Honeywell Emmen
Combustions Controls Center Emmen
Business IT & Business improvement

Universiteit Twente
School of Business, Public Administration & Technology
Department: Information Systems & Change Management

A PRODUCT-ATTRIBUTE APPROACH TO TOTAL DATA QUALITY MANAGEMENT

Assuring high data quality in the product-attributes of the Oracle database at Honeywell Emmen

By: Roy Boelens (s0008486)

Start: 1 February 2006
End: 15 July 2006.

Company Contact:
Kor Louissen
Business IT & Business Improvement Manager

University Twente Contacts:
Dr. A.B.J.M. (Fons) Wijnhoven
IS&CM Department, Universiteit Twente
Ir. H.G.A. (Rick) Middel
OOHR Department, Universiteit Twente

Management Summary

Problem description

Data is an important aspect in keeping companies competitive. Within Honeywell Emmen there are however numerous customer complaints, and internal reports of problems with the quality of data in the database. For example, shipments were delayed because of missing data, transfer prices were outdated and bills of materials were not accurate. These problems have lead to the start of this project. The planned implementation of SAP has also contributed to the importance of the product-attribute quality. Interviews with stakeholders from the different departments involved with product-attribute data quality resulted in the identification of three main causes of the problems:

- No structured way to analyze and improve product-attribute data quality,
- No structured way to analyze and improve the division of responsibilities in the product-attributes manufacturing process, and
- Lack of a structured way to communicate problems about quality of product-attributes

To eliminate these causes a literature research has been conducted and based on these findings a method has been designed. This method was adapted to handle the product-attributes Honeywell Emmen uses, and the redesigned method was tested on two of these product-attributes (*Country of Origin, Product Family Code*).

Recommendations

Based on this data quality theory and the tests of the designed method the following recommendations are formulated:

- Form a data quality improvement team, with:
  - A team leader, who is responsible (process owner) of the product-attributes and has the power to implement solutions
  - A team engineer who can facilitate the team meetings
  - A data manufacturer who has extensive knowledge about the database system
  - A financial team member, who understands financial consequences of proposed solutions
  - Product-attribute data suppliers, selected on the chosen product-attribute to handle
  - Product-attribute data users, selected on the chosen product-attribute to handle
- Introduce the designed method as a tool the data quality improvement team can use.
- Institutionalize the data quality team and assign data quality responsibilities to the team leader.
- Facilitate the data quality improvement efforts and make hours available for the team members.

Motivation

The recommendations are based on the Total Data Quality Management methodology designed by Wang [22]. The goal of this method is to improve the quality of data. The method is based on the Total Quality Management principals of, customer satisfaction, continues improvement, teamwork and participation. Using the method should provide a structured way to analyze and improve the quality of the product-attributes. The description of the data manufacturing process provides a way to analyze and improve the division of responsibilities in each product-attribute data manufacturing process. Working in a team with all involved departments can shorten communication chains and therefore increase the quality of communication.

Consequences

Implementing the data quality improvement team costs the chosen team members time and therefore money. Cycles of the method can however be planned in periods of less workload. The two group meetings of the team both take an average of two hours. The time required to implement the solution depends on the agreed actions that have to be taken. After the actions are implemented and the involved employees follow the described data manufacturing process the quality of the product-attribute will be, and remain at, the level the users of the product-attributes require. The quality of the database will therefore increase when more product-attributes are handled with the method

Management Samenvatting

Aanleiding
Data is een belangrijk aspect in het verkrijgen van concurrentievoordeel. Bij Honeywell Emmen zijn er echter verschillende klachten van klanten en interne berichten van problemen binnengekomen die betrekking hebben op de kwaliteit van data in de database. Bijvoorbeeld, klachten over vertraagde leveringen, omdat gegevens over de oorsprong van de producten ontbrak, transfer prijzen die al aangepast hadden moeten worden, en fouten in materiaal lijsten. Deze klachten en problemen hebben geleid tot het starten van dit project. De geplande implementatie van SAP heeft er ook toe bijgedragen dit project nu te starten, omdat data kwaliteit ook erg belangrijk is bij het implementeren van een nieuw database systeem. Interviews met betrokken partijen hebben geleid tot de identificatie van drie hoofdoorzaken van het probleem:

- Geen gestructureerde manier van analyseren en verbeteren van de kwaliteit van de productparameters
- Geen gestructureerde manier van analyseren en verbeteren van de verdeling van verantwoordelijkheden rond de productparameters
- Geen gestructureerde manier van het communiceren van problemen met productparameters

Om te zorgen dat deze oorzaken weggenomen worden is er een literatuuronderzoek uitgevoerd. Gebaseerd op deze uitkomsten is een methode ontwikkeld die de productparameters die Honeywell Emmen gebruikt kan verbeteren, en deze methode is vervolgens getest op twee productparameters *Country of Origin, Product Family Code*).

Aanbevelingen
Gebaseerd op de theorie van Total Data Quality Management en testen in de praktijk binnen Honeywell Emmen kunnen de volgende aanbevelingen worden gedaan:

- Vorm een data kwaliteitsverbeteringteam met:
  - Een team leider, die verantwoordelijk is voor de kwaliteit van de productparameters
  - Een team '*engineer*', die de meetings van het team kan faciliteren
  - Een '*data manufacturer*', die kennis heeft van het databasesysteem
  - Een financieel teamlid, die kennis heeft van financiële gevolgen voor de voorgestelde oplossingen
  - Een data leveranciers, (wordt geselecteerd als bekend is welke parameter behandeld wordt)
  - Een data gebruiker, (wordt geselecteerd als bekend is welke parameter behandeld wordt)
- Introduceer de ontwikkelde methode, als *tool* voor het data kwaliteitsverbeteringteam.
- Institutionaliseer het data kwaliteitsteam en geen de team leider verantwoordelijkheid voor de kwaliteit van de data.
- Stel resources beschikbaar voor datakwaliteit verbeteringsinitiatieven en maak uren beschikbaar voor de team leden.

Motivatie
De aanbevelingen zijn gebaseerd op de *Total Data Quality Management* methodologie ontwikkeld door Wang [22]. Het doel van deze methode is het verbeteren van de kwaliteit van data. De methode is gebaseerd op de volgende *Total Quality Management* principes: Klanttevredenheid, continu verbeteren, en *teamwork* en participatie. Het gebruik van de methode zorgt voor een gestructureerde methode om the kwaliteit van de productparameters te analyseren en verbeteren. De beschrijving van het proces waarin de data verzameld en ingevoerd wordt, zorgt voor een goede mogelijkheid om de verdeling van de verantwoordelijkheden te herzien. Het toepassen van de methode met een data kwaliteitsverbeteringteam zorgt ervoor dat communicatie afstanden verkleinen en dat team leden elkaar makkelijker weten te vinden als er problemen in een bepaalde productparameters zijn.

Consequenties
Het implementeren van het data kwaliteitsverbeteringteam zal de gekozen teamleden kostbare tijd kosten. De *cycles* van de methode zouden echter wel om drukke periodes heen gepland kunnen worden. De twee groepsbijeenkomsten waaruit de methode bestaat zullen beide gemiddeld twee uur duren. De tijd die vervolgens nodig is om de oplossing volledig te implementeren hangt natuurlijk af van de acties die ondernomen moeten worden. Als deze acties eenmaal zijn ingevoerd en de betrokken medewerkers voeren hun taken uit volgens het beschreven data verzamel en invoer proces, dan kan Honeywell Emmen ervan uitgaan dat de kwaliteit van de productparameters op het gewenste niveau van de gebruikers blijft. De kwaliteit van de database zal hierdoor na elke *cycle* van de methode verbeteren.

Preface

In February 2006 Honeywell Emmen offered me an internship to write my thesis. This unique offer gave me the opportunity to apply the theoretical knowledge learned at the University of Twente. Within Honeywell, I was placed in the Business Improvement and Business IT department, with the project initiator as my supervisor.

The colleagues at Honeywell offered me a warm welcome, and helped me with my thesis anyway they could. I learned to know the company in Emmen and many new people. Especially the finance department supported me through the whole process by providing a good working atmosphere. I would like to thank all colleagues at Honeywell Emmen for these pleasant six months.

I would also like to thank Kor Louissen. He offered me this opportunity within Honeywell. I would also like to thank him for the support during the project, and for the effort put in to get the recommendations into practice.

Next to the people at Honeywell, I would also like to thank Fons Wijnhoven and Rick Middel of the University of Twente for their support and advice. The meetings in Enschede and Honeywell Emmen provided me with valuable feedback and directions.

Last, but not least, I would like to thank my girlfriend, and my dad and brother for their advice and for bringing me to work every day.

Emmen, 15 September 2006.

Roy Boelens

Index

Chapter 1: Honeywell Emmen BV

This first chapter explains the basic characteristics of Honeywell. This is done in a top down approach. The general characteristics of the worldwide company are described first and then we will move down to Honeywell Emmen BV. By reading this chapter you will learn where the organization of Honeywell stands in the world and what kind of products Honeywell produces worldwide. The structure of this chapter is as follows. Section 1.1 explains the history of Honeywell and the coming about of Honeywell Emmen BV. Section 1.2 explains the company wide vision and mission of Honeywell. Section 1.3 describes the products and services Honeywell offers. Section 1.4 first describes the company wide organizational chart and then the position of Honeywell Emmen BV and their organizational chart. Section 1.5 explains something about the culture and people working within Honeywell. Section 1.6 discusses the Oracle information system in Honeywell Emmen.

1.1 History of Honeywell

The first Honeywell Company in the history can trace its roots back to 1885. An inventor named Albert Butz patented the furnace regulator and alarm. He formed the Butz Thermo-Electric Regulator Co., Minneapolis, on April 23, 1886, and a few weeks later invented a simple, yet ingenious device that he called the "damper flapper". This device was the first to control the temperature in a room by controlling the air available to a furnace and therefore the warmth produced by the furnace. It was not long before The Consolidated Temperature Controlling Co. incorporated acquired Butz's patents and business. By 1893 the company had renamed itself to Electric Heat Regulator Company. The first company ads ran in 1895 featuring the now famous thermostat. In 1898, the company was purchased by W. R. Sweatt, who, by 1916, had changed the name of the company to Minneapolis Heat Regulator Company (MHR) and expanded its product line. Four years later, MHR patented the first electric motor approved by Underwriters Laboratories.
Meanwhile, in 1904 a young engineer named Mark Honeywell, was perfecting the heat generator as part of his plumbing and heating business. Two years later, he formed the Honeywell Heating Specialty Co, incorporated, and specialized in hot water heat generators.
In 1927, Minneapolis Heat Regulator Company and Honeywell Heating Specialty Co. merged to the Minneapolis-Honeywell Regulator Co., and became the largest producer of quality jeweled clocks.
In 1942, the company invented the electronic autopilot (C-1), which proved to be critically important to the U.S. war effort.
In 1953, the company introduced the T-86 "Round" thermostats, which replaced chunky, rectangular models. One of the world's most recognizable designs, it remains in production today and adorns the walls of more households around the world than any other thermostat (Figure 1).
The company then made several acquisitions in several areas and the company's name was officially changed to Honeywell Inc. in 1963, even though it had been casually referred to as such for nearly 40 years. [I]*
In the end of the ‘90s Honeywell merged with the Allied Signal concern. This new Honeywell is now one of the biggest and leading companies in the world. Worldwide Honeywell has around 120.000 employees and a sales of $25.6 billion dollars in 2004 [II] Therefore it has been listed in the Fortune top 100 of “Most admired companies” and Honeywell is on of the 30 selected companies to represent the Dow Jones industrial average.

Figure 1: The famous thermostat from Honeywell [I]

* Numbers are used to refer to literature sources: [1,2,3], Roman numbers are used to refer to web sources: [I,II,III] and letters are use to refer to interviews: [A,B,C]

Honeywell in the Netherlands

In 1934 Honeywell opened the first office in the Netherlands (Amsterdam). This was the first not North-American office of the Honeywell Company. Honeywell was also one of the first American companies that founded a Dutch office. The Netherlands were famous for their infrastructure and were the central point of sea transport, which is why right from the start in the Netherlands the office was used to support the sales activities in all of Europe. When Honeywell took over Brown Instruments in the United States it was the beginning of the industrial division. After 1945 this division was intensely involved in the post-war rebuilding of the Netherlands.

When in 1963 large amounts of gas were discovered near Slochteren, Honeywell decided to open a factory in Emmen, the Honeywell Combustion Control Center. Up until today they produce gas controls and safety equipment for the international market.
In the period since then as a result of more mergers and takeovers several more offices opened in the Netherlands. In Den Bosch (Honeywell Safety Management Systems), Purmerend (Security House), Apeldoorn (Honeywell Building Solutions) and in Weert [II]. Today Honeywell has around 1300 employees in the Netherlands. Of which around 500 work in Honeywell Emmen [III].

1.2 Vision and mission

The Honeywell company has the following three corporate missions:

- Build a world that's safer and more secure,
- more comfortable and energy efficient,
- more innovative and productive.

Next to these missions Honeywell has identified five initiatives. They are the business goals which lead the company [2]:

- Growth

  Honeywell's first initiative is Growth. Honeywell tries to pursue organic growth through four fundamental strategies, or "pillars":
    - Doing a superb job for customers every day in quality, delivery, value, and technology
    - Superior Sales and Marketing
    - Globalization
    - Developing robust, funded technology roadmaps for new products and services, all supported with a strong commitment to Design for Six Sigma (DFSS)
- Productivity

  From without the history of the company productivity is one of the strengths of Honeywell. Main goals were to minimize material costs, indirect costs and labor. Nowadays the focus is more on lowering functional costs and real estate costs. This is because the costs in these areas are still considered as high.
- Cash

  Goals in this area are improvement of profit, turnaround, and of course shareholders value. There is also a focus on the working capital. This because working capital is a symptom of the effectiveness of operating practices.
- People

  People are what Honeywell differentiates from the rest. Therefore Honeywell tries to focus on consistently providing good feedback to employees, reinforcing the Twelve Behaviors (see next page), and rewarding achievements. Honeywell sees these factors as essential to building a company culture of sustainable performance.
- Enablers (DigitalWorks and Six Sigma)

  These enablers support all activities of the whole company.

To pursuit these five core initiatives Honeywell uses twelve behaviors that every individual within the company should have. These behaviors are also embodied in every project, process, and product of the company. Each behavior differentiates levels of performance. Therefore the behaviors are also used to measure every employee's performance. The behaviors are:

- Growth and Customer Focus
  Focus on growth and on improving the relationship with the customer
- Leadership Impact
  Think like a leader. Conceptualize, plan and execute.
- Gets Results
  Commitment to reaching business goals.
- Makes People Better
  Positively influence direct colleagues and other employees.
- Champions Change and Six Sigma
  Support the continues improvement mindset and think of long term goals.
- Fosters Teamwork and Diversity
  Work as a team and respect diversity.
- Global Mindset
  View the business from all relevant perspectives and think globally.
- Intelligent Risk Taking
  Risks are necessarily for making profit.
- Self-Aware/Learner
  Know you own strengths and weaknesses.
- Effective Communicator
  Listen to and be listened to, this does not always mean agreeing.
- Integrative Thinker
  Decide and take actions by applying common knowledge and intuition.
- Technical or Functional Excellence
  Be aware of advances in your field of excellence.

Both the five initiatives and the behaviors together form the vision of the company.
The vision and the mission of Honeywell ensure that Honeywell stays competitive and ready for the future.

1.3 Products and services

The Honeywell Corporation has divided its products and services in four main product categories:

- Aerospace
- Automation and Control Solutions
- Transportation Systems
- Specialty Materials

The sales in those categories are divided as follows:

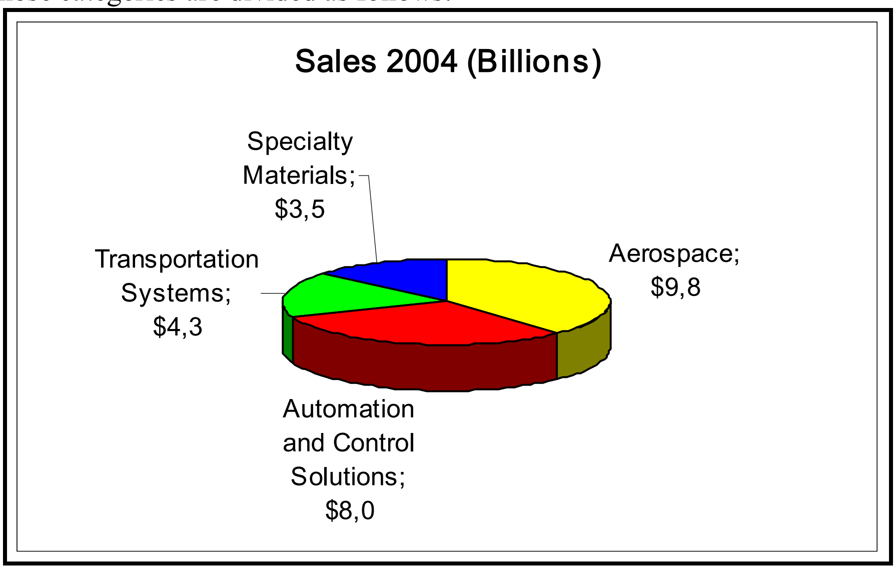


Figure 2: Honeywell Global Sales in 2004 [II]

The four product categories are shortly described below:

1.3.1 Aerospace

Aerospace is Honeywell's biggest sales market with a sales volume of about $9, 8 billion. It is seen as the world's premier supplier of aircraft engines and systems, avionics, and other products and services for airliners, regional and business aircrafts, military aircrafts, and even spacecrafts. Aerospace has around 40.000 employees worldwide. The headquarter of these operations is in Phoenix, Arizona. Main customers are Airbus, Boeing, General Electric, Lockheed Martin, U.S. Department of Defense, U.S. National Aeronautics and Space Administration (NASA), and leading airlines and airport authorities.
Main products in this category are:

- Aircraft engines
- Auxiliary power units
- Integrated avionics and aircraft control systems
- Aircraft safety systems
- Precision guidance and navigation
- Environmental control systems
- Electric generation and distribution systems
- Wheels and brakes
- Repair & overhaul
- Spare parts
- Support & services for space and communications facilities
- After-sales avionics support solutions
- Aircraft lighting
- Aircraft cabin entertainment and passenger productivity products/systems

1.3.2 Automation and Control Solutions (ACS)

ACS has a sales volume of about $8 billion. ACS is one of the global leaders in providing product and service solutions that improve efficiency and profitability, support regulatory compliance, and maintain safe, comfortable environments in homes, buildings and industry. Products and services in this category are used around the world in more than 100 million homes and buildings as well as in 24 of 25 top oil refineries. ACS has more then 40.000 employees worldwide in over 100 countries. The headquarter is situated in Minneapolis, Minnesota. The main customers are Alcoa, AstraZeneca, BASF, Boeing, Brinks, ChevronTexaco, DaimlerChrysler, DuPont, ExxonMobil, General Motors, PDVSA, ConocoPhillips, Procter & Gamble, Qatar General Petroleum, Sasol, Sinopec, Stora Enso, Sydney Airport, TotalFinaElf, Weyerhaeuser, and building and homeowners.
The main products in this category are:

- Controls for heating, cooling, ventilation, air conditioning, humidification, industrial process automation, video surveillance, people and asset tracking and access control equipment
- Security/fire alarm and industrial safety systems, home automation systems
- Advanced software applications for home/building control and industrial optimization
- Sensors, switches, control systems and instruments for measuring pressure, air flow, temperature, electrical current and more
- Home water products
- Combustion control solutions

Honeywell Emmen is a small part of this product category, because it produces gas valves and electronics to regulate these valves (controls for heating).

1.3.3 Transportation systems

The transportation systems product category has a sales volume of about $4, 3 billion. They are the world's leading innovator of automotive turbochargers. The transportation systems product category also offer some of the best-known consumer automotive product brands and manufacture world-class braking material for major auto manufacturers around the world. There are more then 14000 employees working in 18 countries all over the world. The headquarter of this product category is in Torrance, California. Main customers are: Advance Auto Parts, Auto Zone, BMW, Bosch, Canadian Tire, Continental Teves, CSK Auto, DaimlerChrysler, Ford, General Motors, Honda, Nissan, Renault, TRW, Volkswagen, Volvo, Wal-Mart. The main products they buy are:

- Garrett® turbochargers
- Prestone® antifreeze/coolant
- Autolite® platinum spark plugs
- FRAM® automotive filters
- Holts® car care products
- Bendix® and Jurid® brake pads

1.3.4 Specialty Materials

With a annual sales of about $3, 5 billion this is the smallest product category. But the products in this category belong to the highest-performance specialty materials in their market. They include fluorocarbons, specialty films and additives, advanced fibers and composites, customized research chemicals, and electronic materials and chemicals. There are about 8500 people working in 50 locations. The headquarter in this category is located in Morristown, New Jersey.
Customers in this area are Intel, Motorola, IBM, Samsung, Infineon Technologies, Carrier Corp., Haier Group, York International, Dow Chemicals, BASF, U.S. Military, Pfizer, GlaxoSmithKline, Mohawk, Shaw.
The products that are sold in this product category are:

- Environmentally-friendlier refrigerants and blowing agents
- Barrier films for pharmaceutical, beverage and food packaging
- High-strength fibers and composites for bullet-resistant vests, marine rope and cut protection
- Chemical and metals used to build semiconductor
- Packaging solutions for integrating computer chips
- Sapphire substrates for LED lighting
- Customized research chemicals and reagents for drug discovery
- High-performance lubricants
- Luminescent coatings and pigments
- Nylon for residential and commercial carpeting

1.4 Structure

To convert the vision and mission into daily operations, Honeywell has the matrix structure as presented in Appendix 1. On the horizontal axle, the company is divided in the four main product divisions: Aerospace, Automation and Control solutions, Transportations systems, and Specialty materials. This is done, to concentrate the knowledge of the products together and to serve the customer the best possible way in a high changing environment. The supporting departments follow on the other axle of the matrix. They also concentrate their knowledge and they support the other departments with their centralized knowledge. In this organization structure the managers of the product divisions and the functional divisions can use their specific knowledge in their own area of interest. One of the main disadvantages of this structure is that employees can experience dual authority. This intensifies the need for meetings and communication and it can slow down formal authorization processes [7].

In Appendix 1, the location of the Honeywell Emmen company is shown as a red square in the Environmental & Combustion Control product group. This is because Honeywell Emmen produces gas valves. The organizational structure of Honeywell Emmen is shown in Appendix 2.

Honeywell Emmen consists of six main functional departments [II]. The departments are: Finance & administration, R&D, Marketing, Sourcing, Human recourses and the Operations department. The Operations department is further divided into ten sub-departments. These sub-departments all have a direct connection to daily production. All the functional departments are supported by the supporting departments of the global Honeywell company.

My place is marked with the red square. That is within the Business Improvement & Business IT department. The manager is Kor Louissen. He is my contact and he assigned this project to me.

1.5 Culture

The fact that Honeywell is founded in America has his impact on the way the business is run in the Netherlands. The most important aspect of this American culture that can be found in Honeywell Emmen is the language. The official language for Honeywell Netherlands is English. So the language in formal reports and procedures is English. Also all the communications with other Honeywell companies is in English. This is necessary because in communications there is almost always a company in a foreign country involved. Internal and informal communications are sometimes also in English, because in the finance department an employee of a different nationality (Polish) is present. But in general the informal language is, of course, in the mother language; Dutch.
Honeywell shares are traded at the Dow Jones stock exchange. This means that the focus in the company is very result driven. Monthly reports have to be distributed to the shareholders and stock markets. This means that the culture within Honeywell is masculine.
The power distance between management and employees within Honeywell Emmen is, in contrary to American standards, small. Employees speak to their superiors just like they speak to other employees. Doors are open to everyone. So the management is very open to employees, where in America boundaries exist between management and employees.
Next to language, masculinity, and power distance, other important factors in determining the strength of the organizational culture are, according to Daft [7], the rites and ceremonies, stories, symbols. None of these factors have been found in the company. The only factor of importance in the strength of the culture in Honeywell Emmen is the symbol. Symbols are signs of expressing the companies core beliefs, accomplishments, mission and vision. In Honeywell Emmen there are symbols in almost every office. From Six Sigma mission posters to awards for selling more then 10 million products for example.
It can be stated that the company has a culture that has his weaknesses and his strengths. The American way of working is only partly visible in daily operations, mainly because of the difference in power distance between the American culture and the culture in Honeywell Emmen.

Another aspect that influences the daily working climate is the fact that Honeywell has undergone several reorganizations [B]. This influences the working climate in the fact that employees are more stressed. The same work has to be done by less people, so employees have to work more. This causes difficulties in planning meetings and also in planning other activities than the activities that they are responsible for.

1.6 The Oracle System at Honeywell Emmen

Oracle is the system that most Honeywell companies use to control their operations. Oracle is a product of the Oracle Corporation. Oracle uses the System/R model of IBM and the relational database uses the SQL language to process data queries. Oracle is (one of) the most used database systems around the world [VI].

Oracle is a so called relational database management system (RDBMS). The system makes sure that all the necessary data in the database is accessible and can be further processed by the applications that need the information.

The Oracle RDBMS consist of several modules. Each company can decide which of these modules it wishes to implement. These modules altogether can support the whole business process. The Oracle RDBMS has two main modules which support most of the business, Oracle Manufacturing and Oracle Financials. These modules both communicate with each other and use the database as a central storage place.

Oracle Manufacturing can describe the complete production system, from purchasing to the sale of products. The modules that can be implemented are:

- Material Requisition Planning
- Inventory Accounting
- Work In Process
- Bill of Material
- Costing
- Purchasing
- Human Resources
- Order Management

Oracle Financials handles the financial data flow from the manufacturing module. The main module General Ledger is able to make financial reports and keeps track of all accounts. Modules in this system are:

- General Ledger
- Accounts Receivable
- Accounts Payable
- Fixed Assets

The decision to use the Oracle database system was taken in 1992 by Honeywell Inc. in Minneapolis. The plan was to integrate as much data of all Honeywell corporations over the world as possible into one database, so that the headquarter could always have all data available. The database warehouse is located in the Application Center in Newhouse.

In Honeywell Emmen the system was implemented in 3 phases. The first phase embodied the implementation of the kernel of the Oracle pack, the Oracle Manufacturing system. The modules that were implemented are: Material Requisition Planning, Costing, Inventory Accounting, Work in Progress en Bill of Material. These are seen as the kernel of the system. In the second phase the modules Purchasing and Accounts Payable are implemented. In the final phase the Order Management module was implemented.
The database in Honeywell Emmen is part of a larger database. The structure is presented in figure 3. The highest level in which data for Honeywell Emmen is stored is the NL1 level. In this level the data of all the products and materials of all the other Honeywell companies mentioned in figure 3 is stored. This general database level is also connected to other Honeywell Oracle databases in other parts of the world.
The NLV level is implemented to validate the orders the different affiliates make with Honeywell. It checks if certain product-attributes are correctly filled in and stops orders if differences between the data exist.
The lowest level is the organization level. For Honeywell Emmen this is level called 567. Products that are sold in Emmen are stored in this part of the database.

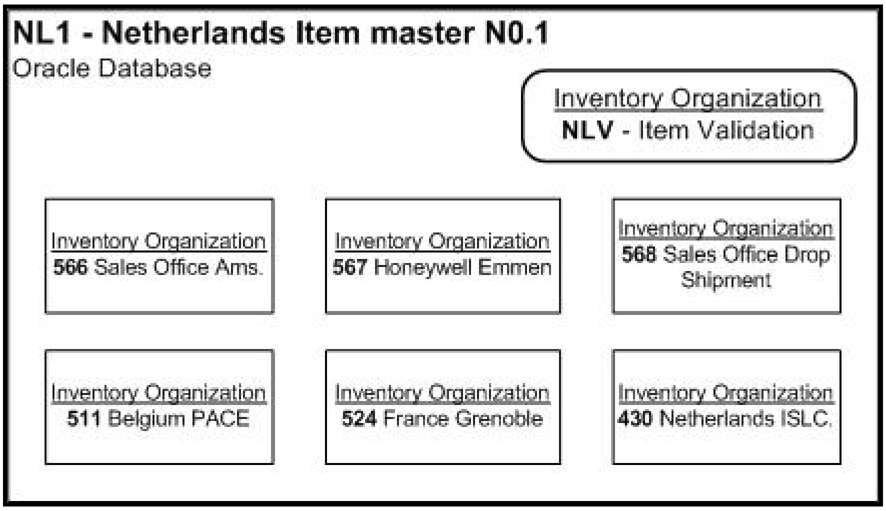


Figure 3: Oracle Database structure in the Netherlands [13]

This first chapter gave an overview of the Honeywell company and its operations. The next chapter states the problems Honeywell is facing and presents the method that is used to solve these problems.

Chapter 2: Problem definition and approach

This chapter explains the problems Honeywell Emmen faces and explains the approach that is used to solve these problems. In section 2.1 the problem is first identified and the stakeholders involved are analyzed. In section 2.2 the problem is further analyzed to find the root causes of the problem. In section 2.3 the problem definition is given together with research questions that will be answered. In the final section the causes are analyzed to come to an approach that will solve the problem. The project planning is also presented here. To identify the main problem and its causes, several interviews have been conducted [B, C, D, and E]. Based on these interviews the research method and planning were made.

The use of the word 'quality' in this report means quality on all relevant quality dimensions for data customers.
The term 'data' is used to refer to both data and information. Data usually refers to information at its early stages of processing, and information refers to the product at a later stage. But unless specified otherwise, this report uses 'data' interchangeably with 'information'.

2.1 Problem identification

The Oracle database in Emmen (level 567) contains data of about 22.000 products. These products are:

- Products that Honeywell Emmen produces and sells (different types of gas valves and electric components, see subsection 1.3.2),
- Raw materials (for example aluminum),
- Parts (for example processed aluminum), and
- Finished goods (For example gas valves from other Honeywell companies that have to be repacked and sold to Dutch customers)

In Oracle, Honeywell uses about 150 product-attributes to represent the information of these products. The number of these product-attributes is large because Oracle can be used in many different types of organizations ranging from producers of defense systems for the Ministry of Defense to mass producers of electronic components (as experienced in the Oracle Gebruikersgroep Oost Nederland (OGON) platform [A]). These businesses all have to use different product-attributes. Another reason why the number of these product-attributes is big, is that the Oracle system bases its calculations on these product-attributes. This means that calculations in material planning, financial data, purchasing, and the other systems are dependant on the product-attributes. The product-attributes are therefore very important.
The term product-attributes refers to three sorts of product data in the database (for simplicity reasons this report only uses the term product-attributes):

- Product parameter: standard in Oracle, all products have this type of product-attributes. These parameters are divided into groups: For example the parameter *Minimal order quantity*[*] belongs to the General planning group.
- Product categories: A category can be assigned to a product to classify products. Categories can be assigned by all departments responsible for filling in product-attributes.
- Product Descriptive Flexfield: A field that can be extended to capture extra information not otherwise tracked by Oracle. This field can be customized to capture additional information unique for the departments and for Honeywell.

