bol.com

UNIVERSITEIT TWENTE.

Operational data quality

A case study about the quality of net purchase prices in the catalog of bol.com

Lieke Lubbers

December, 2013

Master thesis Industrial Engineering and Management

Operational data quality
A case study about the quality of net purchase prices in the catalog of bol.com

Author
Lieke Lubbers
Student number: s0180955
l.s.m.lubbers@student.utwente.nl
Industrial Engineering and Management
Track Production and Logistic Management
University of Twente

Supervisors University of Twente
Dr. A.B.J.M. Wijnhoven
Industrial Engineering & Business Information Systems
University of Twente

Dr. M.E. Iacob
Industrial Engineering & Business Information Systems
University of Twente

Supervisor bol.com
Mr. K. Hiddinga
Business Analyst Operations & Fulfillment
Bol.com, Utrecht

Publication date
December, 2013

University of Twente
Drienerlolaan 5
7522 NB Enschede
Tel: +31 (0)53 489 9111
Fax: +31 (0)53 489 2000

Bol.com
Keulsekade 189
3534 AC Utrecht
Tel: +31 (0)30 00 8888

Management summary

The focus of this research is on the quality of net purchase prices in the catalog of bol.com. Because of the fact that the net purchase price is the input factor for many processes, like calculation of the selling price for the web shop, calculation of the stock value and selection of the preferred supplier, an incorrect net purchase price in the catalog has big consequences. This shows the relevance of correct net purchase prices in the catalog. Another consequence of incorrect net purchase prices is that the forecast of daily revenues is not accurate. In 2012 the company is taken over by Ahold, a stock market listed company, this emphasizes the importance of accurate forecasting of daily revenues. The problem is that the net purchase prices, as they are stored in the catalog, are not reliable. The last years a difference is noticed between the net purchase price that is agreed with buyer and supplier, and the net purchase price on the order invoice. This is caused by the fact that the entry and handling process of net purchase prices is not sufficiently validated. The central question in this research is how the data entry and handling process can be redesigned such that the quality of net purchase prices will improve.

A literature study on the topic of data quality assessment and improvement methodologies has shown that TDQM (Total Data Quality Management) is a suitable method to apply in this research. However the four basic stages of this methodology cannot be followed blindly, to apply the method in practice some contextual redesign is required. Net purchase prices pass different systems and it is essential to understand the complete set of processes and how the quality of data at each processing stage is impacted by its quality at previous stages. The TDQM methodology should help assessing data quality not only at the final stage but also at any of the preceding intermediate stages. To reach this a TDQM framework that can be applied to operational systems need to be developed. For developing a TDQM methodology which can be applied in an operational context, the four basic stages, called definition, measurement, analysis and improvement, are used, but the details and execution of the different stages differs from the traditional TDQM methodology. The methodology is complemented by steps from the BPE approach. The BPE (Business Process Engineering) approach is suitable for the operational context but provides for some steps vague definitions and actions. The combination of these two methodologies, adapted to the needs of the specific environment, results in an improved TDQM methodology which can be applied in an operational context.

The first three stages of the improved TDQM methodology, called definition, measurement and analysis, are applied to get more insight in the entry and handling process of net purchase prices and to find the critical areas in this process. For assessing the quality of the net purchase prices in the catalog cases of mismatches between net purchase prices in the order system of bol.com and the prices on the invoice are collected. Together with the buyers these cases are analyzed and evaluated. The result of these stages is a list with ten root causes of price differences. These causes are all types of production errors and influence the intrinsic, contextual and representational data quality dimension. An additional finding, resulting from interviews, is that the culture among the buyers is also an issue which results in unreliable net purchase prices in the catalog. The buyers are not aware of the fact that they are, together with the Product & Content department, responsible for correct net purchase prices in the catalog and the influence of their tasks on other processes is not sufficiently known by them.

In the improvement stage only five out of ten causes are addressed, these are the five causes owing to incorrect or delayed operations by the buyers. The desired situation is based on insights obtained from the analysis of the current situation and by interviews with employees from the Product & Content department, the Controlling department and with buyers from the Buying & Merchandising department. To bridge the gap between the current and the desired situation, three types of changes are necessary: changes in the IT system, changes in the process/working method, and changes in culture. These changes are translated into concrete solutions and recommendations. In short the solutions and recommendations are:

- Implement a dashboard for monitoring the net purchase prices during the handling process.
- Implement the possibility for buyers to enter the number of products as an end condition of purchase deals.
- Suppliers need to send a new price list to at least two buyers and this need to be stored on a central disc which is only accessible for the buyers who are doing business with the supplier.
- Buyers should block time in their agenda to make sure that they enter manual offers before the date from which the offers are valid.
- Existing reports about mutations in net purchase prices in the catalog need to be used to check if the manual offers are entered correctly. A workflow is created which shows how buyers need to handle when entering manual offers.

After the solution generation process and the formulation of the recommendations, the most important question is what the worth or merit is of the proposed solutions. For evaluating the monitoring tool a survey is conducted among the users of the tool. The user evaluations are based on the concept of Task-Technology Fit. From the survey can be concluded that the characteristics of the monitoring tool match with users' task needs. The TTF perspective assumes that a better fit between technology functionalities and task requirements results in better performance.

The implementation of the solutions can start independently of each other when the requirements are met. The first requirement that must be met is that the culture among the buyers need to change. The interaction between the buyers and the technology is a critically important contributor to the failure or success of change initiatives. Only when the buyers are aware of their responsibility for correct net purchase prices in the catalog and realize that the proposed solutions can help them improving these quality, the solutions can be implemented successfully. For the long-term solutions where the IT department is involved an additional requirement must be met before implementation can start, this is the estimation and scheduling of the IT time needed.

Preface

This master thesis is the result of my graduation project for the study Industrial Engineering and Management at the University of Twente. For my graduation project I performed a research on the quality of net purchase prices in the catalog of bol.com. Important to note is that the findings have emerged during writing this thesis. In the meantime changes have been implemented and improvements are visible. This means that at the time of reading this thesis the problems encountered are no longer relevant.

Using this opportunity, I would like to thank several people who supported me during my graduation project. First of all I want to thank my external supervisor Kristian for giving me the opportunity to do this research, for sharing his innovative ideas and for his guidance. Our conversations have brought this research to a higher level. I also want to thank all my colleagues of bol.com for providing me a comfortable, fun and motivating work atmosphere. Everyone was always willing to help me and to answer my questions.
Furthermore I would like to thank my supervisors of the University of Twente, Fons Wijnhoven and Maria Iacob, for their supervision, feedback and for sharing their knowledge on my research topic. I appreciate all the time and effort they both put into this research.
Last but not least, special thanks to my parents for their never-ending support and confidence.

Lieke Lubbers
Utrecht, December 2013

Table of contents

Abbreviations

BIT	Bol Information Mart
BPE	Business Process Engineering
BPMN	Business Process Model and Notation
CEO	Chief Executive Officer
CM1	Contribution Margin
DQ	Data Quality
DSRM	Design Science Research Methodology
EAN	European Article Number
EDI	Electronic Data Interchange
ETL	Extract Transform Load
FTP	File Transfer Protocol
IS	Information System
KPI	Key Performance Indicator
LIS	Leverancier Information System
LSM	Local Sourcing Module
OCR	Optimal Character Recognition
PO	Purchase Order
POS	Product Offer System
RUMBA	Reasonable, Understandable, Measurable, Believable, Achievable
SCP	Supply Chain Portal
SDD	Seller Dashboard
TDQM	Total Data Quality Management
TQM	Total Quality Management
TTF	Task-Technology Fit
UI	User Interface
VAT	Value Added Tax

List of tables and figures

List of tables

List of figures

Chapter 1: Company introduction

This chapter gives a quick overview of bol.com. First, the history and the mission, vision and strategy are described. Furthermore the market they operate in is discussed and the competitors and the organizational lay out are described.

1.1 Background

Bol.com is the number one online retailer in the Netherlands. It offers non-food products and delivers directly to customers' homes in the Netherlands and Belgium, reaching almost 4 million customers (Ahold, 2013). The company is founded in 1999 by the German multinational Bertelsmann and started as an online book seller. Nowadays their product range has expanded to over six million products in nine different categories: Books, Music, DVD & Games, Toys, Baby, Health & Beauty, Computer & Electronics, Living, Pets, Garden & Tools.

Bol.com continues to attract visitors to its online store. In February 2011 the company has launched Plaza, which gives customers access to products from other retailers and it enables individuals to exchange second hand products. Other improvements that are implemented the last years are the introduction of the new mobile bol.com store. This store offers more product information and search options to help customers with finding what they are looking for. Furthermore the payment options have been expanded (Ahold, 2012a). In May 2012, the company has been taken over by Ahold and a couple of months later they have introduced their first joined service to offer customers greater shopping convenience (Ahold, 2012b). From that moment customers have been able to pick up bol.com orders in Albert Heijn stores. These strategic improvements support the mission statement.

Bol.com-CEO Daniel Ropers describes the mission statement in original language as follows: "We willen gewoon de beste winkel zijn**"** (Retaildetail, 2013)

The vision statement outlines where the company wants to be in the future. The vision statement in original language according to bol.com CEO Daniel Ropers:

"Onze visie is om Bol.com te ontwikkelen tot hét platform waar de consument alles kan vinden wat zijn of haar hart begeert, omdat we voor elke koopvraag van de klant het beste antwoord online hebben" (TwinkleMagazine, 2012)

Besides the enormous growth of products and customers, there is also a rising number of awards. All effort to serve customers quickly and efficiently has been awarded several times, these awards are based on customer appreciation. Furthermore the website has a position in the list with most visited websites of the Netherlands (Bol.com, 2013).

1.2 Key figures

To get an idea of the size of the organization and the yearly sales some key figures are given in the table on the next page. In 1999 the company started with yearly sales of 10 million Gulden. Since that time an enormous growth in sales is noticed: between 2002 and 2008 the sales numbers tenfold and a growing percentage of 14 percent is achieved in 2010 (Bol.com, 2013). Also an enormous growth in product assortment can be noticed, the number of unique products is nowadays more than six million compared to 70.000 products in 2000.

Description	Figures
# Employees (May 2013)	±550 employees
# Products in assortment (May 2013)	>6 million products
# Product categories (June 2013)	9 product categories
# yearly customers (May 2013)	4 million customers
# Sold products (2011)	>17 million products
Yearly sales (2011)	355 million Euro

Table 1.1 Key figures (Bol.com, 2013; Volkskrant, 2012)

1.3 Competitors

Despite the fact that bol.com is the number one online retailer in the Netherlands, it needs to stay competitive. Their competitors can be divided in two different types: competitors with a comparable business model and competitors that offer comparable products in specific categories. A competitor with a comparable business model is Amazon.com, this company is also offering online products. Bol.com stays competitive because it focuses on the Dutch speaking customers while the focus of Amazon.com is not on these customers. Furthermore bol.com CEO Daniel Ropers states that bol.com can benefit from their economies of scale in the Benelux, this in contrast to Amazon.com which is a new player in the Benelux (Retaildetail, 2013).

The second type of competitor competes offline in specific categories. The business model of these companies is completely different but they offer comparable products. For example ToysXl is competing in the toys category and Praxis is competing in the garden & tools category.

1.4 Organizational structure

The organizational lay out is shown in the figure below.

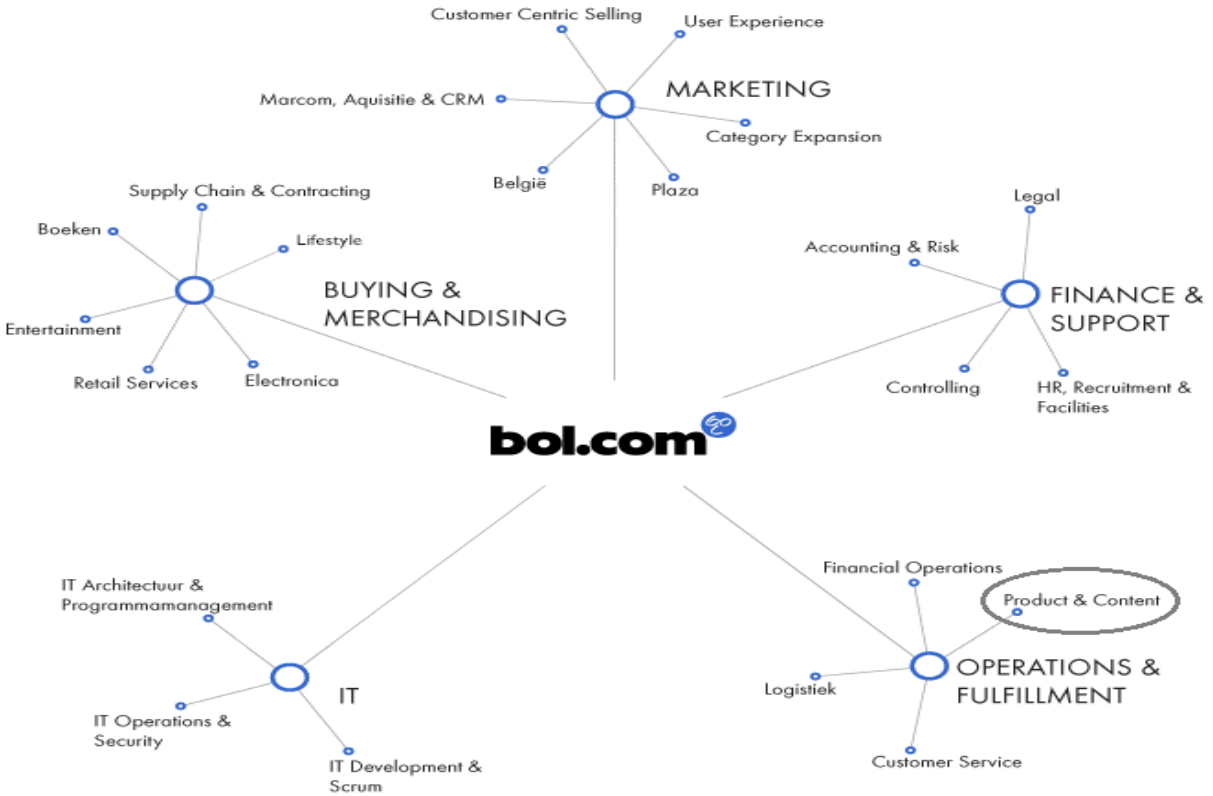


Figure 1.1 Organizational structure (Bol.com, 2013)

As can be seen in the figure, there are five different departments (Finance & Support, Operations & Fulfillment, IT, Buying & Merchandising and Marketing), each controlled by their departmental director. Furthermore all the departments consist of several sub-departments.
This research is executed within the Product & Content department, sub-department of Operations & Fulfillment. The responsibilities of this department are maintaining and improving the product catalog. Besides they are responsible for the correct information on the website, like titles, authors, product characteristics, images, prices and delivery times (Intranet-bol.com, 2013).

1.5 Preliminary conclusion

Since the start of bol.com, in 1999, an enormous growth in product assortment and sales numbers can be seen. Such an enormous growth is associated with changing requirements for the used information systems and processes, furthermore it may influence the current working procedures. This changing environment is one of the motives for this research.

Chapter 2: Research design

This chapter starts with the motivation for the research and the formulation of the problem statement. In view of the problem statement, the research goal is formulated and a conceptual model shows the causal relationship between the variables. Subsequently the research questions are defined and an overview of the research methodology is given. This chapter concludes with a reading guide which explains how this thesis is constructed.

2.1 Motivation

This research is executed within the Product & Content department. As stated in section 1.4 this department is part of the Operations and Fulfillment department and their main tasks are maintaining and improving the product catalog. Besides they are responsible for the correct information on the website, here we are talking about information about titles, authors, product characteristics, images, prices and delivery times.

The focus of this research is on the quality of net purchase prices in the catalog. Before continuing, the glossary below will clarify some words.

- Net purchase price = Purchase price exclusive VAT. The supplier invoices this price to bol.com. Standard retail discounts that are agreed in a contract are already included. Other kind of discounts and bonuses afterwards are not included, these are calculated in the net net purchase price. Moreover the net net purchase price is influenced by the exchange rate. The quality of the net net purchase price is another research topic and is out of scope of this research.
- Supplier offer = Collection of monetary conditions (among other the net purchase price) and delivery time for which bol.com can buy a product at the supplier.
- Selling price = Price that is used for the web shop. It is the price for which customers can buy products at bol.com. This price is higher than the purchase price to cover the costs. The contribution margin (referred to as CM1) is the marginal profit per unit sale.
- Selling offer = Collection of monetary conditions (among other the bol.com selling price) and the delivery time for which the customer can buy a product on the website.
- Offer system = The system which collects all the supplier offers of one specific product. All the supplier offers are stored in a normalized way, subsequently the offers of one product are compared. Based on among other the purchase prices and the delivery times the preferred supplier is calculated and finally the selling price for the web shop is calculated.
- Catalog = Assortment of all available products.

Net purchase prices are stored in the offer system. The information in the offer system can be received in different ways, for example by use of automatic catalog feeds of the suppliers or the data can be entered manually. More about the input methods can be read in section 4.1.2. The net purchase price is the input factor for many consecutive processes, like calculation of the selling price for the web shop, calculation of the stock value, selection of the preferred supplier and calculation of the contribution margin.

The buyers are responsible for correct net purchase prices in the system. Net purchase prices are agreed with buyer and supplier. The buyers make contract deals with the suppliers. Changes in net purchase prices needs to be correctly entered in the system, these price changes can occur daily, weekly, monthly etc. In principle these net purchase prices should match with the net purchase price on the invoice sent by the supplier when bol.com has placed an order.

In the last years a difference is noticed between the net purchase price that is agreed with buyer and supplier and the net purchase price on the order invoice. Because of the fact that the net purchase price is the input factor for many processes, an incorrect purchase price in the catalog has big consequences for the consecutive business processes. This shows the relevance of correct net purchase prices. Another consequence of incorrect net purchase prices is that the forecast of daily revenues is not accurate. In 2012 the company is taken over by Ahold, a stock market listed company, this emphasizes the importance of accurate forecasting of daily revenues.

2.2 Problem definition

The problem is that the net purchase prices, as they are stored in the catalog, are not reliable. The cause of this problem is that the input of purchase conditions is not sufficiently validated in the current situation. As stated before these incorrect data influences many other processes. The problem statement can be formulated as follows:

The data entry and handling process is not sufficiently validated. This results in unreliable net purchase prices in the catalog.

Before starting the discussion about data it is important to note that in literature often a distinction is made between data and information. According to Eppler (2006) data designates raw, unconnected, quantitative or qualitative items. A piece of data is just a registered item without context. Data in this sense becomes information when it is related to other data. In other words, when linking various sets of data to form one coherent statement the resulting entity can be called a piece of information and this piece of information forms a message. However, this distinction depends on the point of view of the author. Pipino (2002) states that data and information are often used synonymously (Pipino, Lee, & Wang, 2002). In this thesis the term data will be used, but may refer to both terms.

From the problem statement above the following main research question can be derived:

How can the data entry and handling process be redesigned such that the quality of net purchase prices in the catalog will improve?

2.3 Research goal

The goal of this research is to give recommendations for improving the data handling process. According to Van den Berg et al. (2008) a process consist of all the activities which take place between the demand for a service or a product and the actual delivery. For executing the activities among other systems, procedures and workflows can be used. Recommendations can be in the field of procedural improvements and/or in the field of system improvements. The meaning of 'improving' in this context is that the data entry and handling process needs to be validated. The whole data

process, from the input till the invoice matching procedure, needs to be closed. In this way errors in invoice matching cannot be assigned to incorrect net purchase prices in the catalog. To clarify, when talking about matching it is about matching the purchase price on the order invoice and the purchase price as agreed with buyer and supplier.
When the data entry and handling process is validated sufficiently the quality of the data concerning net purchase prices in the catalog will improve too. If the purchase prices in the system are correct and accurate they can be a reliable input for the consecutive processes. Because this research is part of the overall program 'Margin Management', the positive effect on the accurate calculation of the contribution margin (CM1) is the most important consequence. The difference between the forecasted contribution margin and the actual contribution margin provides insight into the reliability of the net purchase prices.
The figure below shows the causal relationship between the different variables.

