IQ Principles in Software Development

Dipl.Kfm. Michael Mielke

DB Bildung (DZB / TQM Team)
Teamleiter Informationsmanagement TQM
Solmsstrasse, 18
60486 Frankfurt / Main
michael.mielke@bahn.de

Abstract: This paper shows how we developed a 7-Steps-Method to determine, measure and improve DQ in our team. Based on an actual business process description, we worked out the critical process points. Then we interviewed all Team Members to get a basic idea of how our Information Customers rate Data Quality at all and within the sub-processes looking at the critical points. After we collect all that data we could set up a system that allows us to measure DQ in a Customers focus. This System is the basis for all our further work. Currently we are somewhere between Step 4 and 5, this paper will show consolidated and anonymous data till Step 3. We plan to continue publishing our results when further steps are taken

1 Problem Description

We are an internal consulting group of 80 team members. Our organizational structure looks like a matrix. Beside our boss we have a sales-team, a production-team, a service and IT-team and 5 different customer focus consulting groups. All tools they used to organize their work were developed internally by their own. There are two main databases, which are used to manage the team, one focuses on projects the other is focusing on the skills of our staff. Both databases are maintained basically looking more on data collection than on DQ. Most of the users are concerned about systems accessibility, reliability of data, usability, understand ability and so on. On the other hand the users do not update the current software system timely, which is the most important for reports

2 Problem Analysis

To avoid very high complexity we based our metrics on the 16 dimensions Richard Y. Wang published. We also determine the user groups of our organization and the basic business processes. The service group contains: Secretary, Accounting, Quality Management, DQ-Management, and IT-Management. In our team philosophy IT and Services is not a department. We look at it as independent services providers addressing all department members. Each service unit is lead by a Service-Unit-Manager who assists, consults and coaches also the Head of Department.

3 Measure Customer Focus

To answer all questions of our clients we set up a framework and designed a calculation model following our 7-Steps-Method.

1. Questionnaire your customers to get an idea which DQ Dimensions are the most important for them.
2. Analyze the business processes looking at the data streams that are involved.
3. Determine together with the involved users the critical points in their business processes
4. Questionnaire the involved users again, focusing on the process, sub-process and critical control points about their DQ requirements.
5. Determine metrics to measure if you fit the user requirements based on processes and critical control points
6. Measure and analyze
7. Questionnaire your customers, redesign your metrics if necessary and start with step 1 again.

We applied a basic Quality Function Deployment Technology known as the relationship matrix to priorities the 16 DQ Dimension invented by Rich Wang and others. We decide for our work to focus on the first five dimensions, that's the dimensions with the highest-ranking values. The following figure shows the framework we are using as well as some sample results to illustrate the ranking results.

Value rows to columns where 0 means not important, 1 means less important, 2 means more important. For example: Accuracy is less important than Objectivity comes out to value 1. All members of our organizational unit took place in the questionnaire, we then accumulate the result sets based on which process they belong to and computed the average value. We picked the first 5 Dimensions for each business process, according to our questionnaire we focused on what the member of our business unit where most interested in. As a result we have got 12 dimensions, then we put weights to each value as follows Rank 1 = 5 Points, Rank 2 =4 Points, Rank 3 = 3 Points, Rank 4 = 2 Points and Rank 5 = 1 Point.. In our next step we accumulate this information about our "clients" requirements on a control card and provided it to developers and business analyst so they where able to check their designs against these requirements.

439

Row is (0,1,2) related to column where: 0 means not important, 1 means less important and 2 means more important. Example: Accuracy is less important than Objectivity put in Cell C2 1	Accuracy	Objectivity	Believability	Reputation	Accessibility	Access security	Relevancy	Value-Added	Timeliness	Completeness	Amount of Information	Interpretability	Ease of understanding	Ease of manipulation	Consise represantation	Consistent represantation	Cross-added value	Standardization on Base of 10 10*Cross-added-value divided by max Cross-added-value
Accuracy		1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	**17**	6,07
Objectivity	2		1	2	2	2	2	2	0	0	0	1	1	1	1	1	**16**	5,71
Believability	2	2		2	2	2	1	2	1	1	1	1	1	1	1	1	**19**	6,79
Reputation	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	**14**	5,00
Accessibility	1	1	1	2		1	1	0	0	0	0	0	0	0	0	0	**6**	2,14
Access security	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	**14**	5,00
Relevancy	2	1	1	2	2	2		2	2	1	1	1	1	1	1	1	**19**	6,79
Value-Added	1	1	2	2	2	2	1		0	0	2	1	1	2	2	2	**20**	7,14
Timeliness	2	2	2	2	2	2	2	2		2	2	2	2	2	2	2	**28**	10,00
Completeness	2	2	2	2	2	2	2	2	2		2	2	2	2	2	2	**28**	10,00
Amount of information	1	1	1	0	1	1	2	2	2	2		2	2	2	1	1	**20**	7,14
Interpretability	0	2	1	1	1	0	2	1	1	0	0		2	1	1	0	**13**	4,64
Ease of understanding	2	2	1	1	2	2	2	2	2	2	2	2		2	2	2	**26**	9,29
Ease of manipulation	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	**14**	5,00
Consise representation	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	**0**	0,00
Consistent represantation	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		**0**	0,00