For example a type of gas valve has a product parameter attribute *Item number*: VK8115F1001B (which is in the Inventory group), a product category attribute *PO CommCde*: 567L.EMMN, and a descriptive flexfield attribute *Duty Tariff Code*: 84818059.

[*] For clarity, the product-attributes used in this report are printed in *Italic*.

During a product lifecycle the product undergoes changes. For example materials can change, but also suppliers or *Country of Origin* can change. In Honeywell the changes are divided into two main categories depending on there impact:

- Changes that impact the customer in some way.
  These changes are prepared by the engineering department and recorded by the Product Data Management department. These changes can for example be based on customer whishes, new laws or other influences. For example: newer versions are developed, cheaper, stronger, or more environmental materials are used, and packaging of the product is changed. All these changes are recorded in so called 'Engineering change orders' (ECO's). The Product Data Management department makes up this ECO. The link to this ECO is then send out to all involved departments and in this e-mail the departments are requested to fill out or update the product product-attributes they are responsible for, if needed.
- Changes that do not have impacts for the customer
  These changes are made by the responsible departments themselves. For example if a supplier is changed by the purchasing department, or if the cost price is changed by the finance department, or external changes like changes in taxation, delivery times of several suppliers and duty tariff codes. These changes have no effect on the physical state of the product and therefore do not require an ECO.

All these product changes have to be recorded in the product-attributes so that The Oracle System can use them for: For example, financial calculations, material requirements planning, for printing shipment labels and invoices, and for other activities. The next subsection identifies the involved stakeholders in this process.

2.1.1 Stakeholders

There are several stakeholders involved in inputting, changing and using the product-attributes. First there are the internal stakeholders, represented by different departments within Honeywell. These departments are involved in filling in or changing the product-attributes. Each of the departments is responsible for filling in their own group of product-attributes. So when a new product is produced or bought, or the product-attributes of an existing product are changed during its lifecycle, the departments are asked to input or update the product-attributes they are responsible for.
The internal stakeholders are:

- Product Data management department
  This department defines new items and sets basic product-attributes as the product number and description. Data management also sends out engineering orders to other departments if changes have to be made in the product-attributes.
- Purchasing department
  This department fills in the product-attributes that are related to purchasing, for example country of origin of the bought products, the default buyer and the Commodity code.
- Customer Logistics department
  This department fills in the product-attributes related to transporting products to customers. For example information for customs and duty tariffs.
- Availability Management (Planning) department
  This department fills in most of the product-attributes. For example the safety stock amount and the minimum order quantity.
- Finance department
  This department fills in the product-attributes related to the financial control. These are for example standard cost price, the bill of material, transfer price, and the sales account number.
- Quality department
  The quality department fills in product-attributes related to the quality of a product. They are the product-attributes: receipt required, inspection required and receipt routing.

The stakeholders that use the product-attributes and require the product-attributes to be of high quality are the internal stakeholders described above together with external stakeholders.

These external stakeholders are:

- Customers
  They receive the products with their shipment data, and they receive invoices with data that comes from Oracle. There are also customers (other Honeywell companies mainly) that have direct access to Oracle to make it easier for them to order products and see product details.
- Statistical Bureaus
  They require data from Honeywell to use for export figures and other statistical data. This requires product-attributes as the country of origin and the duty tariff code to be correct.
- Suppliers
  Suppliers receive orders with data that comes from Oracle. There are also suppliers that have direct access to Oracle to simplify the order process.
- Public Offices
  For example customs offices, they require documents with shipments that prove origin to determine the tax percentages for products that cross the borders.

To clarify the whole process of filling in and changing data in Oracle, figure 4 has been drawn. The process starts with a need to changes an existing product or to input a new product in the database. This need can be initiated by a internal stakeholder or an external stakeholder. The process ends with certain product-attributes of a product being filled out or updated.

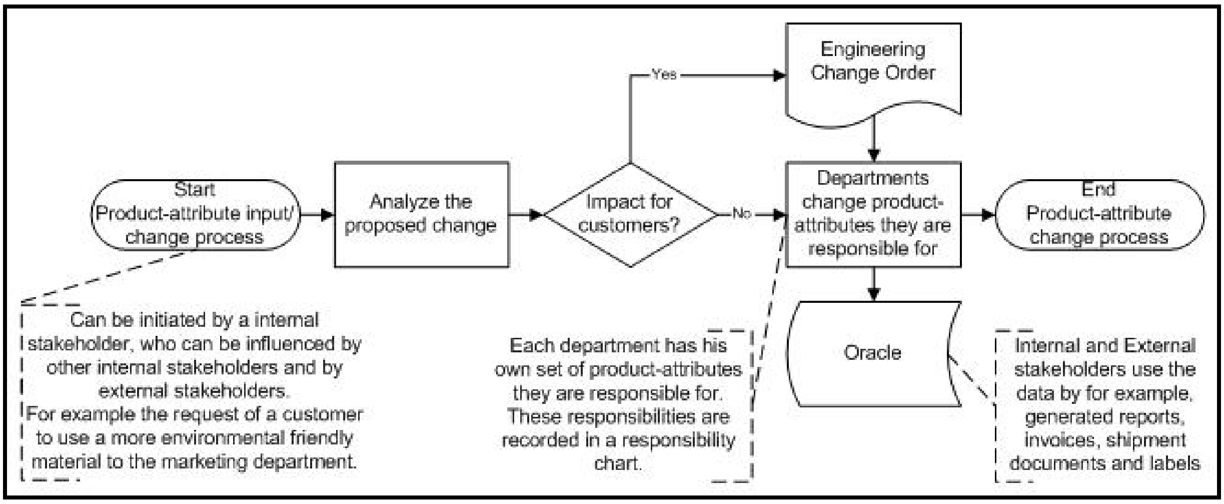


Figure 4: Product-attribute input and change process

The next subsection explains the identification of the main problem that these stakeholders have.

2.1.2 Problem Summary

The initiator of the project, the Business IT & Business Improvement (BIT&BI) manager, explained that he noticed that stakeholders, who use the product-attributes, have to deal with more and more problems [B]. For example shipments were delayed because the use of missing *country of origin* and *duty tariff codes*. Also customers reported errors in scanning *EAN bar codes* that Honeywell printed. Another problem that occurred was that there were differences in *transfer prices* between Honeywell Emmen and Honeywell Brno.

The departments involved first tried to solve these problems ad hoc by changing product-attributes when they were needed or when the problems occurred, and managed to do so for several months. But they came to the conclusion that more actions were required, because the problem concerned not only multiple departments, it also influenced contacts with important customers.

The BIT&BI manager identified these problems and took responsibility by taking the action to start a structured analysis of the problems with the quality of the product-attributes.

Another reason for addressing this problem now is the implementation of SAP, planned in 2008. It is of high importance to start this new system with a database that consists of high quality data. If the start of the design process and the implementation process is based on a database full of low quality data, the developed and implemented new processes could contain mistakes [21].

There is also more pressure from within the management of the Environmental Combustion and Control (ECC) sub-category of the Automation and Control Solutions product category. The management of ECC also recognizes the risks involved in migrating to SAP with existing problems in data quality. They are setting up a quality division as we speak. This team should check the quality in the databases and notify companies when problems occur from an overview position.

From this subsection the problem scheme in figure 5 has been identified:

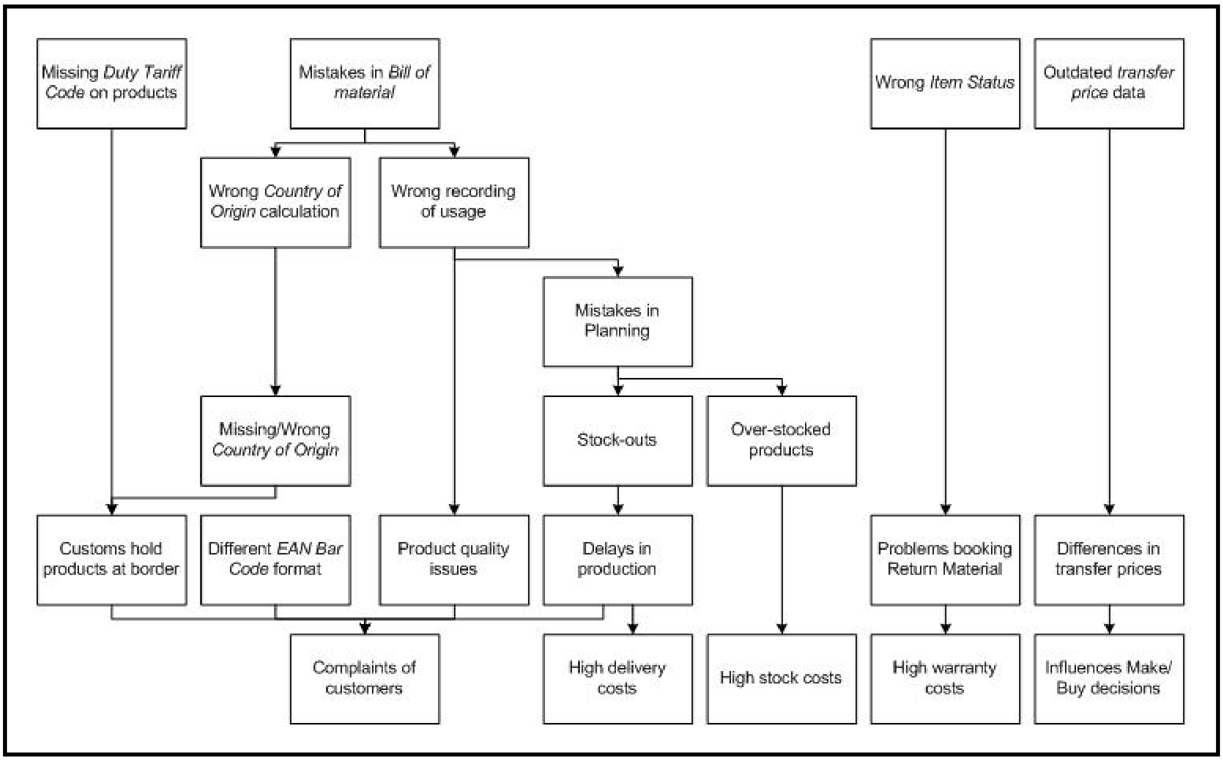


Figure 5: Problem scheme

From this figure can be stated that there are several impacts that effect the company. Four of the identified problems directly lead to complaints of customers and therefore a lowered customers satisfaction. Also the cost are higher, because of problems in stock, delivery, and warranty. There are also reports of wrong transfer prices which can lead to wrong decisions in make or buy questions. According to Redman [17] these impacts are typical impacts of low data quality within companies. As shown in the problem scheme, these problems started in several product-attributes: *Duty Tariff Code*, *Country of Origin* and *Item Status* for example. The BI&BIT manager also has doubts about the quality of the other product-attributes. Redman also recognizes, next to the experienced impacts, several other impacts. The impacts could cause problems in the future, if their underlying product-attributes are also of low quality.

It can therefore be stated that the main problem is the quality of the product-attributes in the database. In some product-attributes this quality is to low and in other product-attributes the quality is expected to be low. To get a better understanding of the underlying aspects of this problem, the next section identifies and analyzes the causes of the problem.

2.2 Causes of the problem

As can be stated from the previous chapter the main problem that has to be dealt with is the quality of data in the Oracle System. Interviews with a supplier of data, a user of data, and the BIT&IT manager [B,C,D] showed that there were multiple causes for the low quality of the data of the product-attributes. The BIT&IT manager helped to order and analyze these causes. Appendix 3 shows the results of this evaluation. In the Appendix three main causes can be identified (marked as squares with bold lines):

1. No structured way to check and improve the quality of the product-attributes
2. No structured way to evaluate responsibilities for inputting and updating the product-attributes
3. No structured communication between involved stakeholders about problems with the product-attributes

To explain these main causes, the paths from the causes back to the main problem is explained next.

2.2.1 No structured way to check and improve the quality of the product-attributes

The first main cause of the problem is the lack of a structured way to check and improve the quality of the product-attributes. A consequence of this is, that there is no check to see if all data has been entered, after and sending ECO. This can lead to missing data in Oracle. Because there is no structured way to check the quality, there might also be no, or a wrong, understanding about, what important requirements are for data customers. For example in the *EAN bar code* problem there was a misunderstanding about what type of code should be delivered. This can then lead to wrong ideas about the quality of data and therefore the quality can be lower that expected.
When a method is found or developed that is able to measure the quality of the product-attributes, that accounts for requirements of the data customers, and that has guidelines how to analyze and improve this quality, this cause of low data quality can be addressed and possibly eliminated.

2.2.2 No structured way to evaluate responsibilities

The second main cause that was identified is that there is no structured way to evaluate the responsibilities the employees have in imputing and changing data. Because of this, there is a possibility that the responsibilities are divided wrongfully. This can lead to employees not knowing they are responsible for maintaining a certain product-attribute. It can also lead to responsibilities given to employees that do not have the knowledge to complete the task, or to employees that do not have the time to complete the task. Another main cause of these factors is the recent changes in employees due to reorganizations [VIII]. These changes could lead to shifts of tasks to other or new employees. These employees have to make time and acquire the knowledge to be able to complete the tasks they are given.
The consequences of not having enough time and knowledge, not having clear who should do the task, and not knowing what the consequences might be of not doing the job, can all lead to mistakes in inputting or changing the data for the product-attributes or not updating the product-attributes. This then leads to a lower quality of data.
So when a structured way to evaluate the division of the responsibilities, and also to addresses problems with the division of responsibilities is found or developed, the above causes could be prevented and eventually the quality of the product-attributes should improve. Limiting employee changes can also help prevent these causes, but Honeywell Emmen has only little influence on global strategies. This global strategy is determined by the top management of Honeywell and Honeywell Emmen is only a small player in this management (see Appendix 1 and chapter 1).

2.2.3 No structured communication between involved stakeholders

The last cause was identified in two interviews [C,D] and also by the BIT&BI manager [B]. There is not enough communication between involved departments about problems with data quality. This can have consequences in a lot of areas. Not enough communication can lead to lower quality in the supply of data, because wrong data can be supplied without knowing. This can directly lead to mistakes in the input of the data, because the data itself is of low quality.
Another consequence is that the required format for the data is not clear. Employees do sometimes not know the format in which data has to be entered, for example in the *EAN bar code* problem. Sales affiliates could not scan codes that Honeywell Emmen used, because Honeywell Emmen used a different coding format for some products. Other causes of the format not being clear could be that the Oracle Input screens are not clear enough and that there are no restrictions to the entry of data formats

that are not wanted. All these causes are likely to, again, lead to mistakes in input and change of data, which causes lowered data quality.
Lack of communication can also lead to responsibilities not being clear to employees. This can in turn lead to employees not knowing that they are responsible for a certain task. This could then lead to not updating the data when required or not imputing data that is required. Again this leads to loss of data quality.
Lack of understanding the process in which the data for a product-attribute is created is also a consequence of not enough communication. Not knowing the process can lead to a lack of insight in possible consequences of not taking required actions. This then can lead to not updating required product-attributes and loss of quality.
So when communication is improved between involved employees, and with data suppliers and external users, a lot of these identified causes can be eliminated. This improved communication between stakeholders should eventually lead to problems being solved and increased quality of the product-attributes.

2.2.4 Ensure high data quality in the product-attributes

The goal of this report is therefore to ensures high quality of the product-attributes in the eyes of the data users. From the analysis in this chapter this can be achieved by eliminating the three main causes of the problem with data quality.
A structured approach to handle all the causes of the problem is therefore necessary to be able to meet the quality the users of the product-attributes require. So this approach should be able to handle the following aspects:

- The quality level of the product-attributes in the database can be analyzed and, if needed, can be improved to match the needs of the users.
- The responsibilities for inputting and maintaining the product-attributes can be analyzed and, if needed, improved.
- If there are problems with the quality of the product-attributes they can be easily communicated between involved stakeholders.

In the next section the problem definition is given together with research questions that need to be answered to be able to eliminate the above discussed causes.

2.3 Problem definition and research questions

From the identification of the problem in section 2.1 and from the cause analysis in section 2.2 the following problem definition can be stated.

> *What structured method should Honeywell Emmen implement to assure the quality of the product-attributes is increased to the level the users of the product-attributes require?*

To be able to answer this question, the following research questions need to be answered. These research questions should eliminate the causes of this problem discussed in the previous section:

1 What should be the definition of product-attribute data quality for Honeywell?
2 What methods does the literature provide, to be able to increase the quality of the product-attributes?
3 What method, or adapted method, can Honeywell Emmen use to increase the quality level of the product-attributes to meet the desired level of the users?
4 How should Honeywell Emmen assure the method does improve the level of quality in the product-attributes?

The next section explains how the problem definition and the research questions are answered and what the planning will be.

2.4 Research method and planning

In this section the problem definition and the research questions are examined and the steps that will have to be taken to come closer to the ideal situation, and to reach the goals of the report, are explained. The planning, that follows these steps, can be found in Appendix 4.

The problem definition described above obviously asks for a method that can assure the quality of the product-attributes. This method should be able to answer all the research questions. To find or design a method, the first step that can be taken is to perform a literature research. To search for a suitable method and to try and implement it. The problem however is usability of most of these theories in academic management research [1]. Van Aken therefore suggests to complement this so called "Organizational Theory" with a more prescription-driven research approach. He calls this "Management Theory". This "Management Theory" will have to be field-tested and based on grounded technological rules. To assure that the method found or designed in this report is also usable in practice, the method will be tested in practice.

In chapter 3 literature research therefore first tries to find a method in literature. As explained, this method should be able to answer the research questions described above. This method can then be adapted or used as a basis to design a more specific method from, for the situation in Honeywell (chapter 4). This chapter is based on interviews and own practical experiences within Honeywell. A fifth chapter then tests the method in practice. These tests follow the reflective cycle as described by Van Aken [1]. This means that a first case to use the method on is chosen, the steps of the method are planned, and the method will be used. After this first cycle of the method, the method is evaluated using a survey (see Appendix 5) and interviews (see interview section). From this evaluation new design knowledge is developed and the method will be improved. This cycle is repeated on a different case, to further improve the method. This method of testing can be compared to the concepts of software development, in which the technology is first tested by the originator of the software, the so called α-testing. And secondly by a third party, the so called β-testing [1]. However this β-testing is not handled in this report, because of limited financial resources and limited time.

In the last chapter the findings of this report are summarized and possible directions for further research are also given. The following table summarizes the research method, and the planning is shown in Appendix 4.

The research method used in this report		
"Organizational Theory"	Find a method in academic literature and develop it further. **Research Question 1 and 2**	Chapter 3
"Management Theory"	Adapt the method to the situation within Honeywell **Research Question 3**	Chapter 4
Field testing	Test the method within Honeywell and develop new design knowledge to improve the method **Research Question 4.**	Chapter 5
Concluding chapter	Reflect on the designed method and the used research approach	Chapter 6

Table 1: The research method, based on Van Aken [1].

Chapter 3: Theoretical foundations

This chapter searches for a structured method in the literature that is able to deal with the problem of data quality analyzed in the previous chapter. As explained in the previous section, additions are made to the method if the method is not able to answer the research questions stated in section 2.3. The concluding method is the basis ("Organizational Theory" [1]) on which the method for Honeywell Emmen ("Management Theory" [1]) is based in chapter 4.
The first section therefore searches for a method in literature. This method is then shortly introduced and in the next sections the method is critically analyzed. Each phase of the method is handled separately to be able to asses if each aspect of the problem described in chapter two is handled in the best possible way. In section 3.8 the important aspects from each phase are summarized.

3.1 Sources

As stated in the previous chapter, the problem in this report concerns the quality of an object, in this case quality of the product-attributes. The research area in which answers may be found is therefore, quality management. From the research questions and section 2.2 however can be concluded that the quality problem also has causes in the areas. For example, organizational theory (division of responsibilities) and communications theory. Therefore, a broad methodology is needed to reach the goal of eliminating the causes of low quality. In quality management, the Total Quality Management (TQM) framework is the broadest approach to handle quality problems. This approach has evolved from quality inspection techniques of Shewart in 1931 to an integral methodology to handle quality related issues in 1990, the TQM framework [5].

The three concepts that characterize the TQM methodology are: Customer satisfaction, continues improvement, and teamwork and participation. These concepts are connected to the causes of the problem described in chapter two:

- Low quality of data leads to frustration and lowered satisfaction for customers according to Redman [17]. According to the study of Kovac, Lee, and Pipino [14] the basic requirement in the customer hierarchy of needs is the total data quality (no errors in the data) and the delivery reliability of the data (always get the requested data). Improving the quality of data is therefore one important aspect that leads to higher customer satisfaction.
- To remain at a high level of quality the processes must reviewed frequently (at least once a year). This is because customer requirements may change, custom regulations may change, and also because employees may change. Changes in employees can cause responsibilities to be wrongfully divided.
- Teamwork and participation are connected to the cause of not having a structured way of communicating between stakeholders. Teamwork is a way to improve communication between involved stakeholders departments and can also improve participation in designing better processes.

An important pioneer in the TQM area is W. Edwards Deming. He became famous because of his change cycle, the Deming-cycle. This cycle consists of four steps:

- Plan
- Do
- Check
- Act

The TQM framework is based on this cycle [5]. A lot of quality authors have contributed to the TQM framework by adapting the methodology to implement the Deming cycle in specific situations.

The Total Data Quality Management (TDQM) methodology is such a methodology. This methodology is at the moment the only approach which focuses on data quality related issues and provides a method and tools to measure and improve the data quality [19]. Richard Wang is its founding father and his research has been quoted by many other authors in the field of data quality [14, 17, 19, 21, and 24]. Therefore the methodology is used as a guideline for this project. The methodology and its tools are critically evaluated to see if they are able to give answers to our research questions and to see if they are able to eliminate the causes of our problem.

3.2 Total Data Quality Management introduction

The basis of the TDQM methodology is the same as the basis of the TQM framework. The same underlying concepts that characterize the TQM methodology (see previous section) are also the underlying concepts that characterize the TDQM methodology. Within TDQM these concepts can be found in, the central role the data users (customer satisfaction) play in the methodology, and in the first basic steps that have to be taken by an organization in using the TDQM framework. These four steps emphasize the importance of teamwork and participation and they emphasize the importance to institutionalize the methodology to assure continues improvement. The four steps are [22]:

1 Establish a data quality team that will use the TDQM methodology
Forming a data quality team is an important factor for the success of the methodology. A data quality team is, if the team consists of the right people with the right sources, able to greatly improve the communication. (Because the forming of a team is crucial for the success of the method, the next section will first go into more detail on how this team should be formed.)
2 Teach data quality assessment and data quality management skills to the data quality team
To be able to implement the methodology, guidance is needed. Therefore it is necessary to test the methodology in practice with team members. Testing the methodology teaches members how it works and also gives valuable information about practical problems and possible shortcomings of the methodology.
3 Clearly articulate the data product in business terms
The general data product was described in section 2.1. When a specific product-attribute is being improved with the TDQM methodology, the first phase (define) will describe the data product more clearly. This improves the knowledge about the product-attribute and form a solid basis for further analysis.
4 Institutionalize continues data product improvement
Once the report is finished and it has proven to be valuable to the improvement of data quality, the organization should institutionalize the designed method and the data quality team to assure continues data improvement.

Once these four basic steps have been taken, the organization can use the TDQM methodology to improve the data quality. To improve the data quality the TDQM methodology uses an adapted version of the Deming quality improvement cycle as explained in the previous section. In stead of the Plan, Do, Check, Act definitions, the TDQM uses the definitions: Define, Measure, Analyze and Improve as steps of the cycle (see table 2). These definitions more clearly represent the phases that are necessary in the process of improving data quality.
The next part of this chapter examines the TDQM methodology in more detail. Section 3.3 first describes the data quality team that has to be formed to be able to execute the methodology . The next sections each cover a phase of the TDQM methodology (see Table 2). All these sections identify the theory the TDQM methodology of Wang describes in [22] and critically review this theory, keeping in mind the research questions that have to be answered and the main causes that have to be handled.

Deming Cycle (Deming '50)	TDQM Cycle (Wang '80)	
Plan	Define	See 3.4
Do	Measure	See 3.5
Check	Analyze	See 3.6
Act	Improve	See 3.7

Table 2: The Deming cycle and the TDQM Cycle [22]

3.3 Data Quality Team

As explained in the previous section, the establishing of a data quality team is the first basic step to make to use the TDQM framework. Selecting the right people with the right resources is an important factor as described in the previous section. According to the theory of Wang [22] the team should consist of:

- A TDQM champion preferably a senior executive,
- A data quality engineer who is familiar with the TDQM methodology, and
- Other members who are data suppliers, manufacturers, users (consumers), or data product managers.

Planning meetings with this team and using the method described in the next sections can then provide a structured way of communication between the involved stakeholders about the problems with the product-attributes. The theory of Wang does however not elaborate on the sort of team that should be set up. Because teamwork (and participation) is an important concept of TQM, and it can help in eliminating the cause of no structured way of communication, this subject should be further analyzed. In literature mainly four sorts of teams are mentioned [5, pp. 281-291].

- Steering group
  Steering groups are responsible for the setting up and the implementation of quality policies in organizations. Steering groups try to translate the quality policy of the company into a quality program.
- Autonomous team
  Autonomous teams generally contain between six to eighteen employees. The team is responsible for a specific task. The term 'empowerment', which stands for giving responsibility and power to employees, is often named when talking about autonomous teams. Autonomous teams can be given several gradations of empowerment, ranging from only giving the team possibilities to propose improvements to giving the team power to control their tasks themselves.
- Quality circles
  The term 'quality circle' originates from Japan. In the early 1960s, in companies in Japan, small groups of employees with the same work area came together on voluntarily basis to discus quality related problems and to try and improve quality. The teams were trained in using quality improvement tools. They identify, analyze and present solutions to the management. The management can then decide what to do. The implementation of the solutions will be done by the teams themselves, possibly with help from the management. Unique features of these teams are [26]:
    - Circles are small, between 4 to 15 employees
    - Circle members come from the same shop or work area
    - Circle teams meet one a week in company time, in a special training room
    - Circle members work under the same supervisor, he or she is also the leader of the team
    - Circle members are trained in the rules of quality circles
    - Circle members can decide themselves which projects to address
    - Circle teams present the proposals to the management, who then can decide which action to take
- Quality improvement teams
  The quality improvement teams are a development from the quality circles. Quality improvement teams systematically analyze quality related problems and try to find causes and solutions. The team only consists of four to eight people to improve the effective functioning. If the team is formed to improve processes the employees usually come from different departments to account for as much different views as possible. In some cases the team members can change depending on the problem that the team handles at that moment.

The choice of the sort of team and its members depends on the wishes of the organization, the goals of the team, and the sorts of problems the team had to deal with. Wang his idea of a team is closest to the quality improvement team described above. With the difference that Wang does not describe the possibility to change team members when handling other problems. This is because Wang designs the team to handle one specific data product.

Wang also does not mention anything about the autonomy of the team. A quality improvement team has the power to choose their own projects and to also implement them (based on the features of its predecessor, the quality circles). Because of the presence of a senior executive in the members list proposed by Wang in the beginning of the sections, and because of the presence of an improvement phase, one can conclude that Wang most likely sees the data quality team as a team not only to advise the management, but also to implement the solutions. This is in line with the description of the quality improvement team.

3.4 The TDQM Define phase

For the first phase of the TDQM methodology, three steps are important according to Wang [22]:

- determining the data products characteristics (3.4.3)
- determining the requirements for the data products (3.4.4)
- determining the process that produces the data product; the data manufacturing process (3.4.5)

The problem within Honeywell is not located in one data product, but in the product-attributes. This difference is first explained in subsection 3.4.1, and the extra step that is needed because of this difference (subsection 3.4.2).

3.4.1 Data Products and product-attributes

The first step in the define phase of the TDQM methodology is described above as defining the characteristics of the data product[*]. So the TDQM methodology handles one data product at a time. However, the problem in this report concerns the quality of the product-attributes, not just one data product. From the problem identification in chapter two can be concluded that there are many different product-attributes that the method should be able to handle. These product-attributes are all used by different users for different purposes. For example a product label is a data product which can contain product-attributes like: *Item number*, *product description*, and *EAN Code*. Another data product is a report for a statistical agency which can contain product-attributes like: *Item number*, *Country of Origin*, and *Duty tariff code*. The basis of the data products is a selection of product-attributes (see figure 6).
The data quality problems described in this report are however not connected to one specific data product. They are connected to several product-attributes. To therefore go trough an improvement cycle of the TDQM methodology in this report, the method should be able to handle one product-attribute at a time. So a cycle of the TDQM method in this report does not handle a data product, it handles a product-attribute that can be part of several data products. The TDQM method of Wang does not account for this difference. So therefore this difference (data product in the TDQM theory and the product-attribute in the to be designed method) will have to be taken into account to be able to answer the research questions from chapter two.

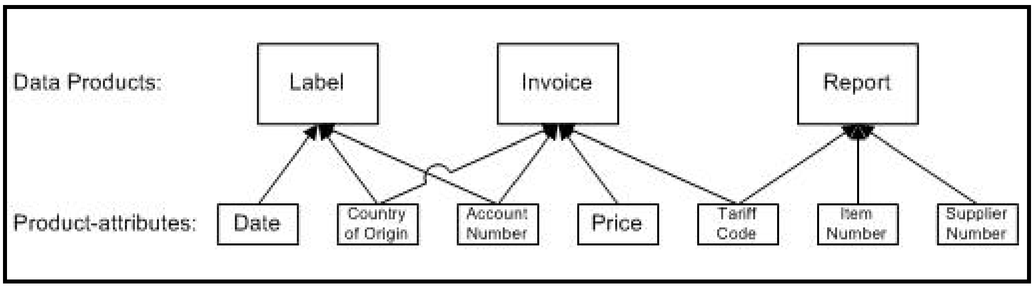


Figure 6: Data Products and Product-attributes

The first consequence of this difference leads to an introduction of a new step. The selection of a product-attribute to handle. Wang does not elaborate on this possibility in his method. Therefore

[*] To clarify this difference the terms data product and product-attributes will be underlined in this subsection.

subsection 3.4.1 discusses this subject. The next subsections cover the original three steps of the first phase of the TDQM methodology. Subsection 3.4.2 describes how Wang handles the characteristics of his data product and how the product-attributes described above fit in. Next, subsection 3.4.3 describes how the requirements of the data products, or product-attributes, can be determined. And the last step, the description of the data manufacturing process, is described in the last subsection, 3.4.4.

3.4.2 Product-attribute selection

The first consequence of the difference between data products, which Wang uses to go trough a TDQM cycle, and product-attributes which this report uses to go through a cycle, is that, in the method that is designed, a step is needed in which the product-attribute that is handled by the method is selected. The TDQM methodology of Wang does not elaborate on the possibility that one product-attribute will have to be chosen out of a large number of product-attributes. The selection of a product-attribute will also influences who the involved employees are. Therefore this selection should be done before the team members can be selected. This means that the team should consist of permanent members, who can then choose non-permanent members depending on the selected product-attribute. From the team members described in the beginning of section 3.3, the team champion and the team engineer would be the permanent members. The other members would then be chosen depending on the product-attribute selected.