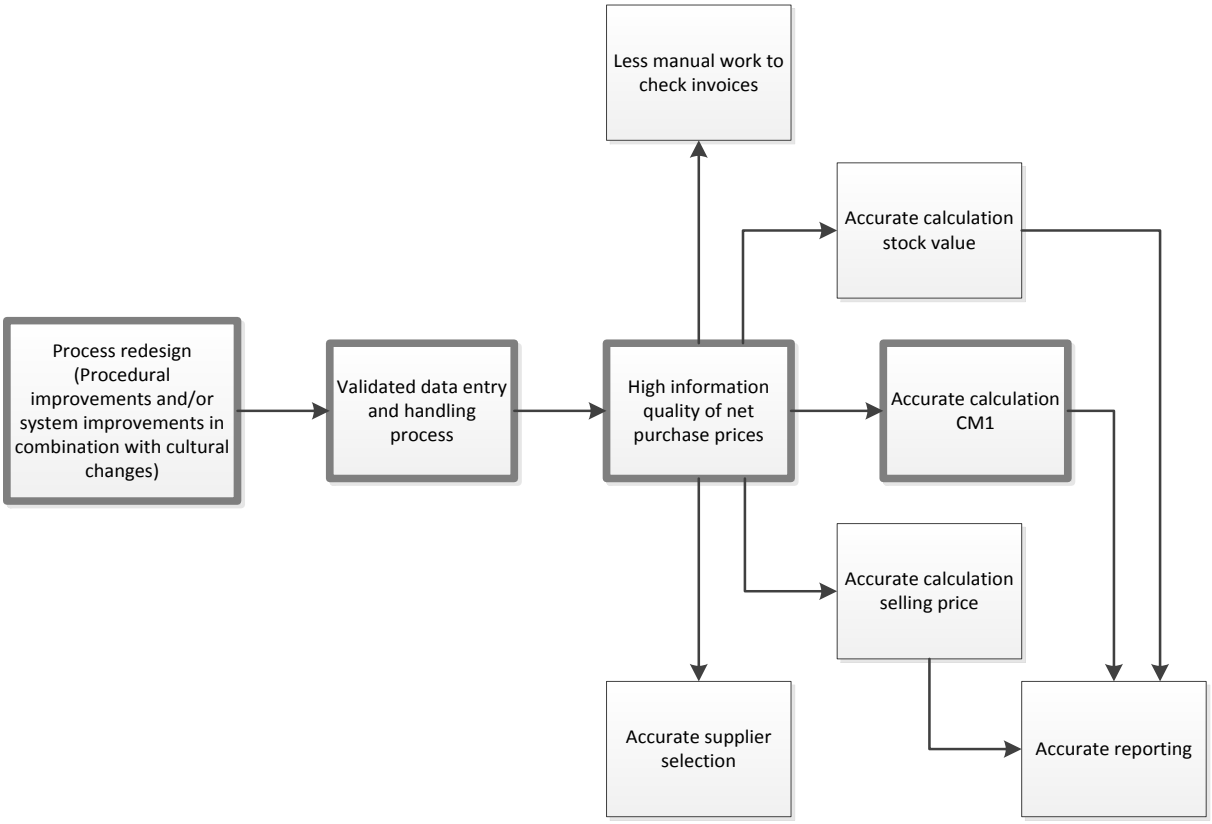


Figure 2.1 Causal relationship between the variables

2.4 Problem owners

The problem owners are the people that are affected adversely. In this research the buyers are the problem owners because they are judged on their daily contribution margin. In the current situation the contribution margin is inaccurate and this is the consequence of unreliable net purchase prices. Besides the financial impact, the work pressure is high. When mismatches occur buyers need to check manually all the mismatches between the price on the invoice and the net purchase price that is agreed. This means that buyers need to do a lot of manual work nowadays. In practice invoices are not always checked because it takes too much time. Furthermore the controlling department is affected adversely, for them it is not possible to calculate and control the contribution margin accurately.

2.5 Sub questions and methodology

To be able to answer the main research question, various sub questions need to be formulated. The answers on these sub questions are supportive for answering the main research question.

The sub questions are:

- What is the current state-of-the-art in data quality and data quality assessment methods?
- What type of data errors can be found in literature?
- How can the data quality systematically be improved and in what way can the effects of unreliable data be eliminated?
- How should the process be specified to assure a reliable and validated data handling process?
- What solutions are available to assure a validated data handling process and do they meet the requirements of the desired solution?
- What steps need to be taken to implement the solutions?

As described in section 2.3 the goal of this research is to give recommendations for improving the data entry and handling process. These recommendations can contain both procedural and system improvements. To reach this goal this research is divided into different stages. The used methodology is based on the Design Science Research Methodology (DSRM) of Peffers et al. (2008). The DS process consists of the next stages:

- Problem identification and motivation
- Definition of the objectives for a solution
- Design of a solution
- Demonstration
- Evaluation

The essence of the stages described in the Business Process Methodology of Weske are comparable with the stages mentioned above (Weske, 2007). It depends on the type of research which stages need to be completed. Four types of design science research are: a problem-centered approach, an objective-centered approach, a design- and development-centered approach and a client-/context-centered approach (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007). The idea for this research resulted from observation of the problem what means that this research can be classified as problem-centered and starts with the first stage called problem identification. The different stages are explained below in more detail and can be seen in figure 2.2.

In the first stage, a literature study is done on the topics of data quality, possible data failures, and data and process assessment methodologies inclusive useful techniques, tools and metrics. Furthermore a case study about the current data handling process of bol.com is executed. The data problems that occur in this process and the causes of these errors need to be analyzed. The literature study combined with the case study will give insight in the current situation.
Then the objective for a design of solution must be defined. The redesign of the process should bring the current situation closer to the desired situation. Based on the information from the first stage, requirements for redesigning the process are formulated.
In the next stage solutions are designed. The input factors for the design are the requirements formulated in the previous stage.

The stage which follows contains the validation. In this stage an estimate is given of how well the redesigned process, which can consist of both procedural improvements and system improvements, meets the desired situation. The redesigned process needs to improve the quality of net purchase prices in the catalog with the result that the CM1 can be calculated more accurate. The validation stage is supported by a literature study about existing evaluation methods. After the validation the implementation plan is described.

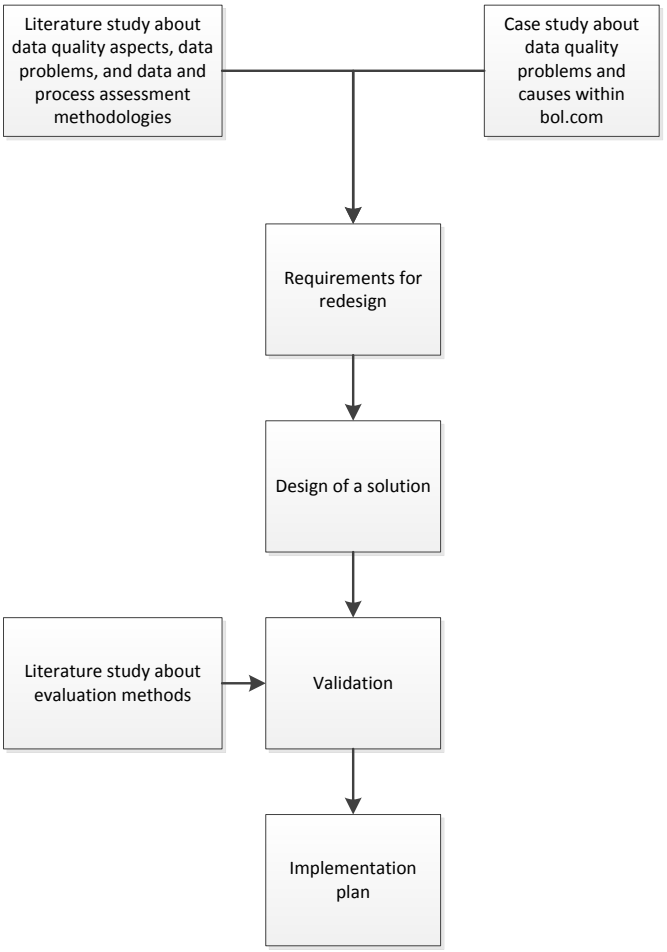


Figure 2.2 Schematic overview of the research methodology and the structure of the thesis

2.6 Reading guide

The first chapter of this thesis gave an overview of the company and in the second chapter the design of the research is outlined. The focus of the third chapter is on relevant literature about data quality aspects, data errors and data assessment methods, while the forth chapter describes the case study about data quality errors within bol.com. The results from the case study are linked to the data quality errors from literature. Chapter five is about changes that are required to reach the desired situation, these requirements are based on the literature study and case study from the chapters before. Furthermore solutions are generated. Chapter six elaborates on the validation of the redesigned process. After the validation stage the implementation plan is described in chapter seven. The research methodology as presented in figure 2.2 is discussed in chapter three till chapter seven, at the end of these chapters a preliminary conclusion emphasizes the highlights of each chapter. Finally in the last chapter overall conclusions are drawn, recommendations for future research are given, and the implications for theory are described.

Chapter 3: Literature study

This chapter gives the state-of-the-art in data quality and data quality problems. Also data quality assessment methods that can be found in literature are discussed and a framework is developed for assessing and improving data quality in an operational context.

3.1 Concept of data quality

A considerable body of literature exists on data quality and the importance of data quality has long been recognized by researchers (Bai, Nunez, & Kalagnanam, 2012). According to several researchers data must be seen as a strategic business asset. Redman (2008) states that data is not just 'the lifeblood of the Information Age', it is the means by which they distance themselves from their competitors (Redman, 2008). Research by Jiang and Zhao has shown that nowadays corporations pay more attention to the quality of data, however in practice many companies still rely blindly on the data in their systems. A lack of data quality has a high economic and social impact on the organization. Poor data quality causes inaccurate information, which is associated with wrong decisions (Jiang & Zhao, 2012). Several studies have shown that poor data quality causes high costs and operational inefficiencies, furthermore research has shown that the high costs of clean data has been ranked as one of the top barriers to strategic business plans. These negative impacts caused by flawed data are called the data quality risks (Bai, et al., 2012; Görz & Kaiser, 2012). The high impact on the organization, both positive and negative, emphasizes the importance of high data quality.

Before continuing the discussion about data quality it is important to know what kind of views can be found in literature and which view regarding the definition of data quality is adopted in this thesis. Different researchers have answered the question how the term data quality needs to be interpreted. Eppler (2006) states that data quality is "the fitness for use of information; information that meets the requirements of its authors, users, and administrators". This statement has exactly the same message as the definition of Wang and Strong (1996). According to Wang and Strong data quality is "data that are fit for use by data consumers". The International Association for Information and Data Quality (IAIDQ, 2010) is using the definition of Larry English on their website. They describe data quality as "the degree to which information consistently meets the requirements and expectations of all knowledge workers who require it to perform their processes".
One component is returning in all the definitions, this is the component which contains the contextual evaluation. The quality of data can only be evaluated in the context of the users, this means that it depends on the user how data quality is assessed. For one person a data element can be appropriate, while the same data element can be really inappropriate for another person. In this thesis the concept of 'fitness for use' is adopted because the task performance of data users is directly influenced by the data.
A completely different definition of data quality is given by Agosta (2002). He states that data quality is a misnomer. Data in itself is meaningless, it is basic raw material. It is what you do with the data that has value. Data is the content; and when it is structured, then it has value as information and the quality of information can be assessed (Agosta, 2002). In view of Agosta there is a distinction between data and information but, as said before, in literature data and information are also used synonymously. In the first three definitions of Eppler, Wang and Strong, and English the terms are used interchangeably and with these three definitions in mind, this research is continued.

3.2 Data quality dimensions

A lot of research is done about what aspects are important for describing data quality. Wijnhoven (2012) discusses the quality of data based on five viewpoints, these viewpoints are derived from the philosophy of knowing, also called epistemology.

- The empirical philosophy of Locke. According to Locke, data needs to be a true representation of reality. The focus is on representational veracity, completeness and meaningfulness of primary data. Failures in this context result in a lack of correspondence with reality and are related to the integrity of data.
- The rationalistic philosophy of Leibniz. From this perspective the focus is on logic and causal reasoning. It is all about derivative data, this in contrast to the empirical philosophy of Locke which focuses on primary data. Failures in this context result in logical inconsistencies, with the consequence of wrong conclusions and decisions, and loss of causal relationships.
- The Kantian philosophy. Kant argues that different perspectives needs to be integrated to produce good data. Failures in this context are related to a lack of perspectives and lack of integration of the different perspectives.
- The dialectical philosophy of Hegel. From this perspective the focus is on integrating different perspectives, both propositions and counter-propositions, resulting in a synthesis of the opposing views. Failures in this context are the result of a lack of understanding of motives of the data producer and a lack of synthesis of opposing views.
- The pragmatic philosophy of Singer. According to Singer, the quality of data should be expressed in terms of how it improves the decision making and in what way it contributes to the construction of a solution. Data is only valuable when it results in progression. Failures in this context are related to a lack of usefulness and practicality. This pragmatic perspective supports the idea that data needs to fit users' requirements (Wijnhoven, 2012).

Besides the abovementioned viewpoints, based on an epistemological approach, the quality of data can also be described from a customer viewpoint. Ultimately it is the data user who judge if certain data is fit for use, this emphasizes the importance of taking a customer viewpoint of quality. The concept of 'fitness for use' is nowadays widely adopted in literature.

To understand what data quality means for data users, Wang and Strong (1996) have developed a framework that contains the aspects of data quality. A two-stage survey is conducted to collect data quality aspects from data users. In the first stage data users were asked to list all data quality aspects that are important from their viewpoint. In the second stage the items in the list are ranked and categorized. The 'fitness for use' concept is reflected in this empirical approach, this in contrast to a theoretical approach or intuitive approach based on the intuitive understanding of the researcher. The resulting framework is shown in figure 3.1. As can be seen Wang and Strong detected four data quality dimensions: intrinsic data quality, contextual data quality, representational data quality and accessibility of data quality. Each quality dimension consists of several data quality aspects (Wang & Strong, 1996). Definitions of the data aspects can be found in appendix A.

- Intrinsic data quality implies that data has quality in its own right. From a traditional point of view, which does not take into account the customer viewpoint, this dimension includes the accuracy and objectivity aspects. From the survey of Wang and Strong can be concluded that accuracy and objectivity are not sufficient for data to be considered as high quality, believability and reputation are also an integral part of intrinsic data quality.

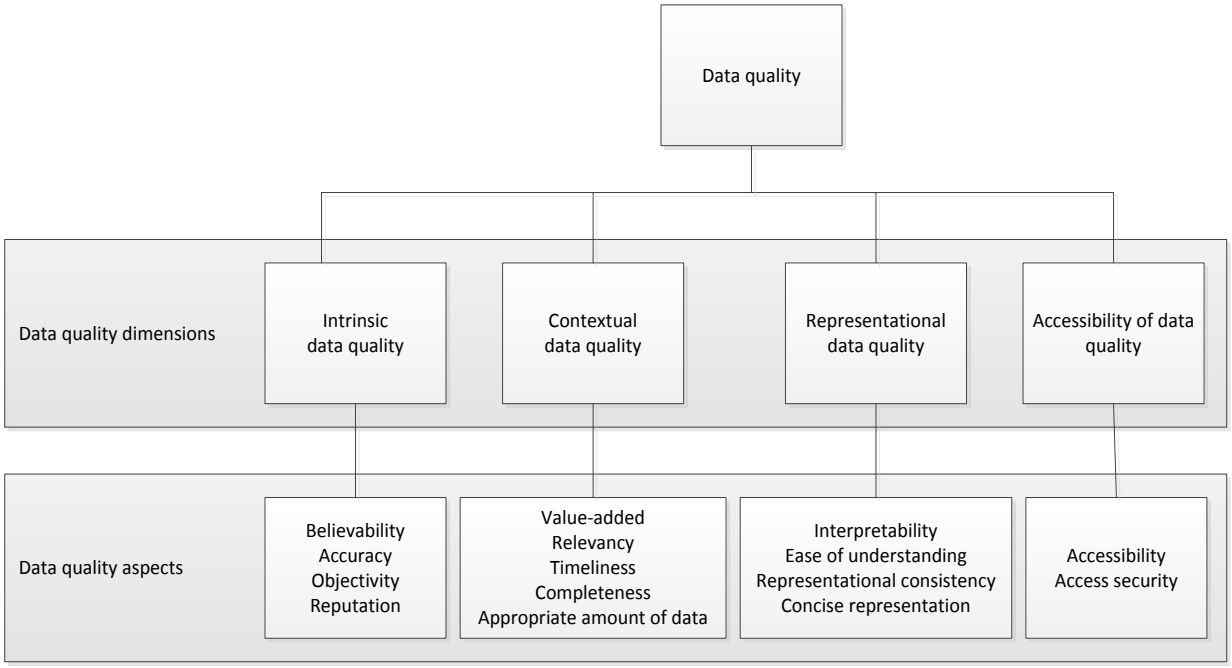


Figure 3.1 Data quality dimensions and aspects (Wang & Strong, 1996)

- Contextual data quality highlights the requirement that data must be considered within the context of the data user's task. The data user judges if a data element is relevant for completing a certain task.
- Representational data quality includes four aspects related to the format of data (concise and consistent representation) and related to the meaning of data (interpretability and ease of understanding).
- Accessibility of data quality is a dimension which is highly recognized from a traditional point of view. According to the survey of Wang and Strong also the data users recognize its importance. The last two dimensions emphasize the importance of computer systems to store and provide access to data in such a way that it is accessible but secure.

Other research on dimensions and aspects of data quality shows a lot of overlap, however there are some aspects which can complement the framework. Batini and Scannapieca (2006) use the same classification of data dimensions and aspects as the ones mentioned in the framework (Batini & Scannapieca, 2006). Furthermore Eppler (2006) conducted a survey on employees in a knowledge-intensive company. This research shows that reliability is also an important aspect, however this aspect is already part of the framework because it is part of the accuracy aspect according to the definition of Wang and Strong. Lee et al. (2002) developed their quality aspects from a theoretical approach, this in contrast to the empirically developed aspects. The aspects they can add to the framework of Wang and Strong are unambiguous and uniqueness (Lee, Strong, Kahn, & Wang, 2002). Here again the unambiguous aspect is already added to the framework, according to the definition of Wang and Strong this aspect is part of the ease of understanding. In the framework of Wang and Strong the uniqueness aspect can be added as part of the representational data quality dimension.

From this section can be concluded that it depends on the point of view which data quality dimensions and aspects are regarded as important for describing and assessing data quality. The framework of Wang and Strong, represented in figure 3.1, shows important dimensions and aspects

according to data users. The five different philosophies which are discussed above are not based on the customer's point of view, however each philosophy correspond with one or more dimensions in the framework. Referring to the quality dimensions and aspects in the framework of Wang and Strong, the following connections between the customer's viewpoint and each of the epistemological perspectives can be seen:

- The view of Locke corresponds with the data quality aspects accuracy, believability, completeness and objectivity. When assessing data quality from an empirical philosophy the focus is on the intrinsic data quality dimension.
- The rationalistic view of Leibniz corresponds with the accuracy aspect, the completeness aspect and the representational consistency aspect in the framework of Wang and Strong, these are part of the intrinsic, contextual and representational data quality dimensions.
- The idealistic perspective is related to the aspects completeness, timeliness, relevance and representation. When assessing data quality from an idealistic philosophy the focus is on contextual and representational data quality dimensions.
- The dialectical view of Hegel corresponds with the completeness, relevance and ease of understanding aspects in the framework of Wang and Strong, part of the contextual and representational data quality dimension.
- The focus of the pragmatic view of Singer is on the contextual data quality dimension.

In the remaining part of this research the dimensions in the framework of Wang and Strong are used as base because these data quality dimensions are the result of collecting data from users, and as stated in section 3.1 the customer's viewpoint, and so the 'fitness for use' approach, is adopted in this research. In chapter four is explained which data quality dimensions and aspects play a key role in this research.

3.3 Data quality problems

In many companies poor data quality is not uncommon. Most of these companies are not aware of the impact of low data quality on their organization. According to Strong et al. (1997) a data quality problem is defined as any difficulty encountered along one or more quality dimensions that make data completely or largely unfit for use. Eppler (2006) as well as Eckerson (2002) have enumerate a number of data quality problems that may occur in an organization (Eckerson, 2002; Eppler, 2006). Typical data quality problems are listed below.

- Obsolete or outdated data. This means that data is not updated according to recent changes.
- Manipulation of stored data. Stored data can be deleted with the consequence of missing relationships. This can result in unclear causal effects in a diagnosis. Furthermore data can be modified. Successive processes may be influenced in a negative way, it can for example lead to confusion. Two forms of manipulation can be distinguished, unintended manipulation is an error with misinformation as a result and intended manipulation results in disinformation.
- Spelling or typing errors. The result is valid data which is not correct. For example when typing a phone number and two numbers are reversed, the result is still a valid phone number but it is not the number that was originally meant. Even validation routines cannot catch small typing errors which represent a valid value.
- Incorrect data entries because of lack of input validation. There are insufficient validation routines to check data that is entered into a system. A lack of adequate validation can result in useless data.

- Duplicates or multiple data sources. Mismatches among sources of the same data can occur, furthermore multiple data sources can result in conflicting recommendations while analyzing data.
- Mismatched syntax or formats. These problems may occur when organizations try to integrate data from multiple systems. Corresponding fields in the different systems may use different sequence of characters , for example first-middle-last name versus last-first-middle name, resulting in a syntax error. Another example of this type of problem has to do with data conversion. ETL (Extract, Transform, Load) tools are used to pull data from multiple systems, transform the data in such a way that it fits operational needs and load it into another system or database. Failure to perform this conversion can result in errors or lost data.
- Wrong data coding or tagging. When data is insufficient or inadequate categorized, it could be difficult to navigate or trace data. The data can be 'lost' in the database.
- Lack of referential integrity checks. Often the integrity of data that is loaded in the end systems is not checked adequately. If during the data handling process for example tables are changed or updated, integrity problems can be created which are not detected.
- Poor system design. When organizations rush to design and implement new systems, sometimes developers take shortcuts. This may result in incorrect data when a lack of attention is paid to for example data types, character encoding or scalability.

Referring to the five philosophies for describing data quality, the first and the third till the seventh problem are in the field of Locke. The second problem is in the field of Hegel. The philosophy of Leibniz is related to the eighth problem and the last problem is in the field of Kant.
The problems that are shown in the list above are related to both the data itself and the data process, this emphasizes the fact that data can be seen as a product as well as a process. The list above is not complete but it shows that a data quality problem can easily arise. To get a clear overview of the problems various researchers have classified the data problems. It is important to recognize the different categories of data quality problems because it can give a direction for a possible solution.

3.3.1 Classifying data quality problems

Some researchers categorize the quality problems according to their dimension. According to this view, problems can be divided into content, time and format problems (Eppler, 2006). In short, content problems are related to biased data, time problems are related to outdated data and format problems are related to edited data. The result of editing or putting together data is that the same data may lead to different interpretations. This approach of classifying problems is narrow in scope. Using this classification not all the problems mentioned in the previous subsection are covered, for example most problems related to the data process are not in scope of these classification.

The IAIDQ (2011) has developed a classification of problems based on symptoms and causes. According to this approach data quality problems can be classified as operational problems, conceptual problems or organizational problems. Table 3.1 summarizes the three different types of problems and the associated symptoms and causes.

