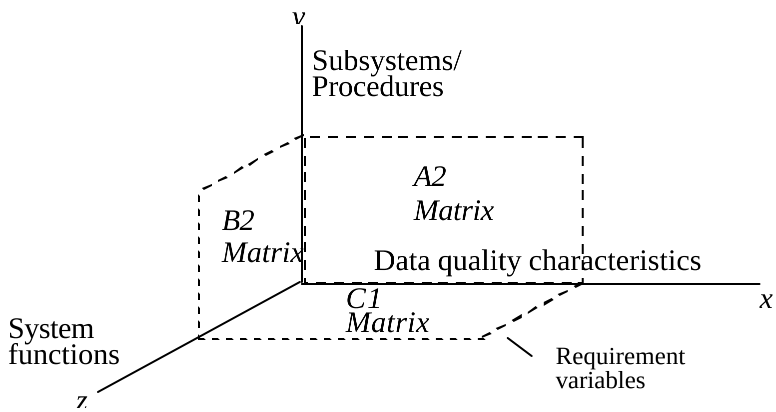

Fig. 2. Allocated Procedural Support

Now we are able to see the correlation that exists between data characteristics and subsystems / procedures allocated to realize them *(A2 matrix)*. Because of the way we used, the result of these designed subsystems / procedures will be qualitative data that will satisfy the customer requirements.

Every designed project has to take into account the level of necessary resources. The new step of the analyzing process is given by the replacing of the system functions with the other dimension consist of the cost of quality (Figure 3).

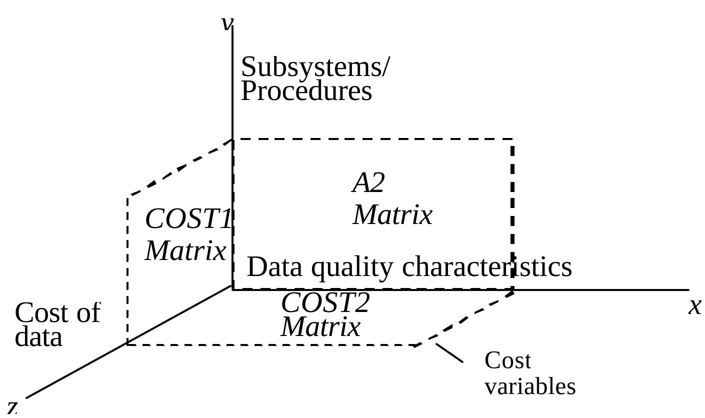

Fig.3. The Cost of Data Quality

This tree-dimensional picture praises costs in two appearances: first, the relationship that exists between costs and subsystems / procedures *(COST1 matrix)* and secondarily, between costs and the obtained result, data quality characteristics *(COST2 matrix)*. The first aspect shown represents an easy support for applying the activity based costing methodology.

All these spatial structures finalize the prototype support system. The team involved in the design of data product will evaluate it and will simulate different alternatives. This isn't a difficult approach because of the flexibility delivered by the prototype support system. In the same time, the result of this methodology allows the suitable maintenance.

Because of the wide range of expertise required for large systems, it becomes evident the need to decentralize the approach by decomposing the system into meaningful subgroups with lower interaction. The geometric nature of the method represents a natural medium in which to perform this decomposition and manage the interaction.

The Formalization of the Method

A very important aspect of using DQFDLS is the capability to translate the formal relationships in a mathematical form. The correlative method we provide here can be used as a real instrument in order to simulate different alternatives of the designed project. Finally, the best one will be select.

Every spatial structure presented above is corresponded with a set of tree correlated matrixes. To advance from one stage to the other means to change only a spatial dimension. From mathematical point of view this is similarly with the keeping of one matrix like support for the new representation and to add to it the two new derived matrixes. The first set of matrixes is shown in Figure 4.

Let $F_1 ¸ F_{n2}$ be system functions, $Char_1 ¸ Char_{n1}$ be data quality characteristics and $Des_1 ¸ Des_{n3}$ customer desires. Therefore $n1$ denote the number of data quality identified, $n2$ denote the number of system functions and $n3$ the range of the set of customer desires. To individualize a variable we will use $i=1 ¸ n1$ for data quality characteristics, $j=1 ¸ n2$ for system functions and $k=1 ¸ n3$ corresponding to customer demands. .

