

TOTAL DATA QUALITY MANAGEMENT: A STUDY OF BRIDGING RIGOR AND RELEVANCE

Fons Wijnhoven, University of Twente, Enschede, Netherlands,
a.b.j.m.wijnhoven@utwente.nl

Roy Boelens, SG Automatisering, Emmen, Netherlands, r.boelens@gmail.com

Rick Middel, University of Twente, Enschede, Netherlands, h.g.a.middel@utwente.nl

Kor Louissen, Honeywell, Emmen, Netherlands, kor.louissen@honeywell.com

Abstract

Ensuring data quality is of crucial importance to organizations. The Total Data Quality Management (TDQM) theory provides a methodology to ensure data quality. Although well researched, the TDQM methodology is not easy to apply. In the case of Honeywell Emmen, we found that the application of the methodology requires considerable contextual redesign, flexibility in use, and the provision of practical tools. We identified team composition, toolsets, development of obvious actions, the design of phases, steps, and actions, and sessions as vital elements of making an academically rooted methodology applicable. Such an applicable methodology, we name “well articulated”, because it incorporates the existing academic theory and it is made operational. This enables the methodology to be systematically beta tested and made useful for different organizational conditions.

Keywords: Total data quality management, applicability of theory, design science.

1 INTRODUCTION AND RESEARCH DESIGN

Data is computer stored information (Stamper, 1973) and its quality can substantially influence on three strategic vectors of organizations, customer interaction, asset configuration, and knowledge leverage [Venkatraman & Henderson, 1998]. Therefore, data is an important resource for organizational competitiveness [Redman, 1998]. Incorrect data at the point of customer interaction may result in significant perceived quality problems by clients, whereas high quality data gives opportunities of serving clients better. For asset configuration, high quality of data may improve inventory management, production and resources planning. In business networks, poor data quality may result in accumulating errors through the value chain, and high quality data may reduce human coordination costs. High quality data enables business analysts to find proper insights in production and service processes and propose ways of improving them.

The literature presents different views on quality [Garvin, 1988], which are reflected in the concept of data quality. Zahedi [1995] emphasizes two relevant views on data quality, i.e. a product based view to data quality, which defines quality as the difference in the quantity of some desired ingredient or material in a product and reality, and a production oriented view on data quality, which defines data quality as process conformance to requirements. In total quality management (TQM), both views have been integrated by Juran as fitness-for-use. This is probably the most fruitful perspective on data quality, because data is primarily important for users. The current total data quality management

Data quality category	Data quality dimensions
Intrinsic data quality	Believability, Accuracy, Objectivity, Reputation
Contextual data quality	Value-Added, Relevancy, Timeliness, Completeness, Appropriate amount of data
Representational data quality	Interpretability, Ease of understanding, Representation consistency, Concise representation
Accessibility data quality	Accessibility, Access security

Table 1. Data quality dimensions [Wang & Strong, 1996: 20]

TDQM theory is an implementation of the idea of fitness-for-use [Wang & Strong, 1996]. The main quality dimensions of this approach are summarized in Table 1. TDQM is specifically a problem for larger firms, who own large data bases. In such organizations, as a consequence of required specialization and differentiation [Mintzberg, 1983], many people have responsibilities for data and many people may have different data use needs. Additionally, the coordination costs [Gurbaxani & Wang, 1991] related to discussing data requirements and quality actions within Honeywell are often perceived as high and the intangible nature of data management makes an assessment of its benefits difficult [Sveiby, 1997]. Consequently, an organization of this kind needs efficient procedures and methods to coordinate data quality, without concrete value deliveries.

Many companies are aware of data quality issues, and TDQM has become a common theme in organizations. Surprisingly enough, the literature in the field of Total **Data** Quality Management is scarce. Only a few papers in academic journals report on data quality ([Ballou et al, 1997; Pipino et al, 2002; Redman, 1998; Wang & Strong, 1996] and only one paper includes a full discussion of a TDQM methodology [Wang, 1998]. This TDQM methodology has not resulted in a wide spread use, and we assume that considerable problems exist in the *applicability* of this body of knowledge. Applicability of knowledge can be studied in at least three ways, i.e. 1) by an action research study, which stresses the contextual local relevance of the knowledge, 2) by an evaluation study, which stresses the academic quality and rigor of the knowledge and its application, and 3) by a design science study, which aims at integrating the relevance and rigor in one study [Hevner et al, 2004; Romme, 2003; Van Aken, 2004]. This article chooses for the last option. Design science sees the relation between academic insight and practice from two ways; i.e. application of academic insights should deliver concrete value for practice, and applying academic insights should give feedback to

academic work [Hevner et al, 2004; Van Aken, 2004]. To assure utility, both Van Aken [2004] and Hevner et al [2004] suggest assessing the designed methodology in practice and using the evaluation of these tests to refine the methodology. To assure utility of the TDQM methodology, this paper follows the logic of the design science framework as given in figure 1.