Problem type	Symptoms	Causes
Operational problem	Data is missing, inaccurate or invalid	Problems with data capture or transmission
Conceptual problem	Data is missing, inaccurate or invalid	Data is not well defined or not suitable for intended use
Organizational problem	Persistent operational or conceptual problems	Disconnects between organizations that collect and use the data

Table 3.1 Classification of data quality problems based on symptoms and causes (IAIDQ, 2011)

Operational data quality problems occur if data is incorrect, inaccurate or invalid to such an extent that well thought decisions cannot be made. The accuracy and validity of data are playing a key role in describing data quality according to the empirical philosophy of Locke, as explained in section 3.2. Concerning the three types of problems, the pure operational problems are the easiest to solve. Solutions for solving this kind of problems can be found in modifying the data collection method to capture correct and accurate data. For example by using a new technology for data collection or performing checks at the entry. Furthermore these problems can be solved by downstream quality checks, however solving the problem at the source is the most effective.
Conceptual data quality problems often arise when data is used for purposes other than the one the data was originally designed for. Examples of problems are when definitions of data elements are imprecise, when the user does not understand how to interpret the data or when coding is based on local interpretations and can vary from site to site. These problems influence the contextual and representational dimensions of data quality and are the focus of the idealistic philosophy of Kant, as explained in section 3.2. The symptoms of conceptual problems are often the same as the symptoms of operational problems, however the solutions are not similar. Solutions can be found in the field of redefinition of the data element, rethinking the use of data, or utilizing additional data sources to complement the previously collected data.
Organizational data quality problems do have similar symptoms as those of the operational and conceptual problems, but the problems persist over time. Problems can occur when data creators and data users operate in different organizations. It is not a technological or definitional problem and for that reason harder to solve. In literature there are no clear solutions for solving organizational data quality problems, but most of the researchers have recognized the value of communication and agreement between data creator and data user to ensure data quality (IAIDQ, 2011). The dialectical philosophy of Hegel focuses on the agreement and understanding of data creator's motives.
The downside of this classification approach is that the three problem types are very general. Problems related to data itself instead of the process, for example obsolete or outdated data problems, are not covered in these classification.

A more complete and systematic approach of classifying problems is developed by Strong et al. (1997). These researchers classify data problems along the life cycle of data. The life cycle consists of data production, storage and maintenance, and information use. The table on the next page shows the data quality problems classified along the data life cycle according to Strong et al. It shows also a general direction for a solution. The data quality problems listed in section 3.3 are covered in the eight problems mentioned by Strong et al. (1997).

Problem	Stage in life cycle	Description	Consequences	Solution
Multiple sources	Production	Multiple sources of the same data produce different values and lead to confusion and less credibility	Use of data declines. Users distrust data. Maintenance is more difficult and costly	Develop common definitions and consistent procedures
Subjective production	Production	Data is produced using subjective judgments, leading to bias	The objectivity of the data decreases and data is difficult to evaluate	More training, better rules, expert systems
Production errors	Production	Systematic errors in production leads to lost information	Data searching and correcting increases	Process improvements, incentives, controls
Too much information	Storage and maintenance	Large volumes of stored data make it difficult to access data in a reasonable time	Excess time is required to extract and summarize data	Analyze data needs, develop regular, frequently extracted subsets of relevant data
Distributed systems	Storage and maintenance	Distributed, heterogeneous systems lead to inconsistent definitions, formats and values	Data can no longer be easily combined, due to the format differences and incompatibilities	Integrate systems in one platform or reduce the amount of systems
Changing task needs	Storage and maintenance as well as use	As users tasks or the organizational environment changes, the data that is relevant and useful changes.	Mismatches develop between available data and what is needed for tasks	Anticipate changes in tasks and revise processes and systems before the mismatch becomes a crisis
Security and privacy requirements	Use	Easy access to information may conflict with requirements for security privacy and confidentiality	Mechanisms for security block or delay access, so the information provides less value	Develop consistent policies and procedures for secure information
Lack of computing resources	Use	The IT infrastructure is insufficient and limits access to information	Knowledge workers are demotivated and cannot work productively	Develop technology upgrade policies so users know when to expect more resources

Table 3.2 Classification of data quality problems along the life cycle of data (Strong, Lee, & Wang, 1997)

3.3.2 Data quality problems related to data quality dimensions

The systematic approach of classifying problems along the life cycle of data makes it possible to relate problems to the data quality dimensions and aspects which are mentioned in section 3.2. In the figure below the eight data problems, classified along the life cycle as can be seen in table 3.2, are linked to the data quality dimensions they influence. Both, the quality dimensions and the data problems, are the result of research by Wang and Strong.

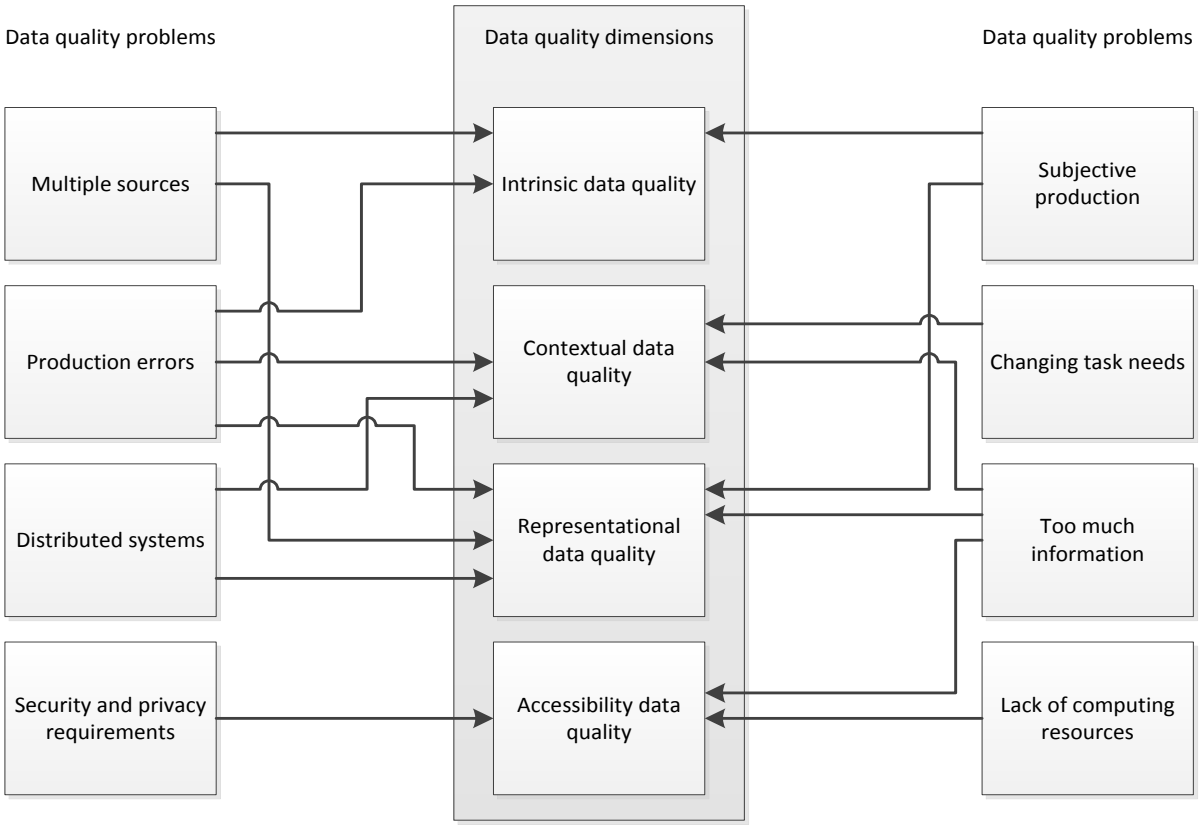


Figure 3.2 Relationship between data quality problems and data quality dimensions (Strong, et al., 1997)

The problems that are described in literature, combined with the quality dimensions they influence and the directions for solutions provide a basis for evaluating the data quality problems at bol.com. The next chapter represents a case study about data quality problems which occur during the data entry and handling process within bol.com, these problems can be linked to the data quality dimensions found in literature and to the problems summarized in table 3.2. The table also provides a direction for solutions, this could be a starting point for improving the data quality although the directions for solutions in the table are very general. A more systematic way for assessing data quality and consequently improving the quality of data is necessary. In literature a wide range of methods can be found for assessing and improving the quality of data.

3.4 Methods for assessing and improving data quality

By using assessment and improvement methodologies we strive to assure the data quality in present and in future. This concept of 'data quality assurance' is described as all those planned and systematic actions necessary to provide adequate confidence that a data product will satisfy a given set of quality requirements (Wang, Storey, & Firth, 1995).

Batini et al. (2009) summarizes and compares the various data quality methodologies which exist in literature. For convenience in this section only the abbreviations of these methodologies will be used, a table with the full names can be found in appendix B. The methodologies differ on the following points:

- Methodological stages and steps. In the most general case a data quality methodology consist of the state reconstruction stage, the assessment and/or measurement stage and the improvement stage. In this context the assessment term means that measurements are compared to reference values. The improvement stage can consist of cost evaluation, assignment of process responsibilities, strategy and technique selection, design of improvement solutions, process redesign, improvement management and improvement monitoring. Which steps are completed depends on the methodology. The selection of the methodology and the corresponding steps are carried out in the end of this section.
- Strategies and techniques. In general there are two types of strategies for improving the data quality, called the data-driven strategy and the process-driven strategy. Data-driven strategies improve the quality of existing data while process-driven strategies improve the data quality by redesigning the process which create, modify and store the data. Referring to the main research question, in this research we strive to improve the data quality by redesigning the process. This is also visualized in figure 2.1. Main techniques to support the process-driven strategies are process control, by inserting checks and control procedures, and process redesign.
- Data quality dimensions. As can be read in section 3.2 the most important quality dimensions are provided by Wang and Strong (1996) and most of the researchers agree with these dimensions. The dimensions support the 'fitness for use' concept.
- Costs. The costs of poor quality can be divided in process costs, which represents the cost for repeating processes because of data errors, and opportunity costs, which refer to missed revenues. In this research we will not evaluate the costs of poor data quality in detail. This because every data quality problem consist of different degrees of errors and so they do have a different impact on the costs. A typing error for example has a higher impact when the error is made before the comma than when the error is made behind the comma. Only cases will be evaluated and based on these cases only a rough estimate of the missed revenues can be made.
- Types of information systems. Based on the degree of integration their exist a monolithic information system which do not share data, a distributed information system which addresses also data exchanges between systems and workflows, and a data warehouse which represents a centralized collection of data retrieved from various databases. The data quality method which can be applied to data warehouses (the Data Warehouse Quality Methodology) focuses on higher quality of existing data by use of a data-driven strategy. On the other hand, data quality methodologies for distributed information systems focus on techniques for process improvement by use of process-driven strategies (Batini, Cappiello, Francalanci, & Maurino, 2009). In this research we are talking about a distributed information system. The process-driven strategy that is used for these kind of information systems corresponds with the aim of the main research question.

All the methodologies focus on different data quality issues. The differences in focus can be recognized at a glance by dividing the methodologies into four categories. For classifying the

methodologies, two of the above mentioned dimensions can be used. Since the aim of this research is to improve the data entry and handling process, it is necessary that the focus is on both the assessment and the improvement stage. A methodology which does not focus on the improvement stage is unsuitable. The first dimension is therefore about whether or not to address the improvement stage. The other dimension is about whether or not to address the cost issues, also called the economic issues. As stated before in this research we will not evaluate the costs of poor data quality and the opportunity costs. The result is shown in figure 3.3. Four quadrants are shown and for this research a methodology within the fourth quadrant is suitable. Important to note is that another meaningful dimension is the used improvement strategy, so a process-driven strategy or a data-driven strategy. However if this dimension is used as one of the two dimensions the methodologies which does not focus on the improvement stage, but only focus on the assessment stage, are not part of the figure. The resulting figure will not give a complete overview of all the methodologies and the differences in focus cannot be recognized at a glance. This is the reason that this dimension is addressed later. In figure 3.3 the following quadrants are shown:

- The first quadrant shows methodologies which focus on the improvement stage, and logically also on the preceding assessment stage, and which address economic issues.
- The second quadrant shows a methodology which focuses not on the improvement stage, so the assessment stage is the last stage for this methodology, and which addresses economic issues.
- The third quadrant shows methodologies which focus not on the improvement stage, so the assessment stage is the last stage for this methodology, and which does not address economic issues. The methodologies from the second quadrant and the third quadrant are the so-called audit methodologies.
- The fourth quadrant shows methodologies which focus on the improvement stage, and logically also on the preceding assessment stage, and which does not address economic issues.

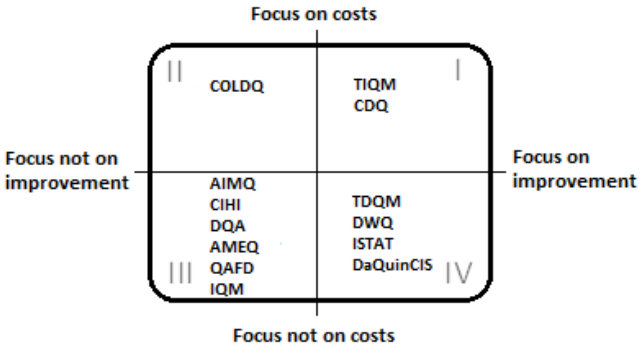


Figure 3.3 Classifying DQ methodologies

Referring to the main research question the aim of this research is to redesign the process in such a way that the quality of the net purchase prices in the catalog of bol.com will improve. To reach this goal a methodology which focuses on the improvement stage is suitable, furthermore the cost issues will not be addressed, from this point of view only the methodologies in the fourth quadrant are interesting. There are four methodologies in this quadrant, TDQM, DWQ, ISTAT, DaQuinCIS, of which only TDQM is suitable for executing this research. TDQM is process driven, it covers the most important steps in the assessment and improvement stages (for example identification of critical areas, process modeling and improvement management), it is based on the 'fitness for use' approach because this approach addresses the process modeling step and the data quality measurement step, and it focuses on distributed systems (Batini, et al., 2009). All these requirements are playing a key role in this research. The other three methodologies, DWQ, ISTAT and DaQuinCIS, fail in satisfying at least one of these requirements. DWQ and DaQuinCIS adopt a data-driven strategy instead of a process-driven strategy, and ISTAT focuses on the assessment stage and the improvement stage but

not in detail. Important steps like identification of the critical areas, improvement management and improvement monitoring are not addressed while TDQM does address these steps. Full names of these methodologies can be found in appendix B.

3.5 TDQM methodology for operational systems

TDQM, in full name Total Data Quality Management, is derived from the Total Quality Management (TQM) concept. This is an approach for long-term success which focuses on continuous improvement of products and processes. TDQM is established in the 1980's and from that moment on a lot of researchers pay attention to this method, it is among other things explained in Wijnhoven et al. (2007), Batini et al. (2009) and in Eppler (2006). Although a lot of research is done on this methodology, it is not easy to apply in practice. This because the four basic stages cannot be followed blindly, to apply the method in practice some contextual redesign is required.
In this research we strive to improve the quality of net purchase prices in the catalog by redesigning the process. Net purchase prices pass different systems and it is essential to understand the complete set of processes and how the quality of data at each processing stage is impacted by its quality at previous stages. The TDQM methodology should help assessing data quality not only at the final stage but also at any of the preceding intermediate stages. To reach this a TDQM framework that can be applied to operational systems need to be developed. Although TDQM is discussed by many researchers, the appliance in operational context is not yet addressed.

For developing a TDQM methodology which can be applied in an operational context, the four basic stages, called definition, measurement, analysis and improvement, are used, but the details and execution of the different stages differ from the traditional TDQM methodology. The methodology is complemented by steps from the BPE approach. This approach focuses especially on business process engineering (BPE), which is the term for designing, developing and managing business processes, organizations and systems (Van den Berg, Franken, & Jonkers, 2008). Van den Berg et al. presents the BPE approach which consist of methods and tools with the aim to assist in translating business objectives in innovative business processes. The BPE approach is suitable for the operational context but provides for some steps vague definitions and actions. The combination of these two methodologies, adapted to the needs of the specific environment, results in the framework described below. The needs can be discovered by for example interviews with the stakeholders. Figure 3.4 shows the different stages, as can be seen the execution of the different stages is an iterative process.

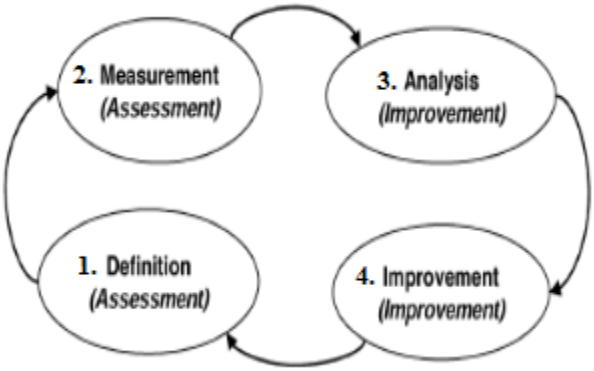


Figure 3.4 Stages TDQM (Batini, et al., 2009)

- Stage 1: definition. In the definition stage the most important data dimensions are identified. The most important dimensions could be the ones which have the highest impact on the quality problems and on the organization. A score sheet can be used to discover which data quality dimensions are the most important in view of the different stakeholders. The stakeholders are the data producers, data manufactures, data consumers and data managers. The next step is to define the data quality requirements from the perspective of

the stakeholders. Questions about the expected quality and perceived quality of the selected data quality dimensions are answered, and these quality dimensions with the biggest gap between expected and perceived quality need improvement. The last step is to define the process model. Knowledge of the process helps understanding why certain quality dimensions are more important than others. The following three approaches for modeling exist, process-oriented modeling, resource-oriented modeling, and data-oriented modeling. The various modeling approaches differ in sequence of modeling steps, one model starts with the process, while another model starts with the actors and the last model starts with the items. It is helpful to define process models on different levels. A model on a high level helps visualizing the relationship between the operational systems while a model on a lower level helps understanding the details of an operational process. Depending on the aim of the researcher/analist, modelling the process on different levels of detail is a valuable addition to the TDQM methodology to make it applicable in an operational context. The process can be modeled with for example MS Visio or by using a BPMN (Business Process Model and Notation) modeler like Bizagi. From above can be stated that for completing the definition stage successfully both the stakeholder approach and process model approach need to be applied.

- Stage 2: measurement. In the measurement stage data quality metrics are defined, using these metrics the quality of certain data dimensions can be measured. Three quantitative measurement forms are: the simple ratio, the min or max operation, and the weighted average. These quantitative metrics need to be reasonable, understandable, measurable, believable and achievable, also called the RUMBA guidelines (Wijnhoven, Boelens, Middel, & Louissen, 2007). In the operational context it can be hard to use objective metrics for measuring requirements, to overcome this issue metrics can also be subjective for measuring expectations. This distinction between objective and subjective metrics is described by Eppler (2006) and is a valuable addition for the operational context. Another distinction between objective and subjective measurement is that subjective measurement is based on qualitative evaluations by stakeholders while objective measurement is based on quantitative metrics. The next step is to measure and present the data. Tools for presenting data are for example pie charts or Pareto diagrams, which display the percentage of an error compared to the total number of errors. The result of the measurement stage is that problem areas arise.
- Stage 3: analysis. In the analysis stage the data quality problems found in the previous stage are analyzed and the root causes of discrepancies are identified. For identifying root causes all the operational systems and the associated processes need to be examined. This means that not only the final process is examined but also any of the preceding intermediate processes and systems. A cause and effect diagram can help analyzing the root causes. Other commonly used tools for analysis are interviews, models, or case studies. A critical step that is missing in the standard TDQM methodology is identifying which causes have the highest priority for improvement. This can be identified by relating the causes to the objective of the research, it shows the impact of the causes. A tool for identifying the priority is the risk assessment matrix, this is a tabular illustration of the impact and frequency of errors (Rausand, 2011).
- Stage 4: improvement. In the improvement stage first of all the size of the changes is determined, these can vary from doing nothing till redesigning of the complete business

processes. Furthermore the impact of the changes are determined and a project team is assigned depending on the size of the redesign process. Subsequently requirements for redesign are defined. These requirements are formulated with help of insights in the current situation, insights in the desired situation, and in case preconditions are formulated these must also be taken into account. Also the size of the changes are determined. These important steps are missing in the standard TDQM methodology but are part of the BPE approach. After this step strategies and techniques for improving are selected. When selecting solutions it is important to have insight in the required resources and to understand what the impact is on the process and the organization. A score sheet can be used to score the different alternatives. Subsequently the selected solution need to be translated into specific actions, these actions can be presented in a Project Action Plan. Team members need to be assigned to execute the action and to keep track of the status of the actions. Last but not least the improvement needs to be managed and monitored.

The abovementioned TDQM methodology is made applicable for operational processes by adapting, improving and adding several steps. In summary, critical steps that are added are:

- Identification of the causes which have the highest priority for improvement, as part of the analysis stage. Also a new tool for these identification is introduced.
- Definition of the requirements for redesign as part of the improvement stage.

Furthermore the methodology is adapted and/or improved on the following points:

- The process modelling step, as part of the definition stage, is described in more detail. Depending on the aim of the researcher processes can be modeled on different levels; a high level model which shows the relations between the operational systems and a low level model helps understanding the details of an operational process.
- A new way of measuring and/or assessing data quality is introduced. The subjective metric could be useful in case it is hard to use objective metrics for measuring the quality of operational processes. A subjective measurement is based on qualitative evaluations by the stakeholders.
- The analysis stage is defined in more detail. For identifying root causes all the operational systems and the associated processes need to be examined, meaning not only the final process.

The improved TDQM methodology, suitable for operational processes, is used in the next chapters. The aim of the case study in chapter four is to analyze the problem areas in the data entry and handling process of bol.com. For executing this case study the first three stages of the TDQM methodology for operational systems are completed. When these problem areas are known the required types of changes, which are necessary for reaching the desired situation, are described in chapter five. In the same chapter the remaining steps of the improvement stage are discussed in detail.