To find which product-attributes are important to be handled first the TDQM methodology does not provide guidance. The TDQM methodology does however stress the importance to match the requirements to data quality of the data users with the actual data quality level. Customer satisfaction is also one of the three main concepts of the TQM framework. From this can be concluded that the most important product-attributes are those product-attributes that are important to the users of the information. The product-attributes that have recently caused problems to data customers are therefore important, because problems with data for external data customers can lead to lowered customer satisfaction [17]. A consequence of this assumption is that the permanent members of the data quality team should be employees that are well aware of problems that occur with product-attributes.

To find out which product-attributes could cause problems in the future, the article of Redman can be used [17]. This article describes the typical impacts of low quality data on enterprises. With this knowledge, the permanent data quality team members can better decide on which attributes to select first to evaluate with the developed method.

Redman [17] analyzed data from several cases and found a wide array of quality problems. He categorizes these issues as follows:

- Issues associated with data ‘views’, such as relevancy, granularity, and level of detail.
- Issues associated with data values, such as accuracy, consistency, currency, and completeness.
- Issues associated with the presentation of data, such as the appropriateness of the format, ease of interpretation, and so forth.
- Other issues such as privacy, security and ownership

Few enterprises measure all of these aspects. Most case studies use only the accuracy of data as the quality dimension. These studies have reported error rates of between 0.5 and 30% [17]. Comparison between figures from different cases is difficult, but Redman expects data error rates that are measured to be between 1-5% in a typical enterprise (Error rate = number of erred fields/number of total fields). Data that is not measured can also contain errors, so management should also expect that problems exists in there. Another issue, according to the article of Redman, is that in lots of databases inconsistencies exists between databases or levels of data in databases.

The quality problems can have several impacts on organizations according to Redman [17]. To each of these impacts, organizations can connect product-attributes that could be able to cause these impacts. By making sure these product-attributes are of the quality the users of these product-attributes require, the organizations can make sure future problems are prevented. Possible impacts of the problems that

Redman recognizes are divided into the three organizational levels: operational level impacts, tactical level impacts, and strategic level impacts. They are summarized in table 3.

Typical Impacts:	
Operational Impacts:	• Lowered customer satisfaction
	• Increased cost: 8-12% of revenue in the few, carefully studied cases. For service organizations, 40-60% of expense.
	• Lowered employee satisfaction
Tactical Impacts:	• Poorer decision making: Poorer decisions that take longer to make
	• More difficult to implement data warehouses
	• More difficult to reengineer
	• Increased organizational mistrust
Strategic Impacts:	• More difficult to set strategy
	• More difficult to execute strategy
	• Contribute to issues of data ownership
	• Compromise ability to align organizations
	• Divert management attention

Table 3: The impacts of poor data quality on the typical enterprise [17]

Operational Impacts:
According to Redman poor data quality at the operational level leads directly to external data customer dissatisfaction, increased costs and lowered employee job satisfaction.
The customers of an organization, for example, expect the details associated with their orders to be correct. If data is wrong, customers have to put in effort (time=money) to solve the mistakes in the data. It is of crucial importance that customers are served optimal.
Other external data customer might be:

- Suppliers (sometimes customers at the same time)
- Public Offices
- Statistical Bureaus

Each organization can then determine with product-attributes are important to the external data customers.

Besides external data users there are internal data users. Internal data users are own employees or departments that use information from other employees or departments. The information used by internal users needs also to be of high quality. If this quality is however low, then errors caused by this low quality have to be traced back to their roots and have to be solved. This leads to higher costs. Some research has been done to estimate the costs of poor data quality. These estimates have proven difficult to calculate according to Redman. His report speaks of studies that have estimated that about 8-12% of the revenues are lost due to poor data quality. More informal studies in service organizations have estimated expenses of 40-60% of the total costs of these organizations to be a result from poor data [17]. So product-attributes in which mistakes are more difficult to find, and take longer to repair then in other product-attributes are important in this impact.

Another impact of low quality of data that was found by Redman is lowered employee satisfaction. If errors constantly reoccur employees get frustrated. So for this impact, all product-attributes in which a lot of reoccurring mistakes happen are important.

Tactical impacts:
Poor data quality has tactical impacts that include poor decision making, difficulties in reengineering and increased organizational mistrust. There is no evidence to support that data needed and used by managers is any better than data used by customers or employees. So this can influence a decision making process. Important product-attributes for decision making are the product-attributes concerning costs and product-attributes that influence planning.

Poor data makes it also more difficult to reengineer [17]. Engineering processes aim to put the right data in the right place at the right time to better server a customer. But you cannot simply serve customers when the data is not correct. This is a direct consequence of wrong data in the same product-attributes as described above in the section about product-attributes that affect the customer.

Organizational mistrust is also increased by poor quality. Wrong data that moves from one department to another creates trust issues for that department. The data is not treated as valuable anymore. Important product-attributes in this impact are product-attributes in which data is supplied by another department then the department that uses the data.

Strategic impacts:
On the strategic level there is, according to Redman, less evidence of the impact of poor data quality, because there have not yet been studies in this direction. There are, however, consequences that stem directly form the operational and tactical level. These consequences can be based on the results from the product-attributes mentioned in those two levels.
The first consequence is that, since the selection of a strategy is by itself a decision making project, poor data influences the strategy of a company. Poor data can also lead to issues concerning data ownership and this can lead to difficult political situations and conflicts between employees/departments. Management attention is then diverted from customers to internal matters and alignment of departments and organizations toward common goals becomes more difficult.

Once the permanent team members have selected the product-attribute to handle, the rest of the team members can be selected and invited. The characteristics of the product-attribute can then be described.

3.4.3 Data product characteristics

The first step in the define phase according to Wang is to define characteristics of the data product. According to Wang data product characteristics can be defined at two levels. At the higher level, the data product has to be described in terms of its functionalities to the data customers (users). At a lower level the attributes and their relationships can be identified. According to Wang it is also important to describe each product-attribute that the data product consist of. This is because the attributes make up the data product and determine the way it is produced, utilized and managed. The complete data product with its attributes and relations can be represented in a entity-relationship model.

The difference in data product and product-attributes, between the theory of Wang and the method designed in this report, has influence on the description of the characteristics. In this report the characteristics of the product-attribute can only be described on one level. The theory of Wang, as described above, uses two levels to describe the data product. A high level and a lower level. But because the focus in this report is on a specific product-attribute, which is a part of the lower level in the theory of Wang, the distinction in level of description can not be made.

From Wang's description of the two levels can be stated that it is important to known what the main functions of the product-attribute are, who the stakeholders of the product-attribute are, and what the product-attribute consists of. This means that the following questions need to be answered in this step:

- What are the functions of the product-attributes? What does it do? Where is it used for?
- Does the product-attribute only have data in the database system, or is there other data connected to the product-attribute? For example data found in MS Excel sheets, text documents, contracts, or declarations.
- Who are the data suppliers, data manufactures, and data users? This changes for each product-attribute.

Once these questions are answered the whole team is aware of the main characteristics of the product-attribute. This knowledge can serve as a basis for the rest of the cycle. The result from this step is a complete description of the product-attribute that was selected.

3.4.4 Data product requirements

Once the product-attribute that will be used for the method is selected and its characteristics are known, the quality requirements for this product-attribute should be determined. To determine the quality requirements the TDQM methodology uses the 'data quality modeling' method [23].

The first part of this step to determine the quality requirements is to establish the 'application view'. This means, to represent the data product in an entity-relationship (ER) model. Its purpose is to identify and describe the different attributes that make up the data product, so that for each attribute the requirements can be determined in the next steps of the method. This is in fact the same step as described in the previous subsection. As explained in this section, this report uses the TDQM method to improve one product-attribute at a time in stead of a data product with several attributes. An ER-model of the product-attribute can not be made. The description of the characteristics of the product-attribute is therefore the result of this step.

The second part of the method is to determine the quality dimensions for each of the product-attributes described in the ER model of the data product. Or, for the method for one product-attribute, determine the quality dimensions for the selected product-attribute. This step leads to the 'parameter view'. The TDQM methodology uses a software tool together with several surveys to determine the important quality dimensions. The tool helps to determine the perceived and expected quality in each dimension that is important to information customers, information suppliers, information manufactures and the managers of the information. Important dimensions are dimensions in which the quality is perceived as low and dimensions in which large differences exist between the perceived quality of the supplier and the expected level of quality of the customer.
Unfortunately the software tool Wang uses is not available, because no information about the product and its availability could be retrieved. The tool determines important quality dimensions out of a large list of quality dimensions by letting team members answer multiple choice questions. Based on the answers on these questions, the tool can calculate important quality dimensions. The tool not being available has therefore two consequences: The set of quality dimensions to use will have to be determined and a new way of determining the important dimensions will have to be determined.

Quality Dimensions
In literature there are a lot of views to choose from when determining which set of dimensions to choose. For example; Abbott presents a product based view to quality. He defines quality as the differences in the quantity of some desired ingredient or material in a product. Crosby defines quality from the production point of view. For him quality means conformance to requirements. Juran sees quality as fitness for use by users [5].
In the view of TDQM, quality data is primarily important for the users of the data. The view of Juran is closest to that. The software tool also uses this view. It uses a set of dimensions carefully selected in the article of Wang and Strong [25]. In this article a large list of possible dimensions is narrowed down to a smaller and more useful list, based on the view of Juran and based on several statistical studies in companies. The dimensions consist of four categories. These categories are based on the conceptual framework the article uses and the remaining quality dimensions were sorted in these categories based on the statistical studies. The categories and their dimensions are:

- Intrinsic Data Quality; denotes that data have quality in their own right.
    - Accuracy, the extent to which data is correct or reliable
    - Objectivity, the extent to which data is unbiased, unprejudiced, or impartial
    - Believability, the extent to which data is regarded as true or credible
    - Reputation, the extent to which data is highly regarded in terms of its source or content
- Contextual Data Quality; Highlights the requirements that data quality must be considered within the context of the task at hand.
    - Relevancy, the extent to which data is applicable and helpful for the task at hand
    - Value-Added, the extent to which data is beneficial and provides advantages from its use
    - Timeliness, the extent to which the data is sufficiently up-to-date for the task at hand

  - **Completeness**, the extent to which data is not missing and is of sufficient breadth or depth for the task at hand
  - **Amount of data**, the extent to which the volume of data is appropriate for the task at hand
- **Representational Data Quality;** Emphasizes the importance of the role of systems.
  - **Interpretability**, the extent to which data is in appropriate languages, symbols, or units, and the definitions are clear
  - **Ease of understanding**, the extent to which data is easily comprehended
  - **Concise representation**, the extent to which data is compactly represented
  - **Consistent representation**, the extent to which data is presented in the same format
- **Accessibility Data Quality;** Also emphasize the importance of the role of systems.
  - **Accessibility**, the extent to which data is available, or easily and quickly retrievable
  - **Access security**, the extent to which access to data is restricted appropriately to maintain its security

So from this list the important quality dimensions can be chosen for the selected product-attribute.

Determining the important quality dimensions
Because of the software tool not being available, the requirements of the data quality and which dimensions are important will have to be determined by hand. The tool determines the important quality dimensions based on the team members answers on 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.

So the team members have to determine the importance of the dimensions. The data quality team should also asses their perceived level of quality and the expected level of quality of those dimensions. These variables are in fact the same aspects the Servqual-Model uses to asses quality of services [5, pp. 296]. The model describes a formula in which the score on quality is based on a formula which uses the three variables mentioned above. The variables can be scored on a scale ranging from 1 till 7.

- 1 meaning the dimension is of low importance, the perception of the quality is low, or the expected level of quality is low.
- 7 meaning the dimension is of high importance, the perception of the quality is high, or the expected level of quality is high.

The formula that the model uses is:

$$Q = \sum Li(Pi-Ei)$$

Where:
Q = quality of the service
Li = importance of the i th dimension
Pi = perceived quality of the i th dimension
Ei = expected quality of the i th dimension

This formula gives a total quality score of the product-attribute as a result. This result can be split up to give a better indication where quality problems might exist in specific dimensions. To better understand the meaning for each dimension the results for each dimension can be published in a chart. To provide a good overview of the results, three types of charts can be made for each dimension:

- Importance of each dimension according to different stakeholders
- Perceived quality in each dimension according to different stakeholders
- Expected quality in each dimension according to different stakeholders

Once all the team members have scored all the dimensions the charts can be made and the team can start to select the dimensions to measure. For example, dimension that score low on importance in the eyes of the suppliers and score high on importance in the eyes of customers (users of data) would be important dimensions to measure.
So without the software tool, the new method can still determine the dimensions that require quality.

The goal of the third part in the data quality modeling method is to operationalize the important quality dimensions that were identified in the second part into measurable variables. These variables are the quality indicators for the dimensions. This means that for example when for a product-attribute timeliness is important, this can be operationalized by determining the variable 'age of the data'. This step leads to the so called 'quality views' of the different stakeholders.

Because differences in quality views can exist, the fourth and last part of this step integrates these views into one quality view. For example, one view may have 'age of data' as an indicator, where another view may have 'creation time' as indicator. In this case, the team can choose which indicator is easiest to measure or which indicator gives the best information.

The result of this step is that the dimensions in which the data customers require quality are known and that objective measurement variables have been found.

3.4.5 Data manufacturing process

The last step in the define phase of the TDQM methodology is to define the information manufacturing process. This is the process in which the data for the product-attributes is being manufactured. It describes the whole flow of product-attribute data from the supplier to the user of the data. The theory of Wang explains that knowledge of this process serves as a basis for better understanding why certain quality dimensions are important. But in fact, clarity of the data manufacturing process can also help finding the causes of the quality problems in chosen product-attribute. Knowledge of the whole process can lead to a better understanding why certain steps are needed. It can also lead to insights in possible unnecessary steps, with extra risks of lowered data quality.
Having a clearly defined process can also help in keeping the quality of the data high, for example if employees change. Employee changes can for example lead to not completing the required tasks, because there is no clear understanding of who has to do what in the whole manufacturing process.

The TDQM methodology uses the information manufacturing system method [4] to model the data manufacturing process. The technique used in this method is very similar to the more common basic flowchart modeling technique. The therefore model the manufacturing process the basic flowchart modeling technique can be used. The components mentioned figure 7 can be used to model the data manufacturing process.

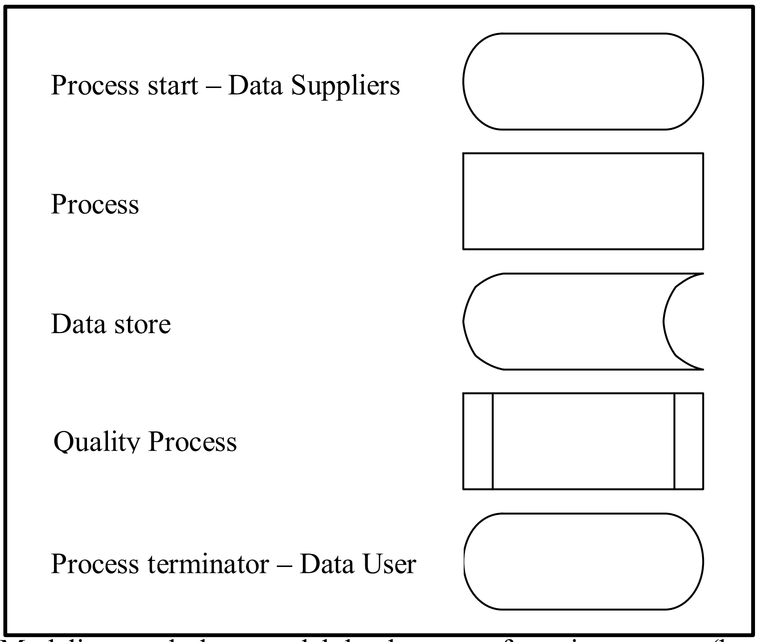


Figure 7: Modeling symbols to model the data manufacturing process (based on [4])

The process start symbol is the point where the process starts, it usually represents a data supplier that has to take action to get data in the database system. The process start symbol does not have a predecessor symbol. The supplier can be either internal or external and potentially supply several different types of raw data. The role of the processing symbol is to show where in the process value is added to the data by manipulating or combining appropriate data units before they can be stored. The data storage symbol models the placement of data units in files or databases where they are available as needed for additional processing. The quality process symbol shows where data quality is being improved in the process. This means that the output stream has a higher quality level then the input stream. The process terminator symbol represents final user of the data that was supplied to the data manufacturing process. The users ultimately judges the quality of the data product. These components can all be used to clearly define the data manufacturing process.

The description of the data manufacturing process completes the define phase. The results are a complete description of the chosen product-attribute, the dimensions in which quality is required are known and objective variables to measure those dimensions are known. Also the process in which the data moves from supplier to the user is known and clearly recorded in a flowchart.

The next section uses the knowledge of this section to establish measurement methods for the important quality dimensions.

3.5 The TDQM Measurement Phase

Once the define phase is complete the product-attribute has been thoroughly described. The main characteristics of the product-attribute are known, the process in which the data for the product-attribute is being created is known, the important dimensions in which users require quality are known and objective variables are determined in these dimensions. In this phase the metrics to measure these variables have to be identified. In translating the variables into measurable metrics the team should keep in mind the dimension of quality the variable was chosen in. Also important, is to keep in mind the business rules and the laws that might have contributed to a dimension being important.
In subsection 3.5.1 examples are given how to measure quality in the quality dimensions given in subsection 3.4.4 by providing metrics. In subsection 3.5.2 guidelines are given on how to determine which metric to choose for the measurement. Wang does not elaborate on this in his theory. But because it can be difficult to translate the dimensions chosen in the previous step into measurable metrics that are able to represent the quality in the dimension, guidelines to choose the metrics are given from the theory of Cooper and Schindler [6] and from the case of IRI [14].

3.5.1 Measurement forms

This subsection uses Pipino [16] and Zahedi [28] to give examples on how to measure the twelve quality dimensions described in subsection 3.4.4.

Accuracy
The Accuracy dimension represents the data being free of error. If one is counting the data units in error, the metric can be defined as the number of data units in error divided by the total number of data units subtracted from 1. This metric is called the simple ratio [16]. In practice, determining what constitutes a data unit and what is an error requires a set of clearly defined criteria. It is possible for an incorrect character in a text string to be tolerable in one circumstance but not in another. This should have also been specified in the define phase.

Believability
Believability is the extent to which data is regarded as reliable. It may reflect an individual's assessment of the credibility of the data source, comparison to a commonly accepted standard, and previous experience. Each of these variables is rated on a scale from 0 to 1, and overall believability is then assigned as the minimum value of the three. Assume the believability of the data source is rated as 0.6; believability against a common standard is 0.8; and believability based on experience is 0.7.

The overall believability rating is then 0.6 (the lowest number). As indicated earlier, this is a conservative assessment. An less conservative alternative is to compute the believability as a weighted average of the individual components.

Objectivity, Reputation

Objectivity is the extent to which data is unbiased, unprejudiced, or impartial. Reputation is the extent to which data is highly regarded in terms of its source or content. Objectivity and reputation are difficult to measure. Stating the origin of the data can for some companies be enough to determine the reputation of the data or the objectivity of the data. The min operation described in the believability dimension can however also be used to determine the result for the objectivity and reputation dimensions.

Timeliness

Timeliness is the extent to which the data is sufficiently up-to-date for the task at hand. Timeliness is defined by customers. They determine how up-to-date certain data needs to be. One must keep in mind that there can be two different definitions of timeliness:

- Availability of information in the system, and
- Availability of information to customers.

The timeliness metric should reflect the availability of the information to the customer, because that is when they can access it and process it.
There are several metrics of timeliness, such as: [28]

- The time interval between two consecutive updates
- The time it takes for information to become available to users between two consecutive updates
- Access time
- Time past since the last update of the data

If the timeliness dimension reflects how up-to-date the data is, with respect to the task it is used for, Ballou [4] can be used to measure the timeliness of the data. Ballou proposes a general metric to measure timeliness. He suggest timeliness should be measured as the maximum of one of two terms: 0 and 1 minus the ratio of currency to volatility. Here, 'currency' is defined as the age plus the delivery time minus the input time. Volatility refers to the length of time data remains valid; delivery time refers to when data is delivered to the user; input time refers to when data is received by the system; and age refers to the age of the data when first received by the system.

An exponent can be used as a sensitivity factor, with the max value raised to this exponent. The value of the exponent is task-dependent and reflects the analyst's judgment. For example, suppose the timeliness rating without using the sensitivity factor (equivalent to a sensitivity factor of 1) is 0.81. Using a sensitivity factor of 2 would then yield a timeliness rating of 0.64 (higher sensitivity factor reflects fact that the data becomes outdated faster) and 0.9 when sensitivity factor is 0.5 (lower sensitivity factor reflects fact that the data loses timeliness at a lower rate).

Completeness

Completeness is the extent to which data is not missing and is of sufficient breadth or depth for the task at hand. The Completeness dimension can be viewed from many perspectives, leading to different metrics. At the most abstract level, one can define the concept of schema completeness, which is the degree to which entities and product-attributes are not missing from a certain data schema. At the data level, one can define column completeness as a function of the missing values in a column of a table. A third type is called population completeness. If a column should contain at least one occurrence of all 50 states, for example, but it only contains 43 states, then we have population incompleteness. Each of the three types (schema completeness, column completeness, and population completeness) can be measured by taking the ratio of the number of incomplete items to the total number of items and subtracting from 1.

Interpretability
Interpretability is the extent to which data is in appropriate languages, symbols, or units, and the definitions are clear. Again it is difficult to measure the interpretability for a product-attribute, because most product-attributes cannot be compared to, for example ISO standards. Interpretability can however be important for the quality in certain product-attributes. Therefore one should make sure that the set of possible values for certain (for example *County* abbreviations) product-attributes, follows the ISO standards to assure interpretability.

Ease of understanding
Ease of understanding is the extent to which data is easily comprehended. The use of ISO norms can also, next to interpretability, increase the ease of understanding for product-attributes. Easy of understanding can then be measured by taking the ration of the number of values that do not follow the ISO standards to the total number of items and subtracting from 1.

Consistent representation
Consistent representation is the extent to which data is presented in the same format. The Consistency dimension can be viewed from a number of perspectives, one being consistency of the same (redundant) data values across tables. A metric measuring consistency, is the ratio of violations of a specific consistency type to the total number of consistency checks subtracted from one.

Accessibility
Accessibility is the extent to which data is available, or easily and quickly retrievable. A similarly constructed metric as the timeliness dimension can be used to measure accessibility. The metric emphasizes the time variable of accessibility and is defined as the maximum value of two terms: 0 or 1 minus the time interval from request by user to delivery to user divided by the time interval from request by user to the point at which data is no longer useful. Again, a sensitivity factor in the form of an exponent can be included.
If data is delivered just prior to when it is no longer useful, the data may be of some use, but will not be as useful as if it were delivered much earlier than the cutoff. This metric trades off the time interval over which the user needs data against the time it takes to deliver data. Here, the time to obtain data increases until the ratio goes negative, at which time the accessibility is rated as zero (maximum of the two terms).

Access security
Access security is the extent to which access to data is restricted appropriately to maintain its security. One way to measure this dimension is using a ration of users who should not be able to access the data, but can anyway, on the total amount of users that have access to the data, subtracted by 1. For sensitive data this dimension should be closely evaluated to keep wrong users from accessing the data.

Relevancy, Value-Added
Relevancy is the extent to which data is applicable and helpful for the task at hand. Value-added is the extent to which data is beneficial and provides advantages from its use. Relevancy and added-value can be measured by using a simple ratio. One should however keep in mind that these dimensions will deal with the product-attribute itself. They can be causes of not maintaining the data.

Amount of Data, Concise representation
Amount of data is the extent to which the volume of data is appropriate for the task at hand. Concise representation is the extent to which data is compactly represented. Both dimensions could be explained by stating that the data quantity should not be too little and also not be too much. A general metric that can find this tradeoff is the minimum of two simple ratios: the ratio of the number of data units provided to the number of data units needed, and the ratio of the number of data units needed to the number of data units provided.

3.5.2 Which metric to choose?

So up until now, the product-attribute has been chosen and its important quality dimensions are selected. The team should now choose a metric, with help from the measurement forms in the previous section, to measure the quality of the information on. It might however be difficult to determine if a metric is appropriate for the measurement of a quality dimension. Therefore the criteria of Cooper and Schindler [6, pp. 231] can be used. They stated that a ‘good’ metric should meet the three main criteria: validity, reliability and practicality.

Validity refers to the extent to which a metric measures what is actually meant to be measured. Two major forms of validity are external and internal validity. External validity of research findings refers to the ability of the data to be generalized across persons, settings and times. Internal validity refers to the ability of the metric to measure what it is supposed to be measured.

A reliable measurement metric is, According to Cooper and Schindler [6, pp. 236], the degree to which a metrics supplies consistent measurement results. Reliability can contribute to validity, but is not a sufficient condition for validity. This can be explained by the example of a bathroom scale. If the scale measures your weight correctly (the scale is known to be accurate), then it is both reliable and valid. If it consistently overweighs you by six kilogram, then the scale is reliable, but not valid. If the scale measures different values in each measurement, then it is not reliable and therefore it cannot be valid.

These two criteria, validity and reliability, are a scientific approach to value if a metric is ‘good’. The last criteria is a more operational criteria. Practicality is defined by Cooper and Schindler [6, pp. 240] as:

- Economy, what are the cost of measuring the dimensions with this metric?
- Convenience, is it easy to measure the dimensions by using this metric?
- Interpretability, is it easy for others then the measurer to interpret the results?

Reliability of a metric can only be assessed after a number of measurements and therefore after the selection of the metric. But to asses at forehand which metric to select the criteria of Kovac, Lee and Pipina might be used [14]. These criteria are called RUMBA and are shown in table 4. The criteria have been designed to be able to asses different metrics in data quality problems.

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 4. The RUMBA criteria [14]

3.5.3 Measurement and presentation of results

Once the metrics are determined the actual 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% [15]. Now that the sample size, the quality dimensions, and the metrics are known, the measurement can then be conducted.

The TDQM methodology does not give examples of how to display these results of a measurement. MS Excel provides several types of charts which can be used to graphically display the results. For example a line chart or a bubble chart. To identify which dimensions cause the problems the results can also be graphically displayed using Pareto diagrams (also called bar charts) [5, pp. 242]. 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 a clear way to show in which dimension the biggest problems exist. An example, in which a non existing product-attribute is valued with a simple ratio metric on the accuracy, completeness, and consistency dimensions, is shown in figure 8.

Number of errors on several dimensions in a non existing product-attribute

Dimension	Errors	Percentage
Accuracy	225	56%
Completeness	125	31%
Consistency	50	13%
Total	400	100%

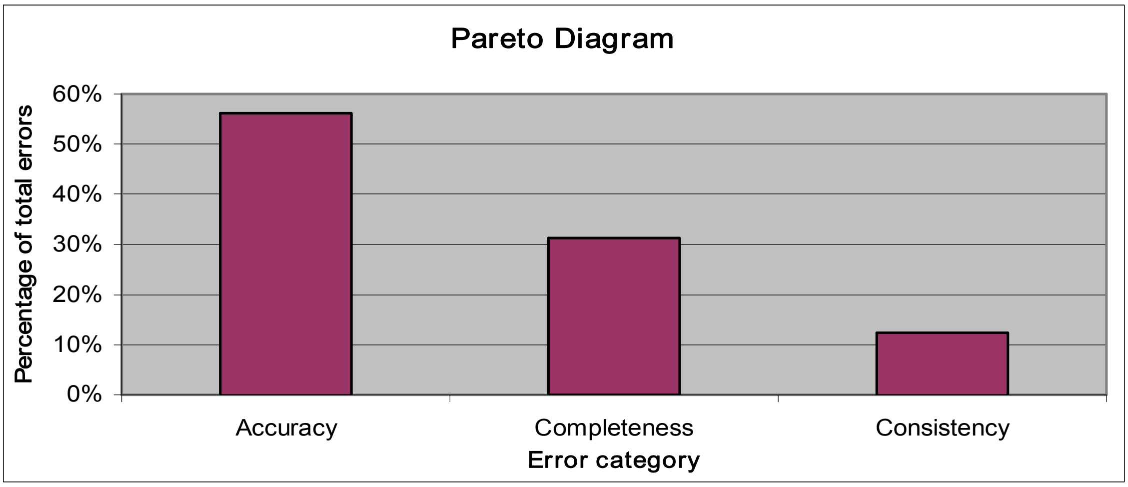


Figure 8: Example of a Pareto diagram

The results of this phase are the results of the measurement. These results together with the knowledge from the define phase can be used as basis the start of the analysis phase that Wang describes.

3.6 The TDQM Analysis Phase

In the analyze phase, the causes of the quality problems in the measured dimensions of the product-attribute have to be found. These causes are then to be improved in the improve phase. The TDQM methodology does not provide much guidance on how to fill in the last two phases of the methodology. In the analysis phase, the TDQM methodology mentions only tools which can make the results of the measurements more clear. For example the use of a Pareto diagram over time to analyze the development of quality in a certain dimension.
The main goal of this phase according to Wang is to find the root causes for the problems in the different dimensions. Root causes can be found with three different methods according to Doggett [8].

- Cause and effect diagram (CED)
  This tool can be used to break down possible causes into more detailed categories so that they can be organized and related into factors that help identify the root cause. See Appendix 14a for an graphical example of the tool.
- Interrelationship diagram (ID)
  This tool can be used to quantify the relationships between factors and thereby classify potential causal issues or drivers of the problem. Appendix 14b displays an example of the tool.
- Current reality tree (CRT)
  This tool can be used to find logical interdependent chains of relationships between undesirable effects leading to the identification of the core cause. Appendix 14c shows an example of the tool.

According to the article of Doggett the CED method is the easiest to use tool for identifying root causes. It does however not identify relationships between cause factors and it also has no formal method to select and evaluate root causes. CRT is the most complex tool for identifying root causes. CRT is the only method that has a mechanism for testing logic which, according to the article of Doggett, contributes to the opinion of users that the method is complex and time consuming. Because of the fact that Doggett ea. see the CED method as the easiest method to use and learn, and because probably not all new team members have knowledge of these methods, the CED method would be best to choose. Also because of the fact that the other methods are more time consuming then the CED method according to Doggett. This method could therefore be used to identify the root causes of the problems in the quality of the product-attribute dimensions for Honeywell.