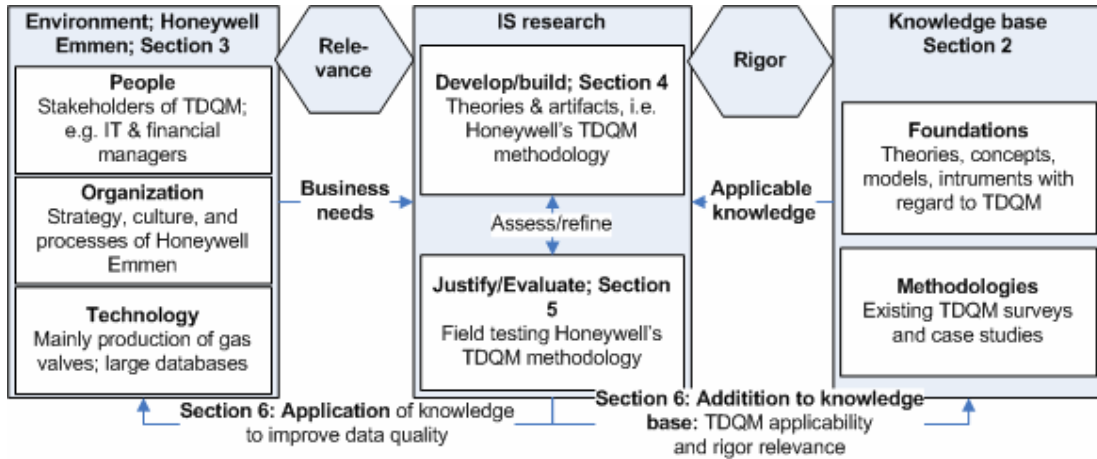


Figure 1: The research plan as an example of design research (based on Hevner et al., 2004, p. 80)

The described TDQM methodology (Knowledge Base) will be applied within the environment of Honeywell Combustion Controls Center Emmen (named Honeywell Emmen in the rest of this paper). A specific method for Honeywell Emmen is first developed (i.e. IS research), based on the existing TDQM knowledge base and the needs of Honeywell Emmen. These needs for Honeywell Emmen are based on interviews with key members in the area of the product-attributes, the business improvement and business IT manager, a financial manager responsible for the status of the products in the database, and a financial controller/planner. An assessment within Honeywell Emmen will then be executed. The results of the assessment are used to refine the method and to find additions to the knowledge base for TDQM with respect to its applicability.

2 THE TDQM KNOWLEDGE BASE

The current literature on data quality reports on concepts to describe the field (e.g. Wang & Strong, 1996) and theories that explain possible impacts and causes of data quality (e.g. Redman, 1998). This study, though, does not focus on this theory but aims at studying a methodology of handling data quality. Such a methodology should help solving problems (as explained by theory), avoiding problems, or achieving certain ambitions. An initial step for such a methodology is the design of the team that has to run the activities.

2.1. The TDQM team

Following Wang (1998), the data product quality team should consist of:

- Team leader: A senior executive able to implement chosen solutions.
- Team engineer: One who should have knowledge about the method and the techniques used in the method. Can also be the facilitator of the team meetings.
- Data suppliers: Those who create or collect data for the data product.
- Data manufacturers: Those who design, develop, or maintain the data and systems infrastructure for the data product.
- Data consumers: Those who use the data product in their work.

- Data product managers: Those who are responsible for managing the entire data product production process throughout the data product life cycle.

This team should be well trained in data quality assessment and management skills.

2.2 The TDQM cycle

TDQM methodologies use an adapted version of the Deming cycle. This cycle has become a key issue in the TQM literature, and it identifies four main steps; Plan, Do, Check, Act [Deming, 1986]. The TDQM methodology uses a version that is directed towards data quality, consisting of the steps; Defining, Measuring, Analysing, and Improving [Wang, 1998].