3.6 Preliminary conclusion

In literature the topic of data quality has been discussed many times and the importance of high data quality has long been recognized by researchers. The data quality problems that may occur within a data process are numerous. To solve these problems and to improve the data quality systematically many data quality methodologies are developed. Referring to the research question only the TDQM

methodology is suitable, although some contextual redesign was required. A TDQM methodology which can be applied in an operational context is developed, based on a combination of the standard TDQM methodology and the BPE approach. The steps of this methodology are followed in the next chapters.

Chapter 4: Case study

In this chapter the data entry and handling process at bol.com is described in detail. For assessing the data quality the TDQM methodology, adapted to the operational context, is used. The first three stages of the methodology are completed with the aim to get more insight in the entry and handling process and to find the critical areas in the process. Subsequently these problems are linked to the literature on data quality problems. The last stage of the TDQM methodology for operational systems, consisting of the improvement stage, is completed in the next chapter.

4.1 Definition stage

The first step in the definition stage is to identify the data quality dimensions and data quality requirements. Subsequently we need to get more insight in the data entry and handling process at bol.com, this helps understanding why certain quality dimensions are more important than others.

4.1.1 Data quality dimensions and requirements

For describing the quality of the net purchase prices we refer to the data quality dimensions and aspects explained in section 3.2. Because not all the dimensions mentioned in literature can be assessed in one time effort and not all the dimensions are relevant in this research, the focus is on the dimensions which play a key role in the data process of bol.com. Interviews with stakeholders are conducted to discover the most important dimensions and aspects for describing and assessing the quality of the net purchase prices. The result of the interviews can be seen in the table below, this table shows the most important dimensions and aspects, together with the requirements. The last column lists the expected requirements, this in contrast to perceived requirements. When the correct monitoring tools are available it can be controlled whether the requirements are met. The stakeholders in this case are business analysts from the Product & Content department and buyers from the Buying & Merchandising department because these departments are together responsible for the correct net purchase prices in the catalog and subsequently the correct selling prices on the website. Also controllers from the Controlling department are stakeholders, they are influenced by the quality of net purchase prices in the catalog because it impacts the CM1. Other data aspects which are detected by Wang and Strong but which are not listed in the table below, for example access security, are during the interviews with stakeholders not marked as relevant for assessing the quality of net purchase prices in the catalog.

Dimension	Aspect	Requirements
Intrinsic data quality	Accuracy	Net purchase prices need to be free of error. This is especially important when entering offers manually.
	Reputation	Net purchase prices need to be from a trusted source. Suppliers who manipulate the data need to be corrected.
Contextual data quality	Timeliness	Net purchase prices need to be updated on time. This is especially important when entering offers manually and when offers are overridden by purchase deals.
	Completeness	Data which is entering the system needs to be of sufficient depth, breadth and scope. This means for

		example that beside the net purchase prices also the units they belong to need to be mentioned.
Representational data quality	Ease of understanding	Net purchase prices need to be easily comprehended without ambiguity. For example when entering manual offers, it needs to be clear if agreed conditions are already processed.
	Interpretability	It needs to be clear to what units the net purchase prices belong to.
	Representational consistency	Net purchase prices need to be presented in the same format and need to be compatible with previous data. Also consistency between the different systems is important.
Accessibility data quality	Accessibility	Lists of net purchase prices need to be available when updating is necessary or need to be easily retrieved on the required time

Table 4.1 Important data quality dimensions in the data entry and handling process of bol.com

4.1.2 Process description

To improve the quality of net purchase prices in the catalog we need to get more insight in the data entry and handling process at bol.com. The data flow from entering the system till the invoice matching is showed in the flow chart in figure 4.1. The figure shows the relationships between the different systems, furthermore it shows the scope of the research.

In the upper left corner of figure 4.1 two different processes can be seen, both processes use the platform of bol.com to sell their products. The two processes are plaza sales and the regular sales.

- Plaza sellers are using the Seller Dashboard (SDD). This dashboard enables individuals to exchange second hand products and it gives customers the access to products from other retailers, the so called plaza offers.
- Regular sales refers to suppliers that deliver their products to the customer via bol.com.

As can be seen in the figure, the plaza sales is out of scope. This research will only focus on the regular sales, this can be explained by the fact that the net purchase prices of suppliers are only relevant in case of the regular sales.

Suppliers can deliver their purchase prices and delivery times in different ways, the difference is in manual or automatic feeds and in large suppliers or small suppliers (in terms of bargaining power).

- Small suppliers use LIS offer (Leverancier Information System) for delivering the offer data into the system. Using this system the suppliers fill a specific format that is sent to them and by use of an automatic feed the data is entering into the system. In case the supplier changes the price or delivery time, this change will be processed automatically. The number of suppliers that is connected to LIS offer is around 155 (last data update is from July 2013).
- Large suppliers use custom feeds to enter their offer data into the system. This means that bol.com adapts their system to the format of the supplier. In case the supplier changes the input data, this data is processed automatically. The number of suppliers that is using custom feeds is around 50 (last data update is from July 2013).

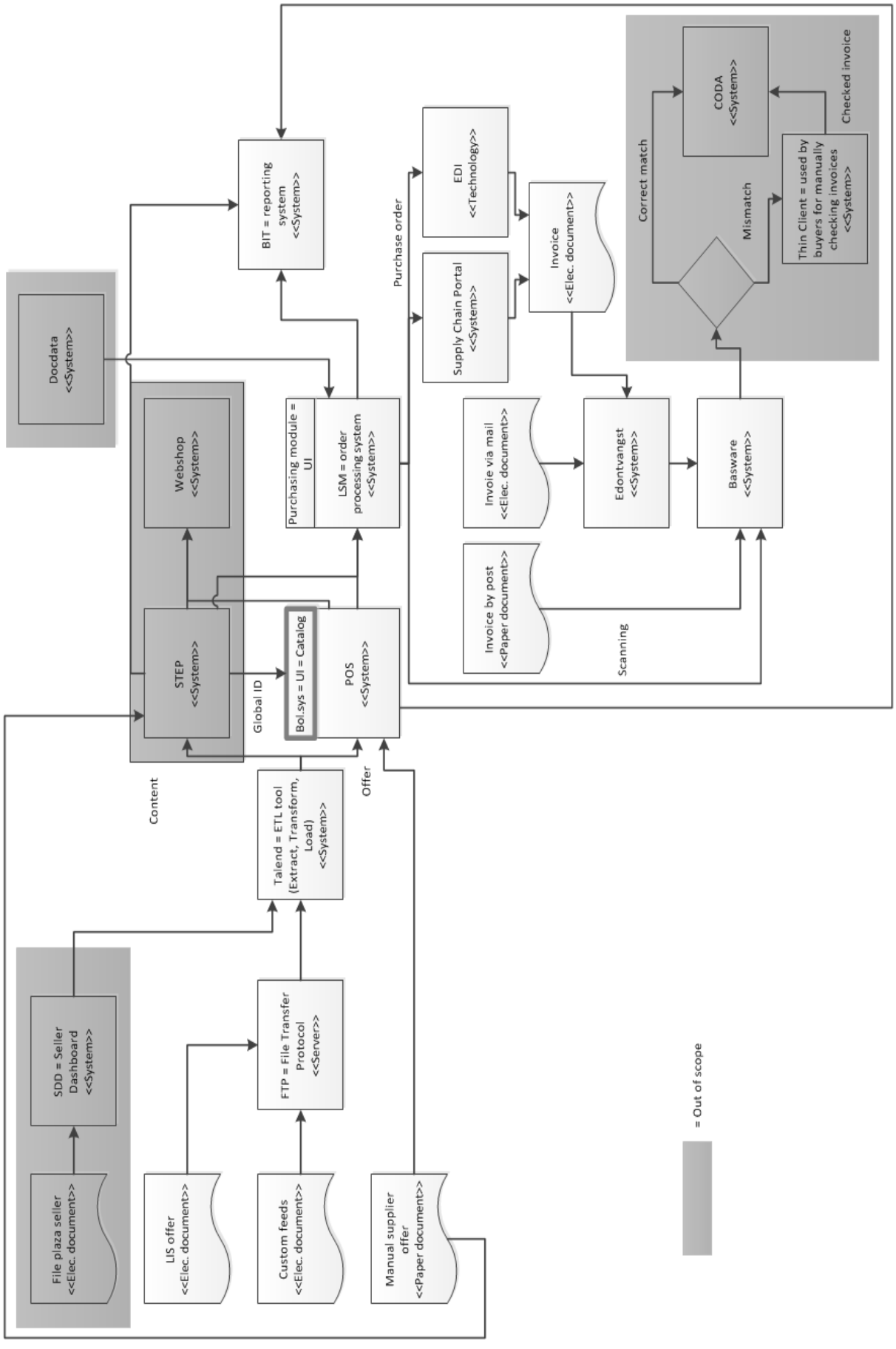


Figure 4.1 Flow chart

- The last category is manual supplier offers. This means that the buyers add the supplier offers to the offer system manually. Manual offers are used when suppliers do not have an automatic connection, reason could be that suppliers are not able to deal with the technology or the contract for an automatic connection is for some reason not yet signed. Suppliers deliver their data by mail. In case the supplier wants to change the input data, this can be done by mail or by phone. Nowadays bol.com is dealing with around 530 suppliers (last data update is from July 2013), given the data above it can be stated that many supplier offers are added manually. This method is more sensitive for errors and the aim is to connect all the suppliers automatically.

Subsequently the function of the FTP server (File Transfer Protocol) is to exchange the files between supplier and the buyer of bol.com. The data delivered by the suppliers is visible for the buyers via the FTP server. This FTP server is used for the data that is delivered through LIS offers or custom feeds.

Afterwards all the input data comes together in the system called Talend. This is an ETL tool (Extract, Transform and Load) where the data passes three processes. First the data is extracted from the different input systems, subsequently the retrieved data is transformed, in this way the data meets the format of the end goal and the last stage is loading the data in the consecutive system. As can be seen in the figure, the consecutive systems are STEP and POS.

- STEP (product catalog) is the system that stores the product content like titles, authors and product descriptions.
- POS (Product Offer System) is the system where the product offers are stored. The product offers consist of the net purchase price and the delivery times and based on these data the preferred supplier is calculated.

The user interface of POS is called Bolsys and this is the catalog of bol.com. Bol.sys gives the users access to the backend systems. This system is used for the next data processes:

- Entering pre-offers; This is the case when suppliers notify that a product will be launched in a couple of weeks and the buyers want to buy the products in advance to possess stock when the product is being released. The data is not yet available for LIS offer or custom feeds so the supplier offer is entered manually in POS.
- Manual supplier offer; This is described on the page before and has to do with suppliers who deliver their data by mail or by phone.
- Supplier offer overrides (also called Purchase Deals); Overriding is the term for manual changes in purchase prices or delivery times in case of temporary Purchase Deals.
- Selling offer overrides: temporary overrides of the list price and selling price in case of Promotion Actions.
- Translations; This is related to for example translating discount codes provided by the supplier in discount percentages or translating tax codes in tax percentages. Suppliers offer the codes but these need to be translated in data to calculate the price for the customer.

The STEP system has its own GUI (Graphical User Interface), this interface is used to communicate with the backend system. In accordance with the principal and as can be seen in the figure, the product content is not in the scope of this research.

After all the product data and offer data is stored in POS and STEP, the data flows to three different directions: the website, the order system LSM and the reporting system BIT.

- Website; The product content, selling price and stock availability or delivery time of the preferred supplier is transferred to the website. From now on the product is available for customers.
- LSM (Local Sourcing Module); This system is used for the order processing.
- BIT; This is the reporting system. The data input for BIT is from LSM, POS and STEP. Examples of reports that are processed are sales reports for each category and reports with price mutations based on supplier number or EAN number.

After an order is placed, the supplier will send an invoice with the number of ordered products and the associated purchase prices. For the suppliers there are different ways to send the invoice.

- Mail; this is a PDF file. This method cannot be used anymore after June 2013.
- EDI (Electronic Data Interchange); This data exchange method is used by the larger suppliers. The tool OCR (Optimal Character Recognition) is used to check if the file includes all the data and can be send to the consecutive system. Around 30% of the suppliers is using EDI for receiving orders (last data update is from July 2013).
- SCP (Supply Chain Portal); This delivery method is used by the smaller suppliers. An invoice is made based on the data of LSM. It is a self-billing system what means that the invoice is generated for the supplier, both suppliers and buyers can make changes on the proposed invoice. These changes are sent back to the order system LSM and are stored in this system. Around 70% of the suppliers is connected to the SCP for receiving orders. This means that for receiving orders all the suppliers do have an EDI or a SCP connection (last data update is from July 2013).
- Manually; Most of the suppliers are sending invoices only manually, this is around 60%. Around 28% of the suppliers is sending invoices both manually and automatic. The reason for sending invoices in two different ways is to make sure that the invoices will arrive at bol.com. In exceptional cases the automatic way of sending invoices is not working. The aim is that all invoices are sent in an automatic way.

All the invoices that are delivered by EDI and SCP flow to the system EDontvangt and subsequently it flows to the system called Basware, this is the system where the matching takes place. The matching is done on the following aspects: the ordered number of products, the net purchase prices, and the EAN numbers. As stated before, the focus of this research is only on the net purchase prices. This part of the matching has to do with checking automatically the net purchase prices on the invoice sent by the supplier with the net purchase prices in the order system LSM. The consequence of the fact that the matching takes place on different aspects is that mismatching percentages does not say anything about the price matching on the invoice.
The invoices that are sent manually are scanned into the system and these invoices also end up in Basware. However these invoices are only stored in the system and these cannot be automatically matched with the purchase price in the order system.
After the matching process the payment process takes place. As can be seen in figure 4.1, the focus of our research is the invoice matching based on net purchase prices and the whole data flow before. The payment process is out of scope.

As stated in section 3.5, depending on the aim of the research process models on different levels of detail are necessary. Essential for the remaining part of this research is to show in more detail the process of entering supplier offers manually, this process is modeled in figure 4.2. As can be seen existing reports, about for example net purchase price mutations in the catalog, are not used by the buyers. When offers are entered manually there is no check if these offers are entered correctly. In a later stadium, section 5.3, we will refer to this figure and the associated information.
For modeling this process a BPMN modeler is used. BPMN is a notation that is readily understandable by all business users (White, 2004).

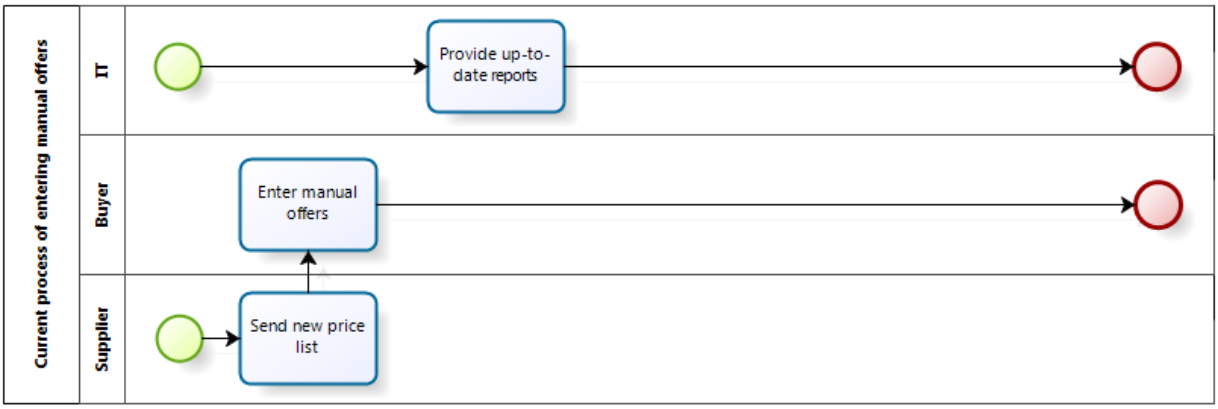


Figure 4.2 Current process of entering manual offers

4.2 Measurement stage

As stated in section 3.5, to apply the TDQM methodology in practice some contextual redesign is required. In this case the measurement stage will be replaced by the assessment stage. According to Batini et al. (2009) the term measurement is used to address the issue of measuring the value of a set of data quality dimensions, while the term assessment is used for comparing measurements to reference values, in order to enable a diagnosis of quality. These reference values can be subjective or objective, this distinction is made by Eppler (2006). A subjective method is used for meeting expectations, this in contrast to an objective method which is used for meeting requirements.

For assessing the quality of the net purchase prices in the catalog of bol.com two stages are completed. The first stage is collecting and analyzing cases of mismatches between net purchase prices in the order system of bol.com, called LSM, and the price on the invoice. The aim of this approach is to discover where in the process errors occur with mismatches as a result. Only comparing the invoice prices with the order prices in LSM is not enough to draw conclusions. The invoice matching is at the end of the data handling process and it is possible that both the invoice price and the order price match while both prices are not as agreed between supplier and buyer. In a case like this the net purchase prices in the order system are not reliable. For this reason the objective assessment method is also applied at the beginning of the dataflow where the data is entering the catalog, this is the second stage. For each supplier the buyers are asked to take a sample of the assortment and to check if the prices in the catalog are the same as agreed between buyer and supplier. This part is called the validation and could be an indication for the current reliability of the net purchase prices in the catalog.
The requirement in both the current and desired situation is that the order price in LSM matches with the net purchase price on the invoice and that both prices match with the price as agreed

between buyer and supplier. Used tools for collecting data are cases and interviews with the buyers from the different product categories.

Cases of mismatches are analyzed from in total 207 suppliers, these suppliers are prioritized based on their purchase value. The cases are collected in three different ways:

- The invoices of suppliers with an automatic connection (EDI or SCP) are matched in Basware. With a query the mismatches based on price differences are filtered. To be clear, price differences are the differences between the net purchase price in the order system LSM, which is filled by the prices from the catalog, and the price on the invoice.
- Cases of mismatches are submitted by the buyers who's responsibility it is to check the invoices. This approach is used for suppliers who sent their invoices manually. These invoices cannot be matched in Basware with the consequence that mismatches in manual invoices can only be discovered if the invoices are checked manually.
- In case buyers argue that there are no mismatches found, and both the net purchase prices in the chain and the suppliers behave as desired, we did checks on invoices by ourselves to find cases to analyze.

All the cases with price differences are discussed with the buyers. The buyers make agreements and purchase deals with the suppliers so in case of price differences they need to know if the prices in the catalog and the order system LSM are correct or if the price on the invoice is correct. In this research the focus is on the data entry and handling process within bol.com so the cases where the invoices are not correct, owing to mistakes of the supplier, are not interesting. The aim is to find the critical areas in the process, so the focus is on the cases where the invoice price is correct and the purchase price in the order system is incorrect. Together with the buyers these cases are analyzed and evaluated, and the critical areas are mapped. The results are shown in the next section.

An alternative way of assessing the quality of net purchase prices in the catalog is to measure the difference between the expected CM1 and the actual CM1. A difference in expected and actual CM1 is the result of incorrect net purchase prices because the contribution margin, the marginal profit per unit sale, is calculated as the selling price minus the purchase price, where the selling price is based on the purchase price. However the difference in expected and actual CM1 does not give a complete picture because incorrect net purchase prices in the catalog can result in a higher or a lower net purchase price on the invoice, and positive and negative differences in net purchase price can cancel each other what means that in theory it is possible that incorrect net purchase prices in the catalog result in an actual CM1 which is equal to the expected CM1. It can be stated that a difference in CM1 is just an indication for incorrect net purchase prices, besides these data is confidential and this are the reasons for not including it as measure.

4.3 Analysis stage

In the assessment stage the root causes of price differences between the order price and the invoice price are discovered. These causes, which are encountered during the case studies and the interviews, are listed below.

1. Manual offers are entered too late. There is accordance between buyer and supplier about the date from which an offer is valid, and this problem occurs when the offer is entered after the agreed date. In this case a timing issue occurs which results in different order prices and invoice prices.

2. Manual offers are incorrect, they are not free of error. This can happen in case of typing errors and has consequences for the reliability of the net purchase prices.
3. Incorrect override in the order system LSM. This means that manual changes are made in the net purchase prices in the order system. When a purchase price is changed in LSM it is only used for a specific order and this price is not used for the same product offered by the same supplier on another order moment.
4. Purchase deals are in the system for a longer time than agreed. This is an incorrect override in Bolsys and this is especially an issue with one-off deals based on the number of purchased units. For example when there is an agreement between supplier and buyer that X number of products can be ordered for 20 percent off, the buyer needs to count manually when X products are ordered and after this the net purchase price needs to change back manually to the regular price. The result is that orders are placed for the deal price while the regular price is valid and is invoiced, a mismatch occurs.
5. The system does not support the possibility to enter different discount percentages for products based on the number of purchased items, also called quantity discounts. In case it is not possible to process these different discount in the feed, it needs to be entered manually with every order. This is a critical area where a lot of mistakes occur.
6. One EAN is coupled to two different products. The different products have different purchase prices. One product associated with a certain purchase price is ordered, while the other product associated with the same EAN but with another purchase price is delivered and invoiced. This results in a mismatch between order price and invoice price.
7. Unclear order units. This means that X number of units are ordered while X number of boxes are delivered or the other way around. One box consists of more units.
8. A double supplier offer in the catalog for one product. Both offers are from the same supplier but differ in purchase price. It happens sporadic that an order is placed for the price of one supplier offer while the supplier is sending an invoice with the price of the other offer. A double supplier offer could happen when for example a newer edition is available; the EAN remains unchanged but the product reference code is changed. This code is on supplier level coupled to an EAN and sporadically new price lists provided by the supplier contain a change in product reference.
9. Mapping issue Basware. This means that mismatches occur as a result of a technical problem in the matching system Basware. For example the wrong price columns are compared what makes it appear that mismatches between invoice price and order price occur, while in practice the prices do match. This problem does not influence the data quality dimensions because in practice there is no difference between invoice price and order price.
10. Other errors. This concerns cases where a mismatch is discovered between the price on the invoice and the order price, but the causes are unknown. Buyers do not know where the error could has arisen and even in combination with the existing reports no conclusion can be drawn. From the interviews with the buyers can be concluded that control is missing during the entry and handling process of the net purchase prices; the prices pass different systems and on different places in the chain these prices can be changed. The risk of missing controls is that data errors remain undiscovered and manipulation of data can occur.

