TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

ZHANG, YIFEI

DESIGNING A SUPPLIER DATABASE TO SUPPORT SOURCING ANALYSIS

Master's thesis

Examiner: Professor Samuli Pekkola
Examiner and topic approved by the Faculty Council of the Faculty of Business and Built Environment on 6 February 2013.

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Degree Programme in Business and Build Environment
ZHANG, YIFEI: Designing a supplier database to support sourcing analysis
Master of Science Thesis, 55 pages, 4 Appendix pages
September 2013
Major subject: Business and Technology
Examiner: Professor Samuli Pekkola
Keywords: master data management, supplier database, sourcing analyses and decisions

Nowadays, more and more business organizations emphasize the importance of competitive suppliers in their strategic benefits. On one hand, with the extension of global industrialization, there are much more suppliers available for a company than ever before, which results in the difficulty for the company to develop a proper sourcing decision. On the other hand, advanced data processing technologies can help companies improve their capability of sourcing data analysis significantly. Thus, how to design an effective supplier database to support sourcing analysis has become an essential topic in recent years.

This thesis aims to design a supplier database based on the theory of master data management (MDM) for processing B2B sourcing data, and the discussions are presented according to the previous research in the sourcing department of Sandvik Mining and Construction (China). The whole design was developed mainly in 5 stages. Firstly, necessary information was gathered from the company as well as some academic literatures. Secondly, pre-research preparation was done by studying all the collected information. Thirdly, a framework of the database was designed out. Fourthly, more detail design was proposed in different modules, focusing on some practical issues in the existing data systems. Fifthly, the outcome of the design was discussed in a comprehensive way.

Although the MDM-oriented supplier database was designed for a specific project in a company, some basic principles of the design can be adopted prevalently. However, it is necessary to claim that this design is not totally completed, still with more practical work to be done in the future. Regarding the further research in this case, more professionals from different backgrounds (i.e. IT, business, management, etc.) should be involved, and the organizational support from the company is important as well.

PREFACE

This thesis aims to develop an MDM-oriented supplier database for the sourcing department of Sandvik Mining Corporation (China), supporting B2B sourcing analyses and decisions. The project provides an excellent opportunity for a business student to approach the academic theories of master data management as well as practical sourcing operation. More importantly, it is a great chance to convert some theoretical methodologies from academic papers into practice in real-life B2B world.

I would like to thank my thesis supervisor Professor Samuli Pekkola for the critical analysis and guidance throughout the execution of the research. In addition, I would also thank Mr. David Gui, Mr. Jing Yang, Mr. Andras Huszta, and many other colleagues in Sandvik Mining and Construction, who have kindly offered me their help and trust for my research in the company.

Tampere, September 2013

Yifei Zhang

CONTENTS

1. INTRODUCTION

1.1. Background

In the recent decade, two notable trends have dramatically changed the general environment of international business-to-business (B2B) market. One trend refers to the worldwide spread of globalization that has almost reached every corner of those populous continents, while the other trend is the rapid development of information technology (IT) which brings the world to an unprecedented era of data. On one hand, globalization drives the B2B networks to become much more complicated and more flexible than ever before. On the other hand, data management as an original IT concept has been introduced to business, and significantly accelerates the development of data analysis in most business areas. Facing to the fierce competition, many companies are trying to data technologies to develop reliable supply networks and rather sophisticated management skills. Benefits of effective supply chain management include lower inventories, lower costs, higher productivity, greater agility, shorter lead time, higher profits, and greater customer loyalty (Chu & Varma, 2011; Stevenson, 2008). Undoubtedly, how to select one or several competitive suppliers is one of the key issues for supply chain management. According to Ghodsypour and O'Brien (1998), in most industries, the cost of raw materials and component parts constitute the main cost of the product; in some cases it accounts for up to 70%. As purchasing activities within a supply chain play a more strategic role and lead the movement from spot purchasing to long-term contractual relationships, sound suppliers selection has become a strategic decision, meaning that it has become a vital source for adding strength to value proposition and for improving the competitiveness of manufacturers (Chu & Varma, 2011; Ha & Krishnan, 2008; Wise & Morrision, 2000). In this situation, China has gradually become a global sourcing center, since its remarkable manufacturing capability, relatively lower costs, as well as its reliable availability of various products.

Different from business-to-customer (B2C) market, purchasing activities in B2B market are much more organized and rational, with less individual effects in the psychological aspect. In other words, in order to select the most suitable suppliers or business partners from large numbers of candidates, the judgment should be made based on certain "evidence" (or with necessary data support). Thus, a multi-functional data system needs to be established in order to store all those large sourcing data and to provide necessary supports for the analysis of both specific sourcing case and some strategic sourcing decisions. Moreover, to evaluate whether a potential supplier is suitable for the buyer company or not does not only depend on its products price, but also many other aspects need to be taken into account, including its technological capability, general environment, transportation, reliability, etc. As a result, the new reliable data system should consist of two vital characteristics. Firstly, it is required to contain as many as business related factors, otherwise, the neglect of certain business affected factors may result in an incorrect outcome of analysis. Secondly, some most valuable information behind massive data could be easily discovered in the system, so that it would ensure the work to be done more efficiently by users. Technically, master data management is considered as an academic approach to realize the characteristics of a database in terms of a structural design.

1.2. SMC in China

Sandvik is a high-technology engineering group with advanced products and a global leading position in its areas, including tooling, materials technology, mining and construction. According to Sandvik Annual Report (2012), mining and construction industries together accounted for more than 50% of the group's total 99 billion SEK invoiced sales in 2012. Since the company was founded in 1862, more than 150 years have passed. (Fagerfjäll, 2012) Nowadays, Sandvik's worldwide business activities are conducted through representation in more than 130 countries. In 1985, it was the first time for Sandvik to enter the Chinese market, until last year, China had already become the third largest market for the company with remarkable growth rate around 9% in 2012. In the same year, the group had 49,000 employees all over the world, with 3,221 employees in China. (Sandvik World, 2013) As the Asia-Pacific market, including Asian countries and Australia, is the biggest market for Sandvik outside Europe, the company's largest global assembly center for mining and construction equipment has been established in Shanghai, China. In addition, the nationwide accelerating industrialization in China has brought a lot of confidence to foreign investors like Sandvik. According to Fagerfjäll (2012), "Sandvik is investing resolutely in China, not to manufacture cheap products, but to participate in the largest collective industrial investment ever made." In recent two years, Sandvik Mining and Construction (SMC) experienced a series of internal reorganization, aiming to separate Sandvik Mining from Sandvik Construction to form two independent subsidiaries. However, as there is certain crossing domain in between, especially for the sourcing business, SMC is still considered as an integrated company in this research.

Since there are some issues mentioned in this paper referring to the inter-department coordination and cross-sectors information management, it is necessary to have a clear understanding of the organizational structure in Sandvik Corporation. Generally, a concrete organizational structure of the company is shown in Figure 1.1.

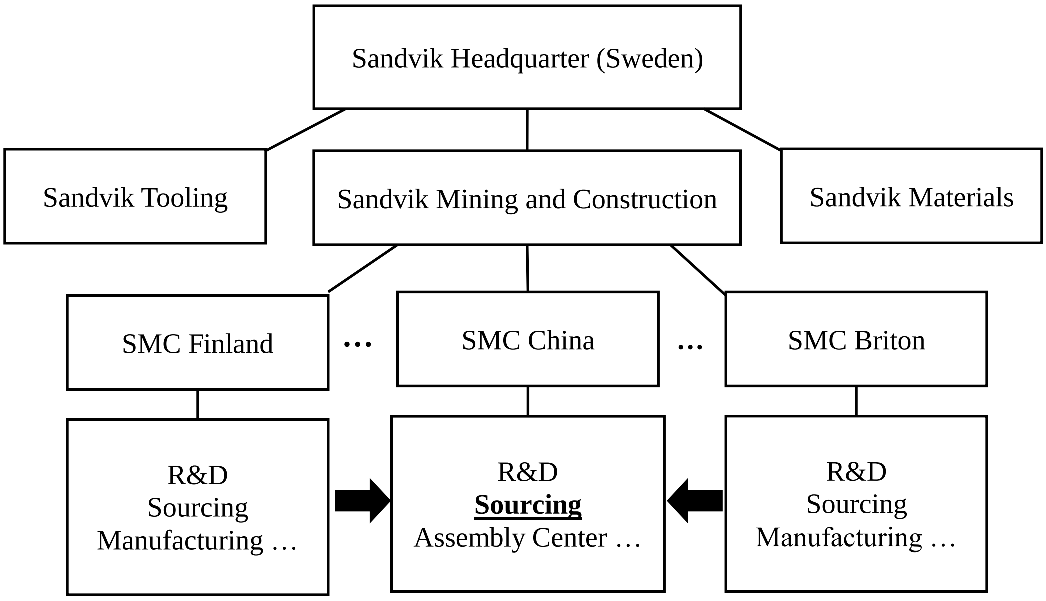


Figure 1.1. *Organizational Structure in Sandvik*

As mentioned before, Sandvik Group consists of three main branches – Sandvik Tooling, Sandvik Mining and Construction, and Sandvik Materials Technology. Each branch comprises of several subsidiaries, which are usually set up based on their geographical divisions. In the figure above, it takes Sandvik Mining and Construction as an example. SMC has a number of sub-branches over the world, and every sub-branch should be relatively independent from each other. However, as the global assembly center was established in China as well as the rapid growth of its domestic market, more and more manufacturing work is transferred to SMC (China). As a result, the workload of the sourcing department in SMC (China) is increasing correspondingly, and a demand arises in the department for a more efficient sourcing data system.

1.3. Objective of the Thesis

After the global financial crisis in 2008, the world economic structure has changed dramatically. With remarkable performance on economic development in recent years, new emerging markets, especially China plays a more and more important role in the global economy. Nowadays, China is not only a global manufacturing center, but also a huge market for various kinds of products. In this condition, many international companies try to improve their global supply networks in order to enhance their competitiveness. Sourcing activities, as the most essential part of the whole supply chain, can affect the company's an organization's future by reducing the operational costs and improving the quality of its end products (Zeydan et al, 2011). Thus, how to develop a reasonable sourcing decision has become the vital issue for every company. On the other hand, more and more advanced information technology provides a considerable approach for the issue. Some techniques are used for managing the content and context of value creation within key supplier relationship (Makkonen & Olkkonen, 2013; Pardo et al, 2011). In this condition, an IT concept of master data management (MDM) would be brought in to the business area in order to improve the efficiency of processing massive sourcing data. In a word, the main objective of this paper is…

… to design an MDM-oriented supplier database which can be used for the general management of supplier information and providing necessary support for sourcing case analysis as well as decision-making.

This paper is structured in seven parts. The first chapter aims to give some brief background information and the main objective of this paper. The second chapter generally describes the previous research information and the research procedure of this paper. The third chapter refers to the analysis on those existing theories, including both data management and business factors. The fourth chapter describes the entire design of the MDM-oriented supplier database. The fifth chapter focuses on the process of designing and its outcomes in an empirical perspective. The sixth chapter presents some discussion as well as supplementation of the whole research, and gives certain advices for further research. In the seventh chapter, the core content of the whole paper is concluded in a relatively brief summary.

2. RESEARCH METHODS AND PROCESS

2.1. Research Methods

According to Gummesson (1993), there are five qualitative data generating methods for case study researches, including using existing material, questionnaire surveys, qualitative interviews, observation, and action science. The advantages and disadvantages of the aforementioned research methods are listed in Table 1.

Table 2.1. *Advantages and disadvantages of the different research methods.*

Research Methods	**Advantages**	**Disadvantages**
Using Existing Material	1. Relatively low cost 2. Theoretical support	1. Limitation of mature theories 2. Effects on the author's creativity
Questionnaire Surveys	1. Standardized survey 2. Real information	1. Subjective testers' answers 2. Questions may be set subjectively
Qualitative Interviews	1. Qualified interviewees 2. Direct communication	1. Hard to find an expert 2. Subjective analysis on the interview report
Observation	1. Easy for operation 2. Flexibility to interact with subjects	1. Highly dependent on the observer's judgment 2. Long duration for the operation process
Action Science	1. Combination of all other methods above	1. Total involvement of the subject may affect the result of the research

Notably, each research method has its own merits and drawbacks, and different methods can usually fit different research conditions. As a result, researchers should carefully choose the methods according to the nature of the study subject (Chen, 2011). Regarding this research, all the five methods mentioned in table 2.1 are adopted to generate qualitative data.

The first approach is self-participation in the real business, which can be scientifically called “action science”. The author of this paper has worked in the sourcing department of Sandvik Mining and Construction (China) for five months as an intern in order to get access to the first-hand information about the subject. Usually, this kind of information is considered to be rather reliable, because it has not been influenced by the third party during its transmission. However, this information gathering method is not a perfect one, because it highly depends on the researcher’s personal experience, which may not present the comprehensive understanding of the fact.

The second approach is observation. To some extent, it may be considered as a part of self-participation or action science, because some really problems cannot be observed outside the company. As an intern in the company, the author had a good opportunity to observe the current issues related to the existing systems, as well as different reactions from employees. However, as is mentioned in table 2.1, the results of the observation are often highly dependent on the observer. In other words, the subjective attitudes or even personal emotion may affect the scientific outcomes of the research. So, in order to guarantee the reliability of the research, this approach is only used to improve the understanding of the general condition instead of specific data gathering.