2.2.1. The define phase identifies three steps

Step 1 determines the data products characteristics. In describing the characteristics of the data product, Wang [1998] uses two levels. The highest level describes the characteristics of the total data product, and the lowest level describes each product-attribute individually. Figure 2 illustrates the difference between data products and product-attributes.

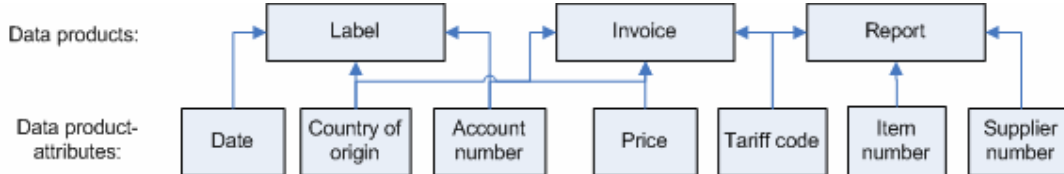


Figure 2: Data products and data product-attributes

Because not all data products and attributes can be assessed in one time effort it is useful to focus on the most important data product or attribute at a time, which may be the products and attributes whose quality problems have highest impact. Therefore Redman [1998] created a list of ten most commonly seen impacts of poor data quality.

Step 2 determines the requirements for the data products. The TDQM methodology proposes to determine which of the quality dimensions for data products (or attributes) [Wang & Strong, 1996] are most important by answering questions about:

- How important each team member thinks the dimension is,
- The perceived level of quality in a dimension, and
- The expected level of quality in a dimension.

Step 3 determines the data manufacturing process

The data manufacturing process consists of data flows from the supplier to the user of the data, including certain processing activities and quality checks. Wang [1998] explains that knowledge of the manufacturing process serves as a basis for better understanding why certain quality dimensions are important and how the responsibilities over this process are divided. Additionally, redundant processes and storages can be detected that often easily lead to errors and inconsistencies. Having a clearly defined process also helps in data quality management by codifying processes, making them person-independently and reliable. A simplified version of such a data manufacturing process is given in Figure 3.

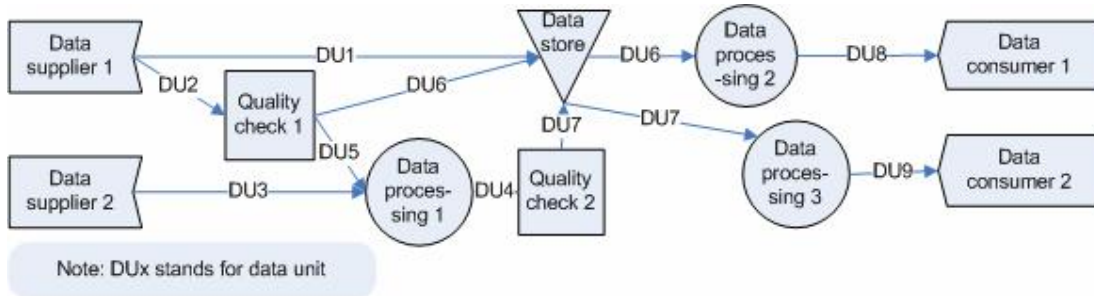


Fig. 3: Illustration of an information manufacturing systems (based on Wang, 1998, p. 64)

2.2.2 The measurement phase

The measurement phase determines the quality of the dimensions identified in the define phase. This phase involves two steps:

Step 1 selects proper metrics. In determining a metric for a data quality dimension (see Table 1), the team should keep in mind underlying business rules, (ISO) norms, and laws that might have contributed to a dimension's importance. Three quality dimensions measurement forms were identified by Pipino et al [2002].

- The Simple Ratio, measures the ratio of outcomes of a selected variable to total outcomes of that variable
- The Min or Max Operation, handles dimensions that require the aggregation of multiple data quality variables
- The Weighted Average. If a company for example has a good understanding of the importance of each variable to the overall evaluation of a dimension, then a weighted average of the variables is appropriate

Because it can be difficult to translate the dimensions chosen in the previous step into metrics which represent quality, the following guidelines (see Table 2) can be used [Kovac et al, 1997].

R	Is the metric Reasonable?
U	Is the metric Understandable?
M	Is the metric Measurable?
B	Is the metric Believable?
A	Is the metric Achievable?

Table 2. RUMBA guidelines for metrics [Kovac et al., 1997]

Step 2 measures and presents data. Once the metrics are determined, measurement can be conducted. If there are for example 20.000 products in the database, a sample size of about 400 products is needed to reach a reliability of 95% with a possible variance of 5% [Moore et al, 1989].