In section 4.2 is described that for each supplier the buyers are asked to take a sample of the assortment and to check if the prices in the catalog are the same as agreed between buyer and supplier. This part is called the validation and gives an indication for the current reliability of the net

purchase prices in the catalog. This sub study has confirmed that there is no control on prices which enter the system.

An additional finding, resulting from interviews, is that the culture among the buyers is also an issue which results in unreliable net purchase prices in the catalog. When positive price differences between order price and invoice price are discovered it is no problem according to the buyers. However the net purchase prices needs to be reliable because it influences consecutive processes like the calculation of the selling price and the calculation of the preferred supplier. The influence of their tasks on other processes is not sufficiently known by them and the buyers are not aware of the fact that they are, together with the Product & Content department, responsible for correct net purchase prices in the catalog.

The causes of the problem found in the data entry and handling process at bol.com can be linked to the problems found in literature classified along the lifecycle. These problems found in literature are given in table 3.2 and it shows for each stage in the lifecycle a general direction for a solution. The causes discovered during the case studies are all production errors and according to literature general solutions can be found in the direction of process improvements, incentives and/or controls. As can be seen in figure 3.2 production errors can have influence on the intrinsic, contextual, and representational data quality dimension.

The figure below shows which causes that are found in practice influence which quality dimensions. To be clear, the causes are linked to the quality dimensions and not to the quality aspects.

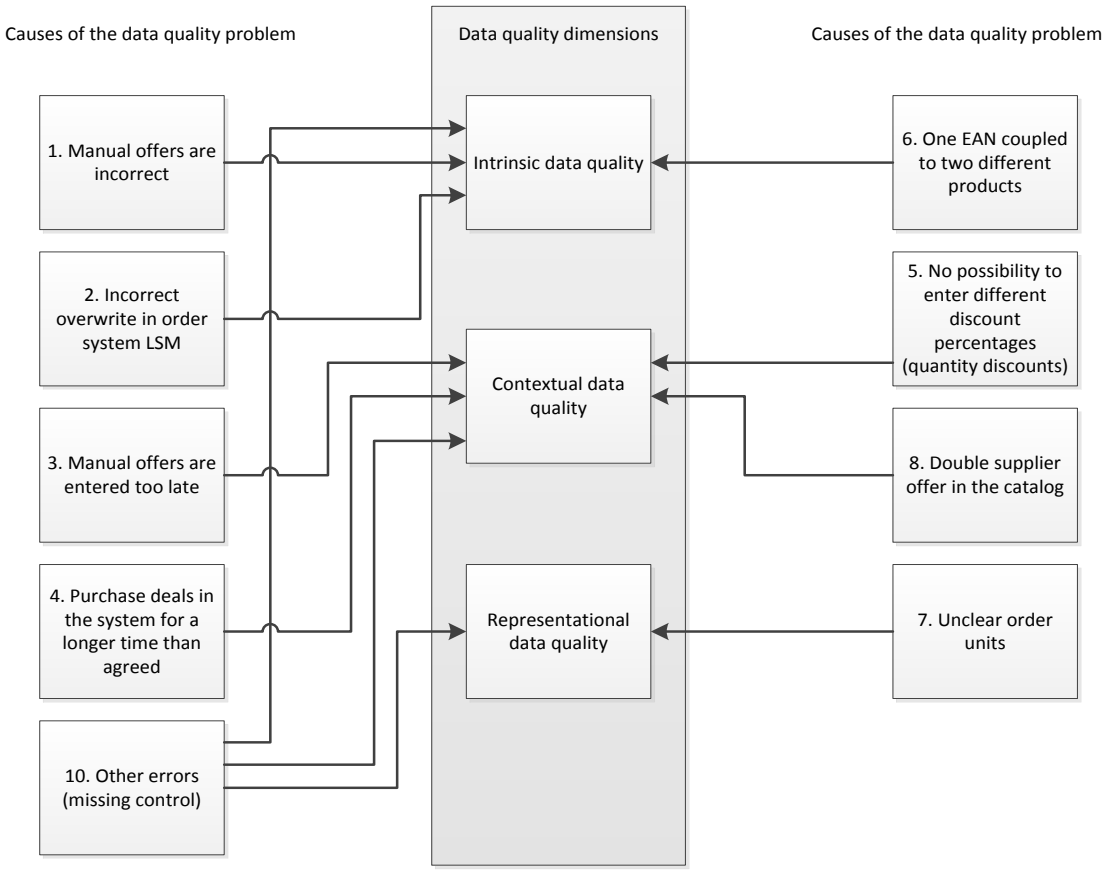


Figure 4.3 Relationship between causes of the data quality problem and data quality dimensions

The mapping issue in Basware (problem 9) is not included in the figure. As stated before, this cause does not influence the data quality dimensions because in practice there is no difference between invoice price and order price. From the figure can be seen that the nine types of production errors in practice also influence the intrinsic, contextual and representational data quality dimension. Furthermore can be seen that the intrinsic and contextual dimensions are influenced by most of the causes of the data quality problem. However, no conclusions can be drawn on which dimension has the highest probability to be of low quality because this depends on the frequency and the impact of each problem.

A tool for identifying which causes have the highest priority for improvement is the risk assessment matrix. A risk assessment matrix is a tabular illustration of the frequency and impact of causes on the problem. It is used to rank the causes according to their significance. Causes with a high frequency and a high impact have the highest need for risk reduction while causes with a low frequency and a low impact have the lowest need for risk reduction (Rausand, 2011).

The impact of the causes, which are found during the case study, cannot be determined because every cause of the data quality problem consists of different degrees of errors and so they do have a different impact. An example which is mentioned before is a typing error, this error has a higher impact when the error is made before the comma than when the error is made behind the comma. Another example is if one EAN is coupled to two different products, when there is a small difference between the net purchase prices of the two products the impact of the error is less high compared to a case with a large difference in the net purchase price between the two products. For this reason the risk assessment matrix with two axes is minimized to one axis, which is not called a matrix anymore. The result is that only the frequency of the causes is shown.

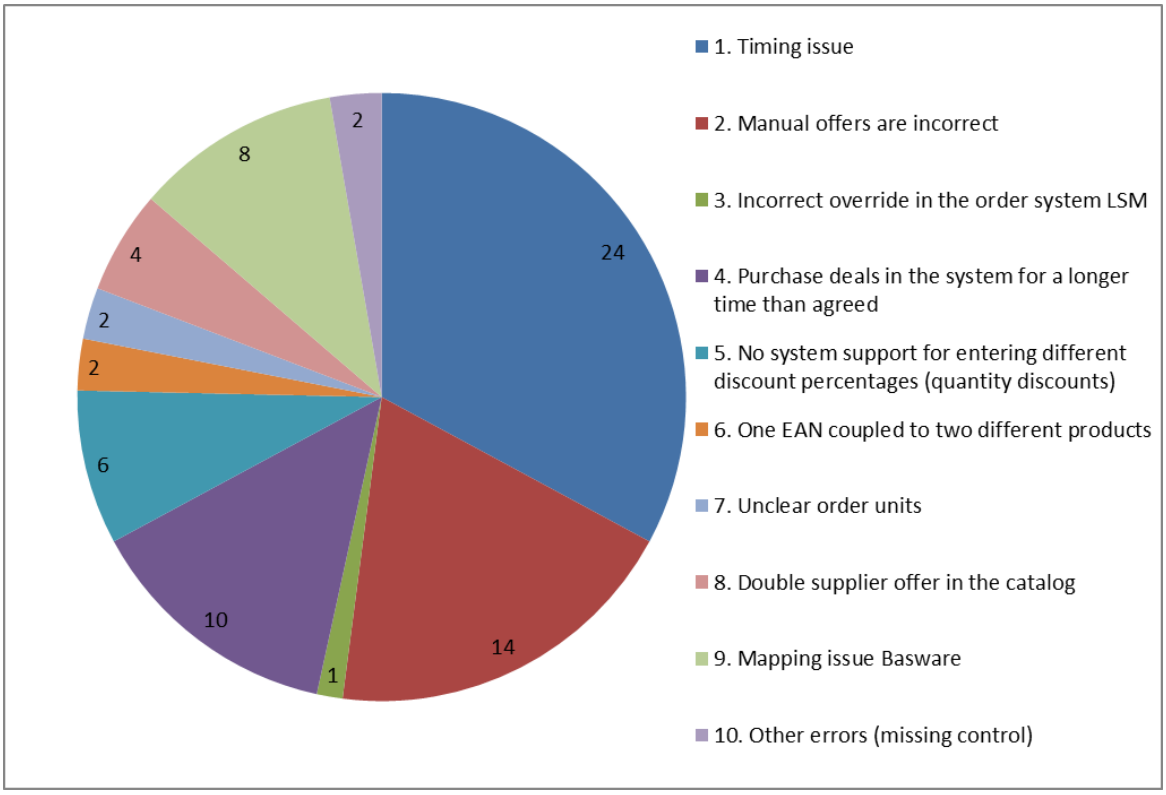


Figure 4.4 Frequencies of causes, based on 207 supplier analysis

Figure 4.4 shows the frequencies of all the causes that occurred at bol.com with the result of a mismatch between order price and net purchase price on the invoice, based on 207 suppliers analysis. For clarity, mismatches as a consequence of incorrect invoice prices at the side of the supplier are not illustrated in this figure. A pie chart is chosen for showing the results because it gives also insight in the frequency of an error with respect to the total number of errors.

The causes discovered during the case studies and interviews do not depend on the categories (Books, Music, DVD & Games, Toys, Baby, Health & Beauty, Computer & Electronics, Living, Pets, Garden & Tools). This means that the solutions that are generated in the next chapters are generic and not specified to categories.

4.4 Preliminary conclusion

For assessing the quality of net purchase prices in the catalog of bol.com, cases of mismatches between net purchase prices in the order system and net purchase prices on the invoice are analyzed. The root causes of price differences are discovered and so the critical areas of the data entry and handling process. The result of the assessment stage is a list of ten causes. The causes are all types of production errors, influencing the intrinsic, contextual, and representational data quality dimension.

Furthermore, from the interviews can be concluded that the buyers are not aware of the fact that they are, together with the Product & Content department, responsible for correct net purchase prices in the catalog. The net purchase prices in the catalog needs to be reliable because it influences consecutive processes like the calculation of the selling price and the calculation of the preferred supplier. The influence of their tasks on other processes is not sufficiently known by them. Given the time restriction, not all the causes can be addressed. The first step in the next chapter is therefore to determine focus. Subsequently required types of changes are formulated and solutions are generated.

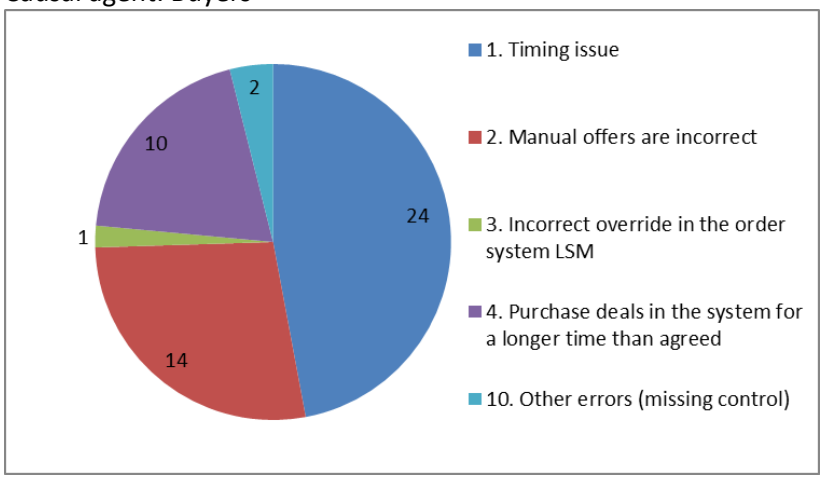
Chapter 5: Improvement

The first three stages of the adapted TDQM methodology are completed in the chapter before, resulting in insights in the data entry and handling process in the current situation and the associated causes. In this chapter the last stage of the adapted TDQM methodology is applied. The first step is to determine focus because not all the causes can be addressed. When the focus is determined the required types of changes are formulated, these are based on the insights in the current and desired situation. Subsequently the types of changes are translated into concrete solutions. The solutions should bridge the gap between the current and the desired situation.

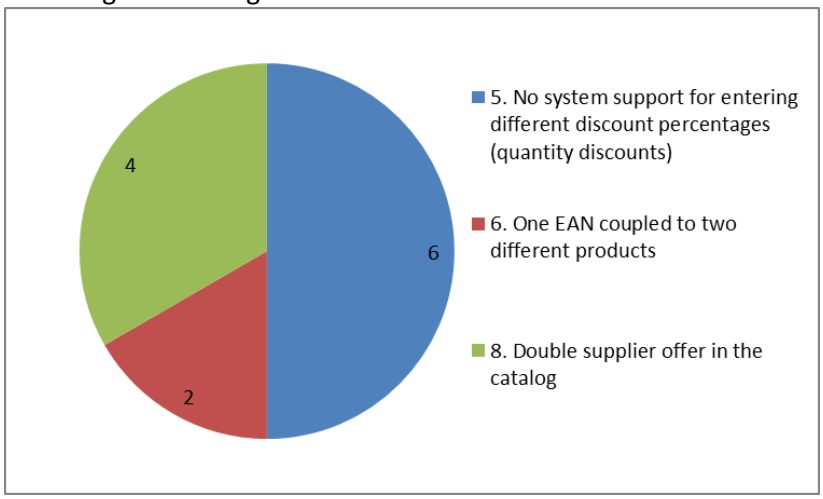
5.1 Required types of changes

To be able to reach the desired situation different types of changes are required. Before going in detail about the types of changes that are necessary, first the focus need to be determined because given the time restriction not all the causes can be addressed. In the chapter before ten causes of price differences between the order price and the price on the invoice are listed, these are illustrated in figure 4.4. In figure 5.1 the same causes are shown classified by causal agent. A causal agent is any entity that is responsible for causes. To be clear, the term includes both persons and systems.

Causal agent: Buyers



Causal agent: Catalog



Causal agent: LSM or Basware

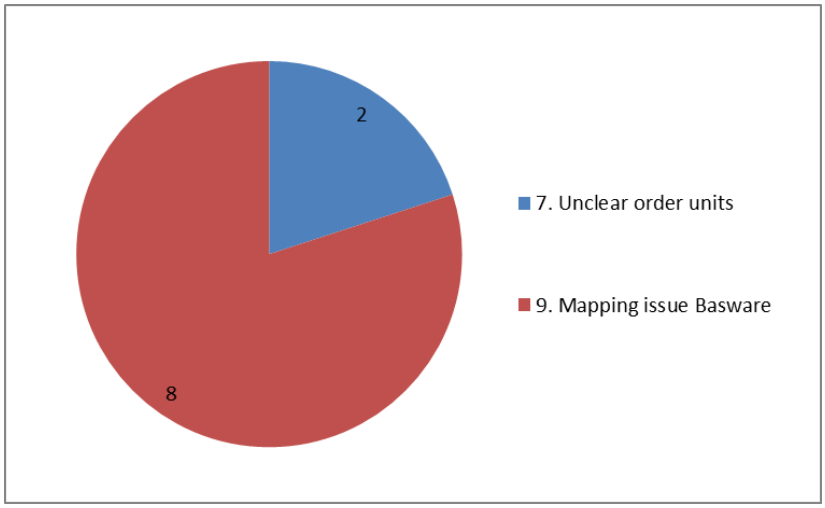


Figure 5.1 Frequencies of causes classified by causal agent (note: figure 5.1 consists of three graphs)

Three causal agents can be distinguished; causes can occur through mistakes of the buyers, causes can occur in the catalog if not the correct tools are available or if the system is not working correctly, and causes can occur in other systems in the data flow. Other systems consist of all the systems after the catalog, called LSM and Basware, as can be seen in the flow chart of figure 4.1.

The focus of this research is on the net purchase prices in the catalog so we are only interested in the causes which do influence the net purchase prices in the catalog. This means that the causes in the lower graph of figure 5.1 are not in scope of this research. On the other hand, the causes in the middle graph of figure 5.1 are in scope of the research but these causes are already in the scope of the IT developers so the solution is not focused on these causes. Remaining are the causes illustrated in the upper graph of figure 5.1. Excluding cause number three, the incorrect override in the order system, all the other causes in the upper graph result in incorrect net purchase prices in the catalog owing to incorrect or delayed operations by the buyers. With the causes illustrated in the upper graph in mind the research is continued. For clarification, the figure below gives an overview of the steps and the findings of the case study, and it shows the focus for the remaining part of this research. In the arrow on the right the buyers are encircled, what means that the focus is on this causal agent and the corresponding causes. In figure 5.1 can be seen that the corresponding causes are cause number one, two, three, four and ten. Referring to figure 4.3 it can be seen that these five causes influence the intrinsic, contextual and representational data quality dimension.

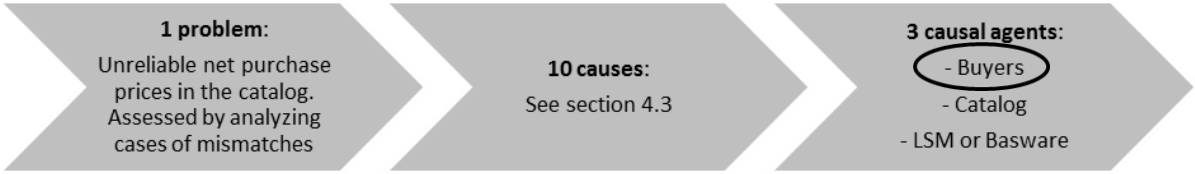


Figure 5.2 Overview of the structure and the findings of the case study, and determination of focus

The required types of changes are based on the gap between the current and the desired situation. The desired situation is already described in section 4.1.1 and the requirements of the data quality dimensions are shown in the table in that section. To be complete, some additional information

about the desired situation regarding the monitoring of net purchase prices is necessary. The grey parallelograms in the figure below illustrate the places in the data handling process where monitoring of the net purchase prices is desired. These are all the places where changes in net purchase prices can be made. For each purchase order line the net purchase prices in the different systems during the handling process need to be visible and related to each other, in this way new insights can be gained about the reliability of the net purchase prices in the catalog and consecutive the accurate calculation of the CM1. To clarify the handling process: first net purchase prices enter the system by means of an automatic feed or manually, in case a deal is agreed between buyer and supplier a purchase deal is entered manually into the system, this deal is overriding the original supplier offer. Both prices are stored in the product offer system POS. Then an order is placed using the order system LSM. Subsequently the supplier can change and confirm the order using the SCP and finally the supplier sends an invoice which needs to be confirmed by the buyer.

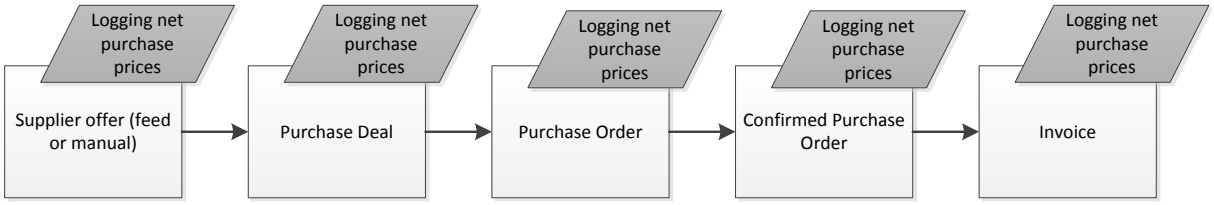


Figure 5.3 Monitoring net purchase prices in the desired situation

Based on the gap between the current and desired situation, in this research three types of changes can be distinguished. These are the following: changes in the IT system, changes in culture and changes in the process/working method. The table below shows for each problem the type of change that is needed to bridge the gap between the current and the desired situation.

Causes (inclusive numbers)	**Required types of changes**
1. Timing issue	Change in process/working method by means of creating a workflow and Change in culture, employees need to feel responsible for the quality of net purchase prices
2. Manual offers are incorrect	Change in process/ working method by means of creating a workflow and Change in culture, employees need to feel responsible for the quality of net purchase prices
3. Incorrect override in the order system (LSM)	Change in IT system, tool needed for monitoring the net purchase prices during whole handling process
4. Purchase deal in the system for a longer time than agreed	Change in IT system, tool needed for counting the number of purchased items
10. No control during the entry and handling process of net purchase prices	Change in IT system, tool needed for monitoring the net purchase prices during whole handling process and Change in culture, employees need to be convinced of the added value of monitoring and need to feel responsible for the quality of net purchase prices

Table 5.1 Causes related to required types of changes

For completeness, below is listed what is meant by the three types of changes and to what issues they are related to.

- IT systems has to do with the technology used for creating, storing, retrieving, manipulating and monitoring data.
- Culture includes the collective mindset, attitude and behavior of employees (Anderson & Ackerman Anderson, 2010).
- A change in process is related to one or more of the activities which take place between the demand for a service or a product and the actual delivery (Van den Berg, et al., 2008). Also task allocation and the associated functions and responsibilities of employees are a part of the process. As stated in chapter two among other procedures and workflows can be used for executing these activities.

The changes in IT system and the changes in the process/working method are discussed below, while more details about the change in culture can be found in the implementation plan in chapter seven.

5.2 Changes in IT system

As stated in table 5.1 two system improvements are necessary to reach the desired situation. First of all a tool is needed for monitoring the net purchase prices during the entry and handling process. Secondly, a tool is needed for counting the number of ordered items with the aim to prevent that purchase deals are in the system for a longer time than agreed. Both solutions are described below.