The third approach is personal interview. Since some people were not active enough in the survey, some face-to-face interviews were carried out. Actually, interviewees performed much more actively in the personal interviews, because communication with the interviewer could inspire their own thoughts. According to Ott, R. L. & Longnecker, M. (2010), the quality of an interview, to some extent, depends on the capability of the interviewer. If they are not well trained, some detail information (i.e. facial expression) may be neglected, and even may result in a bias into the sample data. In terms of this case, although the interviewer is not professionally trained, the content of interviews does not include any complex social or political issues. So, the outcome of interviews is definitely able to reflect some information concerning the research subject. In addition, it is necessary to note that some information would not be provided by the interviewees as confidential business information, and some information was directly provided by the interviewees, which could not be verified by any authority.

The fourth approach refers to questionnaire surveys. The researcher had made some questionnaires and distributed them to the employees in different positions of the sourcing department, and then all the questionnaires were collected and analyzed together. In the questionnaires, there was not any optional answer available for the questions, so that the employees’ minds would not be limited to the certain options. Generally, most people were quite cooperative in the survey, and finished the questionnaires in a relatively short time. The survey indicated that some employees were not rather familiar with the concept of master data management. At the same time, it could definitely reflect the some notable issues in the existing data systems.

The fifth approach is using existing materials. This approach is an important addition to the previous methods, as it mainly focuses on the study of some indirect materials. For example, some subject-related academic literatures should be read before the research, though they may not concern the exactly same issue in the case. Some of these literatures are helpful for the researcher to establish a scientific knowledge framework of the subject rather than providing a specific solution for this case. A lot of theories and methodologies in previous researches are systematically presented in the existing materials, which can be finally converted into some guidance or support of this research.

2.2. Research Process

The research is based on the project of master data management in Sandvik Mining and Construction (China). Generally, the research process consists of four steps and each step follows another, and all the process can be introduced with a timeline, which is presented in Figure 2.1.

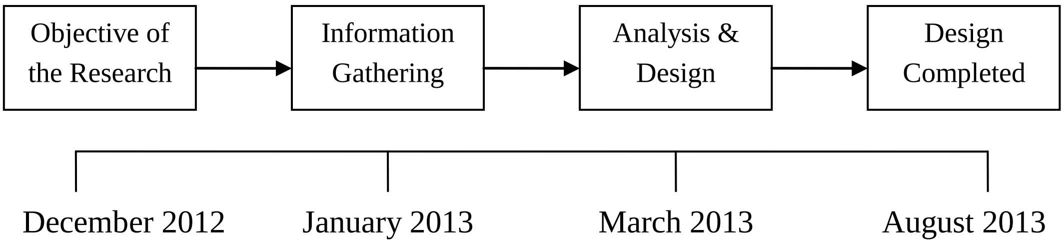


Figure 2.1. *Timeline of Research Process*

In December 2012, the objective of the research was set up as to design a supplier database for the sourcing department of Sandvik Mining and Construction (China). More specifically, the new database should be designed in the perspective of master data management, and be able to improve the general data management situation in the department. The second step started in January 2013, and the main task of this step was to collect all the information required in the research, including some MDM related theories, the drawbacks of the existing data systems, employees' concerning aspects about a new database, etc. Thanks to the management of the company, the researcher was able to work inside the department for a four-month internship since January, which provided the researcher a good opportunity to get access to some important information within the department. For instance, face-to-face interview is an effective approach to take advantage of the "human resource" inside the department to target the issue. A number of experienced professionals in different positions have been interviewed in the research, e.g. Mr. M, a category manager of hydraulic components with 9 years sourcing experience; Ms. S, a sourcing engineer of electrical components with 7 years sourcing experience; Mr. H, a sourcing engineer of steel components with 5 years sourcing experience, etc. (It is necessary to claim that all the full names of interviewees are represented with their initial letters in this paper, since the author has not been authorized to mention their names.) Actually, some surveys about the issues of the existing data systems had been done before each interview, but generally the results from surveys were the same as that in the interview. As a result, this paper mainly focuses on the analyses of interviews, and more detail information about these interviews is discussed in the following chapters. With sufficient information preparation, the research entered the third step in March 2013. In the following months, the researcher analyzed some general structures of databases as well as essential sourcing factors in a B2B market, and then, focused on designing a comprehensive supplier database for the department. Since designing a supplier database was rather a big system project, it was broken down into several sections. Each section emphasized a specific functional module within the whole system, and all the modules were designed one by one. In August 2013, when all functional modules were designed, they were finally combined together and integrated into a complete supplier database. In the last stage of the research, it mainly focused on the coordination among different modules. Despite they were designed separately, they needed to work with each other in order to form a complete system. In the end, several uncertainties in the design were discussed and some advices were given for further researches.

3. THEORETICAL ANALYSIS

This chapter identifies some academic concepts regarding Master Data, Master Data Management as well as Database in the first two sections, building up a theoretical foundation for further analyses. In the third section, the relationship between a sourcing decision making process and the usage of master data is analyzed in order to explore the significance of master data for developing a reliable sourcing decision.

Originally, there was not any academic concept referring to Master Data or the relational fields, because most organizations were not able to process large amount of data with quite limited computing resources. During those periods, "flat data files" (Loshin, 2009) ruled the world. However, with the remarkable development of information technology, human beings are capable enough to accomplish more and more accurate calculations with the help of computers. As a result, many global giants are keen on expanding their business all over the world, and much more data have been collected and analyzed than ever before in order to achieve a comprehensively optimized solution. While various types of data have been consistently aggregating many analysts got lost in the ocean of data, the concept of master data has been proposed in order to improve the efficiency of data processing.

3.1. Master Data

As a relatively new concept, there is not a standard definition given in the academic or business field, while various definitions have been raised in the different perspectives. In the following paragraph, some different definitions of master data are discussed in order to achieve a widely accepted one.

Generally speaking, master data may be roughly defined as the slowly changing fundamental data for transactions (Kokemuller & Weisbecker, 2009). At the level of practical study, master data refer to the critical business information supporting the transactional and analytical operations of the enterprise (Oracle, 2009). More specifically, they are the core reference data that can describe the fundamental dimensions—customer, material, vendor, chart of accounts, etc. (Zynapse, 2010). According to Wolter and Haselden (2006), master data are the critical nouns of a business, referring to four main groups: people, things, places, and concepts. To sum up, a more comprehensive definition is adopted in this thesis, and it is explained as following. Master data, as an academic concept, is defined as "the data about the characteristics of key business objects in a company and is unambiguously defined as well as uniquely identified across the organization. It represents a company's key business objects. These key business objects form the foundation of the company's business purpose and must therefore be used unambiguously across the entire organization." (Otto, 2012; Dreibelbis et al, 2008; Loshin, 2009; Smith & McKeen, 2008). Taking time factors as an example, master data form the basis of business processes, denoting a company's essential basic data that remain unchanged over a specific period of time (Loser et al, 2004; Rosenberg, 1987).

Actually, all the master data are closely related to business intelligence (BI), and most of them are widely used in business analysis that determines the final decision. According to White (2007), "Master data can be used in BI for both historical and predictive analysis." For example, a set of master data that refer to suppliers' addresses are retained in a company's database, and they can be combined with other historical business transaction data to produce certain analytical reports. On the other hand, some new master data should be collected and then combined with some relative historical data to predict the effect of the changes on its sourcing activities. Loshin (2009) claimed that master data objects usually share the following six characteristics. (1) They may be relevant to multiple business areas and business processes. (2) They are effective data extracted from transactions and can be used as analytical system records. (3) They can usually be classified within a semantic hierarchy with different attribution and specialization and so on. (4) They may have specialized application functions to create, update or remove certain instance records, such as creating a new "supplier" record. (5) They are likely to have models for multiple applications or unmolded within flat file structures. (6) They are often managed separately in many systems associated with many different applications. All these characteristics can be used to target sourcing master data objects in the design of a new MDM-oriented database.

3.2. MDM-oriented Database

The purpose of this paper aims to design an MDM-oriented database to effectively support sourcing analyses and decisions. It indicates that the database is developed for optimizing master data management. Thus, two relevant questions should be answered in advance. (1) What is master data management? (2) What is a database?

3.2.1 Master Data Management

Similar to master data, there is not a consensus in the academic field about the definition of master data management, but most definitions share more commonplaces than differences. In this thesis, Master Data Management (MDM) is defined as "a collection of best data management practices that orchestrate key stakeholders, participants, and business clients in incorporating the business applications, information management methods, and data infrastructure to support the capture, integration, and subsequent shared use of accurate, timely, consistent, and complete master data." (Loshin, 2009) It ensures consistency and accuracy of the data by providing a single set of guidelines for their management, moreover, it creates a common view on key company data, which may or may not be held in a common data source (Smith & McKeen, 2008). From a practical point of view, master data management should not be simply considered as a class of information system, but rather an integration of 'application-independent' (Smith & McKeen, 2008) data processing modules. (Otto, 2012)

In order to develop an MDM-oriented data system, it is necessary to understand some different organizational structures of master data management, and choose a proper structural model for this design. According to the famous consulting company Ernst & Young (2012), the sourcing MDM modes can be categorized into 4 groups due to the trend of data centralization in a specific organization. All the 4 sourcing MDM modes are presented in Figure 3.1.

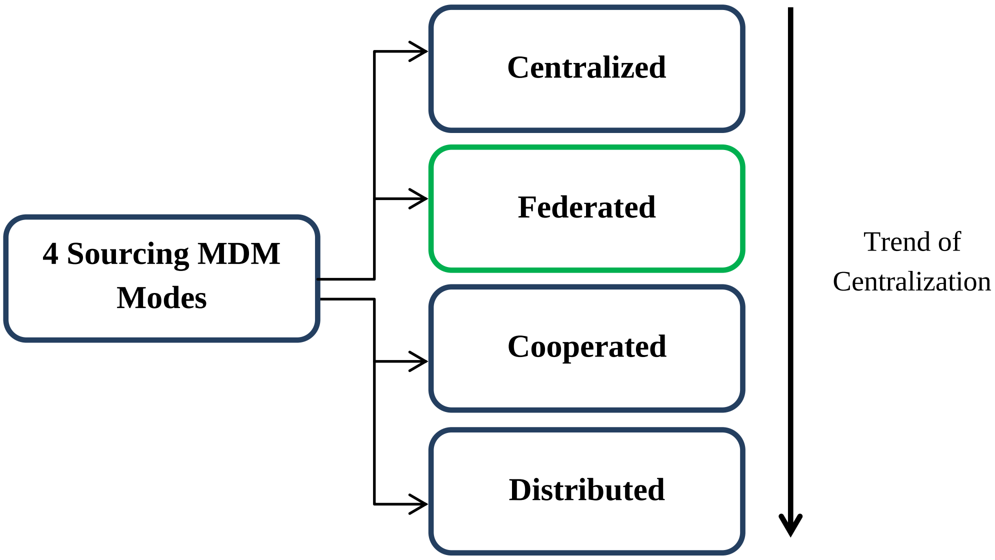


Figure 3.1. *Sourcing MDM Modes*
(Modified from: Ernst & Young, 2012, Recommendation of Sourcing MDM Modes)

A centralized mode means that most data of an organization are controlled by the top management, so lower levels have quite limited rights to operate the data the supplier selection criteria can be categorized into four main groups—cost, quality, logistics, and technology (Erdem & Göçen, 2012). In this mode, it is relatively easy to establish an organization-wide standardized MDM process, and enjoys high efficiency of master data management. On the other hand, the organizational scale is restricted to small ones, because it is an inevitable challenge for the top management of a large company to manage rather a large amount of data. Along with the downward arrow on the right side, more and more rights of master data management are shared among lower levels in an organization. A distributed mode, on the other extreme side (bottom), refers to that the rights of MDM are totally distributed to lower levels (such like individual departments), and the top management is not able to control the data as a centralized company does(Ernst & Young, 2012). In this mode, individual departments or branches can take more efficient management of their own master data, but at the organizational level, it is difficult to combine and standardize the master data from different independent sectors. In other words, this mode is considered better for an operative business rather than strategic development. The middle two modes—"Federated" and "Cooperated" are alike with each other in the organizational structure, but with different extent of autonomous master data management. As their respective positions shown in Figure 3.1., the federated mode emphasizes more centralization, while cooperated mode emphasizes more distribution (Ernst & Young, 2012). According to Accenture Research (2011), "Success in MDM depends on the right balance between central coordination and decentralization/ closeness to the business: a federated model has been proved successful at large and complex company." Regarding this case, although Sandvik Corporation is definitely a big company, its sourcing department of Sandvik Mining and Construction (China) is a relatively small but complex organization. Thus, the federated mode with more centralized requirements is recommended in this paper as an ideally theoretical mode. There are three advantages for this mode. (1) Combination of centralized management and sectors' cooperation, (2) relatively high efficiency of master data management, (3) easy to establish a company-wide standardized MDM process. On the other hand, two disadvantages should not be neglected. Firstly, more advanced management system is required for the coordination between different sectors. Secondly, top management cannot take 100% control of all the master data. In a word, the federated mode aims to manage all data with standardized regulations in an integrated system, but at the same time, it provides certain flexibility for lower-level data management. The detail is discussed in the next chapter.

3.2.2 Database Introduction

According to Daniel (2012), a database is defined as a list of information that a person or entity would want to maintain. Even though this concept is a relatively fresh one for human beings, databases have already existed in our history for a long time. Traditionally, it was common for a businessman to keep a notebook for his/her business records, and the notebook could be seen as an old-version database. Many critical data were kept in the database, and they could be manually retrieved if it was necessary. For example, when a boss has to pay his/her employees' salaries, he/she needs to check the data in the notebook to confirm who has already got the salary.