Several types of charts can be used to display the results. To identify which dimensions cause the problems the results can also be displayed using Pareto diagrams [Zahedi, 1995]. In this chart the horizontal axel displays the errors in the different dimensions, and the vertical axel displays the percentage of the errors in the different dimensions compared to the total number of errors. This is an easy way of showing which dimensions cause most problems.

2.2.3 The analysis phase

The goal of this phase is to find the root causes for the problems in the different dimensions [Wang, 1998]. Three methods to find root causes have become popular nowadays: the cause and effect diagram (CED), the interrelationship diagram (ID), and the current reality tree (CRT). According to Doggett [2005] the CED method is the easiest tool for identifying root causes. This step should answer the questions what the problem is, when it happened, why it happened, and how it impacts the overall goals of the firm.

2.2.4 The improvement phase

Wang (1998) proposes four steps for the improvement phase:

Step 1 solution generations. Wang [1998] proposes to use the information manufacturing analysis matrix designed by Ballou et al. [1997] and presented in Figure 3. This figure may make the lack of quality checks or processes apparent, or identify possible redundant processes and redundant data sources.

Step 2 selects solutions. When selecting solutions, it is necessary to know what impacts the solutions might have, the cost to design and implement the solution, the resources required, and the reduction of risk. Priorities for the solutions can be stated when the importance of each evaluation criterion is known [Zahedi, 1995].

Step 3 develops an action plan. The selected solution will have to be transferred into specific actions. The team should assign team members or other employees to these actions to make sure they are executed. A Project Action Plan can be set up to keep track of all actions and their statuses.

Step 4 checks progress. An obvious step, but only marginally discussed in the literature.

3. HONEYWELL EMMEN'S ENVIRONMENT

Honeywell Emmen is a production factory in The Netherlands which mainly produces gas valves. The data of these gas valves and the data of all its parts is stored in the Honeywell Emmen database system. Honeywell Emmen (and its customers) has experienced problems with the quality of this data in the recent months. The main causes of these problems that are:

- No structured way to analyze and improve product-attribute data quality
- No structured way to analyze and improve the division of responsibilities over the product-attributes
- Lack of a structured way for communication between stakeholders about problems in product-attribute data quality

The problems that Honeywell Emmen experiences are connected to multiple data products which are generated by the database system. The basis of these data products are about hundred different product-attributes.

4. DESIGN OF THE HONEYWELL EMMEN TDQM METHOD

The TDQM method for Honeywell Emmen was designed on basis of the TDQM theory of section 2, and modifications where needed to make TDQM feasible for them.

4.1 The TDQM team

Like Wang [1998] prescribes, Honeywell should first establish a data quality team. Because of the difference in data product handled by the TDQM methodology and product-attribute handled by the method of Honeywell, the following changes to the team described by Wang [1998] have to be made.

First of all, the team members may vary. The method has to be able to handle multiple product-attributes instead of one data product. Therefore, multiple data suppliers and consumers can be involved with each product-attribute. This leads to the need to select those team members based on their role in the chosen product-attribute. Data product managers, however, can not be selected, because the method does not handle a data product but data attributes instead. In this place a new role has to be assigned. A role with much influence in the data lifecycle of the product-attributes in Honeywell Emmen is the financial team member. The financial member has to assure that all the

necessary product-attributes are filled out before making a product 'active' in the database, which means that it can be ordered and produced. These changes lead to the following team members:

- Permanent members:
 - The team champion, i.e. a senior executive, to be able to implement the solutions generated by the team.
 - The team engineer, who has knowledge about the method and the techniques used in the method. Can also be the facilitator of the team meetings.
 - The data manufacturer, i.e. the person who designs, develops, and maintains the data for the product-attribute.
 - The financial team member, who is responsible for an important attribute and who has knowledge about the financial consequences of a proposed solution.
- Non permanent members:
 - The data supplier, who creates and collects data for the product-attribute.
 - The data consumer/user, who uses the product-attribute in his/her work.