5.2.1 Dashboard for monitoring net purchase prices

Primary goal

The proposed dashboard is a tool which gives insights in the behavior of the supplier, the behavior of the buyer, and the way the buyers should interact with the suppliers. This can help the buyers to improve the quality of net purchase prices in the catalog, which is the primary goal of the dashboard, and subsequently the contribution margin can be calculated more accurately.

As shown in figure 5.3, on different places in the entry and handling process changes in net purchase prices can be made. When the net purchase prices in the catalog are reliable, changes in consecutive processes should not be necessary. This means when noticing a lot of price changes between the different systems the prices in the catalog are not reliable.

(User) requirements

The dashboard needs to show for each place in the entry and handling process the value of the price change with respect to the net purchase price in the previous system. This means that for each purchase order line the net purchase prices in the different systems during the handling process are visible and related to each other. For completeness, figure 5.4 shows the four places where changes in net purchase prices can be made. The four places where measurements need to take place are represented by Key Performance Indicators (KPI).

- KPI 1 needs to measure the value of the change in net purchase price of the purchase deal with respect to the price in the automatic catalog feed. In case there is no price agreement, the value of the change in net purchase price is zero.
- KPI 2 needs to measure the value of the change in net purchase price of the original purchase order line with respect to the price of the purchase deal (in case there a price agreement) or with respect to the price in the automatic catalog feed (in case there is no price agreement).

- KPI 3 needs to measure the value of the change in net purchase price of the confirmed purchase order line with respect to the price in the original purchase order line.
- KPI 4 needs to measure the value of the change in net purchase price of the invoice with respect to the price of the confirmed purchase order line.

For each supplier and each KPI the absolute value of net purchase price changes, summed over all the purchase order lines during a certain time period, needs to be showed in the dashboard.

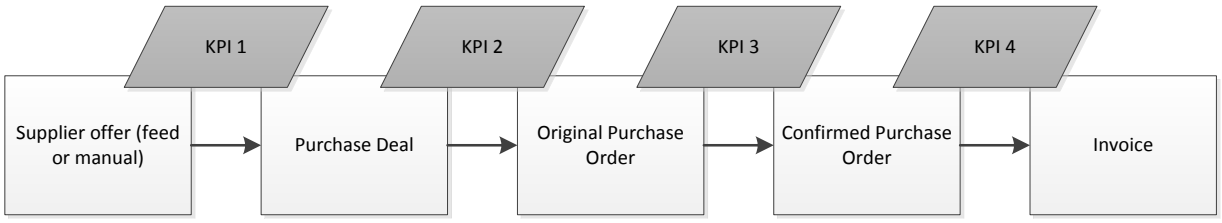


Figure 5.4 KPI's in the entry and handling process of net purchase prices

Although stakeholders of the dashboard consist of business analysts from the Product & Content department, buyers from the Buying and Merchandising department, and employees from the Controlling department, the main users of the dashboard are the buyers. This means that it is important that the monitoring tool meet the buyers' needs. Interviews with selected buyers has resulted in the following desires:

- The dashboard needs to show progression over time.
- Besides the summed value of net purchase price changes on supplier level, this needs also be presented on category level. In case a supplier is supplying products in different categories, different buyers are doing business with this supplier. Both buyers are interested in the performance of that supplier within their own category.

Proposed dashboard and interpretation

The main screen of the proposed dashboard is given in figure 5.5. To be clear, the values are illustrative and are not real values. The table shows for each KPI on supplier level the absolute value of net purchase price changes summed over all the purchase order lines during a certain time period. In this dashboard is chosen for a time period of four weeks. Associated formula for each KPI is:

$$\sum_{all\ purchase\ order\ lines\ during\ 4\ weeks} |\Delta p * q|$$

The colored symbols indicate which summed values are accepted, which values are accepted but needs attention and which values are not accepted. Critical limits need to be set in accordance with stakeholders. For illustration the limits in figure 5.5 are set on 1000 euro and 1500 euro.

Based on the four KPI's, for each supplier a combination of colors is visible. Because 81 different color combinations (3*3*3*3) makes the interpretation too long and unclear, the choice is made that an orange color needs to be interpret as a red color, resulting in total 2*2*2*2 = 16 different color combinations exist. The risk of each combination can be a basis for the ranking of the suppliers. This risk can be determined by the frequency that a certain combination will occur in combination with explanation about what possibly went wrong. The risks of the color combinations, together with an

explanation about what possibly went wrong and the actions that need to be taken by the buyers, should be written in a manual and stored on a place that is all the time accessible for the buyers.

Two examples of interpretations of color combinations:

- Color combination KPI 1: green, KPI 2: green, KPI 3: green and KPI 4: red
  It shows that the supplier is sending invoices with other prices than the net purchase price which is provided in the catalog feed and which is confirmed by themselves. It is a high risk because the supplier is manipulating the data. Action that need to be taken is contacting the supplier to pay attention to the manipulative behavior, and tight control of invoices is required.
- Color combination KPI 1: red, KPI 2: red, KPI 3: green and KPI 4: green
  It shows that the agreed price (purchase deal) is not included in the automatic catalog feed, this is not a risk but in the ideal situation the deal is included in the automatic feed. Furthermore what possibly went wrong is that the purchase deal is incorrect or for a longer time in the system than agreed, this is a logical explanation for the price change in the order system. Repairing action that need to be taken by the buyer is checking, changing and/or deleting the purchase deal. Preventive action that can be taken is contacting the supplier for including the purchase deal in the automatic feed, this to prevent that deals are in the system for a longer time than agreed.

In general can be said that the risk is higher if changes in net purchase prices occur in the end of the handling process (KPI 4). Furthermore can be stated that changes in the purchase deal with respect to the price in the automatic catalog feed (KPI 1) are not a risk. It just shows that the purchase deal is not included in the automatic catalog feed and special attention need to be paid to check the invoices. Including the purchase deal in the feed is a method for preventing that deals are in the system for a longer time than agreed.

The graphs in figure 5.5 show for each KPI the overall trend in the total absolute value of price changes, summed over all the suppliers during a time interval of four weeks. Associated formula for each KPI is:

$$\sum_{all\ suppliers} \sum_{all\ purchase\ order\ lines\ during\ 4\ weeks} |\Delta p * q|$$

This formula represents the absolute value and in the desired situation the graphs for KPI 2, 3 and 4 are declining to zero. As stated before, for KPI 1 changes in net purchase prices are allowed but a declining graph is desired because less manual changes in purchase deals, as a result of automatic catalog feeds which include the deal, can be a method for preventing that incorrect deals are in the system.
Also the graphs show for each KPI the trend of relative values. This means that the absolute value of price changes summed over all the suppliers during a time interval of four weeks is divided by the total absolute purchase value over all the suppliers during a time interval of four weeks. In the desired situation these graphs are also declining. Associated formula for each KPI is:

$$\frac{\sum_{all\ suppliers} \sum_{all\ purchase\ order\ lines\ during\ 4\ weeks} |\Delta p * q|}{\sum_{all\ suppliers} \sum_{all\ purchase\ order\ lines\ during\ 4\ weeks} |p * q|}$$

Figure 5.5 Main screen of the proposed dashboard

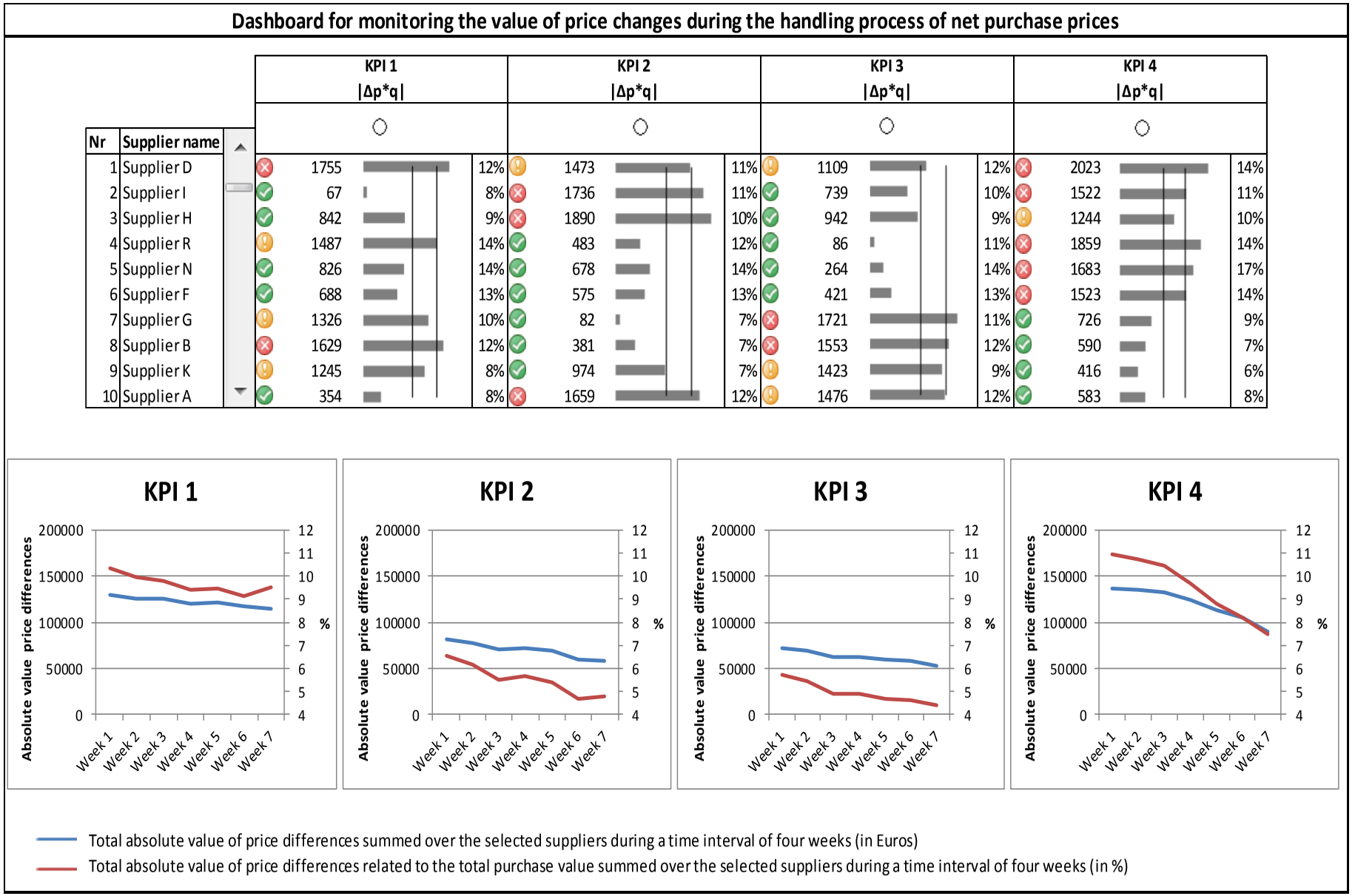
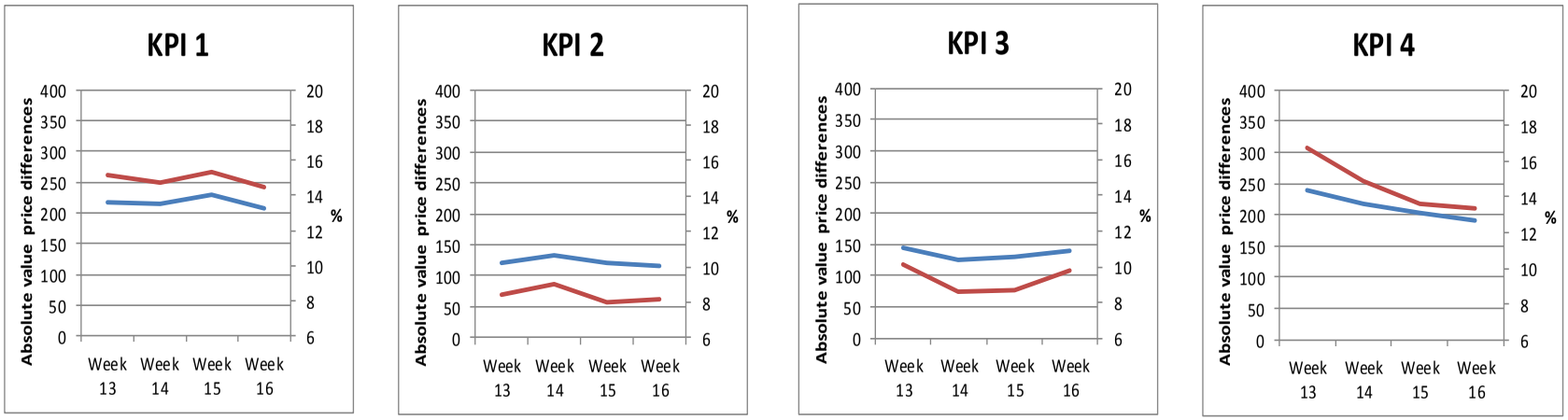


Figure 5.6 Detailed screen of the proposed dashboard

Details for 'Supplier D' 'Category Mobile' 'Week 13-16'

PO date	PO number	EAN	Productgroep	Q	NPP supplier offer	KPI 1 |Δp|		NPP purchase deal	KPI 2 |Δp|		NPP original po	KPI 3 |Δp|		NPP cofirmed po	KPI 4 |Δp|		NPP invoice
3-Apr	527543																
		5308536754355	Mobile	4	20	2,5	2%	17.5	0	0%	17.5	0	0%	17.5	2,5	2%	20
		5312890943304	Mobile	15	22	2	3%	20	0	0%	20	2	3%	22	0	0%	22
		5542998766532	Mobile	5	18.6				0	0%	18.6	1,4	1%	20	0	0%	20
5-Mar	563589																
		5308536754362	Mobile	8	36.5				6,5	4%	30	6,5	4%	36.5	0	0%	36.5
		5312890943386	Mobile	20	42				0	0%	42	0	0%	42	3	4%	45



Detailed screen

To meet buyer's needs, there is the possibility to click on a supplier and filter on among other product group and date for details. Figure 5.6 gives an example of the detailed screen with selection criteria: 'supplier D' 'category mobile' 'week 13 to 16'. For each purchase order line during that time interval the net purchase price in the different systems is shown and the value of the price differences. Buyers can use these cases when contacting the relevant suppliers. Furthermore the graphs show for each KPI the performance of the selected supplier in the specific category.

Looking back to the causes of chapter four, the monitoring tool will contribute to the following causes of mismatches:

- No control during the entry and handling process of net purchase prices. The dashboard makes controlling possible; it gives insights in the behavior of the supplier, behavior of the buyer and the way buyers should interact with the suppliers. It also gives an indication for the reliability of net purchase prices in the catalog.
- Incorrect override in the order system (LSM). Certain color combinations result in detecting this problem (for example when KPI 2 and KPI 3 have a red color). Monitoring will result in more awareness of buyers when overriding net purchase prices.
- Purchase deal in the system for a longer time than agreed. Certain color combinations result in detecting this problem (for example when KPI 1 and KPI 2 have a red color). Monitoring will result in more awareness of buyers when overriding net purchase prices and entering correct end dates of purchase deals.

The tool should help the buyers in executing their tasks. It will only be used when buyers are aware of the added value of the tool and when they realize it helps them improving the quality of net purchase prices in the catalog. Only when this is recognized by them the provided tool will be used properly and the organization can reap the benefits of the IT change. Even more important is that the buyers need to be aware that they are, together with the Product & Content department which provides technical support, responsible for correct net purchase prices in the catalog. Also they need to be aware of the influence of their tasks on consecutive process, including the accurate calculation of the contribution margin. The process of change management is necessary for changing the culture among the buyers, this process is discussed in more detail in the implementation plan.

5.2.2 Tool for counting the number of ordered items

The second improvement in IT system has the aim to prevent that purchase deals are in the system for a longer time than agreed. This is especially an issue with manually entered one-off deals based on the number of purchased units. An example of an one-off deal is when X number of products can be ordered for 20 percent off. At this moment only the date can be entered as an end condition but not the number of purchased units. The problem can partly be detected and prevented by using the monitoring tool.

An additional solution is implementing the possibility for buyers to enter the number of products as an end condition of the purchase deal. A tool should count the number of ordered items and when this number reaches the end condition, the net purchase price should automatically change back to the regular purchase price. The associated formula is given on the next page.

X = end condition
If # ordered products <= X
    Net purchase price = price in purchase deal;
If # ordered products > X
    Net purchase price = original price from feed;

In the current situation the buyer needs to count manually when X products are ordered and after this the buyer needs to change back the net purchase price manually to the regular price. The proposed solution replaces the manual process by an automatic process.

5.3 Changes in process/working method

As can be seen in table 5.1, improvements in the working methods are necessary to minimize the gap between the current and desired situation. The first change in working method, described in section 5.3.1, has the aim to minimize the problem of timing issues while the second change in working method, described in section 5.3.2, has the aim to prevent incorrect manual offers.

5.3.1 Sharing information and planning

A small change in organizational structure can help minimizing the problem of timing issues, this problem is also defined as entering manual offers after the agreed date from which an offer is valid. In practice this is happening when buyers do not have time for entering the manual offers before the agreed date, or when files with new price lists are in the mailbox of the buyer for a long time for example in case of sickness or holidays.

First recommendation is that knowledge/information need to be shared. Solution in this case is that files with new prices lists are sent to at least two buyers. Subsequently it is recommended that price lists are stored on a central disc which is only accessible for the buyers who are doing business with the supplier concerned. Consequence is that the information is not stored on the computer of only one buyer and the information can be read back all the time.

Secondly is recommended that buyers block time in their agenda for entering manual offers. In the contract with the supplier is agreed that the supplier sends a notification in case they want to change or update their price list. Depending on the contract, this notification is send one month or two months before the date from which new offers are valid. When there is agreement about this date, buyers can block time in their agenda for updating the manual offers.

5.3.2 Workflow for entering manual offers

To prevent incorrect manual offers in the catalog, a change in the current working method is necessary. The most important change is that the existing reports about mutations in net purchase prices in the catalog need to be used for detecting exceptional typing errors. Although Eckerson (2002) states that even validation routines cannot catch small typing errors which represent a valid value, large typing errors with a higher impact on the accurate calculation of the CM1 can be detected when the reports are used properly. The figure on the next page shows the desired process of entering manual offers.

As can be seen in the workflow in figure 6.4, the existing reports need to be used for comparing the updated net purchase price with the previous net purchase price. If the price mutation is more than 10%, the updated net purchase price could possibly contain a typing error and it need to be checked if this net purchase price is matching the purchase price on the price list. If there is no match, a correction is needed. The limit of 10% is chosen because small price changes are allowed, suppliers can deliver a new price list with net purchase prices which are higher or lower than the net purchase prices on the previous price list. Small typing errors will not be detected but large typing errors will. In case the limit of 10% seems to be not realistic because almost none of the price mutations do reach this number, the limit can be changed down. The lower the limit the more price mutations will reach the limit, with the consequence that more net purchase prices will be checked and corrected if necessary. Consequence is a higher reliability of net purchase prices in the catalog.

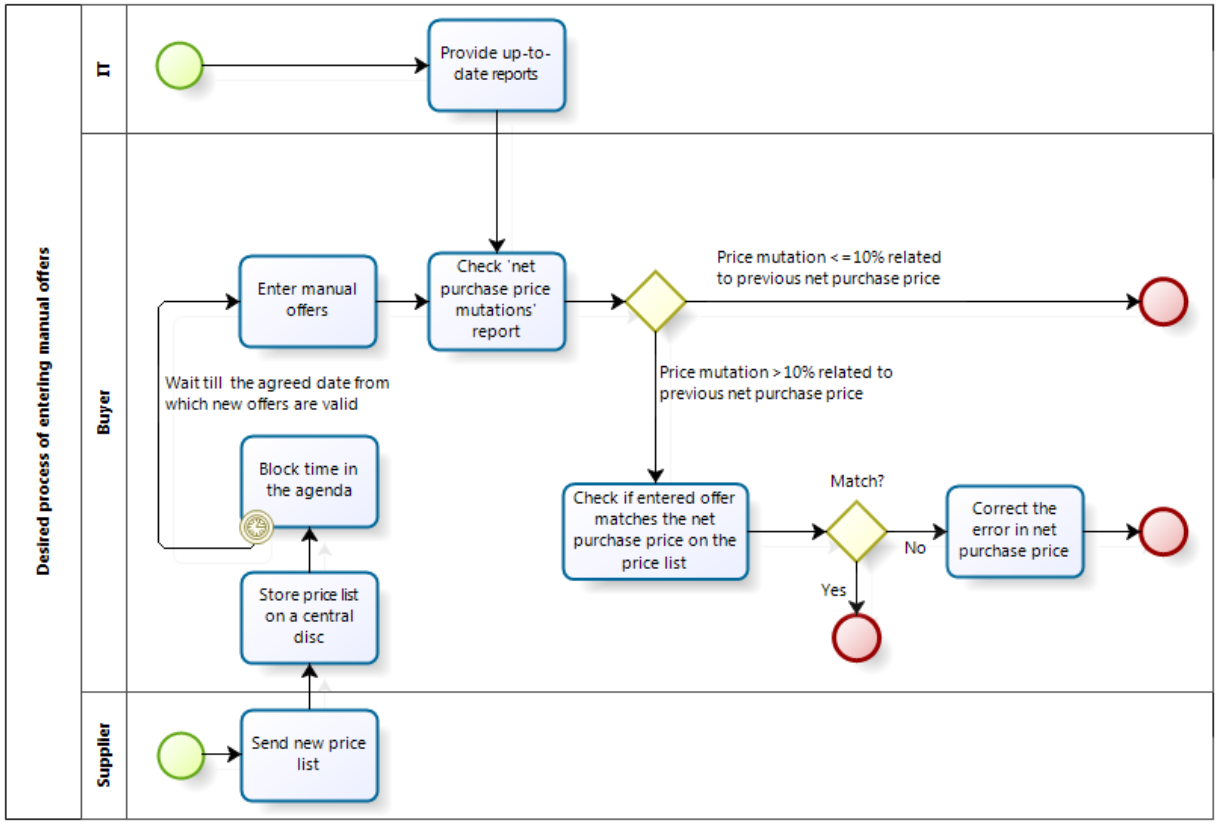


Figure 5.7 Desired process of entering manual offers

The workflow will prevent incorrect manual offers in the catalog. Also the recommendations from section 5.3.1 about information sharing and time blocking, both with the aim to minimize the problem of timing issues, are shown in the figure above. Comparing figure 5.7 with figure 4.2 shows the difference between the current and desired situation.