In a practical perspective, a database refers to a set of facts about an application domain (Halpin et al, 2003). As is shown in Appendix 1, the design of a database life cycle can be divided into four main phases—requirements analysis, logical design, physical design, and implementation. In the first phase, it is necessary to interview both data producers and users in order to determine the database requirements and produce a formal requirements specification. The specification concerns data processing, data relationships, and the software platform for the database implementation. In the second phase, a conceptual model of database is developed from a set of user requirements, and then refined into normalized structured query language (SQL) tables. In fact, many conceptual data models are obtained not from scratch, but the process of reverse engineering from an existing DBMS-specific schema (Silberschatz et al, 2006). The third phase involves the selection of indexes, partitioning, clustering, and selective materialization of data. It focuses on the approaches of storing and accessing those normalized SQL tables on disk, enabling the database to operate with high efficiency. In the fourth phase, the database can be created through implementation of the formal schema using the data definition language (DDL) of a DBMS. Then the data manipulation language (DML) can be used to query and update the database, setting up indexes and constrains as well. More specifically, SQL contains both DDL and DML constructs; for example, the "create table" command represents DDL, and the "select" command represents DML. In addition, as the database begins operation and monitoring, some modifications may be necessary when requirements change or end-user expectations fluctuate (increase or decrease) with the dynamic performance. (Lightstone et al, 2007) As a structural designing proposal, this paper refers to the former 3 phases, especially focusing on the requirement analysis and basic logic design. Certain necessary physical design is included, while specific technical issues in the implementation part are excluded. Detail information is presented in the following chapters.

Additionally, the database proposed in this paper consists of not only the basic database designing framework mentioned above, but also the principle of master data management. In other words, it is a combination of the both aspects. On one hand, the MDM-oriented database is essentially a database, so it should be designed according to the life cycle of a database (Lightstone et al, 2007). On the other hand, "MDM-oriented" refers to a critical feature for the database, which requires the database to emphasize all vital characteristics of master data management (Loshin, 2009). The descriptive term focuses on the unique feature of the database, claiming a quite notable difference between it and other databases. At a practical level, an MDM-oriented database must be capable enough to operate (store, maintain, update, retrieve, delete, etc.) master data (Loshin, 2009), so that it is able to take an effective advantage of the relevant data to support developing a rational business decision. In order to approach an MDM-oriented database for better (sourcing) decisions in a company, above all, it is necessary to understand the connections between a sourcing-decision making process and its relative master data or information management.

3.3. Sourcing Decisions and Master Data

Academically speaking, a sourcing decision is rather a general concept, because it refers to all kinds of decisions related to a sourcing activity. In this paper, it mainly focuses on how to choose a proper supplier by necessary analysis of sourcing master data. According to Razmi et al (2009), "Supplier selection is a complex decision-making process in nature due to different parameters and various aspects which must be regarded." However, a research indicated that although managers considered quality as the most important factor for selecting suppliers, they actually selected the suppliers concerning their costs (Verma & Pullman, 1998). In a comprehensive study, Dickson (1966) presented 23 criteria to select suppliers as following (Razmi et al, 2009).

Table 3.1. *Dickson's supplier selection criteria*
(Source from: Dickson, 1966, An analysis if vendor selection systems and decisions)

No.	Supplier selection criteria
1.	Supplier's suggested net price (including discounts and transportation costs)
2.	Supplier's qualitative capabilities
3.	After sales services
4.	Supplier's delivery capabilities
5.	Supplier's geographical situation
6.	Supplier's financial status
7.	Supplier's capacity and production facilities
8.	Supplier's partnership antecedents
9.	Supplier's technical capacity (including R&D capabilities)
10.	Supplier's organization and management
11.	Future potential purchases from supplier
12.	Supplier's information system (with processing information)
13.	Supplier's operational control (including reporting, quality control, and inventory control system)
14.	Supplier's status in related industry (including credit and leadership)
15.	Supplier's individuals' antecedents
16.	Supplier's organizational behavior
17.	Supplier's eagerness to cooperate
18.	Supplier's policy of guarantee and legal claims
19.	Supplier's capability to meet the product requirements
20.	Effects of supplier's contract on other contracts
21.	Supplier's educational aids corresponding products
22.	Supplier's adaption with the purchaser's procedures and instruments
23.	Supplier's performance antecedents

Even though the 23 criteria presented above have almost referred to all aspects about how to select the most qualified suppliers, certain elements are too abstract to be judged properly. For example, "organizational behavior" is hardly to be evaluated with restricted standards, and to some extent, it is hard to distinguish this item with "operational control". As a result, a series of more concrete supplier evaluation criteria is introduced in Table 3.2. In a comprehensive perspective, the supplier selection criteria can be categorized into four main groups—cost, quality, logistics, and technology (Erdem & Göcen, 2012), more detail information and application of this theory is discussed in Chapter 4.

Table 3.2. *Supplier evaluation criteria*
(Modified from: Erdem & Göcen, 2012, Development of a decision support system for supplier evaluation and order allocation)

Type	**No.**	**Criteria**
Cost	1	**Unit purchase price**
	2	**Terms of payment**
	3	**Cost reduction projects**
Quality	4	**Perfect order fulfillment**
	5	**After sales service**
	6	**Application of quality standards**
	7	**Corrective & preventive action system**
	8	**Improvement efforts in Tech & Quality**
Logistics	9	**On time delivery**
	10	**Order lead time**
	11	**Delivery conditions & packaging standards**
	12	**Flexibility of transport**
	13	**Geographic distance**
Technology	14	**Allocated capacity**
	15	**Flexibility of capacity**
	16	**Flexibility of technology**
	17	**Involvement in new product development**

Comparing to the Dickson's supplier selection criteria (Dickson, 1966) in Table 3.1, the latter supplier selection criteria (Erdem & Göcen, 2012) in Table 3.2 reflects two main differences in term of supplier evaluation fields. One difference is that the two critical aspects—"logistics" and "technology" were taken into account besides "cost" and "quality". The change of supplier criteria came out dramatically in the past 46 years, mainly due to the rapid development of economic globalization in the last a few decades. Nowadays, international business cooperation has become the mainstream in major economies all over the world, and more and more companies have joined in the global market. In such a context, a large number of international corporations would like to search potential suppliers in the "emerging markets". On one hand, the total costs (labor, taxes, material, etc.) in these countries are much lower than the developed markets, providing those international companies a good opportunity to cut down their total costs and keep the core competence. On the other hand, these fast developing countries usually have large populations, which are considered as remarkable potential markets for their products. In other words, these emerging markets have been accelerating the economic globalization, because they can consistently provide international corporations considerable profits. For example, China, as the manufacturing center all over the world, provides large amount of low-cost skilled labors for international companies, at the same time; it has gradually become one of the most important consuming markets for those companies.

Another notable difference refers to the increased requirement on the flexibility of suppliers' capacity, technology, and service (Erdem & Göcen, 2012). Similar to the change mentioned above, economic globalization is the key driving force of this trend. As international business involves more and more relevant parties all over the world, the supply-demand relationship becomes more complicated and flexible than ever before. For instance, when Sandvik's customers in Australia demand more machines for their booming mining industry, the manufacturer will correspondingly increase its orders for the purchase of steel components in China. It means that the local suppliers have no longer just been responsible for the local market, but also related to the globally dynamic changes. To sum up, with the evolution of international business, Dickson's supplier selection criteria (Dickson, 1966) cannot totally fulfill the requirements in 21st century. So, the sector of business data analysis (data categorization in Chapter 4) in this paper is generally designed based on the second supplier evaluation criteria (Erdem & Göcen, 2012), and further information is discussed in the next chapter.

4. DESIGNING AN MDM-ORIENTED SUPPLIER DATABASE

This chapter focuses on the process of designing a supplier MDM-oriented database, which is used to support developing an effective sourcing decision. Master data management is identified as an important approach to improve the efficiency of data processing, incorporating the business applications, information management methods, and data infrastructure to support the capture, integration, and subsequent shared use of accurate, timely, consistent, and complete master data (Loshin, 2009). When a supplier of a certain component is being evaluated, it is notable that thousands of relevant data may directly or indirectly affect the assessment, but only several essential elements are taken into consideration. For example, the salary structure of the supplier's employees, to some extent, may affect its production efficiency as well as flexibility. But in the evaluation process, such an indirectly influential factor is not regarded as master data; instead, some other factors such like "lead time" and "financial condition" are considered as master data. In other words, master data are expected to show the characteristics of key business objects (Otto, 2012), rather than a specific reason of that. Before the introduction of the database designing framework, the present situation and challenges would be discussed in advance.

4.1 Analysis of Present Situation and Challenges

In the following paragraphs, most of the situation and challenges are discussed in a perspective of the MDM research on sourcing, concerning few other aspects in the business. As is mentioned in Chapter 2, all the research-related information in the sourcing department of Sandvik Mining and Construction (China) has been gathered by the author in three major approaches—self-participation, survey, and personal interviews. Comparing to those indirect materials (Gummesson, 1993), such as academic literatures, relevant reports, and surveys from a third party, all the practical information were directly gathered without any potential effect of intermediate links. Since there is not any unnecessary process between information providers and the information receiver, all the direct materials in the research are considered rather reliable.

4.1.1 Sandvik's Sourcing Process in China

In order to point out the existing issues in the sourcing business of Sandvik Mining and Construction (China), it is necessary to understand the whole sourcing process in the company. According to the interview with the sourcing engineer Ms. S (2013), the framework of the current sourcing flow generally consists of 11 steps (as is shown in Figure4.1), and more details are discussed in the following paragraphs.

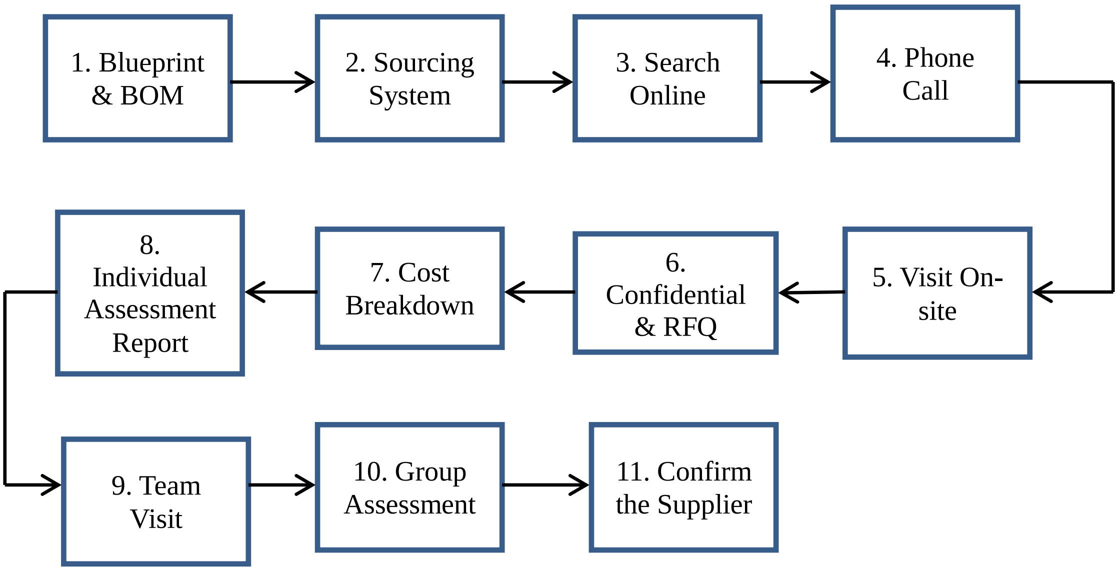


Figure 4.1. *Framework of present sourcing flow in Sandvik Mining and Construction (China)*

In Sandvik, a sourcing project (all the relevant components sourcing for a certain type of machine) is usually divided into five categories—steel, casting, electrical, hydraulic, and VMC (vehicle mechanical components). A sourcing engineer is in charge of the sourcing activities in a specific category, and at a higher level, a category manager is responsible for all the categories sourcing business. In addition, a project manager (responsible for a certain sourcing project), a product quality engineer, and the regional sourcing director (China) are necessarily involved for making a final decision.

(1) In the first step of the whole sourcing procedure, a sourcing engineer needs to obtain a component's blueprint/drawing and its previous BOMs (bill of materials) from a product engineer in a factory. Thus, the sourcing engineer can understand the general costs of a specific component on the sourcing list.

(2) The sourcing engineer searches some previous suppliers' information from the sourcing database (system P), where the content refers to suppliers' business scale, contact information, previous price data and so on.

(3) In addition to those previously cooperated suppliers, the sourcing engineer usually tries to search more potential suppliers on the internet, and adds some of them to the contact list.

(4) The sourcing engineer begins to contact all those suppliers by phone, confirming the information obtained online, and leaves 3-4 suppliers on the list as candidates.

(5) After all the information confirmed via a phone, it is necessary for him/her to visit suppliers' factories, checking their business condition and qualification in person.

(6) If those suppliers are qualified, the sourcing engineer would sign up confidential contracts with potential suppliers for further contact, raising technical requirements

and RFQ (request for quotations).

(7) In order to accomplish more detailed analysis on the sourcing case, the engineer asks for breakdown costs from the potential suppliers, including the cost of material, cost of manufacturing, estimated overheads, and profits and so on.

(8) With the consideration of all the collected information, the sourcing engineer develops an individual assessment report, but there is not a template or certain standards available for the assessment.

(9) Then, the sourcing engineer needs to visit those potential suppliers on-site again with Quality Engineer, Product Engineer and Project Engineer.

(10) After the group visit, a more comprehensive assessment should be developed and submitted. The final supplier is confirmed by the regional sourcing director of Sandvik Mining and Construction (China).