4.2 Honeywell Emmen's TDQM cycle

The TDQ method we designed for Honeywell Emmen, in many ways is close to the theory described in section 2, and summarized in Table 3. This method, though, deviates on four fundamental issues from the theory:

Wang [1998] prescribes three steps in the define phase; determining the data products characteristics, determining the requirements for the data products, and determining the data manufacturing process. To apply the TDQM method on product-attributes (an essential decision based on the large size of the Honeywell Emmen databases), a step had to be added in which a product-attribute and the required non-permanent team members were selected. Experienced permanent team members should have knowledge of past product-attribute problems. Redman's list of ten commonly seen impacts of poor data can be used here to identify key attributes. The selection of the product-attribute influences who the involved stakeholders are. Therefore, the product-attribute selection should be done before the (non-permanent) team members can be selected. This means that the team should consist of permanent members, who can then choose non-permanent members depending on their role with the selected product-attribute.

The designed Total Data Quality Management Method			
Define Phase	Step 1:	Product-attribute and team members selection	Startup Session
	Actions:	Selection of the product-attribute to improve; Selection of team members; Schedule the first meeting	
	Tools 1	Fill in the Standard Project Description Form	
	Step 2:	Characteristics of the Product-attribute	First Group Session
	Actions:	Description of the main characteristics of the selected product-attribute.	
	Tools 2	Product-attribute Characteristics Form	
	Step 3:	Important Quality dimensions	
	Actions:	The members each score the importance, perceived quality and expected quality of each quality dimension; Determine which dimensions need improvement	
	Tools 3	Quality Dimension Score Sheet; Calculation Sheet	
	Step 4:	Data manufacturing process	Measurement Session
	Actions:	Determining the Data manufacturing process and make a flowchart of it	
	Tools 4	MS Visio TDQM Template	
Step 5:	Measurement Metrics		
Measure Phase	Actions:	(Reflect on the results of the previous meeting); Determine the metrics for the quality dimensions; Plan the measurement steps and actions that follow from the first steps	First Group Session
	Tools 5	RUMBA criteria	
	Step 6:	Measurement	Measurement Session
	Actions:	(Reflect on the chosen Metrics); Measure a sample of products and determine the quality; Possibly other data (then Oracle data) should be collected by other DQ Team members; Prepare Pareto diagrams and other graphs	
Tools 6	MS Excel		
Analysis Phase	Step 7:	Describe Specific Problems	Second Group Session
	Actions:	Describe the specific problems found by the measurement	
	Tools 7	Problem Description Form	
	Step 8:	Analysis of the problems	
	Actions:	Analysis of the problem by determining the cause map; Draw the cause map	
Tools 8	MS Visio, MS Excel, or a whiteboard		
Improvement Phase	Step 9:	Solution Generation	Second Group Session
	Actions:	Think of solutions to every cause that is defined on the cause map	
	Tools 9	MS Visio, MS Excel, or a whiteboard	
	Step 10:	Solution Selection	
	Actions:	The team selects possible solutions; The solutions have to be scored based on their impact, costs, and risks. The solution that scores best has to be implemented.	
	Tools 10	Solution Score Sheet	
	Step 11:	Action Plan	Extra Session
	Actions:	Tasks should be identified to implement the solution; The DQ team should assign these tasks to team members or other employees.	
	Tools 11	Project Action Plan	
	Step 12:	Check Progress	
	Actions:	Discuss the progress of implementing the solutions; Solve problems	
	Tools 12	Update the Project Action Plan if needed	

Table 3 The Total Data Quality Management Method for Honeywell

We also had to develop toolsets to facilitate concrete actions in each set. These toolset, consisting of forms, excel sheets, and drawing tools, have been regarded by the Honeywell team as essential for easing discussions in each phase, and for documenting and maintaining consistency during the TDQM cycle. It also made the cycle less sensitive on employee job changes which otherwise may result in substantial lost memory.

We followed the prescriptions of section 2 in many ways, though we had to be much more specific with regards to the last step, i.e. Check Progress, because for Honeywell Emmen real improvements

are aimed at, and a quality cycle was not believed to result in improvements by itself. To be sure that the actions of the action plan are implemented, the team leader can plan a meeting to review the progress of the action plan. Problems can be dealt with and new actions can be initiated if needed.

We also had to plan concrete sessions, each of which made sense in terms of the cycle.

5. ASSESSMENT AND REFINEMENT OF HONEYWELL EMMEN'S TDQM

The designed method has been field tested on two different product-attributes. Before the first product-attribute was selected the permanent team members were selected based on the descriptions in the previous chapter. The manager of the product data management department was chosen as the team leader. This manager is closely related to the product-attributes and is part of the management team of Honeywell Emmen. The team engineer and data-product manufacturer roles were combined into one person, the database administrator. He has extensive knowledge of the database systems, knows the possibilities of the system, and knows the MS Excel and MS Visio tools to analyze the data and present it to the team. The cost administrator was chosen as the financial team member. She has knowledge about the possible financial impacts of a chosen solution and is also involved in the product data lifecycle. She has the responsibility to activate new products on the database.