The CED method designed by ThinkReliability [20] is a well worked out method with clear steps to follow. The method uses the basics of the book by Wilson, Dell, and Anderson [27]. The method is based on three principles:

- Systems thinking, every system has parts that interact.
- Visual communication, combining words, images and shapes
- Cause-and-effect, the basics of the cause and effect principle

The method consists of three steps:

- Problem: what is the problem?
  Determine what the problem is, when it happened, where it happened and how important it is (how it impacts the overall goals). The team has to define the specifics of the problem in the dimensions. The general characteristics of the product-attributes and the process in which the data for the product-attributes is created were already described in the define phase. Specific questions that have to be answered are:
  - What is the problem? Think of this as what would be the name of the issue. In the case of Honeywell the problem would be that the required level of quality in certain dimensions cannot be met in the selected product-attribute.
  - When does it happen?
  - Where did it happen?
  - How did the issue impact one of the overall goals? In the case of Honeywell the overall goal is to improve the quality of the product-attributes. This could again be split up to how to improve the quality in a specific dimensions.
- Analysis: Why did it happen?
  Identify the cause-and-effect relationships and provide evidence to validate the causes.
  The relationships are modeled using a cause map. The start of this cause map is the stated impact on an overall goal. In the case of Honeywell this would mean the quality problem in a specific dimension. In the example in Appendix 14a this step is the "High absenteeism" at the far right of the cause map.
  Next the facilitator of the meeting, in the data quality team this would be the data engineer, should ask the team members the question: Why?
  The answers of the group should then be written down. If it's not directly clear where the cause should be written down, then these causes can first be put next to the scheme to later find a place in the scheme. The result of this step is a cause map of the problem.
  Next step is to complete the map by adding evidence of each cause to the map. This could serve as a base for getting more understanding and support from top management.
- Solutions: What are we going to do?
  Collect possible solutions (ideas) to control individual causes then select the best overall solutions. Write down each specific action to be taken with an owner and a due date. Since this step is part of selecting and implementing it is described in more detail in the improve phase.

Strong ea. [18] give some examples of causes which were identified in other cases. They examined 42 case studies in data quality improvement projects and identified their main causes. This knowledge might be helpful in identifying causes of problems in product-attributes for other organizations as well. The article does not guarantee that these causes are also causes for the problems in other

organizations, but it could give more clarity for the new data quality team, and it could also speed up the process of identifying causes.
These are the causes presented in the article:

1 Multiple sources of the same information will produce different values. Quality dimension which are involved with this cause are: 'consistency' and believability'.
2 Information is produced using subjective judgments, leading to bias. Quality dimension which are involved with this cause are: 'objectivity' and 'believability'
3 Systemic errors in information production lead to lost information. Quality dimension which are involved with this course are: 'correctness', 'completeness', and 'relevancy'.
4 Large volumes of stored information make it difficult to access information in a reasonable time. Quality dimension which are involved with this cause are: 'concise representation', 'timeliness', 'value-added', and 'accessibility'.
5 Distributed heterogeneous systems lead to inconsistent definitions, formats, and values. Quality dimension which are involved with this cause are: 'consistent representation', 'timeliness', and 'value-added'.
6 Nonnumeric information is difficult to index. Quality dimension which are involved with this cause are: 'concise representation', 'value-added', and 'accessibility'.
7 Automated content analysis across information collections is not yet available. Quality dimension which are involved with this cause are: 'analysis requirements', 'consistent representation', 'relevance', and 'value-added'.
8 As information consumers' tasks and the organizational environment change, the information that is relevant and useful changes. Quality dimension which are involved with this cause are: 'relevance', 'value-added', and 'completeness'.
9 Easy access to information may conflict with requirements for security, privacy, and confidentially. Quality dimension which are involved with this cause are: 'security', 'accessibility', and 'value-added'.
10 Lack of sufficient computing resources limits access. Quality dimension which are involved with this cause are: 'accessibility' and 'value-added'.

3.7 The TDQM Improvement Phase

As explained in the previous section, the theory of Wang does not give much guidance to how to actually improve the data quality. The theory only identifies the need to identify key areas for improvement. It advices the use of a method for allocating resources for the improvement of the product-attribute [3]. Wang does however not go into more detail about how to fill in this phase. As explained in the previous section, the cause mapping method therefore continues. As can be seen in the previous section, this method provides a step to actually improve the situation; the solution step. This step ensures that the improve phase of Wang is filled in with a clear and easy way to value possible solutions. So to make sure key areas for improvement are identified.
The solution step consist of three parts:

1 Identify possible solutions.
  Possible solutions are ideas of team members that will prevent the problem form occurring. This step is also commonly referred to as brainstorming or the creative step. For all causes identified in the previous step possible solutions will have to be identified.
2 Select the best solution
  When selecting the best solutions, it is necessary to know what impacts the solution might have and also to have clear what the expected contribution to the improvement of the data quality will be. The cause mapping method therefore uses several criteria which should be considered in choosing a solution:
  - The cost to design the solution
  - The cost to implement the solution
  - The resources required
  - The impact to the overall goals (the data quality)
  - The reduction of risk

The cause mapping method does not provide a way to compare the results of the evaluation of the solutions. The priority matrix is a efficient way to do this and to make a selection of the right solution [5, pp. 247]. To use this tool the team should first decide how important each criteria is. The next step is to score each alternative on these criteria and then calculate the total score for the alternative. Once all alternatives are scored the alternative with the highest score is to be selected.

3 Implementation

In this part the selected solution is transferred into specific actions which have to be taken. The team should assign team members or other employees to these actions to make sure they are executed. An action plan can be set up to keep track of all actions and there statuses. In this action plan the following aspects can be included:

- Action
- Owner
- Due date
- Status
- Completed

Evaluating the results of the implemented solution is difficult and can take time. If for example the solutions is an adapted data manufacturing process, then effects can only be measured by using data from newly input products. To get sample size that delivers representative results can sometimes take over a year, depending on how many new products enter the database.

3.8 Conclusion

This section analyzes if the developed "Organizational Theory" in this chapter is able to answer the first two research questions of this report. The theory described in this chapter is summarized in table 5 to give an overview of Wang's TDQM methodology and the additions that, according to the conditions of the specific problem handled in this report, had been made to its method. The next chapter develops this knowledge into "Management Theory" and adapt the method to the situation in Honeywell Emmen. The method is then field tested in the fifth chapter and evaluated in the sixth chapter.

In the designed method, quality is seen as fitness for use by users (3.4.3). This vision is based on the view of Juran [5]. Therefore the set of quality dimensions on which the quality of the data-products can be measured is also based on research which uses this view [25].
In the data quality requirement step (3.4.3), important quality dimensions can be chosen out of the set of quality dimensions defined by Wang and Strong [25]. This step uses the opinions of the involved stakeholders, including the data users, to identify these important quality dimensions. This step and theory therefore make sure that the data quality team knows what quality means for a certain product-attribute and which quality dimensions are required to be of high quality for this product-attribute according to its users. The first research question (see section 2.3)is therefore answered.

The second research question initiated the search in this chapter for a method that is able to actually improve the product-attribute quality. It is stated in section 3.1 that the method should be broad enough to be able to handle all the identified causes in section 2.2. The broadest quality management methodology which has focus on data quality is the Total Data Quality Management methodology [19]. This methodology was critically evaluated in this chapter and adapted to fit the problem statement in chapter 2. The designed method should be able to improve the quality. Therefore the second research question has been answered. This method can therefore be used as a basis to develop the "Management Theory" on. The next chapter therefore adapts the method to the situation in Honeywell Emmen and chapter five tests the method in practice and improve it if necessarily. Table (table 5) summarizes the TDQM methodology and the additions made in this chapter. A third column is added to this table in the next chapter. This column adds the adaptations that will be made to make the method suitable for Honeywell Emmen.

Figure 9 shows the flow diagram that can be drawn from the method based on the described steps in this chapter. This figure is used in the next chapter to indicate on which part of the method the analysis is focused.

The TDQM Methodology of Wang [22]	Adapted Method based on theory in chapter 3
TDQM Quality Team	**TDQM Quality Team**
Management sets up data product quality team: • Team Champion • Team Engineer • Data Suppliers • Data Users • Data Manufacturers • Data Product Managers **By: Management**	Management chooses permanent team members: • Team Champion • Team Engineer **By: Management**
TDQM Define Phase	**Adapted Define Phase**
	Step 1: Select the product-attribute and select non-permanent team members: • Data Suppliers • Data Users • Data Manufacturers • Data Product Managers **By: Permanent team members**
Step 1: Define Data Product Characteristics: Two levels: High level with functionalities of the whole data product. Lower level with functionalities of each attribute **By: Complete data product quality team**	**Step 2**: Define Product-attribute Characteristics: One level, functions of product-attributes and stakeholders **By: Complete Data Quality Team**
Step 2: Define Information Quality Requirements: Use a tool with quality dimensions from Wang and Strong [25] **By: Complete data product quality team**	**Step 3**: Define Information Quality Requirements: Use the in section 3.4.3 described technique to determine important quality dimensions. **By: Complete Data Quality Team**
Step 3: Define Information Manufacturing System: Model with Ballou, Wang, Pazer, and Tayi [4] **By: Complete data product quality team**	**Step 4**: Define Information Manufacturing System: Model using the basic flowchart technique. **By: Complete Data Quality Team**
TDQM Measurement Phase	**Adapted Measurement Phase**
Step 4: Develop Information Quality Metrics **By: Complete data product quality team**	**Step 5**: Determine measurement metrics: Examples are given in section 3.5.1. Making a selection between metrics can be done based on Cooper and Schindler [6] and Kovac e.a. [14]. **By: Complete Data Quality Team**
Step 5: Measure Data Product Quality **By: Complete data product quality team**	**Step 6**: Measure and present the quality Show results in Pareto diagrams [5]. **By: Agreed team members and other employees**
TDQM Analysis Phase	**Adapted Analysis Phase**
Step 6: Find root causes **By: Complete data product quality team**	**Step 7**: Describe specific problem: Describe the specific problems based on the results from the measurements. **By: Complete Data Quality Team**
	Step 8: Analysis of the problems: Analysis of the problem by drawing a cause map with the whole quality team **By: Complete Data Quality Team**

Table 5: Summary of the development of the method

TDQM Improvement Phase	Adapted Improvement Phase
Step 7: Determine where to invest resources based on [4] and solve the causes **By: Complete data product quality team**	**Step 9**: Generate Solutions: Based on the causes, thinks of possible solutions **By: Complete Data Quality Team**
	Step 10: Solution Selection: Solutions can be scored based on the criteria presented in section 3.7 **By: Complete Data Quality Team**
	Step 11: Action Plan: Assign specific tasks to team members to implement the selected solution **By: Complete Data Quality Team**
	Step 12: Check Progress: Check the progress of the agreed action plan and discuss problems **By: Complete Data Quality Team**

Table 5 (Cont.): Summary of the development of the method

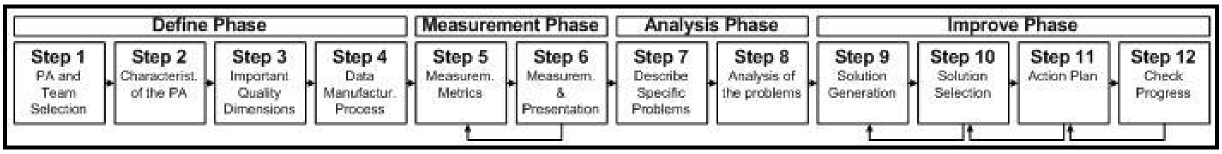


Figure 9: Flowchart of the designed method.

Chapter 4: Design the method for Honeywell

In this chapter the theory from the previous chapter is used to design a method with concrete steps for Honeywell Emmen. The method is designed based on the “Organizational Theory” described in the previous chapter and based on the main causes of the problems within Honeywell found in chapter two. The sections in this chapter follow the theory and steps described in the conclusion of chapter 3 and will develop the “Management Theory”. Section 4.1 therefore describes the quality team. Section 4.2 describes the steps that have to be taken in the define phase and section 4.3 handles the steps in the measure phase. Sections 4.4 and 4.5 describe the analysis phase and the improvement phase, which are based on the cause mapping method. In the last section the method is summarized and a third column is added to the table from section 3.8.

4.1 Product-attribute Data Quality Team (PDQT)

In the previous chapter several sorts of teams where identified. These were:

- Steering Group
- Autonomous Team
- Quality Circles
- Quality Improvement Teams

The problem in Honeywell leads to a team that can be best described by the Quality Improvement Team, because of it should consist of members from different departments, and because the team should be able to change its members. The following characteristics lead to this conclusion.

The forming of a data quality team for Honeywell has to improve the communication between employees involved in the quality of the product-attributes. Because the product-attributes usually have stakeholders in multiple departments (For example *Country of Origin*, this product-attribute has a supplier in the purchasing department and the internal user is the customer logistics department), the team should consist of members from different departments. Wang also recognizes the requirement for a multidisciplinary approach in the TDQM methodology as explained above. He stresses the importance of the presence of team members from different positions in the process that creates the data for the product-attributes.

Another characteristic of the quality improvement team of importance to Honeywell is that a quality improvement team may change its members. Wang does not elaborate on this possibility, because in his report the team handles the problems of one data product. Within Honeywell, different product-attributes will be handled (see section 3.4). Because of this, the stakeholders for the different product-attributes can change. For example the *Country of Origin* product-attribute has the purchasing department as one of the data suppliers, the product-attribute *Cost Price* has the financial department as supplier. Therefore it is necessary that the members of the team can change depending on which product-attribute is addressed.

Team Members

As explained in the previous chapter the team should consist of a TDQM champion or team leader, a data quality engineer, and other members who are data suppliers, manufacturers, users, or data product managers. Due to the fact that the team members should be able to change depending on the product-attribute that is chosen, the team will have to consist of permanent and non permanent members.

Therefore Honeywell should, in the first place, assign a team leader. This is then the first permanent team member. As explained before, this should be a senior executive who acknowledges the need to improve the data quality in the product-attributes and who has the power to implement solutions that are selected by the team.

The next team member is also of key value to the team. The data quality engineer. This member should have knowledge of the TDQM method and should also be able to facilitate the meetings and should be able to document the results of the meetings. This member can therefore also best be a permanent member, since it requires experience with the method and experience to document the results.

Since all the product-attributes in Honeywell are stored in the Oracle database it is of importance to have a database administrator as a team member. The database administrator knows the possibilities and limitations of the system and therefore is valuable in the improvement process. Since all product-attributes are stored in the Oracle database, the system engineer is a permanent member of the team. In the team members list of Wang this role can be best described as the data manufacturer. Someone who controls the database system on which the data is stored. Since the ideal candidate for the team engineer is one of the database administrators within Honeywell, these two functions can be combined.

Wang also mentions the data product manager as a team member. Since the method for Honeywell will not handle a data product and since there is no product-attribute manager, someone closely related to the management of the product-attributes will have to be selected. The person responsible for the financial costing is closest to that. This person has the responsibility to assign a cost price to a product and to make it active on the database. Since this making active requires several attribute to be filled out, this employee can be seen as a sort of manager of these product-attributes. Since this employee is also responsible for the financial costing in Honeywell she can be a good addition to the permanent team members. She can for example also predict the costs of a solution to a problem with a certain product-attribute.

The other members should, according to the theory of Wang be suppliers and users of the data in the product-attributes. As explained above these members can change according to which product-attribute is selected for the method. There are however two restrictions in selecting these members. In the method developed by Wang, these members can be from other companies as they can be the actual suppliers or users of the data. Since Honeywell does wants to keep problems with data quality internally, selecting members outside Honeywell is not possible. The selected members should therefore be able to asses the norms and wishes of these external stakeholders.
The other restriction is the number of suppliers and users that are invited. The quality improvement team literature prescribes a team of between four to eight people. This includes the permanent members. Since the above described four (or three) permanent members will always be a part of the team, four other members can be invited.

The quality team and the members for Honeywell Emmen are shown in the next table:

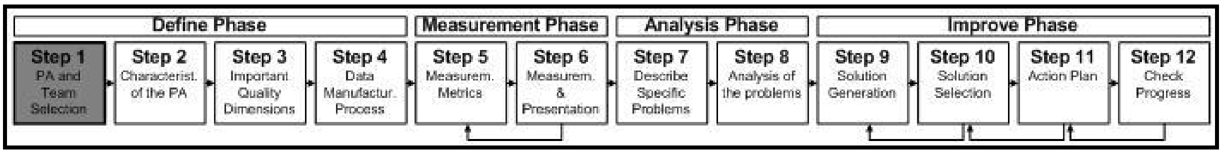
Permanent:	Honeywell Emmen:
Team Leader	Manager of the Product data management department, closest senior executive to the product-attributes.
Team Engineer	The database administrator, has experience with quality improvement projects and also knows the possibilities and restrictions of the database system. Therefore these two roles can be represented by one member within Honeywell Emmen
Data Manufacturing member	
Team Financial member	Knows the financial parameters and works daily with the product-attributes.
Changing	Honeywell Emmen:
Data Suppliers	Who can represent the wishes and understand the possibilities of the suppliers of the data.
Data Users	Who understand the wishes of the users of the data, including restrictions by law.

Table 6: Team members

4.2 Define phase

As explained in subsection 3.4.1 the first consequence of the difference between data products and product-attributes is that the method for Honeywell will start with a step to select the product-attribute to improve. This first step is explained in more detail in sub-section 4.3.1. The second step is the first step in the TDQM method, the description of the characteristics of the product-attribute. The step is explained in subsection 4.3.2. The next step is to define the important quality dimensions (4.3.3) and the last step in the define phase is the description of the data manufacturing process (4.3.4).

4.2.1 Step 1: Product-attribute selection and team members selection



The permanent team members can select this product-attribute based on two criteria:

- Select a product-attribute that have caused problems in the past (4.2.1.1), or
- Select a product-attribute that could cause problems in the future (4.2.1.2).

4.2.1.1 Product attributes that have caused problems in the past

In chapter 2 several product-attributes that have cause problems in the past are mentioned. These are:

- *Country of Origin*
- *Duty Tariff Code*
- *EAN Bar Code*
- *Transfer Price*

4.2.1.2 Product attributes that could cause problems in the future

To find product-attributes that could cause problems in the future Redman [17, see subsection 3.4.1] can be used. Table 3 mentions three levels of impacts which Redman identified and these impacts can be connected to product-attributes which are able to cause these impacts. By using the method on these product-attributes, Honeywell can make sure that these impacts can be prevented. The levels of impacts (operational, tactical, and strategic) will now be analyzed (see also table 3).

Operational level impacts

The first impact that Redman handles in the operational level is lowered customer satisfaction. As explained in subsection 3.4.1 these data 'customers' for Honeywell are:

- Customers (of products Honeywell sells)
- Suppliers (sometimes customers at the same time)
- Public Offices
- Statistical Bureaus

In an interview with an employee of the purchasing department [D] and an interview with an employee of the customer logistics department [E], the following product-attributes have been identified that these external stakeholders use:

- *Country of origin*
- *Duty Tariff Code*
- *Common Coding*
- *EAN Bar Code*
- *Spares only*
- *Project manager*
- *Item Status*
- *Maximum Order Quantity*
- *Minimum Order Quantity*
- *Inc. Order Quantity*

- *Weight Unit of Measure*
- *Unit Weight*
- *SIOP class*

The second operational impact concerns the cost increase as result of low quality of data. Estimating this increase of cost for Honeywell is difficult. There are 150 different product-attributes in which problems and therefore also costs can occur. For each of these product-attributes a different amount of mistakes have to be corrected and a different amount of time is required to correct them. So product-product-attributes in which a lot of mistakes have to be corrected (see 4.2.1.1) and which take a lot of time to determine (have al complex manufacturing process) are important for the costs.

The last operational impact of low quality of data is lowered employee satisfaction. If errors constantly reoccur employees get frustrated. So for this impact, all product-attributes in which a lot of reoccurring mistakes happen are important (see 4.2.1.1).

Tactical level impacts
As explained in subsection 3.4.1 poor data quality has tactical impacts that include poor decision making, difficulties in reengineering and increased organizational mistrust. Important product-attributes for decision making are the product-attributes concerning costs and product-attributes that influence planning.
Product-attribute that concern costs:

- *Standard Cost*
- *Transfer Price*
- *Sales Account*
- *Expense-Account*

Product-attributes that concern planning:

- *Minimum Order Quantity*
- *Maximum Order Quantity*
- *Incremental Order Quantity*
- *Planner Code*
- *Inventory Planning Method*

Poor data makes it difficult to reengineer. Engineering processes aim to put the right data in the right place at the right time to better server a customer. But you cannot simply serve customers when the data is not correct. Therefore the product-attributes that are connected to external stakeholders are important. These are shown in the operational level impacts.

The tactical impact of increased organizational mistrust is also caused by poor quality. Wrong data that moves from one department to another creates trust issues for that department. Important product-attributes in this are product-attributes in which data is supplied by another department then the department that uses the data. Examples of this are the *Country of Origin* and the *Duty tariff code* attributes.

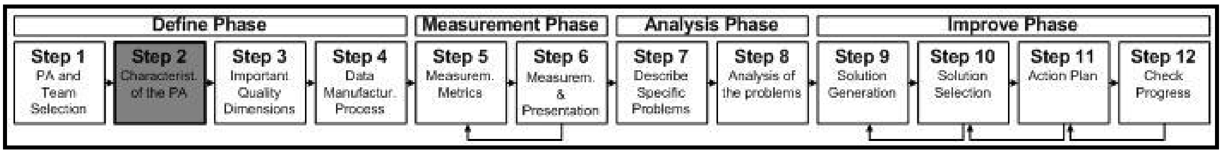
Strategic level impacts
As explained in section 3.4.1 the product-attributes that are important in this dimension are all attributes handled in the other two levels, operational and tactical impacts.

4.2.1.3 Selecting a product attribute

The selection of the product-attribute can be done in a meeting with all the permanent team members, because the non-permanent members can only be chosen when the product-attribute is known. So once the selection has been made, the permanent members can invite involved data suppliers and data customers and plan the next meeting.

To document this first step a standard form has been designed. This form can serve as the first part of the documentation about the product-attribute. In the next steps of the method more documents are created. Together, these documents form a complete description of the product-attribute which can be used by for example new employees. By creating this documentation for each product-attribute a complete description of all processes in which data is created for Oracle can be made.
The first form contains the name of the chosen product-attribute, the names and contact data of the team members, and the planned meetings of the team. Appendix 6 shows this Standard Project Information Form.

4.2.2 Step 2: Characteristics of the Product-attribute



The second step in the method for Honeywell is the first in the TDQM method from Wang [22]. The goal of this step is to clearly define the characteristics of the chosen product-attribute. Clearly, in the way that future new employees can understand its function and know who is involved with the product attribute. Because the first step had to be taken with only the permanent members, the second step has the be planned in a second meeting. This second meeting will be a meeting with all the permanent members, and with the non-permanent members. Handling the product-attribute characteristics with the whole team will contribute to the knowledge of all the individual members about the characteristics of the product-attribute.

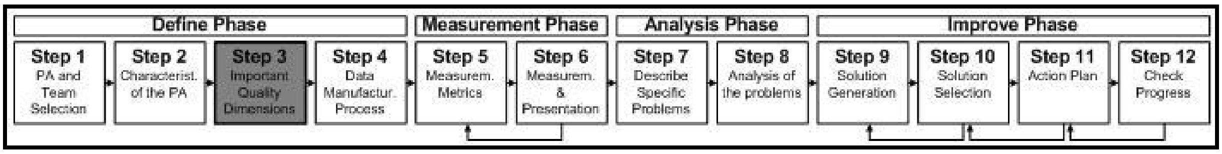
The difference in data product and product-attributes between the theory of Wang and method for Honeywell (as explained in section 3.4 and subsection 3.4.2) influences the description of the characteristics. As explained in those sections the method for Honeywell describes the characteristics of the product-attribute without making difference in level of description.

The document that is designed to store the description of the characteristics handles all the questions that need to be answered according to the theory in chapter 3. The document contains (see Appendix 7: Standard Product-attribute Characteristics Form):

- The name of the product-attribute
- A field to describe the function of the product-attribute
- A field that describes the components of which the product-attribute consists of and the type of these components (for example: Excel sheets, contracts, Oracle data)
- A field where all the suppliers of the data can be mentioned.
- A field where all the manufacturers of the data can be mentioned (employees who change the data in the process in which the data is created)
- A field where all the users of the data can be mentioned

The filled out document can then be stored together with the project information form.

4.2.3 Step 3: Important Quality Dimensions



As explained in subsection 3.4.3 Wang uses in this step the 'data quality modeling' method to determine in which quality dimension the data product requires quality. The method for Honeywell is used to handle a specific product-attribute at a time. Therefore the first step in determining the requirements in the 'data quality modeling' method that the TDQM methodology uses is not necessary for Honeywell.

Unfortunately the software tool the TDQM methodology uses in the second step is not available, because no information about the product and its availability could be retrieved and emails to the distributor were not answered. Because of this, the requirements of the data quality and which dimensions are important will have to be determined by hand. The quality dimensions are based on the view of Juran [5] and are explained in subsection 3.4.3. The data quality team will first have to determine the importance of each dimensions. The data quality team should also asses their perceived level of quality and the expected level of quality of those dimensions.
To gather this data, the Quality Dimensions Score Sheet has been designed (Appendix 8). On this sheet the team members can put in their opinion about the importance of each dimension for the specific product-attribute, their perceived level of quality in each dimension, and their expected level of quality in each dimension. These scores can be filled in during the meeting, on the calculation sheet of the MS Excel document. This part then automatically puts the results in the charts mentioned in subsection 3.4.3. This chart shows the three aspects per dimension (see figure 10 for an example).

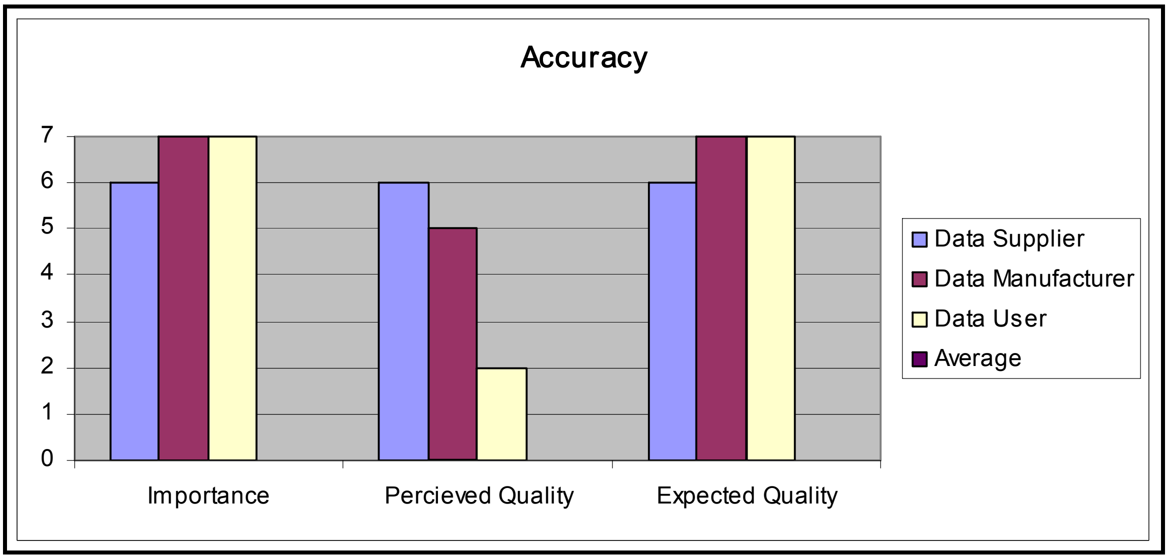
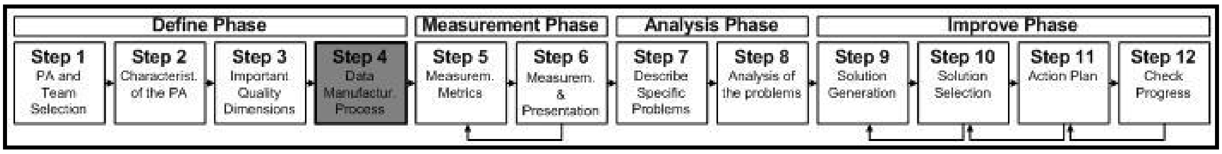


Figure 10: Example of the results of the quality dimension Accuracy, presented in a chart with sample data.

With these charts of each dimension, the team can then determine which dimensions are valuable to measure the quality of in a group discussion. As explained in subsection 3.4.3 the dimensions that are valuable, are not only the dimensions that are valued as important, but also dimensions in which large differences exist in perceived quality between suppliers and users of data (see figure 10).

Once the team has chosen the important dimensions based on the analysis in this phase, the team members should think of a quality indicator in each dimension. The team should then discuss about these indicators and choose one which is gives the best information.

4.2.4 Step 4: Data manufacturing process



The goal of this step is to define the process in which the data for the product-attribute is being created. As explained in subsection 3.4.4 the basic flowchart modeling technique is used to represent this process.

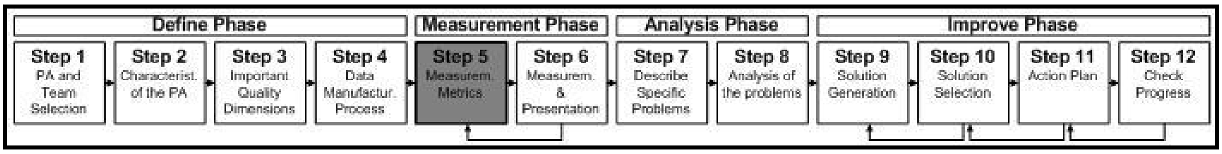
In the team meeting the whiteboard can be used to model the flow from supplier to the customer. After the meeting, a team member can then draw the process with help of MS Visio. A template has been designed which contains forms to model this flow, based on forms in figure 8 (chapter 3).

This step concludes the define phase of the method. The product-attribute to handle is chosen, the characteristics are defined, the important quality dimensions have been established, and the process in which the data for the product-attribute is being created is also defined.
To continue the improvement process the next phase of the method is the actual measurement of the quality.

4.3 Measurement phase

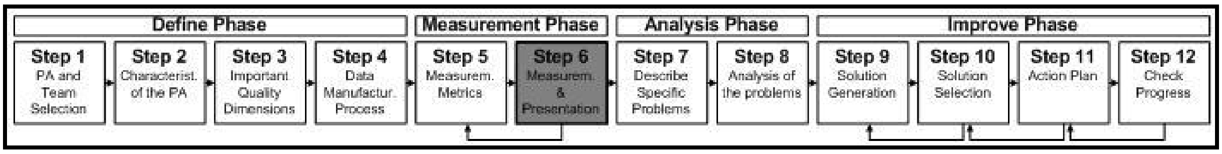
The measurement phase consists of two steps according to the theory described in the previous chapter. The first step is to find a way to measure the dimensions and their quality indicators that were selected in step 3 (4.4.1). And the second step is the actual quality measurement (4.4.2).

4.3.1 Step 5: Measurement Metrics



In step 5 the measurement metrics have to be defined. Section 3.5 gives several examples on how to do this. Section 3.5 also presents RUMBA criteria [14] to value the metrics on. The team should decide together how to measure each dimension. The database administrator is a valuable member in this step, because he knows how to get the required data and because he also knows what can be measured, and what can not be measured.