4.1.2 Challenges and Requirements

According to the definition introduced in Chapter 3, master data usually include the characteristics of key business objects in a company and is unambiguously defined as well as uniquely identified across the organization. (Otto, 2012; Dreibelbis et al, 2008; Loshin, 2009; Smith & McKeen, 2008). In fact, although many essential data have already been used in daily sourcing activities, they are not really well organized. As a result, 3 main problems have been diagnosed in the sourcing flow, and all the issue-related parts are highlighted in Figure 4.2.

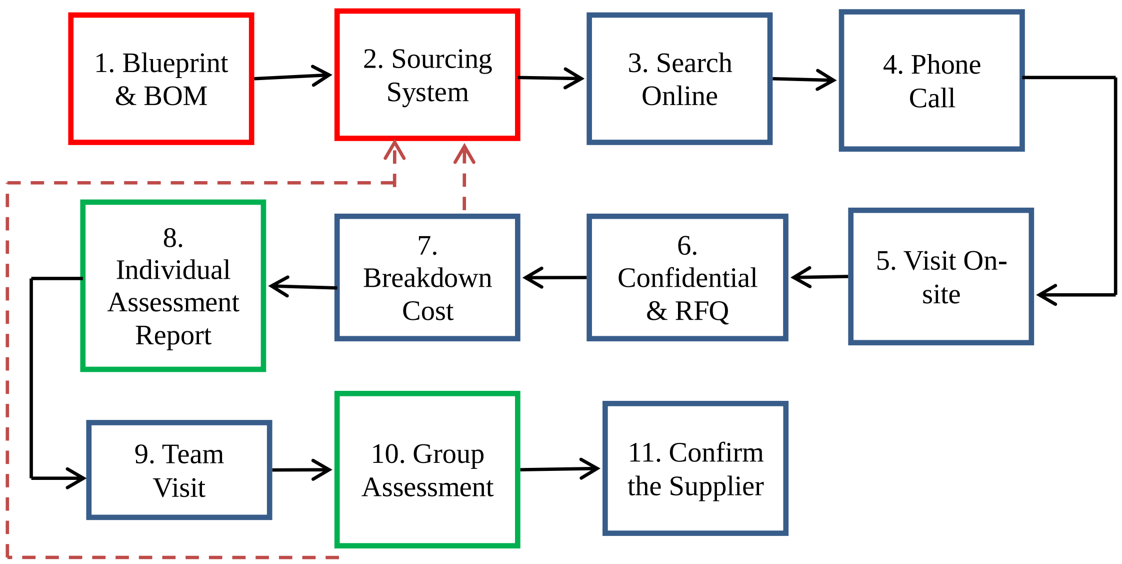


Figure 4.2. *Analysis of the sourcing flow in Sandvik Mining and Construction (China)*

Generally speaking, the three major issues can be concluded as new drawings update, lack of information in the sourcing database, and the management of suppliers assessment. Firstly, it is necessary for a sourcing engineer to collect the latest drawing of a specific component in the preparing stage. However, the fact is that some sourcing engineers do not have a good access to acquire the latest drawings, mainly because the PDC (product development center) in the “home-base” (Finland, UK, Australia, etc.) cannot share all the latest information with the sourcing and production departments in Shanghai, China. This issue is so serious, that some employees mentioned that in the pre-research interviews. As mentioned in Chapter 2, Mr. H is a sourcing engineer of steel components, who has 5 years sourcing experience and has worked in Sandvik for 2 years. In the face-to-face interview, he complained that some cost evaluation couldn’t be done properly due to certain unexpected changes from PDC in the “home-base”, but without necessary information sharing with the sourcing engineers in China. As a result, plenty of time has been wasted in the corporation’s internal information exchange. This problem has come out in recent years, since Shanghai gradually became the largest global assembly center for Sandvik, and a lot of assembly lines were consistently shifted from Europe and Australia to China. Traditionally, Sandvik’s factories in China were only responsible for the domestic market, so the latest drawings could be easily obtained from the local PDC in China. Nowadays, the assembly center in Shanghai is no longer just only in charge of the domestic business, but also needs to take more and more global projects. In a word, a PDC in another country is not able to share the latest drawings in time with those engineers in China, because an efficient global information shared database is not built up in Sandvik.

Secondly, some existing data systems adopted in Sandvik Mining and Construction (China), lack enough efficiency, data integrity, standardization, as well as coordination. There are three different kinds of “sourcing databases” being used in the sourcing department of Sandvik Mining and Construction (China), some data from these sources are overlapped, while certain data are absent in all the three. For example, Mr. M, a category manager with 9 years sourcing experience, discussed his dissatisfaction about the existing data systems in a face-to-face interview. He mentioned that all the three databases contained large amounts of overlapped information, such like Item No., Item Name, Item Description, Supplier Information, and Price and so on. On the other hand, some important analytical information, such like breakdown costs, product deficiency rate, and in-time delivery and so on, were not available in any existing database. The main reason for the confusion of data management in the company refers to the system independence among different departments. More specifically, a department usually adopts its own data management system, such like system P for sourcing, while system A is widely shared in the whole corporation as a data warehouse (According to the agreement with Sandvik, the specific names of their existing data systems are not publicized in this paper.). As a result, the company-wide database does not process all the required data for sourcing analysis, while the independent systems often keep those repeated data without information coordination among different sectors.

Thirdly, in terms of supplier assessment management, two main issues are targeted in Figure 4.2. One is unstandardized supplier assessment process; the other is the disconnection between these assessments and the databases. The former problem results from the lack of formulated supplier evaluation standards, and a relatively subjective assessment method has gradually formed in the department. In other words, the assessment processes are highly affected by the sourcing engineers' individual judgment. According to Ms. S, a sourcing engineer of electrical components with 7 years' experience, there was not any standard set for supplier assessment, so engineers usually just made evaluations based on his/her own skills and experience. However, without certain restricted assessment standards, personal attitudes and even occasional moods of sourcing engineers may affect the result of a rational evaluation. For this issue, a KPI (key performance indicators) template is being developed in the sourcing department, and a standardized supplier assessment process will be carried out in the future. The other important problem concerns the lack of linkages between supplier assessments and the suppliers' data storing systems. In a supplier assessment, especially a group assessment, a lot of information is collected and analyzed in different perspectives, but some critical information has not been effectively stored in the databases as precious wealth in a long term. Ms. S mentioned that a lot of information used for supplier assessment could not be retained in the data systems, so sourcing engineers had to search certain information again and again. For instance, quality related data were not contained in the existing data systems, but a quality engineer was necessarily involved in a group assessment, and certain amount of quality data from him/her would be taken into account for the specific assessment. But it is quite disappointing that most of these data have gone away as the case was complete. As a result, when the higher management aims to improve the sourcing strategy, there is little supportive quality information of existing suppliers available in databases.

To sum up, all the three major issues mentioned above definitely affect the work efficiency as well as further development in the sourcing department of Sandvik Mining and Construction (China), and the situation can be improved through the effort of establishing a MDM-oriented database which can replace the existing data management system. However, it is necessary to realize that the project of data management system modification is only being developed in the sourcing department of the company in Shanghai, with few other Sandvik's departments or subsidiaries involved in. In other words, not all the targeted issues can be solved by designing a comprehensive data system, while other necessary auxiliary efforts are also required for proper organizational management and inter-department coordination. The first problem mentioned above is regarded as an organizational issue rather than a technical issue, so some regulations about on-time data update should be carried out. The second problem can be tackled by diversifying master data storage, and Loshin's theory (2009) about the characteristics of master data objects should be adopted for identifying qualified potential data objects. The third problem can be solved by establishing an information feedback routine, which is shown in Figure 4.2. In such an information loop, master data can be used for both historical and predictive analysis in business intelligence (White, 2007).

4.2 Framework of the MDM-oriented Database

Databases should provide a means of storing, categorizing, searching, and analyzing a large number of similar data, and are central to many business processes (Clark, 2009). The MDM-oriented database proposed in this paper aims to support reasonable sourcing decisions by taking full advantage of all valuable master data. As a result, the current information flow in the sourcing department would be modified.

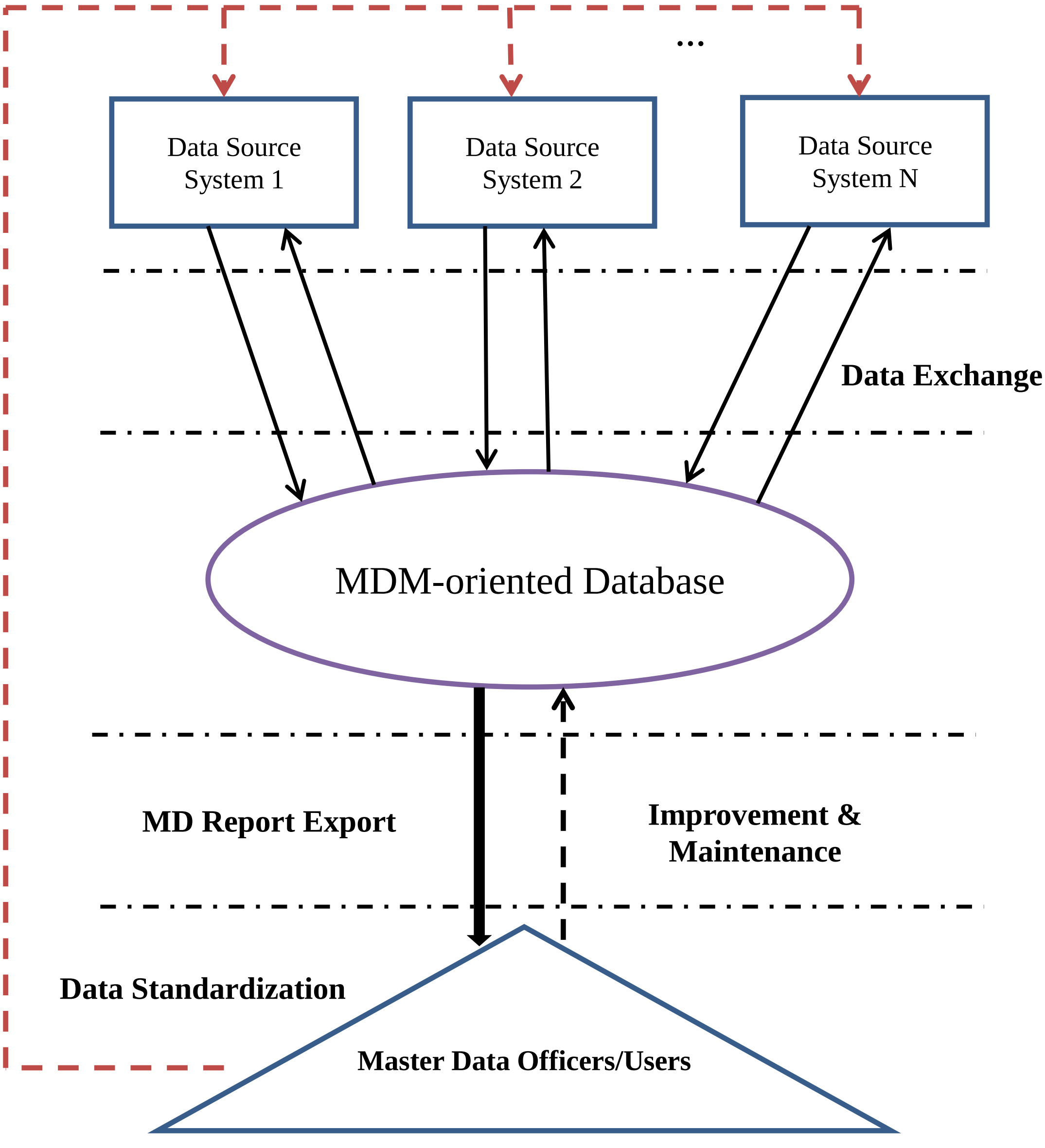


Figure 4.3. *A new structure of data flow within the sourcing department*

As is shown in figure 4.3, the MDM-oriented database is a vital connection between different data sources and the data operators. Actually, the new database tends to become the center of data flow in the whole sourcing process, because the structure of the data system is developed based on the federated mode ((Ernst & Young, 2012). In other words, only if all the master data (from different sources outside) are centralized to the MDM-oriented database, the total efficiency of data processing can be improved by standardized management. Generally speaking, there are three main mutual effects in the structure—data exchange, data export/maintenance, and data standardization. Firstly, the MDM-oriented database absorbs various data from different kinds of existing data sources in order to develop comprehensive master data. With the implementation of the database, more and more defects in the data storage sources may be exposed, and then improved. For example, some overlapped items, such like “due date” and “issue date”, should be stored in a specific database or data warehouse, in order to avoid unnecessary system resource waste. Secondly, after data processing, certain MD information is exported from the database in the form of report or tables, which assist MD officers/users to make more efficient analyses. Conversely, the authorized MD officers are responsible for the database improvement as well as maintenance, so that the MDM-oriented database can be modified according to the requirements. Thirdly, besides the duty mentioned above, the MD officers are in charge of data standardization of those data sources as well. In other words, the data end users are also the data templates designers in the next round, meaning that the whole data flow is closed in a self-modified loop.

In term of the case studied in this paper, the most essential application of the MDM-oriented database is to improve the supplier-related data analysis. How to find out a competent supplier for a certain purchasing item is considered as the key objective in the sourcing business, and some other data regarding a specific case in detail are processed in order to serve the objective as well. As is mentioned in Chapter 3, the supplier evaluation system is built up on the foundation of four main supplier evaluation subjects—cost, quality, logistics, and technology, as well as 17 specific criteria (Erdem & Göcen, 2012). As is shown in Figure 4.4, the data processing workflow within the MDM-oriented database can be divided into six main parts, including data collection, data categorization, data modeling, data consolidation, master data export, and data supervising (Loshin, 2009). More detail information about the design of each part is discussed in the following paragraphs. In addition, it is necessary to claim that the framework of database proposed in this paper is developed for the specific case in the sourcing department of Sandvik Mining and Construction (China), so it may not be suitable for some other cases.