Section 5.1 gives the results of the first cycle of the method within Honeywell and section 5.2 gives the results of the second cycle of the method.

5.1 The first TDQM Cycle: Country-of-Origin attribute

In the starting meeting, the permanent team members selected Country-of-origin as the first data product attribute to consider based on the problems it had caused in the past. At this meeting, also the non-permanent team members from the purchasing and customer logistics departments were selected. The first complete data quality team meeting was also planned. This data was filled out in the Standard Project Description Form.

5.1.1 The first team session

In the first data quality team meeting, the team engineer, the data supplier (i.e. the representative from the purchasing department), and a data user (i.e. someone from the customer logistics department) were present. The team leader and the financial member were not present. Management did not formally approve the team yet, so these team members gave priority to other tasks. Other members also indicated that they had little time to attend the meetings. The third step of the method was done in a group discussion instead of using the calculation sheets, because a prior discussion could have influenced the scores on the Quality Dimensions Score Sheets, and because it saved time. Step five was not completed, because the data supplier and data user had to go to other meetings.

These events lead to the first conclusion that the method has to be flexible, depending on the complexity of the product-attribute handled. Next to this insight the team also made some adjustments to the standard documents, to make them clearer and easier to interpret.

5.1.2 The measurement session

The measurement phase was conducted by the team engineer. Both step five and six were conducted without problems. The results of the measurements revealed problems in the completeness and consistency dimension. The results in the completeness dimension are given in Figure 4. The solutions are sorted in the main product categories Honeywell Emmen uses i.e.

- “Make-products” which are made by Honeywell Emmen themselves

- “Buy-products” which Honeywell buys from other Honeywell companies and sells on the Dutch market
- “Active buy products”, which are buy products that can be bought and sold.

Next to this sortation, the results are also sorted across the different fields in which the data for the attribute is stored, in this case six different fields.

The completeness dimensions revealed missing data in 21% of the 400 sample products, measured on six database levels.

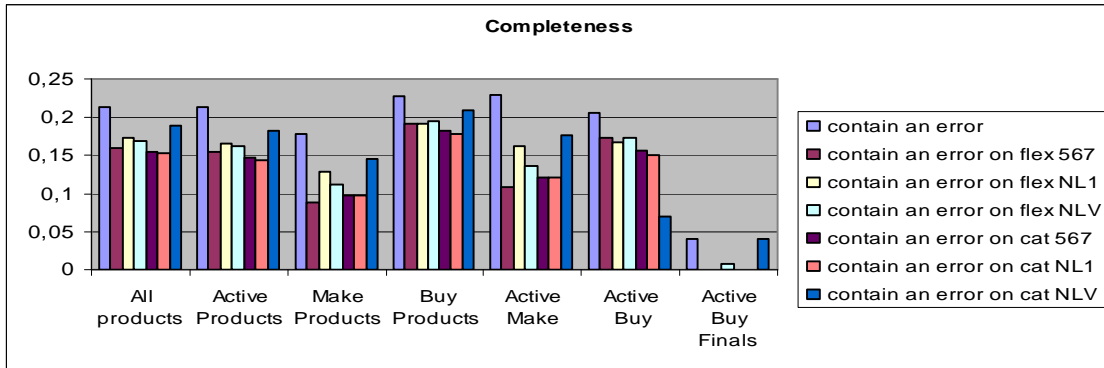


Figure 4: The measurement results in the completeness dimension.

5.1.3 The second team session

The second complete data quality team meeting started without the team leader and the data supplier. Both were on vacation. Step seven until eleven were conducted. The Standard Problem Description Form was filled out, and based on the conclusions of the cause analysis the Project Action Plan was filled out. Since the team had already a clear idea about the causes and the solutions to these causes, the steps did not take much time. The cause map was not drawn and the team quickly decided on which solutions to implement.

5.1.4 Evaluation

The main conclusion that can be stated after the evaluation of the first cycle, based on interviews and an evaluation survey, is that time was a limiting factor in the team meetings, because of lacking formal approval. This increases the need to institutionalize the data quality team.

The evaluation revealed that the knowledge of the team members about the product-attribute improved. The team members indicated that this method provided a structured way to handle the problems with this product-attribute. They liked the fact that it guided the meetings about the analysis of the product-attribute quality. They were also confident that the method would be able to improve the quality of the product-attributes.