4.3.2 Step 6: Measurement



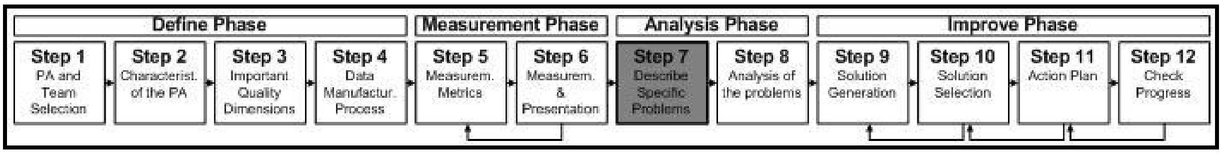
Once the team has agreed on the metrics to measure each dimension on, the second step, the actual measurement, can be conducted. Since this cannot be done during the meeting the team members should plan a new meeting at the end of step 5. In between these meetings the actual measurement can be conducted. Dependent on the workload of this measurement the team can either divide the work of the actual measurement or leave it to the permanent team members. The database administrator should provide the team with the necessary sample products. As explained in section 3.5 a sample size of 400 products is enough in a database of 20000 products. The members should then each measure the agrees quality dimensions and report back the results before the next meeting. The team leader should keep track of this process. The data quality team engineer can then summarize these results in, for example, a Pareto chart (see figure 8 in chapter 3). The results of the measurement conclude the measurement phase.

4.4 Analysis phase

The TDQM method of Wang does not give much guidance on how the fill in the analysis and the improvement phase. Therefore the cause mapping method has been chosen to give guidance to these phases. Because the measurement phase cannot be fulfilled in the first total team meeting, this phase has to start with a new meeting. Therefore the first thing to do is to reflect on the results of the previous meeting. Between the meetings the team members had time to think about the filled in documents and they might have valuable additions.

The next step in this meeting and of the first step of the cause mapping method is to define the actual problem with the quality (4.5.1). The second step is to analyze this problem by drawing the cause map (4.5.2). The next section continues with the other steps of the cause mapping method (solution generation, solution selection, and making an action plan).

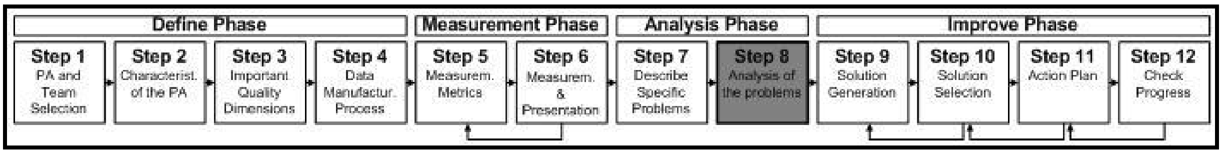
4.4.1 Step 7: Describe the specific problems



Based on the results of the measurement in step 6, the team has to describe the specific problems that follow from these results. There are several questions that need to be answered in this section according to the analysis in chapter 3. The Standard Problem Description Form has been designed to be able to document the answers of these questions (Appendix 9). The fields of this form therefore follow the questions stated in section 3.6.

- The name of the product-attribute
- A field where the problem statement can be filled in
- A field where the question ‘what is the problem?’ can be answered in
- A field where the question ‘where does it occur?’ can be answered in
- A field where the question ‘when does it occur?’ can be answered in
- A field where the impacts to the product-accuracy can be filled in
- A field where the impacts to the users/customers of the data can be filled in
- A field where the need for additional data can be filled in

4.4.2 Step 8: Analysis of the problems

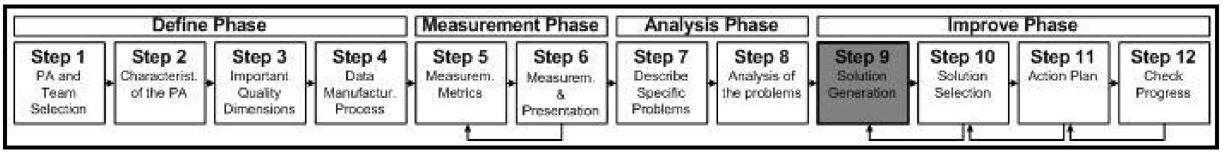


After the description of the problem the team can start with finding the cause of the problem. A cause map is drawn to find the root causes of the problem, as explained in chapter 3. The team engineer should start this step by writing the problem on the far right of the whiteboard (or in Visio with the cause and effect diagram template, or in Excel using the ThinkReliability cause map template) and asking the team members the question: Why?
The answers of the team should then be written down and sorted into categories as shown in Appendix 14a. The result of this step is then a complete cause map.

4.5 Improvement

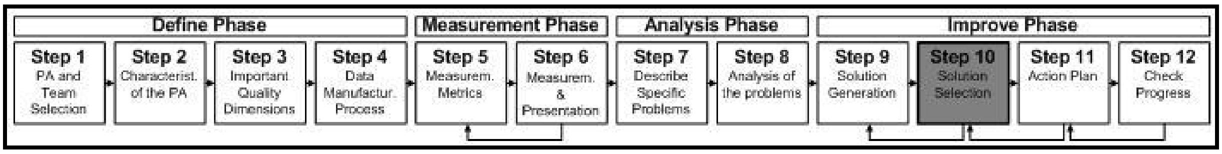
In the improvement phase, as explained in the previous section, Wang does also give no guidance. Therefore the cause mapping method continues. Based on the cause map that was drawn in the previous set, solutions are to be generated (4.6.1), and a solutions has to be selected (4.6.2). Once a solution has been selected the team can make a action plan (4.6.3). To make sure all these actions are executed a new meeting can be planned to check the progress of the action plan or to discus new problems.

4.5.1 Step 9: Solution Generation



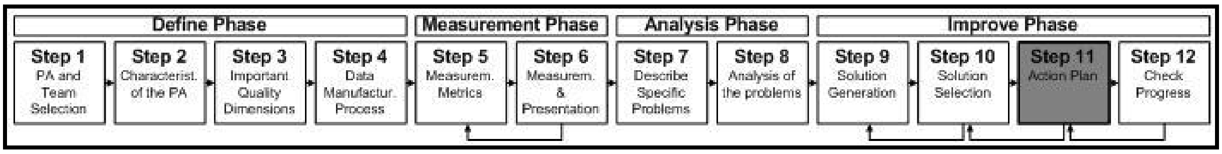
In this step the team has to think of solutions to as much causes of the cause map as possible. The team engineer can put these solutions on the whiteboard next to the causes (or add them to the Visio drawing, or Excel sheet).
This step is meant as a brainstorm session. Therefore as much solutions as possible should be thought of. The next step sorts the solutions and select one (or several).

4.5.2 Step 10: Solution Selection



Out of the solutions that were identified in the previous step, one or several solutions have to be selected. Usually different solutions are connected to each other and can supplement each other. So the first thing to do in the team is to group solutions that can work together.
If a selection has to be made between several (groups of) solutions the priority matrix can be used. To simplify this process, the Solution Score Sheet has been designed (Appendix 10). The team members can individually score the solutions based on the criteria that the cause mapping method has identified. These criteria can also be valued for importance The sheet can then calculate the best solutions based on the scores of each individual member.

4.5.3 Step 11: Action Plan



Once the best solutions has been calculated or selected, the team can decide on the next steps that have to be taken. To make sure all the aspects of the action plan mentioned in section 3.7 are covered in this step, a Standard Task Form had been designed (Appendix 11).
On this form for each action the following fields have to be filled in:

- The name of the attribute
- A brief description of the action that has to be taken
- A owner of the action has to be assigned
- The status of the action can be filled in
- The due date of the action can be set
- A field where can be stated if the actions is completed or not

As explained in the introductions of this section, a twelfth step can be planned to check on the progress of the action plan. In this meeting also problems with the executions of the actions can be discussed and solved.

4.6 Conclusion

This section summarizes all the steps that were identified in the previous sections. From these steps a timeline has been made. As can be stated from the previous sections there are two main group meetings. This is necessary because certain steps can only be completed if certain actions have been completed first. For example measuring the quality of a sample of products. So based on the twelve steps of the method, the following five different sessions can be identified:

The first session is a meeting with the permanent members: the data quality (DQ) team leader, the DQ team engineer, the manufacturing member, and the financial member. In this meeting the product-attribute are selected and the rest of the team members are chosen and informed of the first meeting. The project information form (See Appendix 6) is also filled out.

In the second meeting step two till step five can be completed (see 4.2.2). This takes an estimated 4 hours to complete. These steps can be completed in one session, because almost no work has to be done between the steps (only preparing the graphs). After this session the define phase has been completed and the first step in the measurement phase has also been completed.

Step 6 has to be completed separately, because it requires knowledge of, and access to the database and also knowledge of MS Excel to complete the measurement (see 4.3). Possible other measurements are needed; these have to be done by other team members. For example checking if contracts that state Country of Origin of bought products have to be checked to make sure the data in Oracle is valid. These contracts are stored by the purchasing department and can only be checked by them. This step completes the measurement phase of the TDQM cycle.

Once the results of the measurement are clear the next session can start. In this session the analysis and the biggest part of the improve phase is handled. This session also takes an estimated 4 hour to complete.

The last step should be planned some time after the start of the implementation to monitor the progress of the implementation. This step can be repeated to make sure the solutions gets implemented.
This leads to a schedule of a cycle of the TDQM method for Honeywell as presented in Appendix 12. Appendix 13 gives a summary of all steps and the standard forms that can be used.

Having analyzed all the phases of the method, a third column can be added to table 5 from the conclusion in chapter 3. With the adoptions made in this chapter the method is made suitable for Honeywell Emmen. Therefore the third research question has been answered by this chapter. The next chapter will test the method within Honeywell Emmen, to see if the method is able to actually improve the quality of the product-attributes.

The TDQM Methodology Wang [22]	Adapted Method based on theory in chapter 3	Adapted Method for Honeywell Emmen
TDQM Quality Team	**TDQM Quality Team**	**Quality Team for Honeywell**
Set up data product quality team: • Team Champion • Team Engineer • Data Suppliers • Data Users • Data Manufacturers • Data Product Managers **By: Management**	Choose permanent team members: • Team Champion • Team Engineer **By: Management**	Choose permanent team members: • Team Leader • Team Engineer • Team Manufacturing member • Team Financial member **By: Management**
TDQM Define Phase	**Adapted Define Phase**	**Define Phase For Honeywell**
	Step 1: Select the product-attribute and select non-permanent team members: • Data Suppliers • Data Users • Data Manufacturers • Data Product Managers **By: Permanent team members**	**Step 1**: Select the product-attribute and select the non-permanent team members: • Data Suppliers • Data Users Tool: Standard Project Information Form **By: Permanent team members**
Step 1: Define Data Product Characteristics: Define on two levels: High level (data product) Lower level (attribute) **By: Complete data product quality team**	**Step 2**: Define Product-attribute Characteristics: One level, functions of product-attributes and stakeholders **By: Complete Data Quality Team**	**Step 2**: Define the Product-attribute Characteristics Tool: Standard Product-attribute Characteristics Form **By: Complete Data Quality Team**
Step 2: Define Information Quality Requirements: Use a tool with quality dimensions from Wang and Strong [25] **By: Complete data product quality team**	**Step 3**: Define Information Quality Requirements: Use the in section 3.4.3 described technique to determine important quality dimensions. **By: Complete Data Quality Team**	**Step 3**: Determine Important Quality Dimensions Tool: Quality Dimensions Score Sheet, and the Calculation Sheet, to determine the important quality dimensions. **By: Complete Data Quality Team**
Step 3: Define Information Manufacturing System: Model with method of Ballou, Wang, Pazer, and Tayi [4] **By: Complete data product quality team**	**Step 4**: Define Information Manufacturing System: Model using the basic flowchart technique **By: Complete Data Quality Team**	**Step 4**: Define Information Manufacturing System: Model the process in which the data is being created by first using the whiteboard and later redraw it using Tool: MS Visio TDQM template. **By: Complete Data Quality Team**

Table 7: Update summary of the development of the method

TDQM Measurement Phase	Adapted Measurement Phase	Measurement Phase For Honeywell
Step 4: Develop Information Quality Metrics **By: Complete data product quality team**	**Step 5**: Determine measurement metrics: Examples on how to do this are given in section 3.5.1. Making a selection between metrics can be done based on Kovac e.a. [14] **By: Complete Data Quality Team**	**Step 5**: Determine measurement metrics: Translate the important quality dimensions into measurable metrics. Examples are given in section 3.5.1. **By: Complete Data Quality Team**
Step 5: Measure Data Product Quality **By: Complete data product quality team**	**Step 6**: Measure Product-attribute quality: Show results in Pareto diagrams [5] **By: Agreed team members and other employees**	**Step 6**: Measure the quality of the Product-attributes and show the results in Pareto diagrams [5]. **By: Agreed team members and other employees**
TDQM Analysis Phase	**Adapted Analysis Phase**	**Analysis Phase For Honeywell**
Step 6: Find root causes **By: Complete data product quality team**	**Step 7**: Describe specific problem: Describe the specific problems based on the results of the measurements. **By: Complete Data Quality Team**	**Step 7**: Describe the specific problem: Tool: Standard Problem Description Form **By: Complete Data Quality Team**
	Step 8: Analysis of the problem: Analysis of the problem by drawing a cause map with the whole quality team **By: Complete Data Quality Team**	**Step 8**: Analysis of the problem: Tool: Whiteboard, or MS Visio. **By: Complete Data Quality Team**

Table 7 (Cont.): Updated summary of the development of the method

TDQM Improvement Phase	Adapted Improvement Phase	Improvement Phase For Honeywell
Step 7: Determine where to invest resources based on [4] and solve the causes **By: Complete data product quality team**	**Step 9**: Generate Solutions: Based on the causes, the team thinks of possible solutions **By: Complete Data Quality Team**	**Step 9**: Generate Solutions: Based on the cause map, think of solutions to every identified cause. Tools: Whiteboard, or MS Visio **By: Complete Data Quality Team**
	Step 10: Solution Selection: Solutions can be scored based on the criteria presented in section 3.7 **By: Complete Data Quality Team**	**Step 10**: Solution Selection: Value solutions on to criteria mentioned in section 3.7. Tool: Solutions Score Sheet and Calculation Sheet, to score each solution on, and to calculates the highest score. **By: Complete Data Quality Team**
	Step 11: Action Plan: Assign specific tasks to team members to implement the selected solution **By: Complete Data Quality Team**	**Step 11**: Action Plan: Assign specific tasks to team members to implement the selected solution. Tool: Standard Task Form to be able to monitor the progress of the implementation **By: Complete Data Quality Team**
	Step 12: Check Progress: Check the progress of the agreed action plan and discuss problems **By: Complete Data Quality Team**	**Step 12**: Check Progress: Check the progress of the agreed action plan and discuss problems. Tool: Update the Standard Task Form **By: Complete Data Quality Team**

Table 7 (Cont.): Updated summary of the development of the method

Chapter 5: Testing the method

The first sections of this chapter describe the choices that were made in testing the method. As explained in chapter two, this testing will be incremental to improve the method. Therefore the next sections each cover the results of a cycle of the method. This means that in each sections the experiences of the cycle are explained and the interviews with the members are evaluated. Also the implications for the method are described. The next sections does the same for another cycle of the method, with another product-attribute. The first section first describes the practical considerations.

Like in the previous chapter, figure 9 is used to indicate which part of the method is being analyzed.

5.1 Practical considerations

In this first section the choices that were made in testing the method are explained. As already stated in the second chapter, the method was first presented to key employees which are planned to become permanent team members. In a meeting with the BI&BIT manager the choice was made to first start the team with the members described below. These permanent members where selected based on the description of the required team members according to Wang [22], and based on the experience they have in the product-attributes. The people present were:

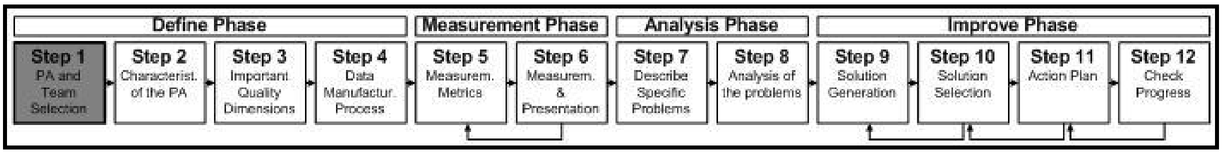
- The planned team leader (Erik Eilering), who is the manager of the product data management department
  His first responsibility is the technical product data, and the adding of new products to the database. Therefore the manager had al lot of knowledge about the products and also the power to implement changes.
- The planned team engineer (Bennie Bruins), who is a business analyst
  The BI&BIT manager sees him as a proactive worker in the data quality area already and therefore feels he should be the person who can organize the meetings and can take care of the progress of the team. Next to that he also has knowledge of the SQL language used to access to the database, and knowledge of MS Excel to be able to analyze the data.
- The financial team member (Roma Porzych), who is the person responsible for costs
  The cost controller is responsible for making new product active and for assigning a cost price to them. She has therefore insight in the important aspects of the products and is aware of possible cost impacts of proposed solutions.
- The BI&BIT manager (Kor Louissen).
  The BI&BIT manager was also present in the presentation. He was there to see how everything was working out and to see the reactions of the team members.

A presentation was given to explain the plan to set up a data quality team, and to explain the designed method that this team can use to improve the data quality. The presentation was also used to explain the intention to adapt the method based on the input after each session, to be able to better suit the whishes of the team, and to be able to better improve the quality.

The first two cycles of the method were conducted, although no formal authorization to the team members was given. This could influence the meetings, by lowering the motivation or because of not feeling responsible for the outcomes. This has to be kept in mind during the analysis of the outcomes for the design of the method, because it could decrease the amount of feedback in the sessions.

5.2 The first TDQM method cycle

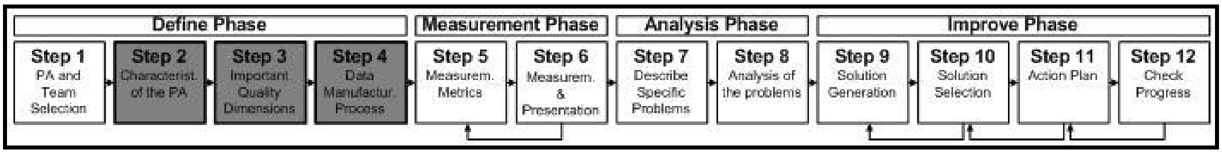
5.2.1 The initiation meeting (Step 1)



The first cycle of the method started with a meeting of the planned permanent team members to execute the first step of the method (the non-permanent members are to be selected in this phase). The people present at the meeting were all the permanent team members, the product data management (PDM) manager, the business analyst, and the cost controller. Because of the previous work that was done, and because of recent problems in the *Country of Origin* product-attribute, the permanent team members selected the *Country of Origin* product-attribute first. This meant that team members who are involved with the product-attribute had to be invited. The team decided that the purchasing clerk would be a valuable member to invite to the team, since she collects the data for the product-attribute from the suppliers. Therefore the role of the data supplier has been filled in. Also a employee from the customer logistics department was chosen, since this department handles the transport of the products to customers in foreign countries and the department also handles the issues with customs. The selected employee is aware of the regulations and requirements which the product-attribute has to meet. So this person can represent the user of the product-attribute. The first whole team meeting was planned one week later, to ensure that all people could be present.

At the end of this first meeting the permanent members discussed shortly about the whole method. The overall opinion was positive, although the team members all thought the method might be too extensive and detailed for some product-attributes. On the other hand they also acknowledged the need for a broad approach, because of the fact that the 150 product-attributes are very different.

5.2.2 The first team meeting (Step 2- 4)



The first meeting had to start without the team leader, because he had to give priority to other tasks. Also the finance member could not come, because of months end closing and because of another task that was due the same day. They both declined the meeting with the same reason; there was no formal authorization, so they had to give priority to other tasks. Also when formal authorization is granted, the team leader should keep in mind not to plan long meetings in months end closing weeks, because the financial team member will not be able to attend the meeting then.
Together with the BI&BIT manager the decision was taken to continue with the meeting as planned, because the other members did invest time in their planning.
So the meeting started with:

- The team engineer and manufacturer,
- A data supplier, from the purchasing department
- A data user, from the customer logistics department

I myself acted as the team leader, together with the team engineer, to make sure the steps where followed correctly. We prepared the meeting by installing a beamer in the room and by printing the empty forms of the steps that we would go through in the first meeting. We started with the second step of the method, the definition of the characteristics of the product attributes. But before this, we asked the members to fill out the first part of the evaluation survey, to establish the knowledge of the product-attribute before the first session.

In filling in the description of the components of the products-attributes in the Standard Product-attribute Characteristics Sheet (Appendix 7), the team directly concluded that space was needed to type in the locations of these components. So that everybody would know where to find the *Country of Origin* components. This is the first adoption to the method that was made directly during the meeting. The other parts of the Standard Product-attribute Characteristics Form (Appendix 7) could be filled out without problems. This second step took about 30 minutes, this was also predicted. However, during the filling in of the form, discussions started about certain components of the product-attribute and problems with these components. This caused the step to take more time then required. But for now, the 30 minutes will remain the estimate.

In handling the third step of the method the following problem occurred. Since the discussions in the previous step revealed a lot of opinions about the perceived and also of the expected quality, filling in the quality dimension score sheet would have been influenced. Because of this, the team decided not to fill out the sheets, but instead to analyze each of the quality dimensions together and discuss which dimensions are important for the product-attributes. The dimensions that were chosen as important for the quality of the *Country of Origin* attribute were:

- Accuracy: the data in the database has to be correct; also the contracts and declarations must be correctly filled in.
- Completeness: the data should be entered in 6 different fields, because of the layers in the database. This can not be changed for now. Therefore the employee filling in the data should fill out the data on all six fields.
- Consistency: Also on this dimension, the employee has to make sure that the data is filled in the same in each field.
- Interpretability: Because of previous mistakes in filling in the data (EC which stands for Ecuador instead of EU for origin Europe), it has to be clear what every abbreviation means.
- Timeliness: Users of the data have to know for sure that the data is up-to-date.

The step took about 45 minutes in stead of the predicted one hour, because the team took the above describe approach in deciding which dimensions were important, instead of the approach using the Quality dimensions score sheet (Appendix 8).

The next step that was taken was “The description of the data manufacturing system”. With help from the beamer and MS Visio, the team engineer drew the cause map based on directions from the other team members. The process of drawing this map revealed many new things for the team members. The decision was made to draw two manufacturing processes, because of the difference in ‘make’ and ‘buy’ products. The data for each process originates from different departments, and also different data is needed. The outcome of the process is shown in Figure 11.

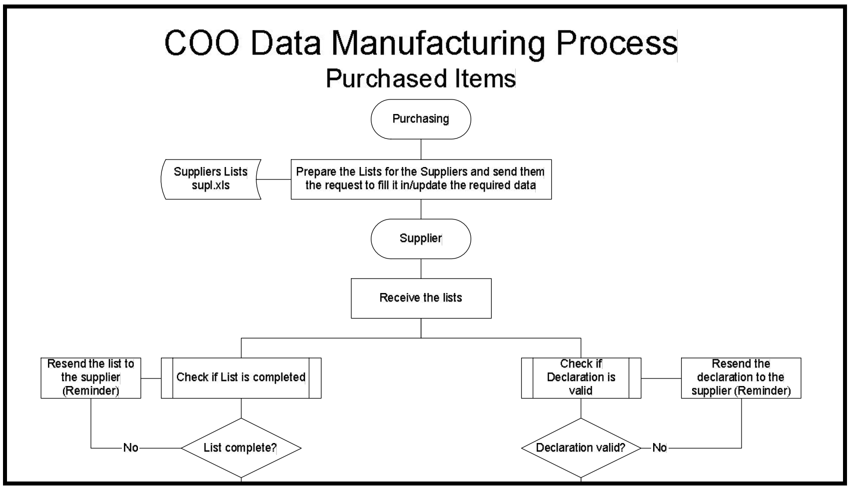
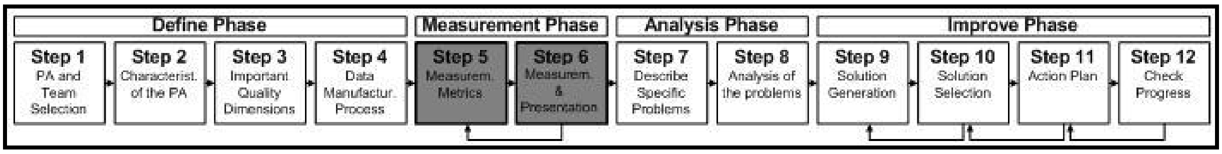


Figure 11: Part of the description of the Country of Origin product-attribute (See Appendix 15)

5.2.3 The measurement session (Step 5- 6)



After the fourth step there was no time left for the fifth step, determining the metrics. The team decided to let the team leader and team engineer decide on which metrics to uses and to let them conduct the measurement prior to the next meeting. The non-permanent team members left the meeting and the team leader and team engineer agreed to meet the next day to work on the next steps. The team leader communicated the documents that were filled in during the meeting to all the team members. Team members could review the documents and prepare feedback for the next session.

The measurement phase was completed by the team engineer and myself. Together we calculated the completeness, consistency, and accuracy dimensions. The dimensions were measured on a random sample of 400 products from the database. The completeness was measure by calculating the ratio of missing *Country of Origin* fields to the total number of *Country of Origin* fields. For each product, there are six *Country of Origin* fields on three different levels of databases. As explained in section 1.6, Honeywell has products in the following three database levels:

- 567, which is the database of Honeywell Emmen
- NL1, which is the master database level for Honeywell Europe
- NLV, which is the validation database.

These levels have a *Country of Origin* field on flexfield and on category (see section 2.1). Figure 12 shows the results of the measurement of the completeness dimension. The results were split up into the different product-categories to increase analyzability of the results.
The consistency dimension was measured across the six different *County of Origin* fields, by using the simple ratio. For a product, all values in the fields must be the same in each field.
The accuracy dimensions was measured by also using a simple ratio. Accuracy was determined by analyzing *Country of Origin* declarations.

For the interpretability we updated the information in the database, so that it was clear which abbreviations belonged to which country based on ISO norms. We also concluded that timeliness could not be measured, since the dates of when the data is entered are not stored in the database. The results of the complete measurement are shown in Appendix 16.

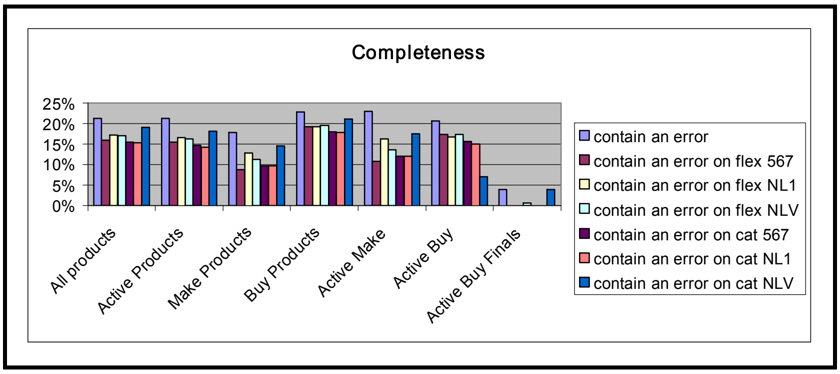
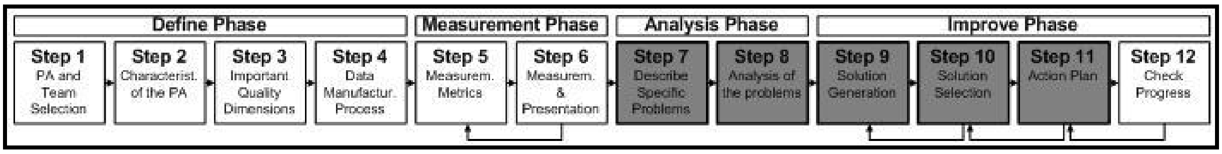
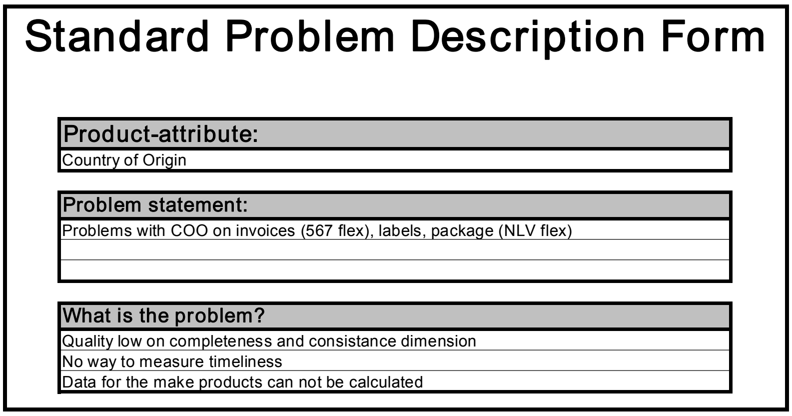


Figure 12: One of the results of the quality measurement (see Appendix 16)

5.2.4 The second team meeting (Step 7- 11)



In the second whole team meeting again the team leader was not present. Also the purchasing clerk was not available. Both were on vacation. The meeting started with the team engineer, the financial team member, the customer logistics employee, and the BI&BIT manager.
The first thing that was done, was looking back at the previous session and shortly discuss the outcomes of that session. Some minor adaptations were made to the data manufacturing systems to improve its clarity. The other documents were all approved.
Then the team engineer continued by presenting the measurement results (Figure 12). The team then discussed about the results and together with the engineer the Standard Problem Description Form was filled out (Figure 13).


Standard Problem Description Form

Product-attribute:
Country of Origin

Problem statement:
Problems with COO on invoices (567 flex), labels, package (NLV flex)

What is the problem?
Quality low on completeness and consistance dimension
No way to measure timeliness
Data for the make products can not be calculated

Figure 13: Part of the filled out Standard Description Form (See Appendix 17)

Three problems were identified:

- Low quality on the completeness and consistency dimension
- No way to measure timeliness and to see of the product-attribute is still up-to-date
- The Country of Origin of the Make products cannot be calculated

In the analysis of the problem that followed, the team decided not to draw a complete cause map. Because in the analysis the team quickly agreed on the main cause: the purchasing clerk, who is responsible for the *Country of Origin* product-attribute, did not now exactly what to do, and therefore mistakes were made and the product-attribute was not maintained. The team agreed that the description of the manufacturing process of the product-attribute provided all the necessary data for this cause to be eliminated.

To be able to see if a Country of Origin in Oracle is still up-to-data, two solutions were identified. The customer logistics employee asked if it was possible to show a 'valid until' data in the Oracle database fields. The team engineer (also database administrator) directly abandoned this idea, because it required Oracle adaptations which would have impact for other Honeywell companies (since they use the same database). This adaptation process would take too much time to get approved and also too much money, according to the team engineer. Therefore this solution was dropped.
The other solution was thought of by the purchasing clerk and was emailed before she went on vacation. She proposed to design a MS Excel sheet in which, for each product in the database, the *Country of Origin* data was entered, and a 'valid until' column was added. Making this sheet available to the involved stakeholders would enable the these stakeholders to check until when the *Country of*

Origin data is valid. The team members agreed that this was the best solution without further discussion.