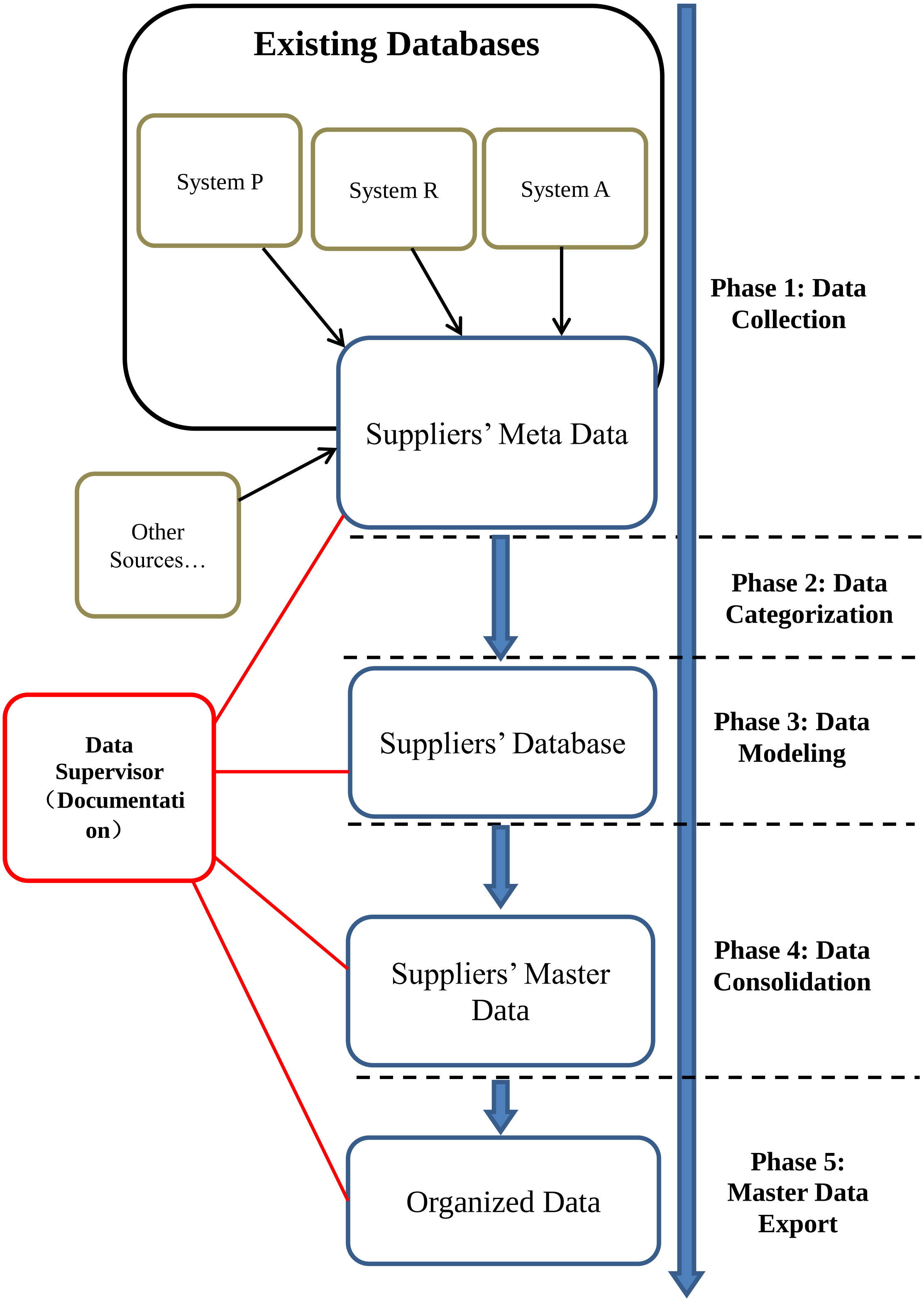


Figure 4.4. *Framework of MD-oriented database for the sourcing department of Sandvik Mining and Construction (China)*

4.2.1. Data Collection

Data collection plays a critical role in the whole data processing workflow, because it provides the necessary premise of all the following practical operations. In other words, it is the foundation of the entire framework, which aims to gather data for further analyses. Especially, in the federated mode of master data management, data collection determines whether all necessary data can be centralized within a comprehensive data processing system (Erdem & Göcen, 2012). Its basic goal is to collect data as much as possible, and to develop some practical data deposition methods that are easy to access and use (Chen et al, 2001). Regarding some relative features of those data, they can be generally divided into two categories—hard data and soft data. Hard data refer to those undisputed facts that are presented in rational, and they are regarded as the primary sector in a data research. Soft data, on the other hand, refer to the behaviorally oriented ones, which are considered less credible. (Phillips, 1999) For example, in a research of a supplier's service evaluation, its "late delivery rate" can be taken as hard data, which originally derive from the statistics of an objective fact. However, in the same case, the evaluating score on the "staff attitude" in a supplier company is usually defined as a kind of soft data, because it highly depends on the individual judgment, thus people in different perspectives may give quite different assessment on it. To sum up, it means that hard data would never be influenced by any subjective decision, while soft data tend to be influenced. Both types of data are often needed in a real research, and hard data are generally considered more essential than the counterpart.

In this case, not only the data types but also data sources and data gathering methods should be taken into account. Otherwise, the comprehensive data collection can hardly be accomplished properly. Regarding the data sources, it can be divided into two major groups—available data sources and potential data sources. As is shown in Figure 4.4, those existing data in the company's databases are currently available, and the data from other sources out of the databases can be defined as currently unavailable. Thus, it is clear that there are two main tasks in this section—to transfer those data from available sources and to gather the data from unavailable sources. According to the federated mode mentioned in Chapter 3, the organizational structure of data management should be relatively centralized (Ernst & Young, 2012). In order to accomplish data centralization, all the data should be transferred into a new data warehouse from those existing databases. In this process, two critical requirements need to be fulfilled. Firstly, all the overlapped data, which from different sources share a same attribute, should be filtered out, retaining only one in the specific aspect. As a result, a lot of memory space can be saved in the integrated system, and the retained data can necessarily correspond with effectual information. Secondly, all the retained data should be modified in a standardized format, if they are in different formats in their original sources. Thus, a framework of data warehouse is developed, which is much more convenient to process those data for further analysis, storage, extraction, and categorization and so on. The primary concept of data warehousing is to effectively access the data stored for business analysis by separating them from those data in the operational systems (Greeff et al, 2004). In this stage, all the existing data are gathered as metadata (Loshin, 2009) from different sources into an integrated data warehouse, and they are formalized in a standard format.

Besides the integration of the existing data, it is also important to gather those unavailable data in the new data warehouse. Since the existing data cannot refer to all the related aspects in the business, new added data are able to provide necessary complementation for the whole system. However, not all kinds of new data are qualified. More specifically, those data have to be either related to the 17 essential supplier evaluation criteria (Erdem & Göcen, 2012) or demanded by the data users. For instance, the latest drawings of certain sourcing components would not be considered useful for supplier evaluation, but experienced sourcing engineer Mr. H claimed that they were necessary for sourcing analysis. As is shown in Figure 4.5, the roadmap of new data collection should be developed in 3 phases—to analyze the business data needs, to target necessary data, and to integrate the data into the new data base.

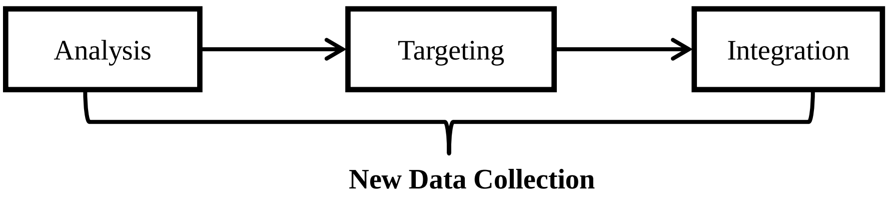


Figure 4.5. *Roadmap of new data collection*

The first phase aims to detect the deficiencies in the current data warehouse. It is essential to make analysis on the effectual business information and the corresponding data, so that those necessary but unavailable data can be pointed out. According to the pre-research survey in the sourcing department, the 8-year sourcing experience project manager Mr. J and 9-year sourcing experience category manager Mr. M both complained the existing price of most sourcing items was too "general", and asked for the information regarding "cost-breakdown" of those sourcing items. Thus, all the required data should be taken into consideration, and added to the data warehouse as many as possible. In the second phase, it is necessary to target those really useful data, filtering out the unnecessary data. The basic principle of establishing this new data warehouse is to provide users the data with the characteristics of key business objects (Otto, 2012). However, it is impossible to contain all the business-related data, because the memory space is never limitless, and the operating efficiency should be guaranteed. In order to target certain qualified new data, potential data objects need to be checked according to the six characteristics of master data objects (Loshin, 2009), the detail information about the characteristics has been presented in Chapter 3. As mentioned before, a survey was arranged for many employees from different positions in the sourcing department of Sandvik Mining and Construction (China), asking them to list the Top 10 needed sourcing data in their daily work and the Top 5 needed data that have not been available in the existing databases. Even though the result of the survey, to some extent, may be affected by individuals' personal experience and emotion, a comprehensive analysis on the whole survey can still develop a rational result about what data should be targeted. The third phase focuses on how to integrate those targeted data into the new data warehouse. Generally, there are two main data input solutions—manual input and automatic input. Manual input is relatively low-efficient, but is more practical and requires less investment. Automatic input, on the other hand, is high-efficient but with less flexibility and more financial investment. For instance, the information of the final approved price list for certain items is saved in the existing databases as a scanned PDF file. Thus, the related data can hardly be used for data processing, unless someone can manually input the data into the databases for further usage. Actually, some IT tools are able to accomplish such data conversion efficiently, such like a Barcode or Radio-frequency identification (RFID), but they require too much monetary investment for the sourcing department. So, employees should take more responsibilities for their items' information, checking and updating the specific data in time. For improving the efficiency in this section, some standardized templates need to be designed for suppliers to fill certain products' information, and the specific sourcing engineer is in charge of checking the data and inputting them to the data warehouse.

4.2.2. Data Categorization

Data categorization aims to improve the efficiency of mass data management by classifying them into different groups based on certain attributes within a semantic hierarchy (Loshin, 2009). Some data objects may refer to more than one attributes in different circumstances, but only the very key attribute should be abstracted after analysis in the context. For example, the name of a supplier company may refer to a series of information, such like the reputation of the company, the business sector it works in, and its liabilities, but for data categorization, it is just treated as a title of the supplier in the database. In other words, different attributes of an item would be broken down for analysis, and then distributed to certain groups.

According to Erdem and Göçen (2012) in Chapter 3, supplier evaluation data can be divided into four main categories – cost, quality, logistics, technology, as well as 17 sub-groups that belong to the four main categories. Although some important business features are categorized in the model, it does not include all master data. In fact, some data cannot be directly obtained from an order or bill. Instead, they can be deduced from existing data. For instance, the item "On Time Delivery" (rate) in the group of "Logistics" cannot be directly extracted from a transaction, because it is not rational to judge a supplier's on-time delivery only depending on one case. On the other hand, the information can be recorded in the database after the transaction, and the on-time delivery rate can be calculated with historical statistics. As a result, many hard data (Phillips, 1999) with limited information about a single transaction should be retained in the system as a record of the case, as well as a source of organized historical statistics.

According to the roles in supplier evaluation system (Erdem & Göcen, 2012), all the sourcing data are divided into two segments -- basic business data (BBD) and supplier evaluation data (SED). Then, most important supplier evaluation data need to be developed from the basic business data. In other words, these two data groups can be seen as two levels of a combined master data management system.