	System functions $\mathbf{F_1 ¸ F_{n2}}$				Customer desires $\mathbf{Des_1 ¸ Des_{n3}}$		
Data quality characteristics $\mathbf{Char_1 ¸ Char_{n1}}$	C1(1,1)			C1(1,n2)	A1(1,1)		A1(1,n3)
		C1(i , j)				A1(i , k)	
	C1(n1,1)			C1(n1,n2)	A1(n1,1)		A1(n1,n3)
Customer desires $\mathbf{Des_1 ¸ Des_{n3}}$	B1(1,1)			B1(1,n2)			
		B1(k , j)					
	B1(n3,1)			B1(n3,n2)			

Fig. 4. Set of Matrixes to Analyze Customer Desires

The next set of matrixes is built accordingly with the previous manner, by recovering the place of customer desires with subsystems / procedures decided $(P_1 ¸ P_{n4})$. Figure 5 presents the new set of determined matrixes, where $n4$ denote the number of subsystems / procedures and p is used to identify an element, $p=1 ¸ n4$.

	System functions $\mathbf{F_1 ¸ F_{n2}}$				Subsystems / procedures $\mathbf{P_1 ¸ P_{n4}}$		
Data quality characteristics $\mathbf{Char_1 ¸ Char_{n1}}$	C1(1,1)			C1(1,n2)	A2(1,1)		A2(1,n4)
		C1(i , j)				A2(i , p)	
	C1(n1,1)			C1(n1,n2)	A2(n1,1)		A2(n1,n4)
Subsystems / procedures $\mathbf{P_1 ¸ P_{n4}}$	B2(1,1)			B2(1,n2)			
		B2(p , j)					
	B2(n4,1)			B2(n4,n2)			

Fig. 5. Set of Matrixes to Analyze Subsystems / Procedures

The last step of analyze involves the cost of data quality. The connection between the last and the actual stage is given by A2 matrix. It will be the base of the new structure. In this case $n5$ is the range of identified categories of cost and c particularizes an element of the added matrixes, where $c=1 ¸ n5$ (Figure 6). Using the double view of projected costs, matrixes *COST1* and *COST2*, it is possible to simulate multiple scenarios and evaluate them in order to identify the best technical solution with the lowest cost level.

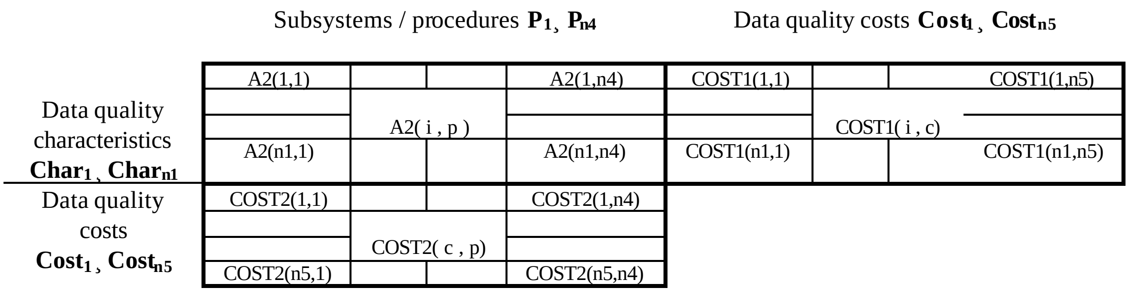

Fig. 6. Set of Matrixes to Analyze the Data Quality Cost

The adherent cost to a data characteristic is computed by summarizing corresponding costs:

$COST1_j = \sum COST1(i , c)$

The level of cost for every subsystems / procedures is computed in the same manner:

$COST2_p = \sum COST2(c , p)$

These sets of matrixes denote the easy maintenance of designed system. Regarding to large systems, this approach delivers also a strong support for decomposing the designed system. This is a very important appearance for the designing team. It suggests also the capacity of using the accumulated experience for some data quality characteristics and to pay more attention for the rest of them.

There are evident benefits of presented method:

- improves focus on customer needs:
  - helps prioritize customer requirements;
  - helps identify competitive gaps (benchmarking);
- reduce late design changes;
- promotes teamwork;
- ensures inter-departmental communication;
- facilitates concurrent engineering.

References

[Aung02] Aungst,S.,Barton,R.,Wilson,D.T.: *Integrating Marketing Models,* Institute for the Study of Business Markets, The Pennsylvania State University, ISBN Report 12-2002.

[Dean98] Dean, E.B.: *Quality Function Deployment from the Perspective of Competitive Advantage,* 1998, akao.larc.nasa.gov/dfc/qfd.html

[Daha02] Dahan,E.,Hauser,J.: *The Virtual Customer,* Journal of Product Innovation Management, September, 2002.

[Fran02] Franke,N.,Von Hipped,E.: *Satisfying Heterogeneous User Needs via Innovation Toolkits: The Case of Apache Security Software,* Massachusetts Institute of Technology (MIT) Sloan School of Management, Cambridge, MA, Working Paper, 2002.

[Rolp00] Rolph, P., and Bartram, P.: *The Information Agenda: Harnessing Relevant Information in a Changing Business Environment,* London, Management Books, 2000.