5.4 Preliminary conclusion

From this chapter on the focus is on the five causes of mismatch between net purchase prices in the order module and on the invoice, owing to incorrect or delayed operations by the buyers. To bridge the gap between the current and desired situation, three types of changes can be distinguished: changes in the IT system, changes in the process/working method, and changes in culture. All the changes aim to minimize or solve one or more causes discovered in the chapter before. The changes

are translated in concrete solutions and recommendations. Some solutions contribute to more than one cause and will have a higher impact on the quality of net purchase prices in the catalog than others. To a certain extent, all of the solutions will contribute to the intrinsic, contextual or representational data quality dimension. Summarizing, the solutions and recommendations are as follows:

- Implement a dashboard for monitoring the net purchase prices during the handling process.
- Implement the possibility for buyers to enter the number of products as an end condition of purchase deals. A tool should count the number of ordered items and when this number reaches the end condition, the net purchase price should automatically change back to the regular purchase price.
- Suppliers need to send a new price list to at least two buyers and this needs to be stored on a central disc which is only accessible for the buyers who are doing business with the supplier.
- Buyers should block time in their agenda to make sure that they enter manual offers before the date from which the offers are valid.
- Existing reports need to be used to check if the manually offers are entered correctly.

After the solution generation process and the formulation of the recommendations, the most important question is what the worth or merit is of the proposed solutions. This is called the validation process and is described in the next chapter.

Chapter 6: Validation

This chapter demonstrates the validation process. A crucial element of validating projects or products is the evaluation (Cleven, Gubler, & Hüner, 2009). In the first section evaluation methods that can be found in literature are discussed and in the section which follows one of the evaluation methods is described in more detail and is applied in practice.

6.1 Evaluation methods

Within a consideration of cost effectiveness the evaluation of design science research artifacts is of major importance. In Cleven et al. (2009) evaluation is defined as the systematic investigation of the worth, merit or significance of an object. An object can represent for instance a potentially construct, model, tool or method. In this research the object of evaluation is the monitoring tool. Evaluation may be external or internal, or a mix of these, and it may be quantitative or qualitative, or a mix of these. Peffers et al. (2008) state that evaluation could take many forms, depending on the nature of the problem and the artifact. Different methods and technologies for evaluation are generally described by Cleven et al. (2008), these are as follows:

- Action research. This represents a qualitative research approach. Research object and researcher influence each other mutually in their development. Furthermore the strong cooperation of researchers and the users of the object is a characteristic of action research.
- Case study. This kind of evaluation method is broadly used for describing and analyzing real world problems. A characteristic of case studies is that uncontrolled influences can emerge from the context an artifact is evaluated in, for example from a company context.
- Field experiment. A field experiment is also carried out in real world surroundings, for example in a company. Independent variables are under control by the researcher and as many characteristics as possible are observed and measured to get insights in the effect of these independent variable on the results, also called the dependent variables.
- Formal proofs. This is an adequate method for Design Science Research evaluation. Rules that have to be considered during the design process are defined, when applying formal proofs.
- Controlled experiment. With this evaluation method several situations are compared that differ in only one independent variable. The effect of the independent variable on the dependent variables is observed. Typical is the comparison of a control group with an experimental group.
- Prototype. A prototype gives the researcher the possibility to assess the suitability of a solution for a certain problem by implementing the solutions generically. It is an adequate method for Design Science Research evaluation.
- Survey. With conducting a survey, information is collected about the suitability of a solution by interviewing representatives of a certain group, for example end users of the object. Questions can relate to subjective and objective issues, but the answers are always subjective and only limited verifiable (Cleven, et al., 2009).

Not all the methods mentioned above are suitable for evaluating the monitoring tool. Evaluating the monitoring tool by means of experiments is at this moment not possible. First of all the net purchase prices in the different systems cannot be linked on the level of a purchase order line. This because of

the limited price history that is available. Invoices can be sent two months later and the net purchase price in the beginning of the handling process that was valid at the moment of ordering is sometimes not in the system anymore, depending of the number of price mutations during that period. Only recent price changes are in the system. Furthermore, from the price history that is available there is no distinction made between original supplier offers which enter the system and purchase deals, this means that the first KPI has no input for measuring. The result is that no conclusions can be drawn about the beginning of the handling process.
Secondly, improvement in performance cannot be shown immediately because there is no baseline for comparison. The current quality of net purchase prices in the catalog cannot be expressed in numbers but, as we did in chapter four, only by evaluating cases. This means that during the first weeks of monitoring no comparison can be made. After for example one month the results achieved can be function as a baseline and the monitoring results after this month can be compared with these starting results for showing the improvements in quality of net purchase prices in the catalog.

From above can be concluded that only the survey is a suitable method for evaluating the monitoring tool. Conducting a survey among the users of the tool results in subjective but valuable user evaluations. Goodhue (1995) defines a user evaluation as an assessment made by an user, along some continuum from positive to negative, about certain qualities of information systems. According to Goodhue user evaluations can be based on the concept of Task-Technology Fit (TTF). The TTF perspective views technology as a means by which a goal-directed individual performs tasks. The focus is on the degree to which system characteristics match user task needs. This perspective suggests that a better fit between technology functionalities and task requirements will result in better performance. Research performed by Goodhue has shown that user evaluations accurately reflect the complex interaction between task needs and their technologies, this strengthens the theoretical and empirical basis about user evaluations of TTF which function as a measure of success (Goodhue, 1995).

Goodhue and Thompson (1995) elaborate on this theory by stating that technologies must not only fit the task they support to have a performance impact but technologies must also be used. Utilization is the behavior of employing the technology in completing tasks. A schematic representation of the combined fit and utilization focus is given in figure 6.1, it shows that both factors influence the performance.

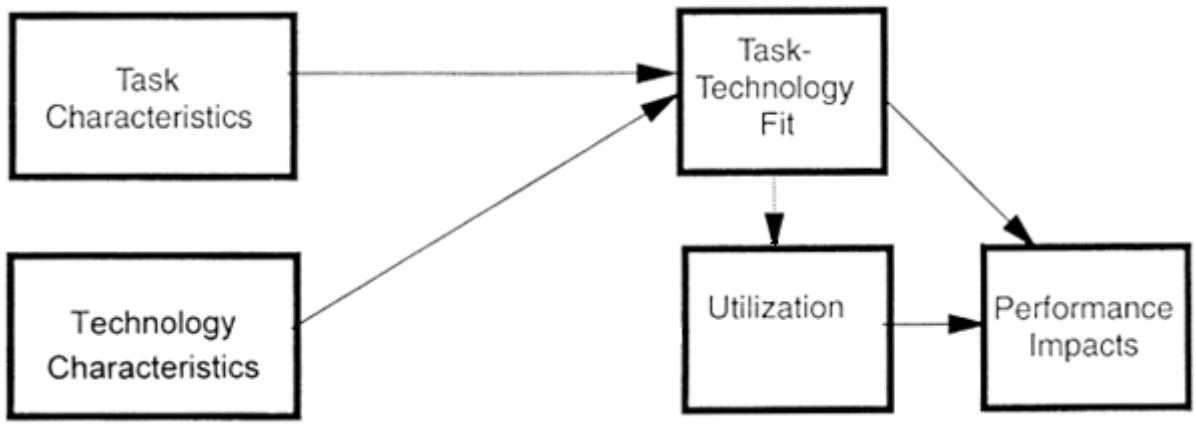


Figure 6.1 Combined Task-Technology Fit and Utilization perspective (Goodhue & Thompson, 1995)

Research performed by Goodhue and Thompson has shown that evidence of the causal link between TTF and utilization is ambiguous, for that reason a dotted arrow is shown in figure 6.1. Furthermore research state that when utilizing is mandatory, which means that use of a system is mandated as part of a job description, the utilization aspect does not need to be considered (Goodhue & Thompson, 1995).

6.2 Task-Technology Fit

In this research user evaluations of the monitoring tool are purely based on the concept of Task-Technology Fit. The utilization aspect is not considered because, as described in the end of section 5.2.1, the first step is that buyers need to be aware of the added value of the tool. They should use the tool because they realize it can help them improving the quality of net purchase prices in the catalog. Change management is necessary for recognizing these responsibilities and this process is described in the implementation plan in the next chapter.

The survey used for evaluating the monitoring tool is based on standard measures provided by Goodhue (1995). The survey is adapted to this specific domain, what means that only the measures that are relevant for evaluating the monitoring tool are selected. Statements about characteristics of the monitoring tool are used as measure. These measures are related to the three interacting process steps:

- Identification of needed data. The dashboard need to present data which can be used by the buyers to support their interaction with the supplier with the aim to improve the quality of net purchase prices in the catalog. The dashboard need to contain the right data at the right level of detail.
- Access to identified data. For the users the dashboard needs to be accessible all the time. Reliability of the system is a key success factor for accessibility. Furthermore the dashboard needs to be easy to use.
- Interpretation of data. The data shown in the dashboard is used for further decisions, so it needs to be clear how the data needs to be interpreted. Data must be accurate enough that it can be interpreted correctly. A clear presentation of the data is needed and the data must be current enough. Currency implicates that the data is up-to-date enough for the buyers' purposes.

To summarize, the seven Task-Technology Fit measures are: right data, right level of detail, accessibility, system reliability, ease of use of hardware and software, presentation and currency. Besides the Task-Technology Fit measures, also measures for task characteristics are important. A distinction can be made between routine tasks and non-routine tasks. For users with non-routine tasks it is more likely that their data need is changing continuously. Combining both measures results in a survey with eight measures. Some measures consist of more than one statement, in total there are thirteen statements. The survey can be found in appendix C. A drawback of the survey is that three out of eight measures consist of only one statement, this has influence on the reliability of the survey. The eight measures are equally important and need to be assessed in a similar way. For this reason the average score of the two statements is taken in case measures consist of more than one statement. In this way an attempt is made to prevent that a measure consisting of two statements is seen as more important than a measure consisting of one statement. It should partly overcome the doubts about the reliability of the survey. A sample calculation can be found in the end of appendix D. Fact remains that the evaluations are subjective, what also influences the reliability of the survey.

Before starting the surveys oral explanation is given about the aim of the dashboard, how the color combinations need to be interpreted, and the proposed dashboards from figure 5.5 and figure 5.6 are explained. Data is collected from five users. They are asked to rate the tool, on a five-point scale, based on fit with their task needs. To prevent that the respondents are influenced by each other, the survey is not conducted in groups but individually. The respondents are four buyers from different categories and one business analyst from the Product & Content department. The buyers are the main users of the dashboard. The dashboard needs to help them in achieving the main goal, which is improving the quality of net purchase prices in the catalog. Besides, the Product & Content department can use the dashboard for checking the behavior of the buyer and the supplier. When the performance of the buyers are under desired performance and no progression is visible, a business analyst can use the monitoring tool for correcting the behavior of the buyer. Ultimately the business analysts from the Product & Content department and the buyers from the Buying & Merchandising department are together responsible for the correct net purchase prices in the catalog, although the responsibility of the business analysts of the Product & Content department is more focused on the technical support.

First of all, from the user evaluations can be concluded that the buyers deal with routine tasks. According to literature, when users deal with routine tasks it is not likely that their data need is changing continuously. This means that the dashboard could be a solution for long-term. Referring to the three interacting process steps described on the previous page, it can be concluded that the identification of the needed data and the interpretation of data have high scores. It shows that in the buyers' perspectives the dashboard shows the right data. It is valuable for executing their tasks. Furthermore the data is presented clearly and understandable. The measures that are part of the second process step, the accessibility of identified data, have a lower score although the average score still refers to agreement of the statement. These scores are based on the performance of the current reporting system, as described in the note in appendix C. All the descriptive statistics can be found in appendix D. To be complete, the respondents did not have additional comments.
The overall conclusion that can be drawn from the user evaluations is that the characteristics of the dashboard match with buyers' task needs. This result is not a surprise because, as described in section 5.2.1, the buyers are involved in the design process of the dashboard to make sure that the dashboard meet the buyers' needs. As stated in Goodhue and Thompson (1995), when users who understand the business task are involved in system design, it is more likely that the resulting system will fit the task need. A better fit between the characteristics of the dashboard and task characteristics will logically contribute to a better performance, although this link is not scientifically proven.

6.3 Preliminary conclusion

In literature numerous methods and technologies for evaluation can be found. Given the constraints it is not possible to evaluate the monitoring tool by means of experiments. Only the survey is a suitable method for evaluating the monitoring tool. Conducting a survey among the users of the tool results in subjective but valuable user evaluations. The user evaluations can be based on the concept of Task-Technology Fit. This perspective suggests that a better fit between technology functionalities and task requirements will result in better performance. From the survey can be concluded that the characteristics of the dashboard match with buyers' task needs, which is not a surprise given the fact that buyers are involved in the design stage of the dashboard.

To what extent the monitoring tool contributes to the problem statement is not answered in this chapter. In the proposed dashboard each KPI shows the value of price changes. A dashboard which presents the number of price changes would be more appropriate for answering the main research question because the number of price changes on the different places during the handling process gives an indication for the reliability of the net purchase prices in the catalog. However from the interviews with the stakeholders can be concluded that the value of price changes is also playing a key role. The result from showing the value of price changes is that suppliers with price changes with a high value will be corrected first, and not the suppliers with a high number of price changes. Ultimately both options have the aim to improve the quality of net purchase prices in the catalog, only the way to reach this goal is slightly different. Despite of the fact that showing the value of prices changes does not match the problem statement in detail, it matches the needs of the stakeholders resulting in a high fit with higher performance as a logical consequence.

Chapter 7: Implementation plan

In this chapter the implementation plan is presented. This plan describes the sequence of implementing the different solutions based on the requirements for implementation, the term of change (long-term versus short-term) and the complexity of change. The complexity depends on the involved actors and systems.

7.1 Requirements, term and complexity of implementation

The table below shows for each solution/recommendation the requirements for implementation and the complexity. As can be seen, in case the IT department is involved the solution cannot be implemented on short term because IT time need to be reserved. Furthermore the most important conclusion from this table is that a change in culture among the buyers is necessary before a change initiative can be implemented successfully. Subsequently the short-term solutions can be implemented immediately and when time is blocked on the agenda of the IT department the implementation of the two long-term solutions can start. When the requirements, mentioned in the third column, are met the implementation of the solutions can start independently of each other.

Solutions/recommendations	**Term**	**Requirements for implementation**	**Involved actors and systems**
Dashboard for monitoring net purchase prices during the handling process	Long-term	Only in combination with a change in culture and IT time is needed	- IT department - Buyers (one of each category) and other stakeholders, to make sure the (user) requirements are included
Implement the possibility to enter the number of products as an end condition of purchase deals	Long-term	IT time is needed	- IT department
Suppliers need to send a new price list to at least two buyers and this need to be stored on a central disc	Short-term	In combination with a change in culture	- Category managers, to communicate the change to the buyers - All buyers, to communicate the new working method to the suppliers
Buyers should block time in their agenda to make sure that they enter manual offers before the date from which the offers are valid	Short-term	Only in combination with a change in culture	- Category managers, to communicate the change to the buyers - Buyers
Workflow for checking manually entered offers	Short-term	Only in combination with a change in culture	- Category managers, to communicate the change to the buyers - Buyers

Table 7.1 Requirements, term and complexity of implementation

7.2 Change in culture

As stated in section 5.2.1, the interaction between the buyers and the technology is a critically important contributor to the failure or success of change initiatives. Buyers need to be aware of their responsibility for correct net purchase prices in the catalog and need to realize that the proposed solutions can help them improving these quality. This means, as outlined in the previous section, that the solutions and recommendations need to be implemented in combination with a change in culture. A structured approach is necessary for changing the culture among the buyers. The field of change management is described extensively in literature. Various models can be used to describe and implement change. For changing the culture among the buyers, with the primary aim to create awareness and responsibility, the change model developed by Lewin is used. Lewin was one of the first psychologists who studied organizational development and his three-stage-model of change consist of the following stages: unfreeze, change and refreeze (Daft, 2008).

Unfreeze

The first part of the change process is usually the most difficult. This stage involves preparing the department to accept that change is necessary. Buyers are made aware of problems. Key for unfreezing is presenting information that shows discrepancies between desired performance and the current state of affairs. Cases of discrepancies between prices in the order system and prices on the invoices, and consequently the discrepancy between the forecasted CM1 and the actual CM1, should be presented. Equally important is that buyers must be made aware of the influence they can have on this problem. Training or an interactive presentation about the systems, the relation between the systems and the influence of their tasks on consecutive processes, can show their role in the process of entering and handling the net purchase prices and makes the buyers aware of their responsibility. These facts need also be written in a manual which can all the time be read back. Unit managers are in the function to present this information. They need to establish a sense of urgency to unfreeze the buyers and to create an openness and willingness to change.

Change

In this stage employees start to believe and act in ways that support the new direction or new working method. Steps that are involved in this stage are: creating a coalition of people with the power to guide the change, create a vision for change that everyone can believe in, and communicate the vision and plans for change through the relevant departments. Unit managers have the power to guide the change among the buyers. In order to accept the change and contribute to making the change successful, buyers need to understand how the change will benefit them and in what way the new monitoring tool can help them in achieving their goal. A way to show the benefit of the monitoring tool for the buyers is to explain cases of different color combinations which could occur on the dashboard, how the buyers can respond to these cases and in what way they should interact with the suppliers. Furthermore successful change involves using emotion as well as logic to persuade employees and to empower employees to act on the plan and accomplish the desired change.

Refreeze

When the changes are taking shape and the employees acquire new attitudes, values or working methods, the organization is ready to refreeze. Positive changes in individual and organizational performance should be made transparent, this can be done by unit managers which provide data and

figures to the buyers to demonstrate the positive changes. Rewarding the success of the change is a key success factor. Furthermore, in this stage changes are institutionalized in the relevant departments meaning that the buyers begin to view the changes as an integral part of everyday business.

It is natural for employees to resist change. Although employee resistance is a complex issue organizations face, it is a critically important contributor to the failure or success of initiated changes within organizations. To increase the chance of a successful implementation of change, several strategies for overcoming these resistance exist (Daft, 2007). The strategies that are relevant in this research are as follows:

- Alignment with needs and goals of users.
  To make sure that change meets a real need will overcome the resistance. Buyers need to use the dashboard to monitor the net purchase prices during the handling process, to monitor the behavior of the supplier, and the dashboard is also giving an indication for the reliability of the net purchase prices in the catalog. This means that the dashboard needs to show the information that is in line with the need of the buyers. Only then the dashboard has an added value.
- Communication and training.
  Users must be informed about the need for change and the consequences of a change. Open communication with the employees who are involved and affected by the change will prevent misunderstanding. Besides, training or an interactive presentation is needed to help the employees understand and cope with their role in the change process. Furthermore buyers need explanation about how to interpret the tables and figures of the dashboard, a manual is here the best option because the buyers can read it as many times as they need.
- Participation and involvement.
  Participation gives those involved a sense of control over the change activity. Understanding is better and the buyers become committed to a successful change process. During the design process of the dashboard buyers are involved, this to be sure that there needs and goals are reflected and that they understand the added value of the dashboard.

Based on the three-stage-model of Lewin in combination with the strategies for overcoming resistance, different steps for describing and implementing change are formulated. The table below gives an overview of the steps that must be taken and the associated tools.

1. Create a sense of urgency for change during a kick-off meeting

- Unit managers need to organize a kick-off meeting with all the buyers. Optional members of the kick-off meeting are the controllers. The kick-off meeting need to be organized as soon as possible; the sooner this meeting, the sooner discrepancies can be tackled.
- During the kick-off meeting unit managers need to present:
  - Discrepancies between desired performance and the current state of affairs
  - Cases of discrepancies between prices in the order system and prices on the invoices
  - An estimation of the difference between expected and actual CM1. Important is to mention that a difference in CM1 is just an indication for incorrect net purchase prices because positive and negative differences in net purchase price in the order system and on the invoice can cancel each other which in theory could result in an actual CM1 equal to the

expected CM1 while there are incorrect net purchase prices in the catalog. - These facts need also be written in a manual and stored on a central location (for example intranet or a shared disc).
2. Make buyers aware of their responsibility in the process of entering and handling the net purchase prices - In the same kick-off meeting unit managers need to give an interactive presentation about the relation between the systems (explain for example the workflow presented in figure 4.1) and the influence of the tasks of the buyers on consecutive processes. - Store this information on a place accessible for all buyers (for example intranet or a shared disc) such that all the buyers can read it back whenever they need to.
3. Establish a coalition to guide the change - Unit managers have the power to guide the change among all the buyers.
4. Create and communicate the change vision with the involved employees - Involve selected buyers in the creation process of the change vision. Two weeks after the kick-off meeting unit managers need to organize a second meeting with one buyer of each category to brainstorm about the change vision. - Unit managers should communicate the change vision by an official briefing, and email to all the involved employees.
5. Change - Involve selected buyers in the design stage of the monitoring tool. Two weeks after the official briefing brainstorm sessions with the selected buyers about the user requirements need to be organized. Business analysts from the Product & Content department need to organize these sessions because they are the initiators of the dashboard. - During these brainstorm sessions the initiators of the monitoring tool need to clarify to the selected buyers how the change (proposed monitoring tool) will benefit the buyers, by explaining cases of different color combinations which could occur on the dashboard, how the buyers can respond to these cases and in what way they should interact with the suppliers. After these sessions all the buyers need to be informed by the business analysts from the Product & Content department by means of an interactive presentation. An alternative option is that the business analysts organize an interactive training for all the buyers. - Business analysts should set up a manual about how the buyers need to interpret the tables and figures of the dashboard. This manual need to be stored on a central location (for example intranet or a shared disc).
6. Evaluate the changes after a month - Business analysts together with the unit managers need to organize a meeting where all the buyers can give their opinion about the changes, positive or negative. To prevent a chaotic situation, it is recommended to split all the buyers into three groups.
7. Communicate the positive changes to the buyers - Unit managers need to send a monthly email to the buyers with the results achieved. Data and figures about among other the performance of the supplier and the buyer, and the more accurate CM1 calculation need to be included in the email to demonstrate the positive changes.