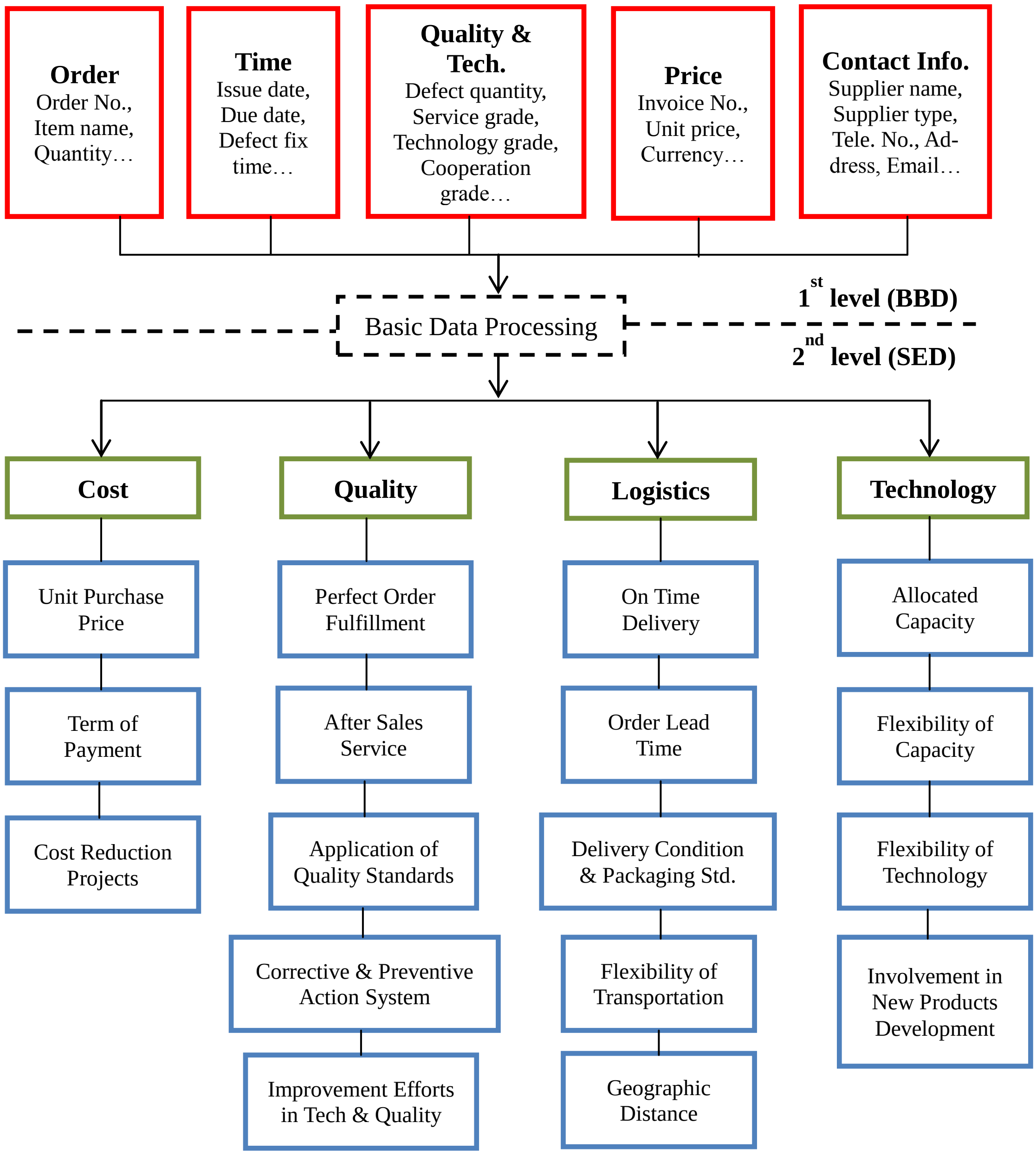


Figure 4.6. *Supplier Data Categorization Model*
(Modified from: Erdem & Göcen, 2012, Development of a decision support system for supplier evaluation and order allocation)

It is necessary to note that there is a new basic data set ("Quality & Technology" in Figure 4.6) added to the first level, because the existing four data sets do not include the essential supplier information about quality and technology. Optimum cost is not the only key requirements of a supplier partnership, which should also refer to joint development, culture, forward engineering, quality and communication (Briggs, 1994; Choy et al, 2002). Within the first level, all the basic business data are divided into five categories – order, time, quality & technology, price, contact information. It means that a single transaction should be broken down into different pieces of data sheets, and each data sheet will be sent to a specific category for further analysis and storage. According to the characteristics of the federated mode mentioned in Chapter 3 (Ernst & Young, 2012), some flexibility of lower-level data management is allowed in this mode. In order to build up an information loop of sourcing workflow, some experienced and professional data assessments made by individual engineers are recommended in this system. For example, "service grade" refers to a grade given by a sourcing engineer to evaluate the general service quality of a supplier. However, the grade would not totally depend on the engineer's subjective judgment. Instead, it must be applied in a supplier rating system, where the supplier's performance would be rated according to the predefined criteria (Choy et al, 2003). Thus, if a strict and rational evaluation system is established, it may reduce the risk of subjective grading behavior. After data processing, those basic business data would be converted into supplier evaluation data at the second level. Then, the developed data are distributed to four categories – cost, quality, logistics, and technology. Both the first level and the second level together have comprised the whole data categorization system.

4.2.3. Data Modelling

According to Loshin (2009), an MDM-oriented system is designed to help users improve their understanding the state of the business from different viewpoints and communicating between service units. A logical data model (LDM) is an important part of the system, because it is considered to be a business abstraction of data specifications. An LDM does not define the physical structure where the data should be stored. Instead, it defines the form, relationships and the identifiers of data. (Loshin, 2009) Data modeling is the process of standardizing and integrating different kinds of data into a universal model of formation.

Regarding the sourcing data discussed in this paper, they should be stored in the same format in the new database. Every single data sheet must be developed for retaining certain information of a corresponding sourcing case. In other words, those data would be updated until the accomplishment of each procurement deal. Generally, there are two reasons for adopting a complete sourcing transaction as a basic unit of the data modeling. Firstly, a complete procurement transaction or deal is considered as a fundamental unit of sourcing activities, and it contains various data that can be analyzed to affect sourcing decisions in the future. Case-based reasoning (CBR) is a well-accepted approach in knowledge management systems, because it is characterized by its capability to keep past experience and match that in various applications (Choy et al, 2005). It includes almost all the factors of necessary sourcing consideration, e.g. cost, transportation, quality, service, time, etc. Secondly, a complete transaction can offer the most reliable information about the sourcing outcome. Experienced sourcing engineer Ms. S (2013) mentioned that a complete procurement bill was more valuable for data storage, since it could provide the final information of the transaction. In other words, the data from a quotation is just an advice of the final deal, but what from a complete transaction is a fixed result that is ultimately accepted by both buying and selling parties.

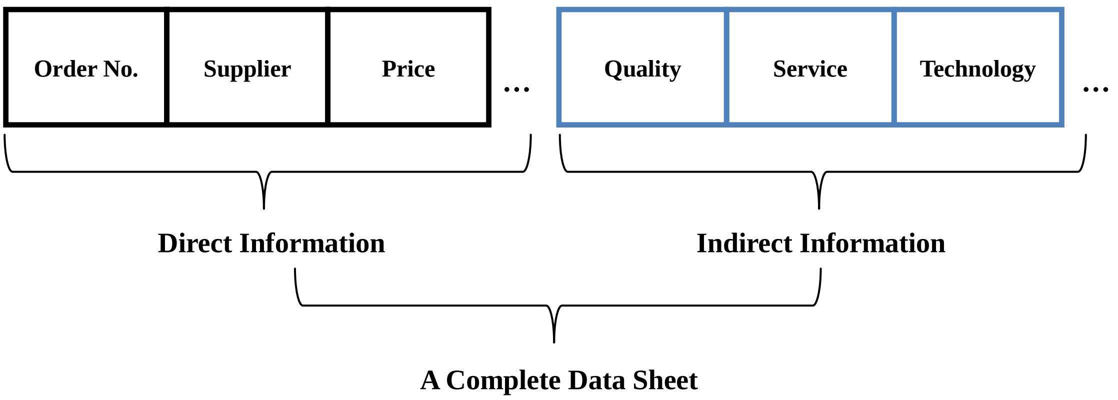


Figure 4.7. *A Model of a Complete Data Sheet*

According to the supplier evaluation theory (Erdem & Göçen, 2012), a single complete data sheet contains two main parts referring to direct information and indirect information. As is mentioned in the basic business data, direct information means all that can be directly extracted from a procurement bill, such like order number, supplier name and specific price of a certain item. On the other hand, indirect information refers to that cannot be directly obtained from a procurement bill, such like the evaluation on a supplier's service, technology and the quality of its products. It is necessary to realize that indirect information is relatively hard to obtain, but it is an essential part of basic business data which would be converted into supplier evaluation data for sourcing analyses. However, if indirect information includes some judgment from the buyer's perspective, it can hardly be absolutely objective. Thus, is it reliable for further sourcing analyses? The answer is yes. Prior researches indicate that the success of an adequate supplier management is measured by supplier performance and buyer satisfaction (Large, 2006; Akamp & Müller, 2012). So, when certain judgments are made by professionals according to the buyer's need, the indirect information is considered quite valuable.

Nowadays, the exchange between buyer and seller involves continual interaction between organizations in a relationship, rather than a single transactional exchange of specified content and duration governed by the market force (Möller, 1994; Makkonen & Olkkonen, 2013). Thus, retaining sufficient sourcing data of suppliers is important for a company, since it may affect the long-term cooperation of them. In terms of the objects of data modeling, it focuses on the basic business data. As is mentioned before, basic business data refer to almost all the business related factors, which can be processed and converted into supplier evaluation data. These business metadata that contain various types of business information can be seen as the foundation of the whole data system. To some extent, they may affect the relationships among organizations, and even change the structure of a certain business sector. In a word, how to take good advantage of the systematically modeled data is a key issue for business development.

4.2.4. Data Consolidation

According to Loshin (2009), data consolidation refers to the process of bringing data instances from different sources and consolidating them into unique records in the master data model via the parsing, standardization, harmonization, and matching capabilities of the data quality technologies. In other words, all the mass data collected from various kinds of sources need to be well organized in the new system, generating master data that can be effectively used in the future. Since data standardization issues have been discussed previously, this part focuses on how to develop certain qualified master data from the metadata for universal usage in the future.

As is mentioned before, the whole master data set comprises of two parts – basic business data and supplier evaluation data, and the latter one can be generated from the former one via data processing. To some extent, the supplier evaluation data are considered as key factors that are able to offer supports for certain sourcing decisions, while the basic business data are the necessary evidence and foundation of the analyses on specific factors. In the literature, 17 to 23 different criteria discussed above are recommended in supplier evaluation (Dickson, 1966; Erdem & Göcen, 2012), however, the most prevalent issues that appear in most studies are cost, quality, technology infrastructure, delivery performance and business issues (Erdem & Göcen, 2012; Wang et al, 2004; Cebi & Bayraktar, 2003). As all the essential issues have been included in the supplier evaluation data (SED) model mentioned in previous parts, the concept of consolidating basic data to develop the SED model is rather reliable in real-life business.

In terms of SED development based on a single transaction, all the 17 items mentioned before (Erdem & Göcen, 2012) can be divided into three main groups for analysis. Group 1 consists of the numeric direct data that can be transferred from a specific bill, e.g. “Unit Purchasing Price”, “Terms of Payment”, “Order Lead Time’, etc. Group 2 consists of the judgment data that can be judged “Yes” or “No” according to the fact of the transaction, e.g. “On Time Delivery”, “Perfect Order Fulfillment”, “Cost Reduction Projects”, etc. Group 3 consists of the scoring data that can be given a grade ranging from 1 to 5 to define the degree of satisfaction of the company on its supplier’s activities, e.g. “After Sales Service”, “Flexibility of Transportation”, “Involvement in New Product Development”, etc. It is notable that the data of group 1 are available in the current database, but the data of group 2 and group 3 are not available and necessary to be set up in the new database. It is clear that the three types of supplier evaluation data are categorized according to their flexibility, and the flexibility of each group is gradually increasing from Group 1 to Group 3. In other words, the direct data and judgment data are considered as “hard data”, while the scoring data are considered as “soft data” (Phillips, 1999).

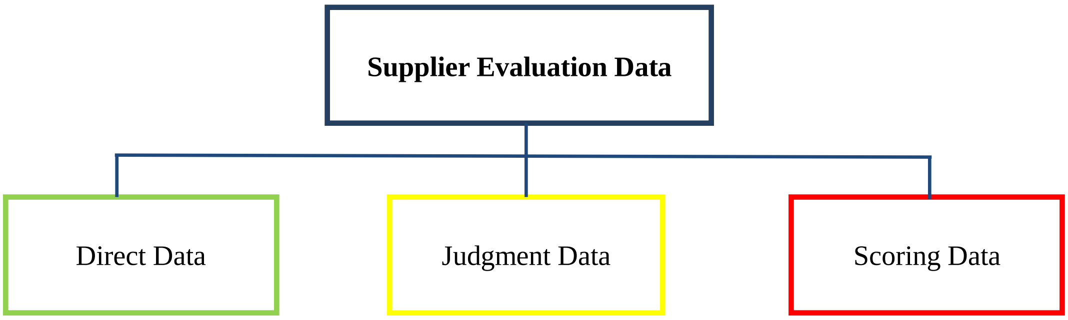


Figure 4.8. *Three Types of Supplier Evaluation Data*

The three types of SED are shown in Figure 4.8, and the different colors of frames refer to the obstacles of operation in the current data system. Similar to the principle of “traffic lights”, direct data in green frame can be used without modification. Judgment data, on the other hand, need to be added to the current database with a two-optional prerequisite. Moreover, scoring data would also be added to the current data system, but with more complicated scoring criteria. According to the federated MDM mode mentioned in Chapter 3, the master data management in this mode should be more flexible than the centralized mode (Ernst & Young, 2012). The classification of three different types of data can ensure the master data management system to be more flexible in practice.

The two-optional prerequisite of judgment data aims to define a specific business item’s condition with a “Yes” or “No” option.

$$J_i = \begin{cases} 1, \text{ if case i can fulfill the requirement.} \\ 0, \text{ if case i cannot fulfill the requirement.} \end{cases}$$

Actually, it is not hard to understand how to take this judgment standard into practice in real-life business. For example, when a sourcing engineer is inputting certain information into the data system after a transaction, he or she may encounter the issue of “On Time Delivery”. In such a situation, the engineer should choose “Yes” or “No” from the options according to the reality of the case. As a result, this kind of information can be quantified in the new database.

Regarding the scoring criteria for the third-type data, it is established based on the academic concept of Analytic Hierarchy Process (AHP) that aims to support multi-criteria decisions (Saaty, 1980; Zhang, 2012). Generally, there are four grading levels set for the criteria, ranging from 0 to 3.

$$S_i = \begin{cases} 3, \text{ better than the majority of suppliers.} \\ 2, \text{ at the middle level among suppliers.} \\ 1, \text{ exist but worse than the majority of suppliers.} \\ 0, \text{ not exist or totally unqualified.} \end{cases}$$

With the criteria above, many items can be quantified with relatively precise grade. As an important feature of a federated MDM mode, the lower-level data management (sourcing engineers' contribution) is successfully included in the system (Ernst & Young, 2012). However, it requires those sourcing engineers to be responsible and objective, so that he or she can consider the big picture of the whole market while grading a specific item after a transaction. For instance, when a sourcing engineer grades a supplier's capability of "After Sales Service" in a certain case, he or she should compare the supplier's performance in this aspect in the case with other suppliers' performance in a similar case. Otherwise, a single grade without comparison in the market is valueless for business analyses and decisions in the future.