5.2 Second TDQM Cycle: The Product-family-code attribute

Prior to the second cycle the method was reviewed and a shorter version of the third step was added. A new Quality Dimensions Score Sheet was developed to be able to document the results of the group discussion about the importance of quality dimensions. Also the scoring scale was reduced from 7 to 3 levels. This was done to improve clarity of the results.

To be able to determine how the method handles a less complicated product-attribute (a product-attribute with only one stakeholder), and to see how much time this takes, the team decided to handle

the Product-Family-Code (PFC) in the second cycle. This product-attribute plays an important role in sales analysis and could therefore influence the decision-making process. The product-attribute is supplied by the financial team member, and also used primarily by the financial team member. Therefore no non-permanent team members had to be invited.

5.2.1 The first team session

The first complete data quality team meeting was planned in the holiday season, therefore, only the financial team member (next to the BIT&BI manager and one of the authors of this paper) was present. During this first session all planned steps (step two till step five) were conducted. In step three the team used the new quality dimension score sheet, which provided more explanation about the dimensions and contained the three point scale (low, medium, high). Step four was skipped, because the data manufacturing process was too simple to describe. A short description of this process was added to the characteristics of the product-attribute. In step five three measurement metrics were chosen to measure the completeness, consistency, and accuracy dimensions.

5.2.2 The measurement session

The measurement session was done by the financial team member with help of one of the authors. The measurement revealed that only five missing data fields were found in a sample size of 400 products. These five products were all new products, which lead to the conclusion that no further actions were required in this product-attribute. The financial team member had no other remarks about the product-attribute and its process. The documents were stored and the project was completed.

5.2.3 Evaluation

Handling this product-attribute proved that the method is able to quickly come to a conclusion about the quality of this relative simple product-attribute. The first session took about one and a half hour and the measurement session took about four hours.

The evaluation did not show new issues. Although the method did not increase the knowledge about the product-attribute (only one stakeholder is involved in the process and she already knew the product-attribute), the method did give insight in the quality of the product-attribute.

6 CONCLUSION: CONTRIBUTIONS TO THE KNOWLEDGE BASE AND APPLICABILITY

Equipping the TDQM team with the designed method has provided Honeywell Emmen with a structured way to communicate about product-attribute problems, and analyzing and improving the quality of the product-attributes. This resulted in for instance the improvements to the completeness quality dimension of the Country-of-Origin product-attribute realized in the first cycle. See Figure 5.

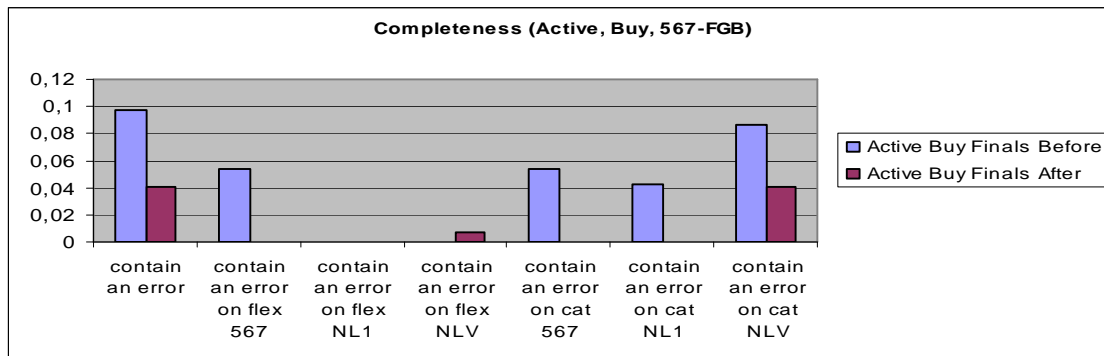


Figure 5: Quality improvements of Completeness of the Country of Origin product-attribute after using the TDQM method.

It is important to note that the TDQM theory was not able to realize these impacts in a direct way, as we had to make the theory applicable by improving it on five points:

- Adjust the TDQM team composition by splitting it in a *permanent and non permanent team* part. The permanent team enables continuity in TDQM and the non-permanent team members help to focus on specific urgent topics.
- We added *toolsets* to support the actions required in each phase of the TDQM, to improve ways of doing the tasks needed and to develop references points for the execution of the cycles.
- Some *obvious actions* were incompletely defined in the theory, but although the lacking actions seem rather obvious, they are essential to make the TDQM a success. It is surprising how vital and complex actions can be that from an academic perspective seem rather obvious. We particularly refer here to the Check Progress activities.
- We also became aware that a good methodology here needs *phases, steps, actions and tools, including a concrete session plan*. These topics are rather frequently mixed up in unclear definitions of methods and techniques.
- In the application of the methodology we also met the urgency of *flexibility*. Situational aspects require a cycle to be shorter or longer, more detailed or more superficial, have other team members and techniques.