Table 7.2 Overview of change steps

7.3 Preliminary conclusion

A critically important contributor to the failure or success of change initiatives is the interaction between the buyers and the technology. Buyers need to be aware of their responsibility for correct net purchase prices in the catalog and need to realize that the proposed solutions can help them improving quality. This means that the solutions and recommendations need to be implemented in combination with a change in culture. The three-stage model of change developed by Lewin can be used for changing the culture. Only when this requirement is met the short-term solutions can be implemented. For the long-term solutions there is an additional requirement before implementation can start, this is the estimation and scheduling of the IT time needed. When the requirements are met the implementation of the solutions can start independently of each other.

Chapter 8: Conclusion and discussion

In this last chapter an answer is given to the main research question. Furthermore recommendations for further research are given and implications of this research for theory are discussed.

8.1 Conclusion

The sub questions which are formulated in section 2.5 are answered in the previous chapters, these answers are supportive for answering the main research question. The central question is:
How can the data entry and handling process be redesigned such that the quality of net purchase prices in the catalog will improve?

The relevance of correct net purchase prices in the catalog is shown by the fact that net purchase prices are the input factor for many consecutive processes, like calculation of the selling price for the web shop, calculation of the stock value, selection of the preferred supplier and calculation of the contribution margin. For assessing the data quality a new TDQM framework is developed, this framework is applicable in an operational context. The first three stages of the improved TDQM methodology, called definition, measurement and analysis, are applied to get more insight in the entry and handling process of net purchase prices and to find the critical areas in this process. The root causes of price differences between the order price and the invoice price are discovered during the case study and during interviews, resulting in the following list of causes:

1. Manual offers are entered too late.
2. Manual offers are incorrect, they are not free of error.
3. Incorrect override in the order system LSM.
4. Purchase deals are in the system for a longer time than agreed.
5. The system does not support the possibility to enter different discount percentages for products based on the number of purchased items, also called quantity discounts.
6. One EAN is coupled to two different products.
7. Unclear order units.
8. A double supplier offer in the catalog for one product.
9. Mapping issue Basware.
10. Other errors. This contains cases where a mismatch is discovered between the price on the invoice and the order price, but the causes are unknown. On different places in the handling process net purchase prices can be changed but control is missing, resulting in the fact that buyers cannot explain where the error could has arisen and even in combination with the existing reports no conclusion can be drawn.

An additional finding, resulting from interviews with the buyers, is that the culture among the buyers is also an issue which results in unreliable net purchase prices in the catalog. Buyers are not aware of their responsibility for correct net purchase prices in the catalog and they are not aware of the influence of their tasks on consecutive processes, like the calculation of the selling price and the calculation of the preferred supplier.

All ten causes discovered during the case study are types of production errors. According to literature, production errors can have influence on the intrinsic, contextual and representational data quality dimension. This has been confirmed by the case study.

Subsequently the last stage of the improved TDQM methodology, called improvement, is applied. First the focus is determined because given the time restriction not all the causes can be addressed. This is done by classifying the causes by causal agent. Three causal agents can be distinguished; causes can occur through mistakes of the buyers, causes can occur in the catalog if not the correct tools are available or if the system is not working correctly, and causes can occur in other systems in the data flow. The focus is on the five causes owing to incorrect or delayed operations by the buyers, these are cause number one, two, three, four and ten.
To bridge the gap between the current and the desired situation, three types of changes are necessary: changes in the IT system, changes in the process/working method, and changes in culture. All the changes aim to minimize or solve one or more causes. The changes are translated in concrete solutions and recommendation, all of these will contribute to the intrinsic, contextual or representational data quality dimension. The solutions and recommendations are:

- Implement a dashboard for monitoring the net purchase prices during the handling process.
- Implement the possibility for buyers to enter the number of products as an end condition of purchase deals.
- Suppliers need to send a new price list to at least two buyers and this need to be stored on a central disc which is only accessible for the buyers who are doing business with the supplier.
- Buyers should block time in their agenda to make sure that they enter manual offers before the date from which the offers are valid.
- Existing reports about mutations in net purchase prices in the catalog need to be used to check if the manual offers are entered correctly. A workflow is created which shows how buyers need to handle when entering manual offers.

For evaluating the monitoring tool a survey is conducted among the users of the tool. From the survey can be concluded that the characteristics of the monitoring tool match with users' task needs. The Task-Technology Fit perspective, the concept on which the evaluations are based, assumes that a better fit between technology functionalities and task requirements results in better performance. A discussion point is the contribution of the monitoring tool to the research question. A dashboard which presents the number of price changes would be more appropriate for answering the main research question because the number of price changes on the different places during the handling process gives an indication for the reliability of the net purchase prices in the catalog. However from the interviews with the stakeholders can be concluded that the value of price changes is also playing a key role, resulting in the proposed dashboard where each KPI shows the value of price changes. Ultimately both options have the aim to improve the quality of net purchase prices in the catalog, only the way to reach this goal is slightly different. Despite of the fact that showing the value of prices changes does not match the problem statement in detail, it matches the needs of the stakeholders resulting in a high fit with higher performance as a logical consequence.

The requirement that must be met before implementation of the solutions can start, is that the culture among the buyers need to change. The interaction between the buyers and the technology is a critically important contributor to the failure or success of change initiatives. Only when the buyers are aware of their responsibility for correct net purchase prices in the catalog and realize that the proposed solutions can help them improving these quality, the short-term solutions can be implemented successfully. For the long-term solutions there is an additional requirement before

implementation can start, this is the estimation and scheduling of the IT time needed. When the requirements are met the implementation of the solutions can start independently of each other.

For conducting this research within the prescribed time of 20 weeks the research is delimited to only the flow of the net purchase prices. Interesting for future research is to extend this research by including the quantities. This means conducting a research on the match between the number of products that are ordered, the number of products on the invoice and, even more important, the number of products that are delivered at the warehouse. Subsequently, possibilities for increasing this matching percentage need to be investigated. Interesting is that also the logistic process at the warehouse is included, it provides a more complete picture of the different processes of bol.com.

8.2 Implications for theory

This research has confirmed that the standard TDQM methodology cannot be applied in practice without doing some contextual redesign. The needs of the business environment need to be discovered for selecting the relevant stages, steps and tools. In this research a TDQM framework applicable for operational processes is developed. In an operational context it is essential to understand the complete set of processes and how the quality of data at each processing stage is impacted by its quality at previous stages. This means that the developed TDQM framework should help assessing data quality at the final stage but, equally important, also at any of the preceding intermediate stages. The developed TDQM methodology is made applicable for operational processes by adapting, improving and adding several steps. As already discussed in chapter three, the critical steps that are added are:

- Identification of the causes which have the highest priority for improvement, as part of the analysis stage. Also a new tool for these identification is introduced.
- Definition of the requirements for redesign as part of the improvement stage.

Furthermore the methodology is adapted and/or improved on the following points:

- The process modelling step, as part of the definition stage, is described in more detail. Depending on the aim of the researcher processes can be modeled on different levels; a high level model which shows the relations between the operational systems and a low level model helps understanding the details of an operational process.
- A new way of measuring data quality is introduced. The subjective metric could be useful in case it is hard to use objective metrics for measuring the quality of operational processes. A subjective measurement is based on qualitative evaluations by the stakeholders.
- The analysis stage is defined in more detail. For identifying root causes all the operational systems and the associated processes need to be examined, meaning not only the final process.

Secondly, this research has proven that the relationship between data quality problems and data quality dimensions also exists in practice. However this is only proven for the ten data errors found in practice, which are all types of production errors. Looking to the life cycle of data, production errors are related to only one stage in this cycle. A detailed discussion about quality problems in the other stages of the life cycle is missing in literature. Problems in the other stages are problems related to data storage and maintenance, and problems related to the use of data. In what way the problems within these two stages influence the data quality dimensions in practice is also unknown. In the field of data quality this is an interesting topic for future research.

Another topic which needs further research is the link between Task-Technology Fit and performance. As already mentioned in chapter six, researchers assume that a better fit between technology functionalities and user tasks will lead to a better performance, but this link is not proven. In this research the assumption is taken that a better fit results in better performance, but the proof of this statement is definitely worth investigating. This topic for further research is not especially related to the field of data quality but refers to the evaluation of models, tools and methods in general.

References

Agosta, L. (2002). The Meaning of Data Quality Retrieved 7/18/2013, from http://www.information-management.com/issues/20021001/5817-1.html

Ahold. (2012a). Bol.com opens new mobile store to give customers access to product range anytime, anywhere Retrieved 7/3/2013, from http://www.ahold.com/Media/bol.com.htm

Ahold. (2012b). Customers able to pick up bol.com orders in Albert Heijn stores Retrieved 7/3/2013, from http://www.ahold.com/Media/bol.com.htm

Ahold. (2013). Our brands: bol.com facts and figures Retrieved 7/3/2013, from http://www.ahold.com/Media/bol.com.htm

Anderson, D., & Ackerman Anderson, L. (2010). Beyond Change Management: How to achieve breakthrough results through conscious change leadership

Bai, X., Nunez, M., & Kalagnanam, J. R. (2012). Managing Data Quality Risk in Accounting Information Systems. *Information Systems Research, 23*(2), 453-473. doi: 10.1287/isre.1110.0371

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for Data Quality Assessment and Improvement. *ACM Computing Surveys, 41*(3), 1-52. doi: 10.1145/1541880.1541883

Batini, C., & Scannapieca, M. (2006). Data Quality: Concepts, Methodologies and Techniques

Bol.com. (2013). Opmerkelijke geschiedenis Retrieved 7/3/2013, from http://banen.bol.com/over-bol-com/opmerkelijke-geschiedenis/

Cleven, A., Gubler, P., & Hüner, K. M. (2009). *Design Alternatives for the Evaluation of Design Science Research Artifacts*. Paper presented at the International Conference on Design Science Research in Information Systems and Technology.

Daft, R. L. (2007). *Understanding the Theory and Design of Organizations*: Thomson South-Western.

Daft, R. L. (2008). *New Era of Management* (Second edition ed.): Thomson South-Western.

Eckerson, W. (2002). Data Quality and the Bottom Line. Achieving Business Succes through a Commitment to High Quality Data *Applications Development Trends* (pp. 1-36): The Data Warehousing Institute.

Eppler, M. J. (2006). Managing Information Quality *Increasing the Value of Information in Knowledge-intensive Products and Processes* doi:10.1007/3-540-32225-6

Goodhue, D. L. (1995). Understanding User Evaluations of Information Systems. *Management Science, 41*(12), 1-18.

Goodhue, D. L., & Thompson, R. L. (1995). Task-Technology Fit and Individual Performance. *MIS Quarterly*, 1-22.

Görz, Q., & Kaiser, M. (2012). An Indicator Function for Insufficient Data Quality. A Contribution to Data Accuracy. 169-184.

IAIDQ. (2010). DQ glossary Retrieved 7/21/2013, from http://iaidq.org/

IAIDQ. (2011). Classifying data quality problems Retrieved 7/21/2013, from http://iaidq.org/

Intranet-bol.com. (2013). Confidential

Jiang, L., & Zhao, J. (2012). *An empirical study on risk data quality management.* Paper presented at the International Conference on Information Management, Innovation Management and Industrial Engineering.

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment. *Management and Information, 40*, 133-146.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems, 24*(3), 45-77. doi: 10.2753/MIS0742-1222240302

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data Quality Assessment *Communications of the ACM, 45*(4), 211-218.

Rausand, M. (2011). Risk Assessment *Theory, Methods and Applications*

Redman, T. C. (2008). Data Driven: Profiting from Your Most Important Business Asset

Retaildetail. (2013). E-tailer bol.com wil 'gewoon de beste winkel zijn' Retrieved 7/3/2013, from http://www.retaildetail.eu/en/in-beeld/item/14831-e-tailer-bolcom-wil-%E2%80%9Cgewoon-de-beste-winkel-zijn%E2%80%9D

Schwartz, R. B., & Russo, M. C. (2004). How to Quickly Find Articles in the Top IS Journals. *Communications of the ACM, 47*(2), 98-101.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data Quality in Context. *Communications of the ACM, 40*(5), 103-110.

TwinkleMagazine. (2012). Bol.com wil startpunt worden voor elke webshopper Retrieved 7/3/2013, from http://www.twinklemagazine.nl/nieuws/2012/02/bol.com-wil-startpunt-worden-voor-elke-webshopper/

Van den Berg, H., Franken, H., & Jonkers, H. (2008). Handboek Business Process Engineering

Volkskrant. (2012). Ahold rond overname bol.com af Retrieved 7/3/2013, from http://www.volkskrant.nl/vk/nl/2680/Economie/article/detail/3252967/2012/05/09/Ahold-rondt-overname-Bol-com-af.dhtml

Wang, R. Y., Storey, V. C., & Firth, C. P. (1995). A Framework for Analysis of Data Quality Research. *IEEE Transactions on Knowledge and Data Engineering, 7*(4), 623-640.

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems, 12*(4), 5-33.

Weske, M. (2007). Business Process Management: Concepts, Languages, Architectures

White, S. A. (2004). Introduction to BPMN Retrieved 10/24/2013, from http://www.bpmn.org/

Wijnhoven, F. (2012). Information failures: A typology, causes and consequences using a Churchmanian epistemological perspective. 1-19.

Wijnhoven, F., Boelens, R., Middel, R., & Louissen, K. (2007). *Total Data Quality Management: A Study of Bridging Rigor and Relevance*. Paper presented at the European Conference on Information Systems

Appendix A: Definitions data aspects

The table below shows the definitions of the data quality dimensions and aspects according to Wang and Strong (1996). These dimensions and aspects are discussed in section 3.2.

Dimension	Aspect	Definition: the extent to which …
Intrinsic data quality	Believability	… data is accepted or regarded as true, real and credible
	Accuracy	… data is correct, reliable and certified free of error
	Objectivity	… data is unbiased and impartial
	Reputation	… data is trusted or highly regarded in terms of their source and content
Contextual data quality	Value-added	… data is beneficial and provide advantages for their use
	Relevancy	… data is applicable and useful for the task at hand
	Timeliness	… the age of the data is appropriate for the task at hand
	Completeness	… data is of sufficient depth, breadth and scope for the task at hand
	Appropriate amount of data	… the quantity or volume of available data is appropriate
Representational data quality	Interpretability	… data is in appropriate language and unit and the data definitions are clear
	Ease of understanding	… data is clear without ambiguity and easily comprehended
	Representational consistency	… data is always presented in the same format and are compatible with the previous data
	Concise representation	… data is compactly represented without being overwhelmed
Accessibility of data quality	Accessibility	… data is available or easily and quickly retrieved
	Access security	… access to data can be restricted and hence kept secure

Table A.1 Dimensions and aspects of data quality (Wang & Strong, 1996)

Appendix B: Abbreviations data quality methodologies

The table below shows the abbreviations and full names of the different data quality methodologies. These methodologies are discussed in section 3.4.

Abbreviation methodology	Full name methodology
TDQM	Total Data Quality Management
DWQ	The Data Warehouse Quality Methodology
TIQM	Total Information Quality Management
AIMQ	A methodology for information quality assessment
CIHI	Canadian Institute for Health Information methodology
DQA	Data Quality Assessment
IQM	Information Quality Measurement
ISTAT	ISTAT methodology
AMEQ	Activity-based Measuring and Evaluating of product information Quality (AMEQ) methodology
COLDQ	Loshin Methodology (Cost-effect Of Low Data Quality)
DaQuinCIS	Data Quality in Cooperative Information Systems
QAFD	Methodology for the Quality Assessment of Financial Data
CDQ	Comprehensive methodology for Data Quality management

Table B.1 Abbreviations of data quality methodologies (Batini, et al., 2009)

Appendix C: Survey

This appendix contains measures for evaluating the dashboard. Statements about characteristics of the dashboard as well as statements about task characteristics are used as measure.

1 = strongly disagree and 5 = strongly agree

Right data

1. The data shown on the dashboard is pretty much what I need to carry out my tasks	1 … 2 … 3 … 4 … 5
2. Critical data, that would be very useful to me in executing my tasks, is not missing	1 … 2 … 3 … 4 … 5

Right level of detail

1. The dashboard shows data at an appropriate level of detail for my tasks and purposes	1 … 2 … 3 … 4 … 5

Accessibility*

1. I can get data quickly and easily when I need it	1 … 2 … 3 … 4 … 5

Easy to use*

1. It is easy to learn how to use the system	1 … 2 … 3 … 4 … 5
2. It is easy to change the settings/filters of the dashboard	1 … 2 … 3 … 4 … 5

System reliability*

1. The system which provides the data is not subject to unexpected or inconvenient down times	1 … 2 … 3 … 4 … 5
2. I can count on the system to be 'up' and available when I need it	1 … 2 … 3 … 4 … 5

Presentation

1. The data that I need is displayed in an understandable format	1 … 2 … 3 … 4 … 5
2. The data that I need is displayed in a readable format	1 … 2 … 3 … 4 … 5

Currency**

1. The data is up-to-date enough for my purposes	1 … 2 … 3 … 4 … 5

Difficult or non-routine tasks

1. I frequently deal with routine business problems	1 … 2 … 3 … 4 … 5
2. Frequently the business problems I work on involve answering questions that have been asked in quite that form before	1 … 2 … 3 … 4 … 5

* Because the dashboard is not yet running, the terms ‘accessibility’, ‘easy to use’ and ‘system reliability’ cannot be measured. The dashboard will be integrated in the existing reporting system called BIT. This reporting system is already in use for generating other type of reports, so the agreement or disagreement of these three terms can be based on the current experience with this reporting system.

** Assume that the net purchase prices, and so the KPI’s, are presented in the dashboard as soon as the invoice is received. When the invoice is received, all the net purchase prices in the different systems during the handling process are known and can be linked to one purchase order line.

Appendix D: Descriptive statistics

The table below shows the descriptive statistics as a result of the surveys. The eight variables in the column on the left correspond with the measures in the survey.

1 = strongly disagree and 5 = strongly agree

Variable	Number of respondents	Mean	Standard deviation	Minimum	Maximum
Right data	5	4.6	0.489	4	5
Right level of detail	5	4.8	0.400	4	5
Accessibility	5	3.8	0.400	3	4
Easy to use	5	4.7	0.471	3	5
System reliability	5	3.8	0.245	3	4
Presentation	5	4.7	0.400	4	5
Currency	5	4.9	0.200	4	5
Difficult or non-routine tasks	5	4.8	0.245	4	5

Table D.1 Descriptive statistics

All the measures are equally important. Some measures consist of more than one statement, in these cases the average score of the two statements is included in the statistics. In this way the eight measures are assessed in a similar way.

For example when one respondent is scoring the two statements belonging to the measure 'easy to use' as follows: a score of four to the first statement and a score of five to the second statement, than for this respondent the average of 4.5 is taken for the measure 'easy to use'.

Appendix E: Literature details

This appendix gives more details about the literature search of chapter three. For executing the literature study different e-books, journals, search engines and websites are used.

Used search engines:

- Scopus, a bibliographic database containing abstracts and citations for academic journal articles.

Search terms:

- Data quality
- Data quality management
- Data accuracy
- Data quality problems
- Data quality assessment methodologies
- Impact data quality

Used journals:

- Journal of Management Information Systems
- IEEE transactions on knowledge and data engineering
- ACM computing surveys
- Communications of the ACM
- European Journal of Information Systems
- Management and Information
- Information Systems Research

These journals all do have a position in the top 50 IS Journals, as can be seen in the table below.

WORLD RANK	TITLE	WORLD RANK	TITLE	WORLD RANK	TITLE
1	*MIS Quarterly*	18	*Communications of the AIS*	35	*Journal of Information Systems*
2	*Communications of the ACM*	19	*IEEE Computer*	36	*The Information Society*
3	*IS Research*	20	*Journal of Strategic IS*	37	*Journal E-U Computing*
4	*Journal of MIS*	21	*Admin. Science Quarterly*	38	*Info Resources Mgmt Journal*
5	*Management Science*	22	*Academy of Mgmt Review*	39	*Interfaces*
6	*IEEE Transactions (various)*	23	*Int'l Journal of E-Commerce*	40	*EM - Electronic Markets*
7	*Harvard Business Review*	24	*ACM Computing Surveys*	41	*Journal of CIS*
8	*Decision Sciences*	25	*Accounting, Management & IT*	42	*European Journal of OR*
9	*Decision Support Systems*	26	*ACM SIG Publications*	43	*Operations Research*
10	*Information and Management*	27	*IT and People*	44	*Int'l Journal of H-C Studies*
11	*European Journal of IS*	28	*IBM Systems Journal*	45	*Journal of the ACM*
12	*Sloan Management Review*	29	*OMEGA*	46	*Australian Journal of IS*
13	*ACM Transactions (various)*	30	*Journal of the AIS*	47	*Org. Behavior and Human Dec.*
14	*Data Base*	31	*Journal of Org., Comp. and EC*	48	*Behavior and IT*
15	*Organization Science*	32	*Human-Computer Interaction*	49	*Scandinavian Journal of IS*
16	*Information Systems Journal*	33	*Information Systems Management*	50	*Computer Journal*
17	*Academy of Management Journal*	34	*Int'l Journal of Man-Machine Studies*		

Table E.1 Top 50 IS Journals (Schwartz & Russo, 2004)

Most of the literature is found by means of backward search. With this search methodology the reference list of an article or book is checked down.