Although most suppliers' evaluation factors can systematically match their numeric description in the database, the information of a single sourcing case is still not enough for a professional sourcing manager. Thus, it is necessary to combine a certain amount of data together and then to produce some statistical outcomes which can reveal more essential issues of the data in a comprehensive perspective. With the approaches discussed above, a "vertical comparison" in the data table could be achieved, meaning that the items of different cases (in a same column in the data table) could compare with each other. However, a "horizontal comparison" is also needed in a real-life practice, and adding certain coefficients or weights to the scored columns is considered as an effective approach to balance the whole numeric data system. To properly evaluate the weights corresponding to different criteria, the AHP methodology needs to be applied via pair-wise arrangement (Erdem & Göcen, 2012; Saaty & Vargas, 2001). In order to support sourcing analyses and decisions, the principle of this methodology is required to possess two key features -- rationality and flexibility. It means that each coefficient would be set within a reasonable range, but its exact value would be decided by users. Thus, whilst the weights of those evaluation factors are constrained at a general level, they can still be modified according to the specific situation of a sourcing case.

Table 4.1. *Coefficients' range for 3 types of data*

Direct Data	Judgment Data	Scoring Data
X	0-3	0-1

According to the AHP theory (Saaty & Vargas, 2001), it is necessary to set certain hierarchy in the system, so some coefficients are proposed in Table 4. There is not any coefficient recommended for direct data, because they do not always contain the comparable numeric information. Although “price” is an obvious exception in this group, it would not be provided with a coefficient. And it should be compared independently, because the number in “price” column includes much more accurate information than a general score, and any coefficient will break the accuracy. For judgment data, the coefficient’s range is 0 to 3. It is because the numeric information available in this group only includes 0 and 1, which multiplies the coefficient can reflect the importance of the evaluation factor in a certain case. Similarly, for scoring data, the coefficient’s range is 0 to 1. It is because the numeric scores are all in between of 0 and 3, and a non-negative coefficient with the maximum value 1 can definitely measure the importance of each evaluation factor in a specific case. So, in the weighted scoring system, the range of any score lies between 0 and 3, with the issue of relative importance taken into account. In a mathematical perspective, a final score of an indirect data can be calculated as

$$S_f = \begin{cases} C_j * S_j, & 0 \leq C_j \leq 3,\ 0 \leq S_j \leq 1. \\ C_s * S_s, & 0 \leq C_s \leq 1,\ 0 \leq S_s \leq 1. \end{cases}$$

S_f is a final score, C_j is a coefficient of judgment data, S_j is an initial score of judgment data, C_s is a coefficient of scoring data, and S_s is an initial score of scoring data. If all the final scores are added up, a total score could be calculated as

$$S_t = \sum_1^n C * S, \quad n= 1, 2, 3, \ldots\ldots$$

S_t is the total score of n scored factors in a sourcing case, C is a coefficient of a specific scored factor, and S is an initial score of a scored factor. Thus, the outcome of the analysis on those numeric evaluation factors can be finally developed as a single number, which ensures a comprehensive comparison between different cases to be possible. In such a variable data management system, certain item’s value can be determined by experienced professional, and the coefficients adopted reflect the importance of the specific item in an evaluation. In a structural perspective, the data consolidation system can improve the flexibility of the federated MDM mode (Ernst & Young, 2012).

4.2.5. Data export

When the applications that access and modify their own underlying data sets (Loshin, 2009) are completed, it is time to export the outcome of data processing. Generally, there are two main data exporting forms proposed in this paper – standard data and visualized data. Standard data, on one hand, refer to the data which are widely stored in the database. They are originally collected from different kinds of data sources, and then are developed through a series of data processing, e.g. data standardization, data modeling, data consolidation, and data storage. In other words, standard data are relatively original data which could be retrieved directly from the database without further processing. They are usually used for checking up certain information of existing a supplier or a transaction, or are considered as a reference for a specific sourcing case analysis.

Visualized data, on the other hand, emphasize that some numeric data can be presented in the form of visualization, e.g. a column chart, a pie chart, and a radar chart. A notable advantage of the visualized data is that they can provide users with a direct impression and understanding of those mass numbers. And different charts are usually fit for different types of data, so there is not the best visualization approach, only the most suitable approach in a specific condition. As is mentioned in previous chapters, the goal of designing the MDM-oriented database is to support certain sourcing analyses and decisions. In other words, the new database needs to be useful for its users, and the data export sector is responsible to provide the users with convenience and efficiency. More details are discussed in Chapter 5.

4.2.6. Data Supervising

According to Loshin (2009), the stakeholder should be adherent to the rule that govern participation and information sharing. Therefore, data supervising, as a relatively independent segment besides the data processing sequence, is added to the whole data system. Although data supervising is not directly involved into the ordinary operation of data collection, categorization, modeling, consideration, and export, it keeps rather close connections with all other segments as a supervisor of the whole data system. In other words, data supervising segment is proposed for ensuring the accurate operation of the whole system, and diagnosing potential system issues in time.

Data supervising can be divided into two essential parts – the inside the system and the outside. The automatic part is responsible for monitoring every single operation of the database, and alerting the professional team if any abnormal issue is detected. On the other hand, the computers of the outside team are consistently connected with the database, and when an abnormal alert comes, the experts need to immediately diagnose the specific module where the issue comes from. In addition, three regulations are recommended for the data supervising team. Firstly, the team is responsible to systematically check the database in a certain period (i.e. a month), in case of certain issues having been neglected by the inside supervising module. Secondly, two or more professionals are required to fix any issue in the system. Concerning that the data system contains a number of important business information, it is necessary to avoid the situation that a single expert is able to change certain information without supervising. Thirdly, the IT team should be in the charge of the director of sourcing department, and any specific change of the database has to be authorized by him or her. To sum up, the data supervising segment is developed based on the federated MDM mode, aiming to improve the efficiency of the whole system by relatively centralized management (Ernst & Young, 2012).

5. EMPIRICAL RESEARCH

This chapter generally consists of two main parts – the process of designing the MDM-oriented database and the outcome of the design. The first part reviews the whole process of the study in an empirical perspective, while the second part briefly introduces its outcome and raises some advices for further improvement. Different from previous chapters, some practical research approaches are discussed in this chapter instead of theoretical concepts. In other words, this chapter focuses on the research experiences acquired from the specific case in Sandvik Mining and Construction Corporation (China) rather than some other general designing theories. Although there is not a fit-for-all approach proposed in reality, the empirical research on this case can still provide some inspiration for those related studies in the field. Regarding some basic pre-designing researches, such as resource analysis and information gathering, certain experiences may be shared universally among different academic sectors.

5.1 Process of Designing the MDM-oriented Database

Designing an MDM-oriented database in the sourcing department is a systematic project, which requires certain knowledge referring to various aspects, e.g. computer science, information management and business-to-business (B2B) administration. As a result, it usually takes a long time for preparation, analysis, development, testing, and improvement. As mentioned in Chapter 3, this paper mainly focuses on the design of a general structure of the MDM-oriented database rather than specific technical issues. So, some implemented phrases like "testing" and "improvement" are excluded in the thesis. Instead, some advices are proposed for the improvement in the future. Generally speaking, the process of the database designing comprises of the following 10 steps.

- Clearly understand the to-be-solved issue.
- Analyze what kind of resources can be utilized for tackling the issue.
- Take full advantage of the resources to gather information about the issue.
- Systematically research on the collected materials.
- Build up a general structure of the original approach.
- Break down "a big picture" into functional pieces.
- Study on each segment one by one.
- Combine all the segments together to develop a final approach.
- Modify some parts if necessary.
- Point out the existing disadvantages of the approach and future advices.

Before starting a research on a specific subject, it is important to have rather clear understanding about what is the issue to be solved in the research. Once the issue is not clearly understood or even misunderstood from the reality, an incorrect target would be set up, and the whole research may be led to a totally wrong direction. In the beginning, the author has already had some face-to-face discussions with Mr. G, who is the manager of Sandvik's construction sourcing department in Shanghai, before the research in order to confirm the correct understanding about the issue. According to Mr. G (2013), the sourcing department has been suffered from the chaos of three independent data management systems for a long time. Thus, the management of the department aims to develop a comprehensive supplier database to support sourcing analysis in the future. The new database would only be applied in the sourcing department in Shanghai subsidiary for a short while as testing, and may be introduced to other subsidiaries if its feedbacks are positive.

After confirming the current situation as well as the goal in the future, it is necessary to think about what kinds of resources could be utilized to achieve the final target. More specifically, the following three related questions should be answered by a researcher him/herself. “What do I need to hit the target?” “What do I have now?” “What else should I acquire?” Actually, this phase is rather important, because no research could be finally accomplished without certain supports and resources outside. As mentioned in Chapter 2, there are generally 5 main approaches to gather information for a research (Gummesson, 1993). In this stage, I developed a plan based on the 5 approaches to gather useful information. For example, as a student in Tampere University of Technology, I could get access to a number of existing literatures in the library as well as many eBooks online. In the company, I was able to interview many employees in the sourcing department during my internship period, in order to understand their requirements and expectations about a new sourcing database.

The third phase refers to the pre-designing preparation, in which all the resources were used to gather different kinds of information for further research. Generally, all the 5 methodologies proposed by Gummesson (1993) were adopted, while most research materials were from three main sources – paper literatures, eBooks and interviewing materials. As mentioned in the last paragraph, those existing academic literatures were acquired in the library or on the internet. On the other hand, the researcher arranged some surveys and face-to-face interviews within the department. The interviews covered about 30% of employees in the sourcing department in Sandvik Shanghai subsidiary, ranging from department managers to sourcing engineers. The topics focused on the disadvantages or problems about the current databases, and the expectations of a new database.

All the original data would be converted into information and even knowledge through study; otherwise, they were worthless for the research. Literatures could be read, while the interviewing and survey materials had to be processed for further analysis. In order to develop concrete research materials, most content of those interviews and surveys had been recorded during the conversations, and then clearly summarized in Appendix 2. For example, those interviews with Mr. M, Mr. H, and Ms. S that mentioned in Chapter 2 were all summarized for further study. Besides those direct materials, it was also vital to summarize certain important knowledge points from those academic literatures about database structure, data mining, master data management, and supplier evaluation, etc.

Then, it is time to figure out a general structure of the to-be-designed database, especially the services it may provide. This phase is rather important for the whole designing process, because a framework of the data system has been proposed. According to the requirements from Sandvik, the new database should be designed based on the principle of master data management. Thus, before building up the framework, it was necessary to make clear the characteristics of master data. David Loshin (2009) proposed that master data core services comprise of 5 main categories as well as 19 sub-categories, and its framework is presented in Appendix 3. Generally speaking, the original idea of the design in this paper was inspired by Loshin's theory (2009) about master data management. However, the theory of Loshin focused on a new and independent master data system, without consideration of its potential connections with those existing data systems. So, "data transferring", as a new service, is added to the framework. It ensures that some data can be directly transferred into the new system from other existing data systems, improving the efficiency of data creation.

In order to construct the new database, a whole project has to be broken down into relatively small segments, then to be completed one by one. In this phase, the designing process of the database was divided into 5 main parts according to its operating workflow (mentioned in Chapter 4). The general structure and functionalities of the MDM-oriented database were developed as a combination of many theories and inspired from the existing data systems in the sourcing department, but the basic principle of the design comes from the federated MDM mode (Erdem & Göcen, 2012). In terms of the designing process, the first step was data collection, which aimed to enrich the database with metadata from different kinds of sources. Then, all the disorganized metadata would be categorized in the second step, where several data groups were established and the corresponding data were distributed to each group according to their own attributes. The third step referred to data modeling, meaning that a standardized data storage model or formation was proposed, and all the data would be stored in the forms. The fourth step was data consolidation, which was also the key part of data processing. In this step, certain numeric attributes were endued to those effective but unquantifiable data (i.e. quality, service, etc.), and then these data could be reasonably evaluated by relatively simple calculation. Moreover, with the numeric attributes, a number of collective data analyses could be realized through some more complicated mathematical methods. The last step was about data export. Generally, there were two main approaches for data export – standard data and visualized data. The standard data were directly retrieved from the data storing sites, and presented in front of a user without processing. Visualized data, on the other hand, referred to the data experienced the process of visualization. In other words, these data were not in the form of number or word, but in the form of figure or chart. Visualization would never change the essence of those data. Instead, it makes the information much more clearly for people to understand.

Entering into the next phase, more practical approaches were proposed according to the specific issues in different steps. Some specific problems needed to be solved, for example, what mathematical formulas were suitable for data consolidation. While all these phases of design were accomplished, different segments should be combined together to form a whole system. At this time, as was shown in Figure 4.4, the position of data supervisor needed to be added to the system. As mentioned before, a data supervisor could be considered as a monitor for the data workflow inside the database, and it was in charge of the centralized data management in the federated MDM mode (Erdem & Göcen, 2012). If an error occurred in anywhere inside the system, the data supervisor would be alarmed and could be authorized to repair the specific error. Regarding the security of the database, a data supervising team should be established. On one hand, the workload for each person could be reduced while a long time operation of the system. On the other hand, personal mistakes or even deliberate damages may also be avoided as much as possible.