We believe that team composition, toolsets, development of obvious actions, the design of phases, steps, actions, and session plans are vital elements of making an academically rooted methodology applicable. A methodology we want to name “well articulated” methodology, when it incorporates the existing academic theory and it is made operational. Such well-articulated methodologies are sources for:

- Beta testing by other companies, so that their development is improved by practical experiences of others, and
- Development of so-called kernel theories (Walls et al, 1992) at the middle range level that explain what choices with respect to the methodology design issues are best given what conditions. The experiences of other companies are important here, and may be collected by surveys and comparative cases studies.

It is obvious that the study we did here is limited to data in databases. The research may be extended to:

- Information, i.e. also data on non-database media. Probably this requires theory from very different fields, like communication theory and interpersonal sense making (Te’eni, 2002).
- Data on the Internet. Knowing the quality here is difficult and people are easily misled by incorrect data on the Internet (Poston & Speier, 2005).

Literature

- Ballou, D.P., Wang, R.Y., Pazer, H., & Tayi, G.K. (1997). Modelling information manufacturing systems to determine information product quality. *Management Science*, 44 (4), 462-484.
- Deming, E.W. (1986). *Out of the Crisis*. Center for Advanced Engineering Study, MIT, Cambridge, MA.
- Doggett, A.M. (2005). Root cause analysis: A framework for tool selection. *Quality Management Journal*, 12 (4), 34-45.
- Garvin, (1988). *Managing Quality: The strategic and Competitive Edge*. Free Press, New York.
- Gurbaxani, V. & Wang, S. (1991). The impact of information systems on organizations and markets. *Communications of the ACM*, 34 (1), 59-73.
- Hevner, A.R.; March, S.T.; Park, J.; Ram, S. (2004). Design Science in Information Systems Research, *MIS Quarterly* 28 (1), 75 -105.
- Kovac, R., Lee, Y.W., & Pipino, L.L. (1997). Total Data Quality Management: the case of IRI. *Proceedings of the Conference on Information Quality*, pp. 63-79. Cambridge (MA).
- Mintzberg, H. (1983). *Structures in fives: Designing effective organizations*. Prentice-Hall, Englewood Cliffs (NJ).
- Moore, D.S. & McCabe, G.P. (1989). *Introduction to the Practice of Statistics*. W.H. Freeman & Company, London.
- Pipino, L.L., Lee, Y.W., & Wang, R.Y. (2002) Data quality assessment. *Communications of the ACM*, 45 (4), 211-218.
- Poston, R. & C. Speier (2005), Effective use of knowledge management systems: A process model of content ratings and credibility indicators. *Management Information Systems Quarterly*, 29 (2), 221-244.
- Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM*, 41 (2), 79-82.
- Romme, A.G. (2003), Making a difference: Organization as design. *Organization Science*, 14 (5), 558-573.
- Stamper, R. (1973), *Information in business and administrative systems*. John Wiley, New York.
- Sveiby, K.E. (1997). *The new organizational wealth: Managing and measuring knowledge-based assets*. Berrett-Koehler, San Francisco (CA).
- Te'eni, D. (2001), Review: A cognitive-affective model of organizational communication for designing IT. *Management information systems quarterly*, 25 (2), 251-312.
- Van Aken, J.E. (2004). Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules. *Journal of Management Studies*, 41 (2), 219-246.
- Venkatraman, N. & Henderson, J.C. (1998). Real Strategies for Virtual Organizing. *Sloan Management Review*, 41 (1), 33-48.
- Walls, J.G., G.R. Widmeyer, & O.A. El Sawy (1992). Building an information system design theory for vigilant EIS. *Information Systems Research*, 3 (1), 36-59.
- Wang, R.Y. (1998). A product perspective on total data quality management. *Communications of the ACM*, 41 (2), 58-65.
- Wang, R. Y. & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12 (4), 5-34.
- Zahedi, F. (1995). *Quality Information Systems*. Boyd & Fraser, Danvers (MA).