Figure 1: Ranking Framework

4 Information Flows in Business Processes

Taking the next step we collect examples of the "information products" the already use like monthly and annual sales reports. We then discussed with the sales department how these information products fits the sales iq-requirements like Relevancy, Timeliness and Objectivity. The basic concept for our approach is, starting on a meta view and going in more deep step by step to determine what exactly the requirements on data are, where the data is stored or provided, how to measure and what values your metrics should have. A monthly sales report for example (including the data you are looking for) has to be in the sales office not later than 5 days after computed at the end of the previous month (Timeliness). These sales reports have to include all sales sorted by main clients and products displaying the difference to the contract-data-base (Relevancy and Objectivity). We expect the monthly-sales-report to fit all these requirements. To compute the key-performance-indicator 2.1 the following calculation method is used:

$$1/3 * P(\text{Timeliness}) + 1/3 * P(\text{Objectivity}) + 1/3 * P(\text{Relevancy})$$

An sales report which is 2 days later provided than required, 4 main clients instead of 5 and with a 20% overall difference between sales-data-base and contract-data-base comes up to: $1/3 * (5/7) + 1/3 * (4/5) + 1/3 * (100 - 20) = 0,2380 + 0,2667 + 0,2664 = 0,7713$

This example report fits its DQ-Requirements with a level of 77,13%

5 DQ Metric System

We adopt a methodology we already have in place for quality management system metrics and allocation. The concept is to use the degree of performance instead of the current values, which allows us to allocate different key performance indicators independent from its origin dimension (miles versus pounds). Allocation of various KPI´s has two perspectives, horizontally within a business process and vertically in conjunction with other business processes. Both views are based on weights, which we estimate with the involved process owners. Our group in total agrees that the value we are generating as is primary the result of a team than on individuals. Therefore we weighted the conjunction in a special way by multiplication. In the conjunction would fail (value of 0) the whole calculated value comes up to Zero
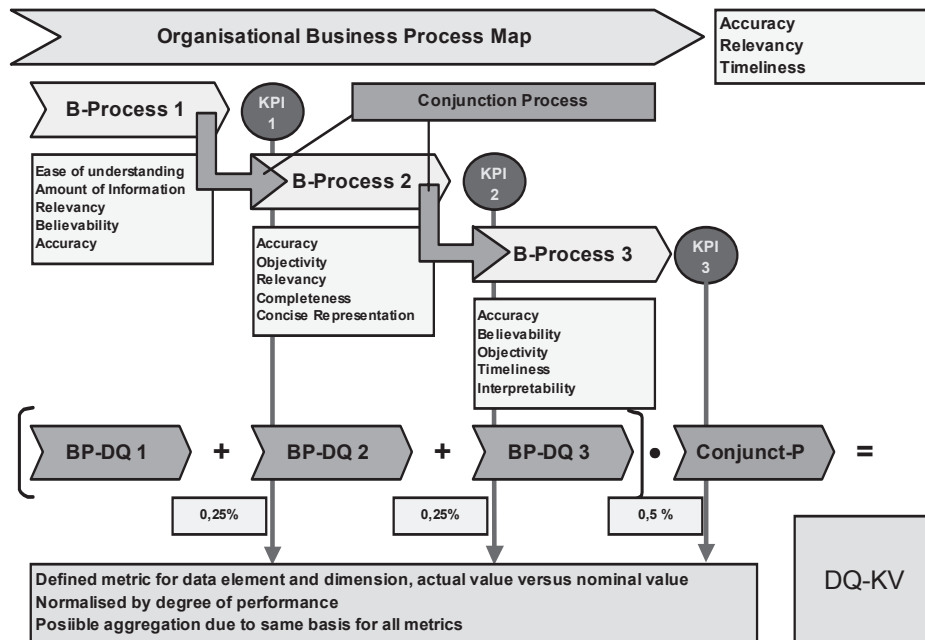
Figure 2: DQ-Metric Meta Model

6 Lessons learned

Getting "customers" data quality requirements into new designed software systems is not easy. It took and still takes us more time to define the measurable requirements than we expected. On the other hand this step-by-step approach already helps us to clarify our information needs and has result in a better cross organizational understanding and less complaints.

When we started the project I was the only one who has a basic idea about data quality, today our whole team knows that data quality is important or even the key for success. The software developers are working more customer-information flow-information quality oriented than before.

Implementing metrics in advance (timestamps, user-id, focus-id..) is a lot easier than trying to measure data quality requirements without appropriate values. It is also less cost intensive while we don not need to change or enhance a system. Last but not least it helps us to focus on the information (Tables, Elements) that are really needed, in most cases we where able to reduce the amount of data elements significantly.

Bibliography

[HLW 1999]   Huang, K., Y. Lee and R. Wang, Quality Information and Knowledge. Prentice Hall, Upper Saddle River: N.J., 1999.

[JMJ 1992]   Juran, J. M., Juran on Quality by Design: The New Steps for Planning Quality into Goods and Services. Free Press, New York, 1992.

[LRW 1991]   Yang, Y. and Y. R. Wang, Data Quality Calculus: A data-consumer-based approach to delivering quality data,   CISL-91-08. Composite Information Systems Laboratory, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, 02139, 1991.

[MM 1999]   Mielke, Michael, Quality Function Deployment, Weka 1999

[MRW 1992]   Madnick, S. and R. Y. Wang, Introduction to Total Data Quality Management (TDQM) Research Program, TDQM-92-01. Total Data Quality Management Program, MIT Sloan School of Management, 1992.

[WZL 2000]   Richard Y. Wang, Mostapha Ziad, Yang W. Lee, „Data Quality",Kluwer Academic Publishers, Nov 2000