To be able to assure that the data for the 'make' products in the database is correct, another solutions had to be found. It is important for customers, to be able to declare that a product has a European origin. Declaring this gives a product the so called 'preferential status', which lowers duty costs for customers. To declare this 'preferential status' the companies need to be able to show, that the product contains about 70% added value from within the European Union (EU). Either by materials from the EU or by labor from the EU. To determine if a product has an added value of 70% cost calculations will have to be made with, as a basis, the *Country of Origin* data of the parts, components and materials that are used in the produced item.
Two solutions to be able to calculate this data were thought of. The first one involved using a specially designed software tool from Honeywell in Germany. But because of the costs to purchase and rebuild this tool, the team decided to first try and solve the problem without this tool. So the second solution was formulated.
The team engineer proposed to try to develop a report from Oracle. This report should be able to produce a list of products and its parts and materials, with costs, and their *Country of Origin*. Based on this information and other cost information (labor), the financial team member would be able to prepare a MS Excel sheet, which states the percentages of added value of all the 'make' products. The customer logistics employee could then, based on this document, make sure the 'make' products would get updated in Oracle, and that right declarations would be make and handed out to the customers. Since the team thought this was the only possible solutions to this problem, they decided to implement it.

So step 9, the solution generation, was done together with step 10, the solution selection and the Solutions Score Sheet was not used. The solutions were too easy to analyze and therefore the sheet was not needed. The two steps took about 1 hour, which is half the time predicted for the steps. But because there were only two solutions that were really handled, and because the tool was not used to calculate the best solution, the actual time required for the steps could still be two hours.

To make sure that the solutions were implemented, the team filled in the Project Action Plan, as also prescribed by step 11 in the method. To make sure that the data that is already in the system also got updated, the team agreed on the action plan shown in Figure 14. The first action that had to be taken was setting op an the MS Excel sheet to be able to see the 'valid until' date of the *Country of Origin* data. The other actions the team decided to take are shown in Figure 14.

Product-attribute Action Plan

	Product-attribute:			
	Country Of Origin			
1	Action:	Get the COO data for finished goods buy		
		Update them in Oracle		
		Make the overview sheet on which valid untill datas are shown		
	Owner:		Due Date:	nvt
	Status:	Compeleted	Completed?	Yes
2	Action:	Get the COO data for material and parts		
		Make a list for R. Moleveld. With products & their COO to update in Oracle		
		Update the overview sheet		

Figure 14: Part of the filled out Product-attribute Action Plan (See Appendix 18)

5.2.5 Analysis

Based on the description of the method in table 7 in the previous chapter the following table shows the steps that were defined based on the theory and the situation in Honeywell and the actual used steps and tools in the first cycle of the method.

Adapted Method for Honeywell Emmen	Steps taken in the first Cycle
Define phase	
Step1: Select the product-attribute and select non-permanent team members	Step 1: *Country of Origin* attribute was selected and next to the permanent team members two other members were selected: a data supplier and a data user.
Step 2: Define product-attribute Characteristics	Step 2: Characteristics were described and the location of the data for the product-attributes was added to the form. Not all team members were present at the meeting.
Step 3: Determine Important Quality Dimensions	Step 3: The tools were not used. The important dimensions were determined by group discussion
Step 4: Define Information Manufacturing System	Step 4: MS Visio was used to model the information manufacturing system of the product-attribute
Measurement Phase	
Step 5: Determine Measurement Metrics	Step 5: Was not done by the whole team. The team engineer and leader determined the metrics, because of limited time of the team members in the first session
Step 6: Measure the quality of the Product-attribute	Step 6: Quality was measured by the team leader and the results were presented in Pareto Charts
Analysis Phase	
Step 7: Describe specific problem	Step 7: The specific problem was described on the designed form. Not all team members could be present at the meeting.
Step 8: Analysis of the problem	Step 8: A cause map was not drawn, because the team agreed on the main cause directly and because of limited time in the meeting.
Improvement Phase	
Step 9: Generate Solutions	Step 9: This step was done by the present team members. For each manufacturing process two solutions were thought of and the best one was directly selected.
Step 10: Solutions Selection	Step 10: This step was skipped, since the solutions were already selected in the previous step.
Step 11: Action Plan	Step 11: The action plan was made and the tasks were assigned to team members.
Step 12: Check Progress	Step 12: Is not yet taken, because no tasks were completed yet.

Table 8: *Country of Origin* Cycle Analysis

From the survey that was taken prior to the first session (only part one) and after the whole cycle (Appendix 19) can be concluded that the general knowledge of the *Country of Origin* product-attribute has improved.

The third step of determining the important quality dimensions caused some difficulties. During the meeting the decision was made not to use the Score Sheets, based on the influenced opinions during the discussion in the second step.
The survey revealed that it was difficult to connect some dimensions to the product-attributes. So therefore a better description of the dimension has been added to the sheets.

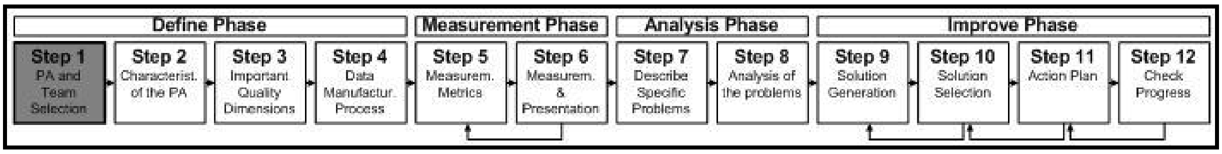
The seven point scoring scale was also discussed. During the session the team members all indicated that a seven point scale could be to detailed. The members indicated that scoring the dimension on a three point scale would be sufficient.

- Low to indicate not important/low perceived quality/low expected quality
- Medium to indicate doubt about the dimension being of low or high importance/perceived quality/expected quality.
- High to indicate high importance/high perceived quality/high expected quality

It can be stated that the main problem that occurred during the meeting was the required time for the sessions. First cause of this is that there was no formal approval of the project yet, but besides that the meetings were hard to plan and not everybody could always be present. In a meeting with the team engineer and the BIT&BI manager after the first session the decision was taken to also make a shorter version of the method.

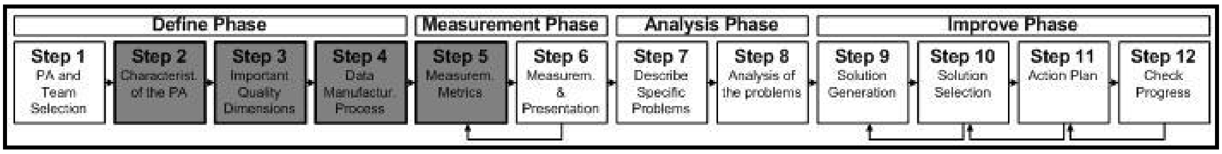
5.3 The second TDQM method cycle

5.3.1 The initiation meeting (Step 1)



In the second cycle of the method the first meeting (step 1) had to be planned in the holyday season. Therefore the team financial member was the only available permanent member. The present members were therefore the financial team member, the BIT&BI manager, and me. The other permanent members were all on vacation. The decision was therefore taken to handle a less complicated product-attribute. A product-attribute with only one stakeholder. This was done in light of the conclusion that the method took a lot of time in the first cycle. Too much time according to the opinions of the non-permanent members. The cycle was therefore conducted to see if the method would be able to handle a shorter version which of the steps could be shortened The selected product-attribute is the *Product Family Code* (*PFC code*). This attribute was selected, because of its importance in sales analysis and importance in decisions about stopping production. So the selection of this product-attribute is not based on past problems, but on possible future problems. Because both, the supplier, and the user of the data, is the financial team member, no other team members had to be invited.

5.3.2 The first team meeting (Step 2- 5)



Although only the permanent team members were invited to the quality team, the first meeting had to start without several members again. The planned team leader still had no formal approval to join the team. Although top management had put in effort to get the approval, the matter was still pending. The main reason for this delay was the fact that the proposal was made during the holiday period and Honeywell operated at minimum capacity. Next to the team leader, also the team engineer could not be present. Because he was on a holyday. So the meeting started with only two people, the financial team member and me in the role of team leader/team engineer.

The second step of the method, the description of the product-attribute characteristics cause no problems. For the third step in the method a simplified Quality Dimensions Score Sheet was designed, based on the usage of the tool in the previous cycle (Appendix 20). This score sheet has, next to the more simple approach, two changes compared to the more extensive sheet.

The first change is the limitation to assessing only the importance of the quality dimension. The choice for this approach is based on the fact that for the less complicated product-attributes, the supplier and users of the data are often the same. If the quality dimensions are determined based on the original score sheet, then the perceived of the supplier and actual experienced level of quality of the user, are determined by one and the same person. The employee who is the supplier of the data, and also the user of the data. This will of course not lead to more insight about the importance of each quality dimension.
The second change is the change in score scale. In the original score sheet all aspects could be scored on a seven points scale. The analysis of the previous cycle revealed that a three point scale could also represent the opinion of the quality team. The result of these changes is shown in Appendix 20.

Usages of the simplified quality dimensions score sheet revealed that it is easy to use and that the time required is a lot less than the predicted 60 minutes. It took about 10 minutes to go through all the dimension. Although the remark has to be made that only two people worked on this step. Working with more people could cause for more discussion and therefore probably causes the step to take about 15 minutes.

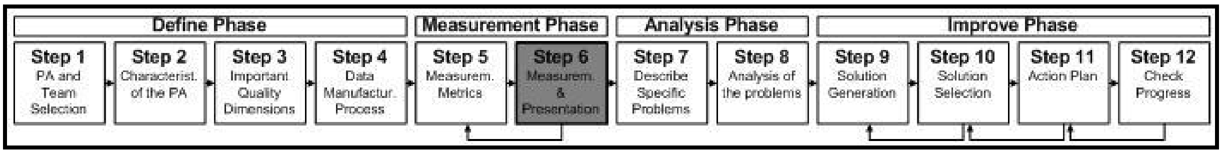
In the fourth step the data manufacturing process is to be defined. The *PFC code* is determined by the financial member when a new product is introduced to the database. She assigns a code to the product based on the category the product belongs to. These product family categories are predefined and can not be changed. The *PFC code* is used in several sales reports which she uses to analyze cost of sales with. Because of the fact that the process is this simple and the fact that only one employee is involved in this product-attribute it is not necessary to draw the data manufacturing system.

In step five, the team agreed to measure the following metrics of the important dimensions that were determined. The quality is analyzed by measuring data from a random sample of 400 products (as explained in section 3.5.2)

- Accuracy: Will be analyzed by comparing product descriptions with the descriptions mentioned by the *PFC Code*. For example product VK4100A1002B. Its product descriptions states that it is a CVI Gas Control. The *PFC Code* states that the product belongs to the C.V.I. category. Therefore the data for this product is accurate. When there are doubts about products and their *PFC Code*, the financial team member assists in determining the correct *PFC Code*.
- Relevancy: Relevancy in this product-attribute is difficult to measure. All *PFC Code* category are used in sales analyses. So all categories are relevant. The financial member did indicate that the categories below *PFC Code* 40 are parts of the other *PFC Codes*. In analyzing sales they are of less importance, but in analyzing cost of sales they are of importance. The main category of products that Honeywell Emmen produces is category 43: The C.V.I. Also the sensors and hydraulics for the USA, Canada, and Japan are important, because they are also produced in Emmen. The other PFC Codes are mostly products which are bought for other Honeywell companies and then sold to customers.
- Value-added: For this dimension the same goes as for the relevancy dimension. The *PFC Code* adds value because of the fact that it is used to analyze sales.
- Completeness: The completeness dimension can be measured by analyzing if all *PFC Codes* are filled in, in the sample of 400 products. All products should have a *PFC Code* since else they will not be accounted for in the sales analyses.
- Interpretability: The importance of this dimension is high, because it is important to have a clear understanding of which category represents which products. The financial member indicated that all the descriptions of the PFC Code categories are clearly described. She also indicated that these categories were defined by Honeywell global, so they can not be changed.
- Consistent representation: Consistency in this dimension means if all the PFC Codes are represented in the predefined 2 decimal category numbers. This is checked by testing if the sample products have one of the predefined category assigned to them.

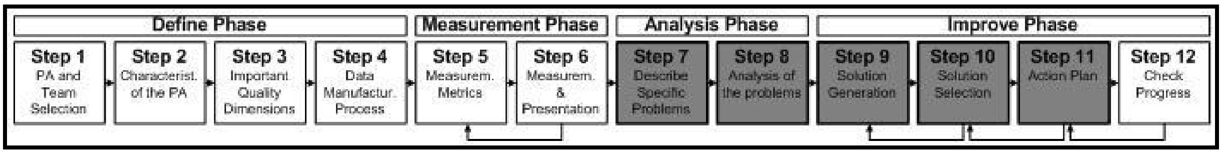
So the dimensions that are to be measured are, Completeness, Consistency, and Accuracy. These dimensions are measured in a random sample of 400 products out of the database.

5.3.3 The measurement session (Step 6)



The measurement was conducted based on the three defined measurement metrics from the previous subsection. Using MS Excel the results are shown in Pareto diagrams. From these results can be concluded that on the completeness dimension only five products were found without a *PFC Code*. This is about 1 percent for the total sample size. In the consistency dimension the same five products give an error, because no data is not a valid format. No errors were found in the remaining 395 products that were checked. So one could say that the consistency is also nearly at 100 percent.

5.3.4 The second team meeting (Step 7- 11)



Because the measurement session did not discover problems in the measured dimensions of the product-attribute, the second meeting was not planned.

5.3.5 Analysis

Analyzing this relative simple product-attribute revealed that the analysis and improvement phases are sometimes not necessarily. This can in fact also be the case when a more complex product-attribute is handled. This all depends on the results of the measurement. Are the results sufficient according to the team? Then they can decide not to continue with the last two phases, because no quality problems exists.

Also the description of the data manufacturing process is exaggerated in this case. Assigning a *PFC Code* to a product does however require knowledge of the products. This can only be attained by building up experience with the products.
The employee responsible for the *PFC Code* attribute assigns a *PFC Code* to the product when a new product is added to the database. Sometimes a production managers is called to check if the category is correct.

Adapted Method for Honeywell Emmen	Steps taken in the second Cycle
Define phase	
Step1: Select the product-attribute and select non-permanent team members	Step 1: *PFC Code* attribute was selected based on its importance in sales and production decisions. Only the permanent team members have to be invited since the financial team member is the data supplier and also the data user.
Step 2: Define product-attribute Characteristics	Step 2: Characteristics were described using the characteristics form. Not all team members were present at the meeting, because of the fact that it was the holyday season.
Step 3: Determine Important Quality Dimensions	Step 3: The important dimensions were determined by filling in the short version of the quality dimension score sheet.
Step 4: Define Information Manufacturing System	Step 4: The data manufacturing process was not described, based on the fact that the process was simple and only one stakeholder represented both supplier and user of the data.
Measurement Phase	
Step 5: Determine Measurement Metrics	Step 5: Together with the financial member this step was completed. All the selected important quality dimensions were analyzed and for three dimensions metrics were defined.
Step 6: Measure the quality of the Product-attribute	Step 6: Quality was measured and the results were presented in Pareto Charts.
Analysis Phase	
Step 7: Describe specific problem	Step 7: From here the cycle stopped, because no problems with the quality could be found. The team therefore decided to store the data created and no further actions were planned.
Step 8: Analysis of the problem	Step 8:
Improvement Phase	
Step 9: Generate Solutions	Step 9:
Step 10: Solutions Selection	Step 10:
Step 11: Action Plan	Step 11:
Step 12: Check Progress	Step 12:

Table 9: *Product Family Code* Cycle Analysis

5.4 Conclusion

Looking back at the three main causes of the problem of low quality, as mentioned in chapter 2, it can be stated that the new method provided a structured way to check and improve the quality of the product-attribute and also a structured way to evaluate the division of the responsibilities of the product-attribute. As can be seen in the results of the evaluation (Appendix 19) the team members indicated that the communication about the problems with the product-attribute also improved and the general opinion of the method was positive. The method provided a clear action plan and the team members directly started with the agreed tasks. The tasks that were completed, were directly communicated to the other team members. So that progress could be reported directly. It can therefore be stated the designed method can improve the quality of the data. The fourth research question has therefore been answered (section 2.2).

Chapter 6: Conclusion and recommendations

This chapter concludes the report by shortly summarizing the report and by stating the recommendations for Honeywell Emmen. The third section explains possibilities for further research that where found in the process of making this report.

6.1 Summary of the report

This report has provided an implementation of an adapted version of Wang's Total Data Quality Management (TDQM) methodology within Honeywell Emmen. The goal of this report was to develop a method, which Honeywell Emmen could use to improve the quality of their product-attributes. This report ensures practical relevancy by following the design science research approach described in the theory of Van Aken [1]. The research method that was used therefore followed the conclusions of the theory of Van Aken. First "Organizational Theory" was created by adapting the TDQM theory to be able to handle a product-attribute. Then "Management Theory" was created by further adapting the method for Honeywell Emmen. The designed method was then tested within Honeywell Emmen. The evaluations of these tests were used to improve the method.

Main changes that were made to the original TDQM method are:

- <u>The data quality team</u>. This team forms the basis of the quality improvement effort. Therefore a more detailed description of the possible forms of this team is given. The use of the TDQM method to improve product-attributes in stead of a data product has lead to a team, which is able to change its members, because of the different data suppliers and users the different product-attributes have.
- <u>Replacement of the software tool developed by Wang</u>, with a method to determine important quality dimensions by hand. Not being dependant on this software tool provides more flexibility to the method, and gives companies for which the software tool is not available also the possibility to use the method. The tools used in this report are all developed in MS Excel and can be obtained by contacting the author of this paper.
- <u>The cause mapping method</u>, to give more guidance in the analysis and improvement phases. The TDQM method proposed by Wang [22] leaves this option open to companies. The cause mapping method was therefore selected to be able to guide the analysis an improvement process.

Looking back at the problem statement and the research questions formulated in chapter 2 the following conclusions can be stated.

Research question 1:

What should be the definition of product-attribute data quality for Honeywell?

Based on the TDQM methodology, product-attribute data quality should be analyzed by the eyes of the users of the data. To assure involvement of product-attribute users, the data quality team contains product-attribute users. They can therefore actively influence they importance of certain quality dimensions in the product-attribute they work with.

Research question 2:

What methods does the literature provide, to be able to increase the quality of the product-attributes?

The method that literature provides is the Total Data Quality Management methodology. This methodology has been explained and supplemented in chapter three of this report.

Research question 3:

What method, or adapted method, can Honeywell Emmen use to increase the quality level of the product-attributes to meet the desired level of the users?

Chapter four described how the Total Data Quality Management methodology was developed into a method that could be implemented within Honeywell Emmen.

Research question 4:

How should Honeywell Emmen assure the method does improve the level of quality in the product-attributes?

Practical relevancy of the method has been assured by testing the designed method. The results of the first test revealed that the method did improve the quality of the product-attribute *Country of Origin*.

Problem definition:

What structured method should Honeywell Emmen implement to assure the quality of the product-attributes is increased to the level the users of the product-attributes require?

The adaptations to the TDQM methodology made the method suitable for Honeywell Emmen. The method assures involvement of data users, to assure quality standards are conform the wishes of these data users. Tests have proved the increase of data quality.

The use of the TDQM methodology within Honeywell Emmen has taught us that the methodology can be used to describe and improve complete data products, but can also be used for describing and improving the underlying product-attributes. The additions made to the methodology ensure that extra guidance is provided in the case the method is used on a product-attribute. The method has been tested on a complex product-attribute with multiple stakeholders from different departments, and also on a simpler product-attribute with only one stakeholder. This revealed that the method has the flexibility to handle both cases.

With this, the remark has to be made that the author of this report played a role in both improvement cycles. Because the cycles of the method were conducted without formal authorization, not all planned team members could be present during the meetings. The author of this report represented these missing team members. Since most of the work has been conducted by the team members themselves, the future cycles of the method will not be influenced.
Formal approval has been given to the proposed team and its members with the finalisation of this report. Therefore the team members will, next to their main tasks, be evaluated based on their presence in the team meetings.

To make sure the method can be valuable to other companies, the method should first be tested outside Honeywell Emmen, but we can however conclude that the proposed data quality improvement team and the designed method answer the problem definition stated in section 2.3 for Honeywell Emmen. The method can help Honeywell Emmen improve their product-attribute data quality in the eyes of the users of the product-attributes.

6.2 Recommendations

Based on the theory of Total Data Quality Management and field tests within Honeywell the following recommendations can be formulated for Honeywell Emmen:

1 Form a data quality improvement team, with:
- A team leader, who is the process owner of the product-attributes within Honeywell Emmen and has the power to implement solutions. This team leader should take the initiative to set up the team meetings and should be responsible for the quality of the quality of the product-attributes.
- A team engineer, who can facilitate the team meetings.

- A data manufacturer, who has extensive knowledge about the database system. The member should be able to extract different reports out of the database, to be able to calculate the quality level in certain quality dimensions.
- A financial team member, who understands financial consequences of proposed solutions.
- A Product-attribute data suppliers, selected on the chosen product-attribute to handle. Someone who usually supplies or enters the data in the database system.
- A Product-attribute data users, selected on the chosen product-attribute to handle. Someone who uses the data in his work, or who has direct contact with external users of the product-attribute data.

2 Introduce the designed method as a tool the data quality improvement team can use.
As explained in this report, the method consists of four phases:

- The define phase
- The measurement phase
- The analyses phase
- The improvement phase

Using a cycle of this method on a product-attribute makes sure all the aspects of the product-attribute are analyzed and improved if necessary.

3 Institutionalize the data quality team and assign data quality responsibilities to the team leader. As explained in section 6.1 Honeywell Emmen did already give formal approval for the data quality team. With this approval Honeywell should also make the team leaders responsible for the data quality of Honeywell Emmen.

4 Facilitate the data quality improvement efforts and make hours available for the team members. The first step in this process has been taken by the institutionalisation of the data quality team. Next to this step, management should make resources available to the team members like:

- MS Excel on the team members personal computers
- MS Visio on the team members personal computers
- Meeting room (with a whiteboard or beamer present), where team members can execute the steps of the method
- Give data suppliers and data users time to participate in the team meetings if a product-attribute is handled which concerns their tasks.

These recommendations are based on the Total Data Quality Management methodology designed by Wang [22] and analyzed in this report. The goal of this method is to improve the quality of data. The method is based on the Total Quality Management principles of, customer satisfaction, continues improvement, teamwork and participation. Using the method should therefore provide a structured way to analyze and improve the quality of the product-attributes. The description of the data manufacturing process provides a way to analyze and improve the division of responsibilities in each product-attribute data manufacturing process. Working in a team with all involved departments can shorten communication chains and therefore increase the quality of communication.

6.3 Implications for further research

This report shows the need for data quality in companies, an often reflected topic [14, 17, 22]. Often action is taken after serious problems have been encountered [14]. Creating more awareness to data quality and its impacts should therefore, be a primary task for researchers in the data quality field.

Within Honeywell Emmen, actions have also been taken, after problems with product-attributes started to effect the customer satisfaction. Implementing the method revealed a lack of time to attend data quality improvement meetings. Although all team members acknowledged the need to improve the data quality, most team members could not get time to spend on the project. So to be able to really improve the data quality situation, senior management support is crucial.

To be able to get management to acknowledge the need to invest in data quality, more research will be needed. This research should analyze the costs and impacts of low quality data [17] and show the positive effects of for example handling the data quality with the TDQM methodology [14]. Because of the growing needs to handle data quality problems, more research should be conducted on implementation of TDQM methods in other companies. These methods should also be thoroughly field tested to be able to be usable in practice [1].

The designed method in this report can also be added to the data quality research. The method broadens the applicability if the TDQM methodology designed by Wang, because it enables companies to improve not only complete data products, but also the product-attributes that make up this data product. The designed method has been field tested within Honeywell Emmen and the results were positive. To however complete the development of a usable "Management Theory", the method has to be β-tested by other companies as well.

To increase the application of the method for Honeywell in general, new research could be set up. The growing awareness of data quality problems and its consequences within Honeywell Europe can contribute to setting up a European research initiative. This research could be conducted in cooperation with the University of Twente, to scientifically test the applicability of the method in other Honeywell companies across Europe. This research would then be an even more valuable contribution to the field of data quality research.

Literature

[1] Aken, Van. J.E., *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, March 2004, pp. 219-246.

[2] Annual report, *Growth at Honeywell*, 2004, Taylor & Ives, Inc.

[3] Ballou, D.P., Tayi, G.K., Methodology for allocating resources for data quality enhancement. Communications of the ACM. vol. 32. no. 3. 1989, pp. 320-329.

[4] Ballau, D.P., Wang, R.Y., Pazer, H., and Tayi, G.K., *Modeling information manufacturing systems to determine information product quality.* 1997, Management Science.

[5] Bij, van der, J.D., Broekhuis, H., Gieskes, J.F.B., Kwaliteitsmanagement in beweging. Van blauwdruk naar contingenties en dynamiek. 1999, Kluwer, Deventer.

[6] Cooper, D.R., Schindler, P.S., Business Research Methods. Eighth Edition. 2003. MrGraw-Hill/Irwin. New-York, America.

[7] Daft, R., *Management,* fifth edition, 2000, The Dryden Press Fort Worth.

[8] Doggett, A.M., *Root cause analysis: A framework for tool selection*. Quality Management Journal. Vol. 12, no. 4, 2005.

[9] Internal document, *Organization Charts Emmen*, Nov. 2005, K. Louissen

[10] Internal document, *De invoer tot en met de uitfasering van producten*, May 2001

[11] Internal document, *Item Define/ Maintenance*. Business Processes, Honeywell Emmen, 2002.

[12] Internal document, *TP-Completion*. Detailed Process and Control Description, Honeywell

[13] Internal presentation, *Item Maintenance Emmen*, Sept. 2002, K. Louissen
Emmen, February 2002

[14] Kovac, R., Lee, Y.W., Pipino, L.L., *Total Data Quality Management: the case of IRI*. In: Proceedings of the Conference on Information Quality, Cambridge, MA, 1997, pp. 63–79.

[15] Moore, D.S. and McCabe, G.P. Introduction to the Practice of Statistics. W.H. Freeman and Company, 1989.

[16] Pipino, L.L., Lee, Y.W., Wang, R.Y., *Data quality assessment*. In: Communications of the ACM, vol. 45, no. 4, 2002, pp. 211-218.

[17] Redman, T. C., *The impact of poor data quality on the typical enterprise*. In: Communications of the ACM, vol. 41, no. 2, 1998, pp. 79-82.

[18] Strong, D.M., Lee, Y.W., Wang, R.Y., *10 Potholes in the Road to Information Quality.* Computer vol. 30, no. 8. 1997, pp. 38-46.

[19] Tayi, G.K., Ballou, D.P., *Examining Data Quality.* In: Communications of the ACM, vol. 41 no. 2, 1998, pp. 54-57.

[20] ThinkReliability Two-Day Cause Mapping Workbook. Effective root cause analysis. Version 7.1, 2006. Novem, Inc.

[21] US department of commerce. *Data Quality management at the United States Patent and Trademark Office*, revised march 1998.

[22] Wang, R.Y., *A product perspective on total data quality management*. In: Communications of the ACM, vol. 41, no. 2, 1998, pp. 58–65.

[23] Wang, R.Y., Kon, H.B., and Madnick, S.E., *Data quality requirements analysis and modeling*. Proceedings of the 9th International Conference on Data Engineering, Vienna, Austria, 1993, pp. 670-677.

[24] Wang, R.Y., Storey, V.C., and Firth, C.P., *A framework for analysis of data quality* research. In: IEEE Trans. Knowledge Data Engineering, vol. 7, no. 4, 1995, pp. 623–640.

[25] Wang, R. Y., and Strong, D. M., *Beyond accuracy: What data quality means to data consumers*, Journal of Management Information Systems, vol. 12, no. 4, 1996, pp. 5-34.

[26] Wijnhoven, F., *Managing Dynamic Organizational Memories*. Instruments for knowledge Management. 1999, Boxwood Press. California, USA.

[27] Wilson, Dell, Anderson, Root Cause Analysis – A tool for Total Quality Management. 1993.

[28] Zahedi, F. Quality Information Systems. 1995. Boyd & Fraser publishing company, Massachusetts.

Web links

[I] http://www.honeywell.com/sites/ honeywell/ourhistory.htm
Our History, Honeywell.com, visited on 6 February 2006

[II] http://www.honeywell.nl/?id=56
Honeywell BV- De geschiedenis, Honeywell.nl, visited on 6 February 2006

[III] http://nl05web/misc/publicaties/257,2
Honeywell Intranet, Company overview, visited on 9 February 2006

[IV] http://www.honeywell.com/careers/meetus.html
Meet Honeywell, Honeywell.com, visited on 6 February 2006

[V] http://www.honeywell.com/sites/honeywell/aboutus.htm
About Us, Honeywell.com, visited on 6 February 2006

[VI] http://nl.wikipedia.org/wiki/Oracle_%28software%29
Oracle (software), Wikipedia.com, visited on 13 February 2006

[VII] http://www.europa.eu.int/comm/taxation_customs/customs/customs_duties/rules_origin/index_en.htm
Europa- Taxation and Customs Union, Europa.eu.int, visited on 14 February 2006

[VIII] http://www.dft.nl/servlet/ArticleByID?ID=2847770&frompage=24uur
Honeywell Emmen naar Hongarije, De Financiële Telegraaf, visited on 6 June 2006

Interviews

[A] OGON session at Thales Hengelo. OGON stands for Oracle users group Oost-Nederland, it is a platform in which Essent Kabelcom, FEI Company, Geesink Norba Group, Thales and Honeywell exchange experiences in the use of Oracle.

[B] Weekly meetings, with Kor Louissen, Business improvement and Business IT manager, product-attribute manager.
Main topic: Product-attributes in general

[C] Interviews, with Marcel Bol, Financial controller, data user.
Main topics: Causes of data quality problems

[D] Interviews, with Janneke Hof, Purchasing Clerk, data supplier.
Main topic: Usage of attributes, Causes of data quality problems

[E] Interviews, with Hans Huy, Customer Logistics Manager, data user.
Main topic: Usage of attributes, Causes of data quality problems

[A] OGON sessions at Thales Hengelo (short summary)

Present companies:

- FEI Company
- Geesink Norba Group
- Thales
- Honeywell

Main topic that was discussed was the connection between Product Data Management (PDM) systems and the Oracle ERP System. How every company handled this connection and how the companies made sure both systems remained up-to-date.
Was the communication one way or both ways, is the connection automated?
The companies did not use the Oracle PDM solution. All of them used a separate system to store and maintain the PDM data. All companies also used a one way communication connection between the systems, form PDM to Oracle.

The main lesson that was learned with respect to this report is that the Oracle ERP System is used in many different companies. Thales produces two customized defense systems a year while for example Honeywell produces over a million similar gas valves a year. Maintaining product-attributes is very different for each of these companies. Also the need for many different product-attributes was confirmed.

[B] Meetings with the Business Improvement & Business IT Manager (summary of important subjects)

8 March 2006:
Meeting about the main problem

What is the main problem Honeywell is facing?
The BI&BIT manager answered that the problems that Honeywell is experiencing now are in the product-attributes. He explained that there were complaints of customers that shipments were delayed because of missing *Country of Origin* and *Duty Tariff Code* attributes. He suspected also problems with Bill of Material
The BI&BIT manager explained that in his vision the main problem was the lack of communication between the departments involved with the product-attributes. This lead to problems with these attributes not being solved.

Who are involved in this problem?
The BI&BIT manager answered that ultimately the customers are effected by the problems. They are the main stakeholders, they need to be kept satisfied. Internally almost every department has a responsibility to fill out or maintain a set of product-attributes. It could be that these responsibilities need to be updated because of the reorganizations in the past and the employee changes over time.

Did these reorganizations have other influences on employees?
The BI&BIT manager answered that they caused employees to stick to their own responsibilities and to avoid risks. He also thought that the changes of functions caused the maintenance of product-attribute data to be undervalued to other tasks.

Other meetings:
Concerned mostly the progress of the project, facilitating questions on how to reach people, organize meetings, evaluating interviews and so on.

[C] Interviews, with a Planner, former Financial controller and , data user. (summary of important subjects)

Main topics: Usage of attributes, Causes of data quality problems

Main questions and answers:

How did the problem with the low quality of the product-attributes came to being?
The planner answered that he thought the problem developed because of not keeping to the responsibility sheets. Employees were not corrected if they had not kept up their data. In some attributes this has eventually lead to not filling in the data at all, because the employee itself did not use the data.

Why wasn't this problems handled before?
The planner answered that there was almost no communication about these problems. In the past Honeywell had a weekly meeting about these issues, but because of changes in employees, not having a clear goal of these meetings, and because no real action were taken after the meetings, these meetings stopped. I was the initiator of these meetings, but after I got transferred to another department I could not plan the meetings anymore.

Who did this team consist of?
The planner answered that there were about ten permanent members. This caused the meetings to be a bit hectic. Some members were not interested, because it did not concern their responsibility.

About the product-attributes:

How is the data for the products enter the system?
The process starts with a change request, or a request for a new product. These request come from marketing managers (from customers), from researchers, or from production managers. New products are added by the product data management department and change request are also handled there. On entering the new product number in Oracle, the other departments get a notification (email) to input or changes the product-attributes they are responsible for.
Some product-attributes are directly set by loading a template. The departments do stay responsible for updating their product-attributes. Sometimes there are also internal changes necessary to the product-attributes. These changes are implemented without further notice.

What product-attributes undergo these changes?
Mainly planner or production attributes, which do not effect the customer.

Where do you think most mistakes are made?
It is hard to say. We've had problems with Country of Origin, Duty Tariff Code, transfer prices and a few other attributes. I think some of these problems are there because a different employee has to enter the data then the employee who uses the data. Supplying employees might not know exactly why the attribute is needed and might also not know how important the attribute is. Because of this, problems can develop. Therefore I think we need to find a way to reintroduce the team meetings.

[D] Interview, with a Purchasing Clerk, data supplier. (summary of important subjects)

Main topic: Usage of attributes, Causes of data quality problems

Can you tell me which attributes suppliers see and sometimes use or supply? In other words: Which attributes are important to suppliers or are important to know of suppliers?
The main attributes that are important to suppliers are the Item Status, Minimum and Maximum order quality. For or own products we need to know the country of origin of the bought products, the same goes for duty tariff code and EAN bar code.

What do you think causes the problems in some attributes?
From my own experience time is the limiting factor. Usually other tasks have priority. I am also responsible of the country of origin attribute, but I don't know exactly where it is used for.

Do you know who to contact if you have questions or problems to fill out the data?
I know you are here to solve the problems with the attributes, so I'd contact you. Or else it would contact Kor Louissen. I also heard that customer logistics has something to do in the process, but I don't know what.

[E] Interviews, with Hans Huy, Customer Logistics Manager, data user. (summary of important subjects)

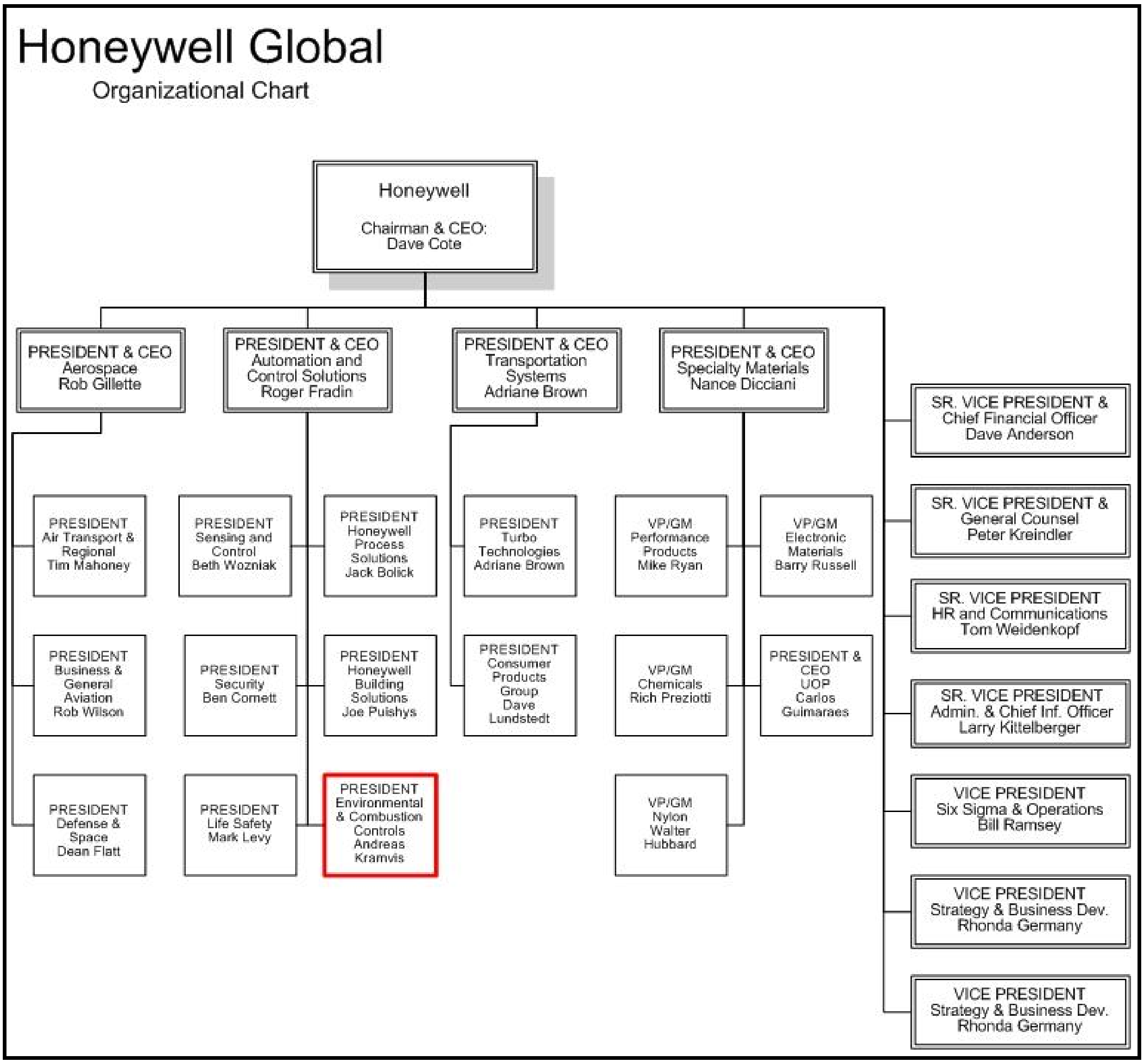
Main topic: Usage of attributes, Causes of data quality problems

Can you tell me which attributes customers see and sometimes use or supply? In other words: Which attributes are important to customers or are important to know of customers?
We have recently received emails from customers who complain that products were held at customs, because of missing country of origin data and missing DTC codes. So these are attributes that customers require to be correct. Also the weight of the products needs to be exact, because of transportation limitations. Order qualities are also important for customers, especially minimum order sizes. We have also had emails about EAN bar codes that contained too many digits.

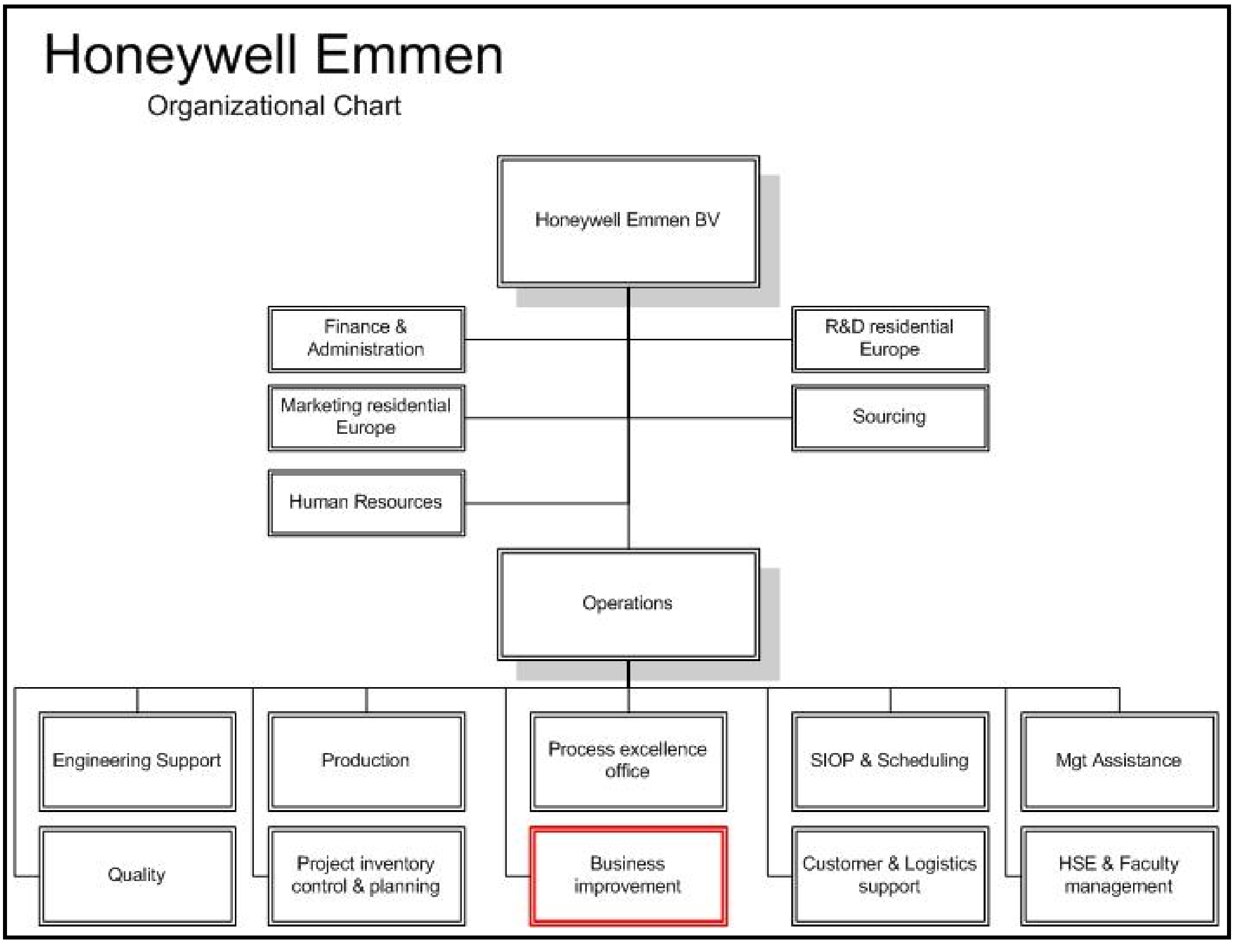
What do you think causes these problems?
For the most part I think, knowledge. Some colleagues do not know why certain attributes are important or do not know how to change them. For some attributes for example 6 different levels in the database have to be changed. If one or two levels are forgotten, the labels that display the data can still be wrong. This is what happened with the country of origin attribute I think. Labels use data from a different place then from where the data is being maintained.
Another problem I think that has to be looked into is the division of the responsibilities. Attributes that they employees use themselves are mostly correct, but when it comes to attributes that have to be supplied for others time is often a problem. Maybe some attributes need to be handled by only one person or a different approach to the problems with the attributes is needed.

Appendices

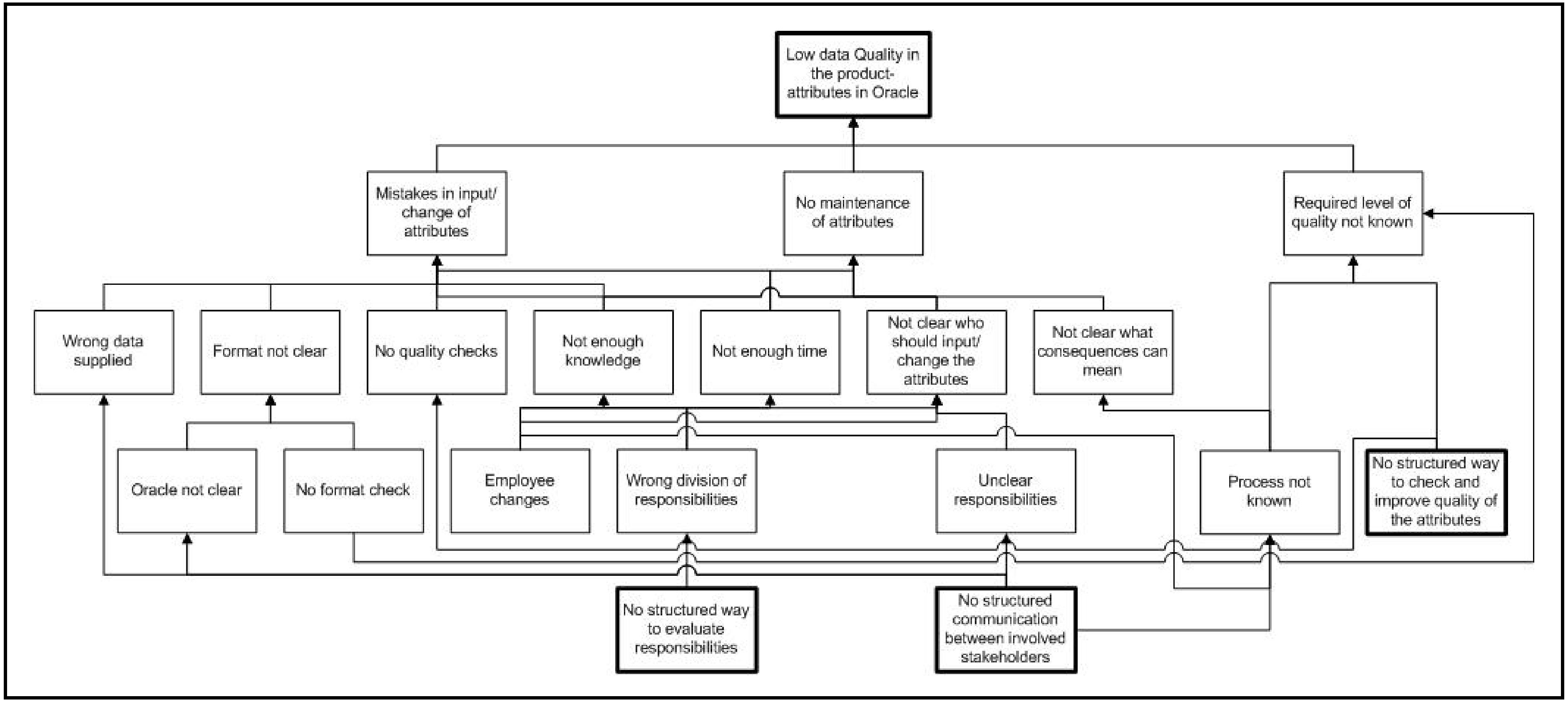
Appendix 1: Organizational chart Honeywell Global [V]



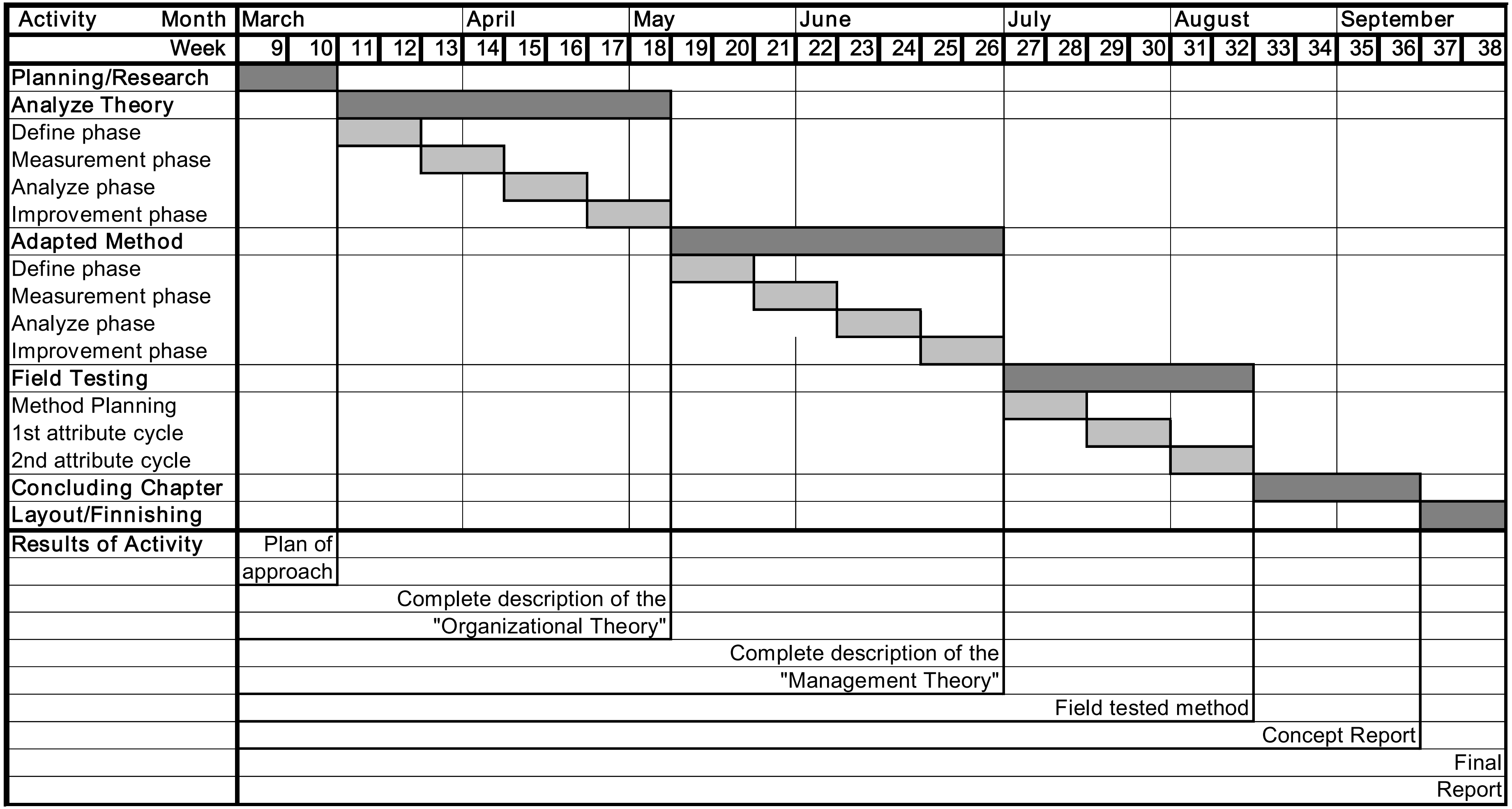
Appendix 2: Organizational chart Honeywell Emmen [9]



Appendix 3: Cause Analysis



Appendix 4: Project planning



Evaluation Survey

This survey is designed to evaluate the new Data Quality Team and the
Total Data Quality Management Method.
The questions are to be scored on a scale from 1 to 7
1 being the lowest score, 4 being neutral and 7 being the highest score

Part 1: General questions

Question 1							Not applicable?	
Do you know how the data is being created for this product-attribute?								
No	1	2	3	4	5	6	7	Exact

Question 2							Not applicable?	
Do you know what the purpose of the data is?								
No	1	2	3	4	5	6	7	Exact

Question 3							Not applicable?	
Do you know who the users of the data are?								
No	1	2	3	4	5	6	7	Exact

Question 4							Not applicable?	
Do you know what the requirements of the users of the data are?								
No	1	2	3	4	5	6	7	Exact

Question 5							Not applicable?	
Do you know who to contact if there are problems with the Product-Attribute?								
No	1	2	3	4	5	6	7	Exact

Question 6							Not applicable?	
How do you experience the communication about the Product-Attribute in general?								
Bad	1	2	3	4	5	6	7	Perfect

Part 2: Questions about the new method

Step 1: Description: Product-attribute and team member selection
Tools: Standard Project Information Form
Time required: 30 minutes

Question 1							Not applicable?	
Is the goal of this step clear?								
Not Clear	1	2	3	4	5	6	7	Clear
Remarks / suggestions:								

Question 2							Not applicable?	
Is (are) the tool(s) provided for this step easy to use?								
Very difficult	1	2	3	4	5	6	7	Very easy
to use								to use
Remarks / suggestions:								

Question 3							Not applicable?	
Is (are) the tool(s) provided for this step complete?								
In-complete	1	2	3	4	5	6	7	Complete
Remarks / suggestions:								

Question 4							Not applicable?	
Is (are) the tool(s) provided for this step clear?								
Not Clear	1	2	3	4	5	6	7	Clear
Remarks / suggestions:								

Question 5		Not applicable?	
Is the time required for this step accurate?			
No			Yes
Remarks / suggestions:			

Part 3: General method questions

Question 1							Not applicable?	
Do you think the method improves the quality of the Product-Attributes?								
No	1	2	3	4	5	6	7	Yes
Remarks / suggestions:								

Question 2							Not applicable?	
Do you think this method and the quality team can improve the responsibilities?								
No	1	2	3	4	5	6	7	Yes
Remarks / suggestions:								

Question 3							Not applicable?	
Do you think this method and the quality team improve the communication?								
No	1	2	3	4	5	6	7	Yes
Remarks / suggestions:								

Question 4							Not applicable?	
What is your opinion of the method in general?								
Bad	1	2	3	4	5	6	7	Perfect
Remarks / suggestions:								

Appendix 6: Standard Project Information Form (1/2)

Honeywell

Data Product Analysis

Product Attribute:

Team Members:	
Team Leader:	
1. Name:	
Function	
Phone:	
Members:	
2. Name:	
Function	
Phone:	
3. Name:	
Function	
Phone:	
4. Name:	
Function	
Phone:	
5. Name:	
Function	
Phone:	
6. Name:	
Function	
Phone:	
7. Name:	
Function	
Phone:	
8. Name:	
Function	
Phone:	
9. Name:	
Function	
Phone:	
10. Name:	
Function:	
Phone:	

Appendix 6: Standard Project Information Form (2/2)

Meetings:	
Pre Planning Meeting	
Present:	
Date/Time:	
Location:	

Group Meeting	
Present:	
Date/Time:	
Location:	

Measurement Session	
Present:	
Date/Time:	
Location:	

Group Meeting	
Present:	
Date/Time:	
Location:	

Progress Meeting	
Present:	
Date/Time:	
Location:	

Appendix 7: Standard Product-attribute Characteristics Sheet

Standard Product-attribute Characteristics Form

Product-attribute:

Product-attribute functions:

Product-attribute components	
	Type:
	Type:
	Type:
	Type:
	Type:

Product-attribute data suppliers

Product-attribute data manufacturers

Product-attribute data users

Quality Dimension Score Sheet

Product-attribute:

Name:	
Function:*	

Data Supplier	Employee who create of collect data for the product-attribute	

Data Manufacturer	Employee who designs, develops, or maintains the data and systems infrastructure for the product-attribute	

Data Customer	Emplyee who uses the product-attribute in their work	

*: mark what is appropriate

The product-attribute will have to be scored on 15 dimensions of quality. For each dimension three aspects will be measured:

- **Importance:**
  How important is this dimension for the quality of the product-attribute?
- **Perceived Quality**
  What level of quality do you think this dimension of the product-attribute **has at this moment**?
- **Expected Quality**
  What level of quality do you think this dimension of the product-attribute **should have**?

The aspects of the quality dimensions can be scored on a scale ranging from 1 to 7

- **1 meaning:**
  - the dimension is of low importance,
  - the perception of the quality is low, or
  - the expected level of quality is low.
- **7 meaning:**
  - the dimension is of high importance,
  - the perception of the quality is high, or
  - the expected level of quality is high.

Appendix 8: Quality Dimensions Score Sheet (2/4)

1 **Accuracy**

The extent to which data is correct and reliable

	low						high
	1	2	3	4	5	6	7
Importance							
Perceived Quality							
Expected Quality							

2 **Objectivity**

The extent to which data is unbiased, unprejudiced, and impartial.

	low						high
	1	2	3	4	5	6	7
Importance							
Perceived Quality							
Expected Quality							

3 **Believability**

The extent to which data is regarded as true and credible.

	low						high
	1	2	3	4	5	6	7
Importance							
Perceived Quality							
Expected Quality							

4 **Reputation**

The extent to which data is highly regarded in terms of its source or content.

	low						high
	1	2	3	4	5	6	7
Importance							
Perceived Quality							
Expected Quality							

5 **Relevancy**

The extent to which data is applicable and helpful for the task at hand.

	low						high
	1	2	3	4	5	6	7
Importance							
Perceived Quality							
Expected Quality							

6 **Value-Added**

The extent to which data is beneficial and provides advantages from its use.

	low						high
	1	2	3	4	5	6	7
Importance							
Perceived Quality							
Expected Quality							

Appendix 8: Quality Dimensions Score Sheet (3/4)

7 **Timeliness**

The extent to which the data is sufficiently up-to-date for the task at hand.

	low						high
	1	2	3	4	5	6	7
Importance							
Perceived Quality							
Expected Quality							

8 **Completeness**

The extent to which data is not missing.

	low						high
	1	2	3	4	5	6	7
Importance							
Perceived Quality							
Expected Quality							

9 **Amount of data**

The extent to which the volume of data is appropriate for the task at hand.

	low						high
	1	2	3	4	5	6	7
Importance							
Perceived Quality							
Expected Quality							

10 **Interpretability**

The extent to which data is in appropriate languages, symbols, and units, and the definitions are clear.

	low						high
	1	2	3	4	5	6	7
Importance							
Perceived Quality							
Expected Quality							

11 **Ease of understanding**

The extent to which data is easily comprehended.

	low						high
	1	2	3	4	5	6	7
Importance							
Perceived Quality							
Expected Quality							

12 **Concise representation**

The extent to which data is compactly represented.

	low						high
	1	2	3	4	5	6	7
Importance							
Perceived Quality							
Expected Quality							

Appendix 8: Quality Dimensions Score Sheet (4/4)

13 **Consistent representation**

The extent to which data is presented in the same format.

	low						high
	1	2	3	4	5	6	7
Importance							
Perceived Quality							
Expected Quality							

14 **Accessibility**

The extent to which data is available, or easily and quickly retrievable.

	low						high
	1	2	3	4	5	6	7
Importance							
Perceived Quality							
Expected Quality							

15 **Access security**

The extent to which access to data is restricted appropriately to maintain its security.

	low						high
	1	2	3	4	5	6	7
Importance							
Perceived Quality							
Expected Quality							

Standard Problem Description Form

Product-attribute:

Problem statement:

What is the problem?

Where does it occur?

When does it occur?

How does it impact the product data quality?

What is the impact for the users/customers?

What additional data is needed?

Solution Score Sheet

Product-attribute:

Name:	
Function:*	

Data Supplier	Employee who create of collect data for the product-attribute	
Data Manufacturer	Employee who designs, develops, or maintains the data and systems infrastructure for the product-attribute	
Data Customer	Emplyee who uses the product-attribute in their work	

*: mark what is appropriate

The solutions will have to be scored on 5 criteria.
For each criteria two aspects will be measured:

- Importance:
  How important is this criteria?
- Score
  What score do you give this solution based on the other solutions?

The importance of the criteria can be scored on a scale ranging from 1 to 5

- 1 meaning:
  - the dimension is of low importance,
- 5 meaning:
  - the dimension is of high importance,

The solutions can also be scored on a scale from 1 to 5 using:

- 1 meaning:
  compared to the other solutions this solutions scores very bad
- 2 meaning:
  compared to the other solutions this solutions scores bad
- 3 meaning:
  this solutions scores average on this criteria
- 4 meaning:
  compared to the other criteria this solutions scores better
- 5 meaning:
  compared to the other criteria this solutions scores the best

Appendix 10: Solution Score Sheet (2/2)

	The cost to design the solution	The cost to implement the solution	The resources required	The impact on the the data quality	The reduction of risk
Importance					
Solution 1					
Solution 2					
Solution 3					
Solution 4					

Notes:

Appendix 11: Product-attribute Action Plan (1/3)

Product-attribute Action Plan

Product-attribute:

1 Action:			
Owner:		Due Date:	
Status:		Completed?	

2 Action:			
Owner:		Due Date:	
Status:		Completed?	

3 Action:			
Owner:		Due Date:	
Status:		Completed?	

4 Action:			
Owner:		Due Date:	
Status:		Completed?	

5 Action:			
Owner:		Due Date:	
Status:		Completed?	

Appendix 11: Product-attribute Action Plan (2/3)

6	Action:			
	Owner:		Due Date:	
	Status:		Completed?	

7	Action:			
	Owner:		Due Date:	
	Status:		Completed?	

8	Action:			
	Owner:		Due Date:	
	Status:		Completed?	

9	Action:			
	Owner:		Due Date:	
	Status:		Completed?	

10	Action:			
	Owner:		Due Date:	
	Status:		Completed?	

Appendix 11: Product-attribute Action Plan (3/3)

Summary

Action	Owner	Due date	Status	Completed?
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				

Appendix 12: Schedule of the TDQM Cycle for Honeywell

Total Data Quality Management Method for Honeywell Emmen - Timeline

Step		Day 1	Day 2				Day 3		Day 4				Day 5	Check
		1/2 h	1st hour	2nd hour	3rd hour	4th hour	1st hour	2nd hour	1st hour	2nd hour	3rd hour	4th hour	1/2 h	
Step 1 Prod.-attr. & Team selection		30												
Step 2 Charact. of the Prod.-attr.	Group Meeting		30											
Step 3 Important Quality dimensions				60										
Step 4 Data manufacturing process					120									
Step 5 Measurement Metrics						30								
Step 6 Measurement							120							
Step 7 Describe Specific Problems	Group Meeting								15					
Step 8 Analysis of the problems										90				
Step 9 Solution Generation											60			
Step 10 Solution Selection												60		
Step 11 Action Plan												15		
Step 12 Check Progress													30	

Appendix 13: The TDQM method for Honeywell Emmen

The steps are displayed in a table with the following rows:

- The required team members are described. Sometimes not all team members can be a part of the step. For example because the members aren't known yet.
- The input of the process is described.
- The type of step is described. For example: A meeting or a group meeting. This has consequences for the location and possible also for the data and time of the step.
- The action that are taken by the step will be described.
- The tools that are used during the step will be described.
- The time required is estimated.
- The output of the step is described.

The Total Data Quality Management Method for Honeywell:

Step 1	**Product-attribute and team members selection**
By:	• Data Quality (DQ) Team Champion • DQ Team Engineer • Finance member
Input:	• Historic experiences • Knowledge about all the product-attributes that are of importance.
Type:	• Meeting
Action:	• Selection of the product-attribute that needs to be improved • Selection of the team members • Schedule the first meeting
Tools:	• Standard Project Information Form (Appendix 6)
Time required:	• 30 minutes
Output:	• The product-attribute • The complete list of DQT members • Time/Data/Location of the first DQT meeting
Step 2	**Characteristics of the Product-attribute**
By:	• Complete DQ Team
Input:	• The product-attribute • Time/Data/Location of the first DQT meeting
Type:	• Group Meeting
Action:	• Description of the main characteristics of the selected product-attribute.
Tools:	• Standard Product-attribute Characteristics Form (Appendix 7).
Time required:	• 30 minutes
Output:	• Product-attribute characteristics
Step 3	**Important Quality dimensions**
By:	• Total DQ Team, permanent and other invited members
Input:	• Product-attribute and its characteristics
Type:	• Group Meeting
Action:	• The members should each score the importance, perceived quality and expected quality in each quality dimension • Determine which dimensions are required to be of high quality
Tools:	• Score sheet on which ratings for the dimensions can be given (Appendix 8). • Calculation sheet which automatically calculates graphs from the ratings given by the team members
Time required:	• 60 minutes / 1 hour
Output:	• Important quality dimensions

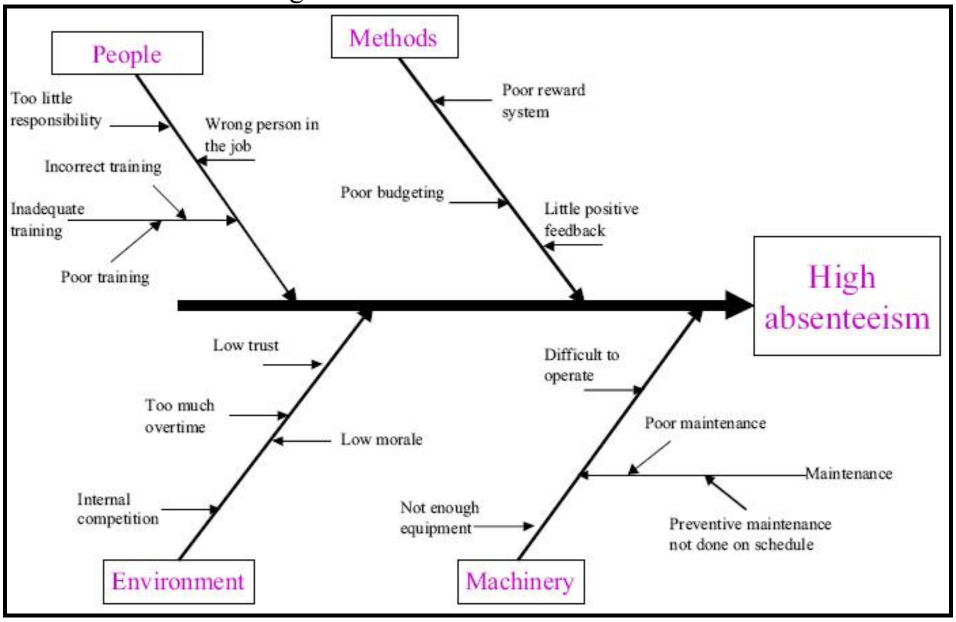
Step 4	Data manufacturing process
By:	• Total DQ Team, permanent and other invited members
Input:	• Product attribute and its characteristics
Type:	• Group Meeting
Action:	• Determining the Data manufacturing process and make a flowchart of it
Tools:	• Whiteboard • MS Visio with the basic flowchart modeling technique described in section 3.2.3
Time required:	• 120 minutes / 2 hours
Output:	• Cleary defined Data manufacturing process
Step 5	**Measurement Metrics**
By:	• Total DQ Team, permanent and other invited members
Input:	• Important quality dimensions
Type:	• Group Meeting
Action:	• Determine the metrics to measure the important quality dimensions on • Plan the measurement steps and other actions that follow from the first steps • Plan the next group meeting
Tools:	• The RUMBA criteria to help determine if the metric is useful (section 3.2.3)
Time required:	• 30 minutes
Output:	• Metrics to measure the important quality dimensions on • Time/Data/Location of the first DQT meeting
Step 6	**Measurement**
By:	• DQ Team Engineer • Possibly other DQ Team members
Input:	• Time/Data/Location of the first DQT meeting • Metrics to measure the important quality dimensions on
Type:	• Measurement sessions
Action:	• Measure a sample of product on the metrics and determine the quality • Possibly other data than Oracle data is required. Which has to be collected by the other DQ Team members • Prepare Pareto diagrams and other graphs
Tools:	• Statistical formula to determine the sample size. (The Sample Size calculator on www.surveysystems.com => Research Aids can also be used) • MS Excel
Time required:	• 120 minutes / 2 hours – or more. Depending on the number of metrics to measure
Output:	• Pareto diagrams of the results of the measurement
Step 7	**Describe Specific Problems**
By:	• Total DQ Team, permanent and other invited members
Input:	• Documents made in the previous meeting • Pareto diagrams of the results of the measurement
Type:	• Group Meeting
Action:	• Reflect on the results of the previous meeting • Describe the specific problems which can be concluded from the results of the measurement
Tools:	• Standard Problem Description Form (Appendix 9)
Time required:	• 15 minutes
Output:	• Description of the specific problems

Step 8	Analysis of the problems
By:	• Total DQ Team, permanent and other invited members
Input:	• Description of the specific problem
Type:	• Group Meeting
Action:	• Analysis of the problem by determining the cause map
Tools:	• Whiteboard • MS Visio or MS Excel to model the cause map
Time required:	• 90 minutes / 1,5 hours
Output:	• Cause map
Step 9	**Solution Generation**
By:	• Total DQ Team, permanent and other invited members
Input:	• Cause map
Type:	• Group Meeting
Action:	• Think of solutions to every cause that is defined on the cause map
Tools:	• MS Visio or MS Excel to add the solutions to the cause map
Time required:	• 60 minutes / 1 hour
Output:	• Cause map with solutions
Step 10	**Solution Selection**
By:	• Total DQ Team, permanent and other invited members
Input:	• Cause map with solutions
Type:	• Group Meeting
Action:	• The team selects a number of solutions to implement and group them into possible solutions. • The solutions have to be scored based on the given criteria in section3.5. • The solution that scores best has to be implemented.
Tools:	• Score sheet to score each solution on. (Appendix 10) • Calculation sheet which will determine the results for each criteria
Time required:	• 60 minutes / 1 hour
Output:	• The chosen solution
Step 11	**Action Plan**
By:	• Total DQ Team, permanent and other invited members
Input:	• The chosen solution
Type:	• Group Meeting
Action:	• Specific tasks should be identified to be able to implement the solution • The DQ Team should assign these tasks to team members or other employees.
Tools:	• Standard Task Form (Appendix 11)
Time required:	• 15 minutes
Output:	• Action plan

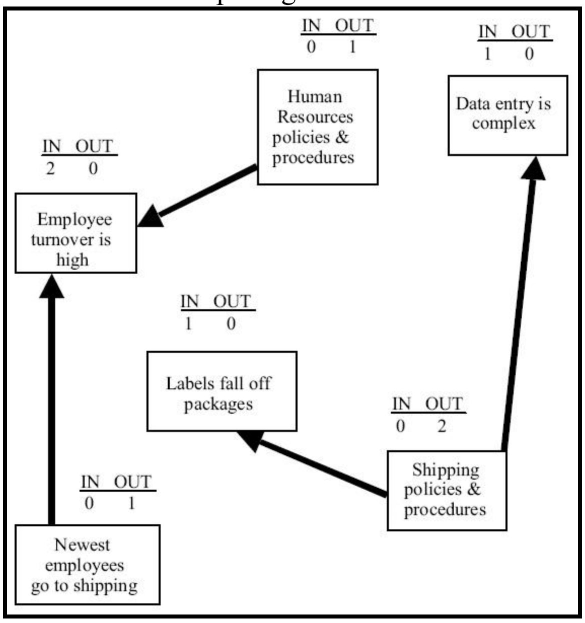
Step 12	Check Progress
By:	• Total DQ Team, permanent and other invited members
Input:	• Action plan
Type:	• Group Meeting
Action:	• The progress of implementing the solution should be discussed • Problems can be dealt whit accordingly
Tools:	• No specific tools
Time required:	• 30 minutes – 60 minutes / 1 hour. Depending on the amount of actions and problems there are.
Output:	• Updated Action plan

Appendix 14: Three Root Cause Analysis Tools [8]

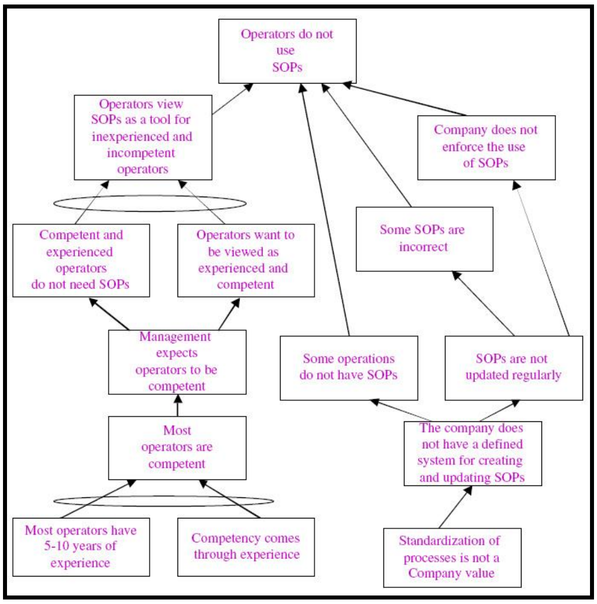
A: Cause-and-effect Diagram



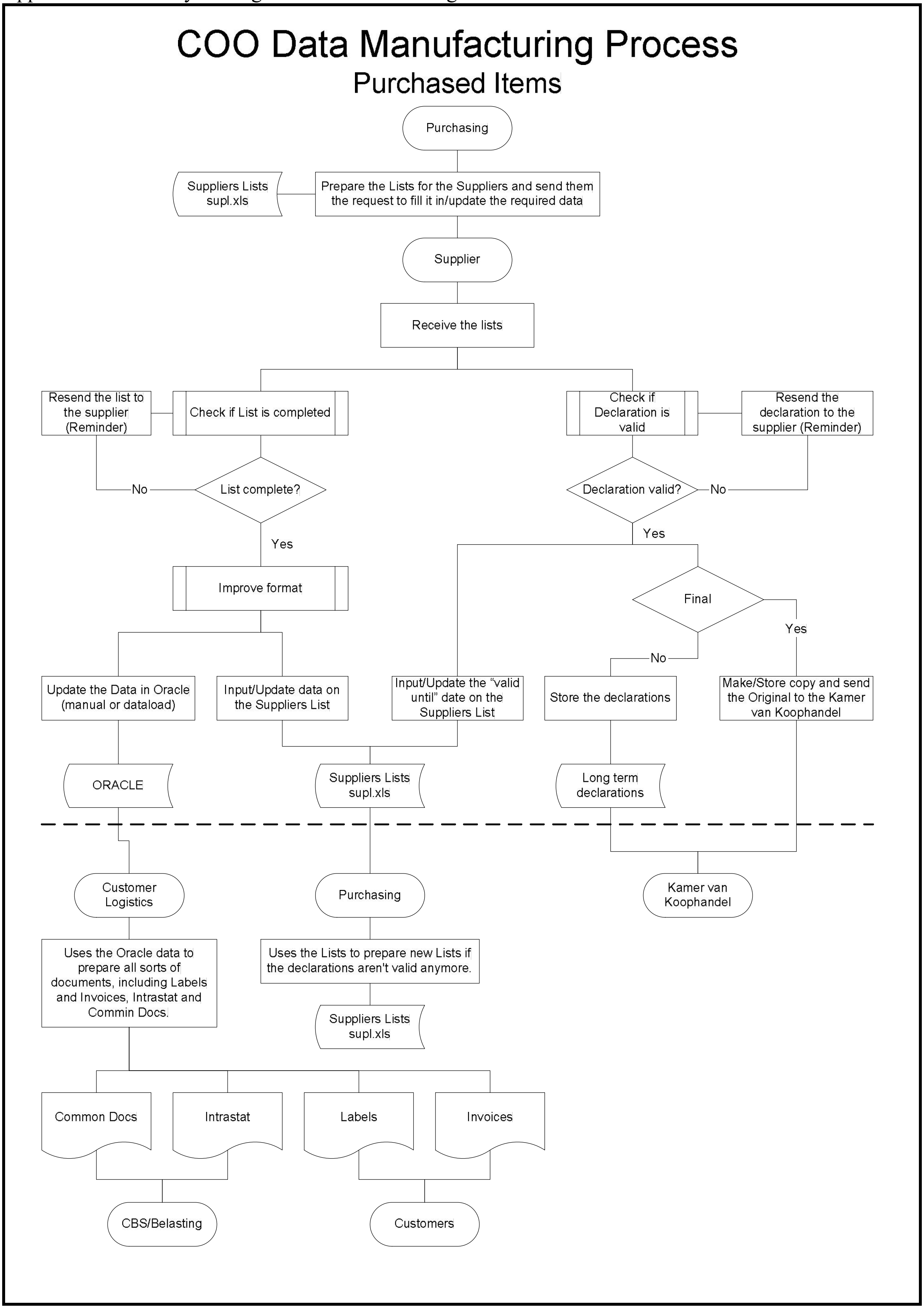
B: Interrelationship Diagram



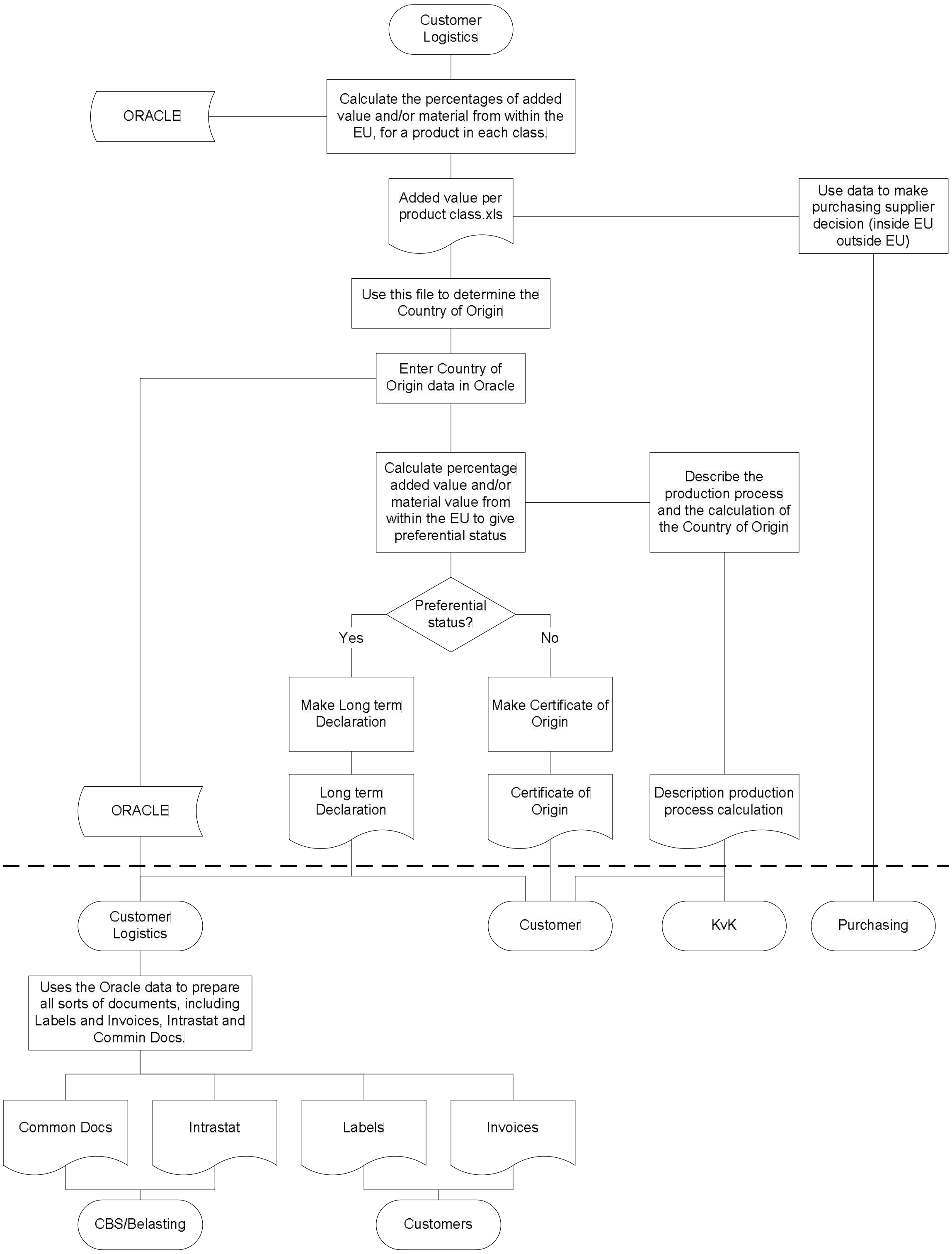
C: Current Reality Diagram



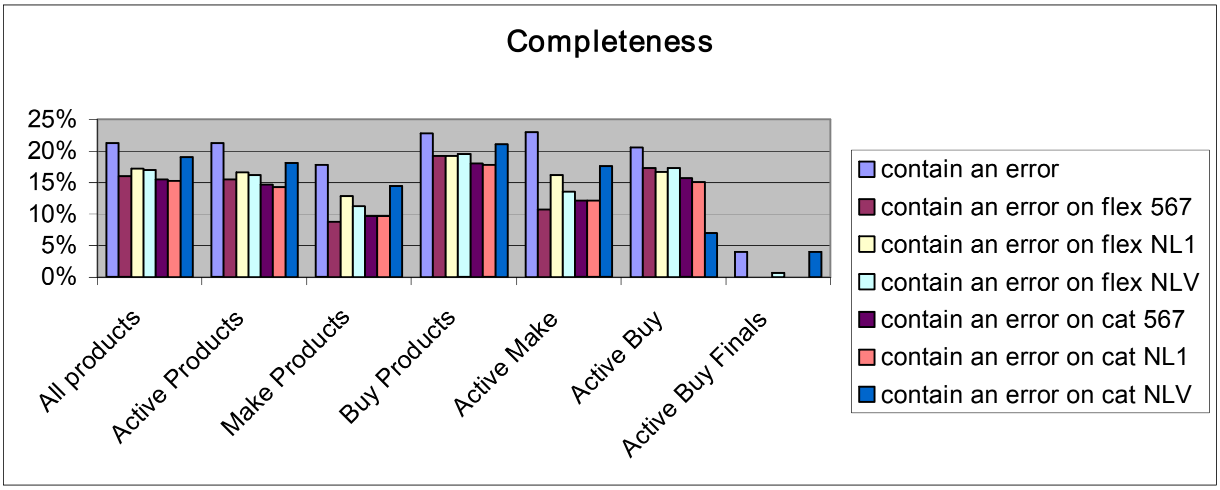
Appendix 15: Country of Origin Data Manufacturing Process

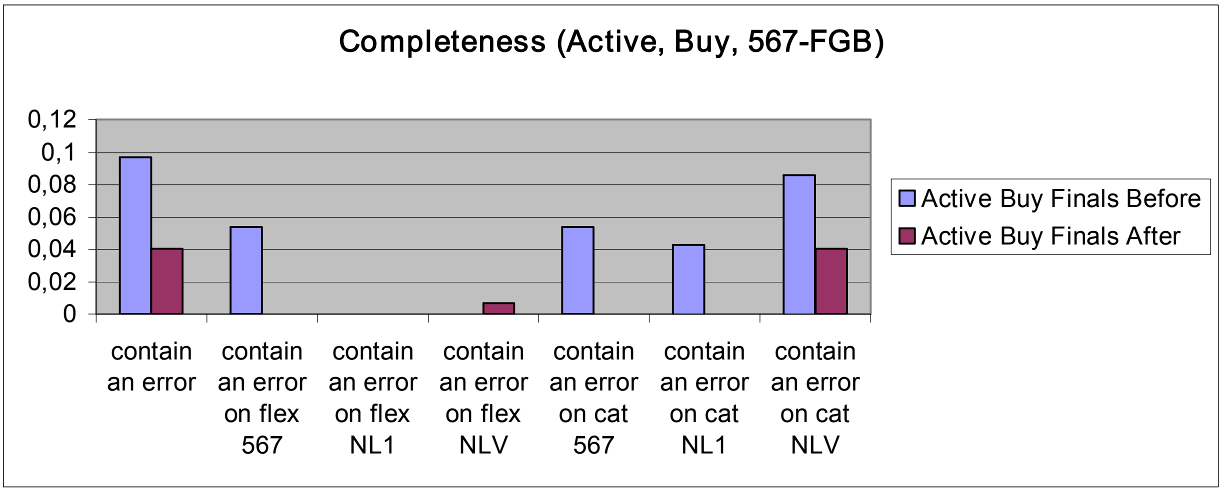


COO Data Manufacturing Process
- Make Items

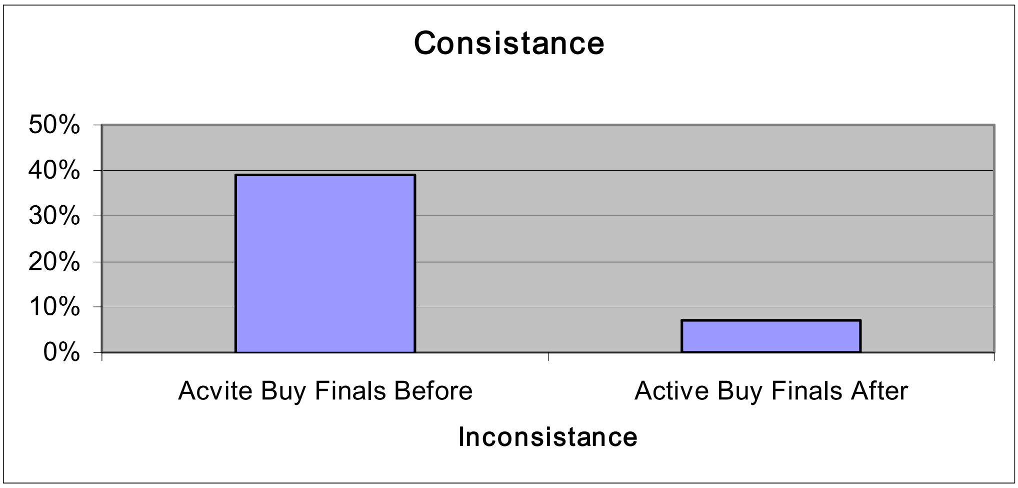


Appendix 16: Country of Origin Measurement results






Consistancy
16%
14%
12%
10%
8%
6%
4%
2%
0%
Inconsistance
All products
Active Products
Make Products
Buy Products
Active Make
Acvite Buy
Acvite Buy Finals


Consistance
50%
40%
30%
20%
10%
0%
Acvite Buy Finals Before
Active Buy Finals After
Inconsistance

Appendix 17: Country of Origin Standard Problem Description Form

Standard Problem Description Form

Product-attribute:
Country of Origin

Problem statement:
Problems with COO on invoices (567 flex), labels, package (NLV flex)

What is the problem?
Quality low on completeness and consistance dimension
No way to measure timeliness
Data for the make products can not be calculated

Where does it occur?			
Invoices	Labels	Reports	Package

When does it occur?
Random

How does it impact the product data quality?
None, or wrong data is printed on labels.
Customs stop products on borders, because data is missing

What is the impact for the users/customers?
Wrong data leads to lowered customer satisfaction
Also internally, mistakes can lead to arguments and frustration

What additional data is needed?
How can the COO for Make products been determined?

Appendix 18: Country of Origin Action Plan Form

Product-attribute Action Plan

Product-attribute:
Country Of Origin

#				
1	Action:	Get the COO data for finished goods buy		
		Update them in Oracle		
		Make the overview sheet on which valid untill datas are shown		
	Owner:	Janneke Hof	Due Date:	nvt
	Status:	Compeleted	Completed?	Yes

#				
2	Action:	Get the COO data for material and parts		
		Make a list for R. Moleveld. With products & their COO to update in Oracle		
		Update the overview sheet		
	Owner:	Janneke Hof	Due Date:	nntb
	Status:	Partly completed	Completed?	No

#				
3	Action:	Make a report that can make the COO calculation for MAKE finals		
	Owner:	Bennie Bruins	Due Date:	nntb
	Status:		Completed?	

#				
4	Action:	Make an Excel sheet from the report of Bennie with the products and their		
		cost percentages.		
	Owner:	Roma Porzych	Due Date:	nntb
	Status:		Completed?	

#				
5	Action:	Make a list of Make items that have COO NL (based on the sheet from Roma)		
		Make a list of Make items that are from Brno (based on the sheet from Roma)		
		Make a list for R. Moleveld. With products & their COO to update in Oracle		
	Owner:	Hans Huy	Due Date:	nntb
	Status:		Completed?	

6	Action:	Make Declarations for Brno		
	Owner:	Hans Huy	Due Date:	nntb
	Status:		Completed?	

7	Action:	Get the COO data for finished goods buy from Brno		
		Update them in Oracle		
		Update the overview sheet		
	Owner:	Janneke Hof	Due Date:	nntb
	Status:		Completed?	

Summary

Action	Owner	Due date	Status	Completed?
1	Janneke Hof	nvt	Completed	Yes
2	Janneke Hof	nntb	Partly Complete	No
3	Bennie Bruins	nntb		No
4	Roma Porzych	nntb		No
5	Hans Huy	nntb		No
6	Hans Huy	nntb		No
7	Janneke Hof	nntb		No
8				
9				
10				

Evaluation Survey Results

Part 1: General questions

Question 1	Do you know how the data is being created for this Product-attribute?	Average Score	
		Before:	2,5
		After:	7

Question 2	Do you know what the purpose of the data is?	Average Score	
		Before:	2,75
		After:	7

Question 3	Do you know who the users of the data are?	Average Score	
		Before:	3
		After:	6,75

Question 4	Do you know what the requirements of the users of the data are?	Average Score	
		Before:	3
		After:	6,5

Question 5	Do you know who to contact if there are problems with Product-Attribute?	Average Score	
		Before:	2,75
		After:	6,75

Question 6	How do you experience the communication about the the Product-Attribute in general?	Average Score	
		Before:	1,75
		After:	5,5

Part 2: Questions about the new method

Question 1	Is the goal of this step **clear**?											
Step:	1	2	3	4	5	6	7	8	9	10	11	12
Score:	7	6	5	5,5	5	7	7	6,75	7	6,75	7	nvt

Question 2	Is (are) the tool(s) provided for this step **easy to use**?											
Score Step:	1	2	3	4	5	6	7	8	9	10	11	12
	7	7	**2,75**	4,75	nvt	nvt	6,5	6	5,5	5,5	6,5	nvt

Question 3	Is (are) the tool(s) provided for this step **complete?**											
Score Step:	1	2	3	4	5	6	7	8	9	10	11	12
	7	5	6,5	5,75	nvt	nvt	7	6,5	6,25	6,5	7	nvt

Question 4	Is (are) the tool(s) provided for this step **clear?**											
Score Step:	1	2	3	4	5	6	7	8	9	10	11	12
	7	7	**2,5**	5	nvt	nvt	6	6,5	5,75	6,5	7	nvt

Question 5	Is the time required for this step **accurate**?											
Score Step:	1	2	3	4	5	6	7	8	9	10	11	12
	5,5	5,5	6,5	5	5	5	6,25	**3,5**	**3,5**	**3,5**	5,75	nvt

Part 3: General method questions

Question 1	Do you think the method improves the quality of the Product-Attributes?	Average Score
		7

Question 2	Do you think this method and the quality team can improve the division of the responsibilities?	Average Score
		6

Question 3	Do you think this method and the quality team improve the communication?	Average Score
		7

Question 4	What is your opinion of the method in general?	Average Score
		6

Question 5	The Method:	Average Score
	Is easy to use	6
	Takes to much time?	5,25
	Is necessarily	7

Appendix 20: Short Quality Dimensions Score Sheet

Quality Dimension Score Sheet

Product-attribute:

Importance:

		Low	Medium	High
1	**Accuracy**	Low	Medium	High
	The extent to which data is correct and reliable			
2	**Objectivity**	Low	Medium	High
	The extent to which data is unbiased, unprejudiced, and impartial.			
3	**Believability**	Low	Medium	High
	The extent to which data is regarded as true and credible.			
4	**Reputation**	Low	Medium	High
	The extent to which data is highly regarded in terms of its source or content.			
5	**Relevancy**	Low	Medium	High
	The extent to which data is applicable and helpful for the task at hand.			
6	**Value-Added**	Low	Medium	High
	The extent to which data is beneficial and provides advantages from its use.			
7	**Timeliness**	Low	Medium	High
	The extent to which the data is sufficiently up-to-date for the task at hand.			
8	**Completeness**	Low	Medium	High
	The extent to which data is not missing.			
9	**Amount of data**	Low	Medium	High
	The extent to which the volume of data is appropriate for the task at hand.			
10	**Interpretability**	Low	Medium	High
	The extent to which data is in appropriate languages, symbols, and units, and the definitions are clear.			
11	**Ease of understanding**	Low	Medium	High
	The extent to which data is easily comprehended.			
12	**Concise representation**	Low	Medium	High
	The extent to which data is compactly represented.			
13	**Consistent representation**	Low	Medium	High
	The extent to which data is presented in the same format.			
14	**Accessibility**	Low	Medium	High
	The extent to which data is available, or easily and quickly retrievable.			
15	**Access security**	Low	Medium	High
	The extent to which access to data is restricted appropriately to maintain its security.			