5.2. Outcome of the Design

Evolved from the data flow structure in Figure 4.3, the new designed MDM-oriented database comprises of three layers in Figure 5.1. Generally, the specific function of all three layers correspondingly refers to standard data storage in the first layer, selected data processing in the second layer, and user interface in the third layer. The dash line arrows in the figure represent the commands from users in the third layer, while the solid line arrows mean that certain data are retrieved from the data storage base in the first layer. It is notable that there are two different data retrieval routines in the figure – direct data retrieval and indirect data retrieval. Direct data retrieval, on one hand, is applied when a user wants to directly fetch certain standard data without any further operation. For example, the interviewed sourcing engineer Mr. H (2013) claimed that “I hope I can always get the latest drawings of any component from the database, and it is very important for my daily work.” On the other hand, indirect data retrieval usually applied when a user aims to get some vital information from those processed data. For instance, the interviewed manager Mr. G said, “As a manager, I would like to see ‘a bigger picture’ from those normal data. In other words, I hope the new designed database can properly deal with those massive sourcing data, and provide me with certain essential information, so that I can make a rational decision on the sourcing case or develop a comprehensive strategy.”

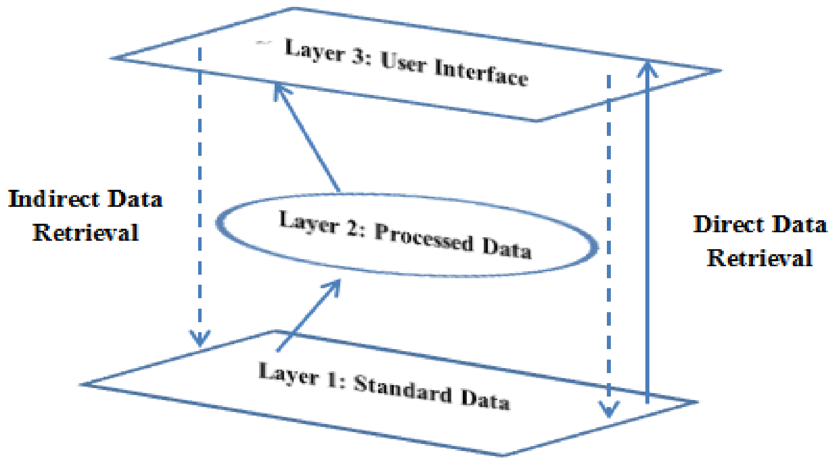


Figure 5.1. *Two data retrieval approaches in the database*

5.2.1. Layer 1: Standard Data

Regarding the structure of the whole system, the first layer can be considered as the foundation of that. In this layer, all the standard data are stored for further utilization, so all other operations in the other two layers cannot be accomplished without the support of this layer. In other words, layer 1 is in charge of providing to-be-processed materials for other layers, while no further work can be done without those materials. To some extent, the diversity as well as quality of data storage in this layer can determine the quality of the whole system, because good and sufficient materials are very important premise of expected outcomes. Actually, most of the standard data in this layer are developed from those data (in Figure 5.2) being used in the three existing data systems.

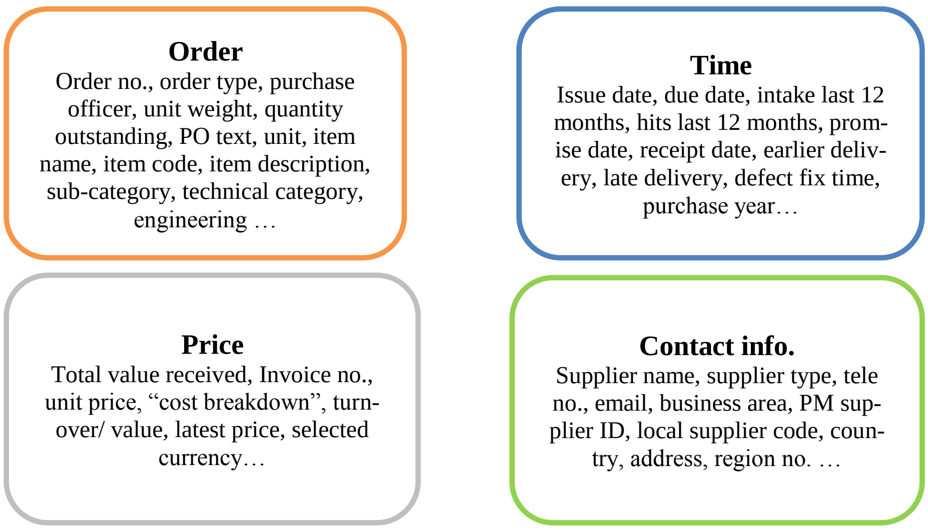


Figure 5.2. *Basic Business Data Categorization in the existing systems*

As discussed above, various sourcing related factors are collected from the existing data systems and divided into four main groups according to their own attributes. According to the researcher's observation, two essential disadvantages – data overlapping and insufficient data diversity were discovered on the data storage layer of the existing data systems. Then, two approaches were proposed for fixing those issues. For the first one, those overlapped items were deleted, so only one item remained corresponding to a specific evaluation factor in the system. For the second issue, more sourcing related evaluation factors were gathered through literatures and interviews, and then added to the existing data storage system. Thus, an even more sufficient factors database has been established, and the detail information is available in Appendix 4. There are totally 68 factors recommended for the new system, with 22 highlighted new factors in addition. It is necessary to note that most of the items in the existing systems are direct data which have been mentioned in Chapter 4. These data usually just describe some detail facts in a specific sourcing transaction rather than being involved into mathematical evaluation. In other words, these kinds of data can only be directly retrieved, while cannot be further processed. As a result, most of those new added items are able to rebalance the functionality of the whole system, and ensure a transaction to be more measurable in a mathematical perspective.

5.2.2. Layer 2: Processed Data

After data retrieval in the first layer, there are two options available for users, one is to deliver the data directly to the third layer, while the other is to process the data in the second layer and then deliver the outcome to the third layer. This part focuses on the data processing operation in the second layer. Generally, the basic principle of data processing in this layer is to analyze data through some mathematical approaches as well as some data visualization technologies. In order to offer managers like Mr. G some comprehensive information from the normal data, it is necessary to do some calculations in regular data processing. Usually, sum and average are the most common calculations in this part according to the observation.

Table 5.1. *An example of data combination*

Order No.	Component Code	Purchasing Data	Supplier Name	Unit Price (Euro)	…
001	DX-700	25.03.2012	Company A	50	
002	DX-700	25.04.2012	Company B	52	
…					
009	DX-700	25.09.2012	Company A	49	…
010	DX-700	25.10.2012	Company C	49	

As is shown in Table 5.1, "Unit Price" can be combined through different cases of the same component, and develop a general analysis outcome based on the historical record.

Since most evaluation factors have already been converted and measured by numbers, the combined analysis could be made briefly via some basic mathematical calculations. Beside sum and average, various calculated results are also helpful for a manager's decision. For instance, the "maximum" and "minimum" of the unit price were considered useful for the category manager Mr. M to understand the historical price range of a specific component. Despite almost every outcome has its own value for further analysis, two main outcomes should never be neglected in this section. On one hand, it is important to show the curve of the historical development of a specific component. For example, a price development curve in the last one year may show the fluctuation of the purchasing price for a specific component in a short history. Thus, the manager can have a better or more direct understanding about the price changes in the past year, and to some extent, may have more capability to predict the new price in the future. On the other hand, it is necessary to calculate the average value of a certain value column (as is highlighted in table 5.1). For an item with a large number of values in the column, assuming the number is i, we can calculate its average value as following.

$$A_i = \frac{\sum_{i=1}^{n} Vi}{i}, \quad i = 1, 2, 3, \ldots\ldots n.$$

In the formula above, A_i is the average of i values corresponding to a single item column, i is the number of recorded cases for the same item, and Vi refers to the value of the item in each transaction. For instance, assuming that there only four transactions of DX-700 had occurred in 2012, as is shown in Table 4-1, its average unit price can be calculated as

$$A_i = \frac{\sum_{i=1}^{4} Vi}{4} = \frac{(50+52+49+49)}{4} = 50.$$

The average unit price in the year 2012 was 50 Euros, meaning that the first two transactions were made higher than the average, while the latter two were made lower than the average. Furthermore, a sourcing manager might see that there was a small price fluctuation in April 2012, but it came to a steady lower level in the third quarter. As a result, the near-coming trend of the price may be probably kept at the lower level. In addition, an experience sourcing manager like Mr. G would notice that the supplier with the highest price was company B, different from other suppliers. So, the price fluctuation might not be a result of market effect, but the issue about the expensive supplier.

5.2.3. Layer 3: Data Output

The third layer of the whole data system is data output, which is the most familiar part for common users. There are two main functions for a user interface – to input and to output. To input means that users are able to send a certain command to the database in the interface, while to output refers to showing the outcome of the user's command in the same interface. In terms of the input segment, the operation should be rather simple in order to reduce users' resistance, since "easy to operate" is one of the top five expectations (in Appendix 2) for the new database. At the same time, various functions should be included in order to satisfy different requirements from users. The output can be either brief or complex according to the demand of users. As is shown in Table 5.1, standard data are always presented in a selectable data table. The table includes a certain number of modulating options, which ensures the users to modify the table according to their needs.

Table 5.2. *An example of a selectable data table*

Component Code	Supplier Names	Supplier's Location	Purchasing Date	Unit Price (Euro)
DX-700	Company A	Shanghai	20. 03.2012	51
DX-700	Company B	Beijing	15. 04. 2012	49
DX-700	Company C	Tampere	18. 06. 2012	54

Regarding the output functionality of the user interface, two key features should never be ignored. The first feature is the reality of the information. It is very important that the reality of certain information should never be changed during the processes of both data analysis and data export. If the reality of data is weakened, the outcome would be valueless, because it can never be connected with the real world without its reality. The second feature is easy to understand. The whole system has been designed for supporting the work in the sourcing department, so the output data should be relatively clear and understandable. When the output information is easy enough for users to realize the essence of the data, their work efficiency as well as quality would be really improved. According to the researcher's observation on many business presentations, data visualization is a quite effective approach to help users understand complicated data. As is shown in Table 5.2, various types of data can be selected in a data table. It aims to increase the flexibility of the data exporting, and is able to reduce the probably negative effect of some unnecessary information included in the database.

Supplier Name	Unit Price
Company A	51
Company B	49
Company C	54

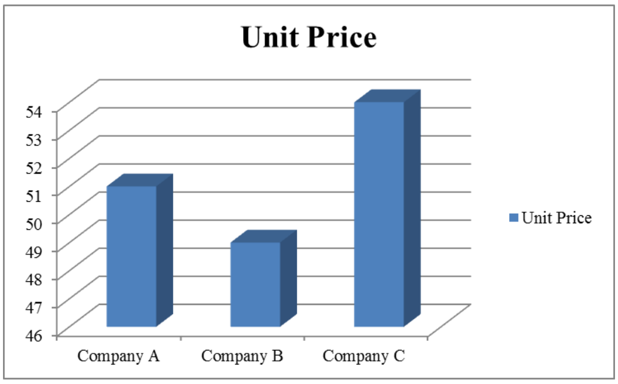


Figure 5.3. *Data visualization of an evaluation factor and plural owners*

In Figure 5.3, only one evaluation item (unit price) is extracted from the database as well as the owners (three companies) of the data, and then a column chart is generated on the right side. Comparing to the numbers in data table, the chart is able to reflect the difference among those objects more visually. So, it is easier for readers to understand the essence of an issue. It is necessary to claim that these figures of visualized data are inspired and developed by the popular software -- Microsoft's Office 2010.

Supplier Name	On Time Delivery	After Sales Service	Flexibility of Tech.
Company A	3	2	1
Company B	0	3	2
Company C	3	1	3

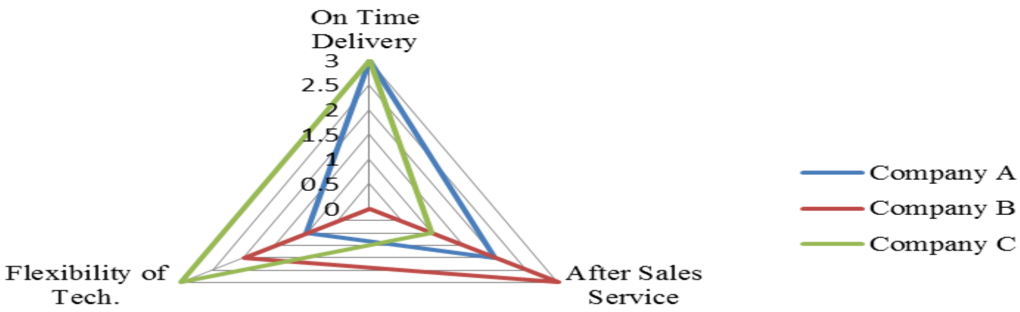


Figure 5.4. *Data visualization of plural evaluation factors and owners*

In Figure 5.4, a radar chart is applied in the visualization of plural evaluation factors and owners. From the lower chart, readers can easily tell the difference of the comprehensive capabilities of those three suppliers, because each area of a triangle in a specific color reflects its corresponding capability. So, company C would be the most competitive supplier just considering the three evaluation factors, while company A in the second place, and company B is the least preferable supplier. However, it is necessary to notice that this visualization approach is just fit for the comparison of those comparable evaluation factors (excluding "price" factor), since they are graded in a standard scoring system, and they can compare with each other in a same order of magnitude. The advantage of this approach is able to reflect a general result of multi-criteria comparison, but its disadvantage is that it cannot distinguish very tiny difference among many candidates as clearly as that in a column chart. To sum up, the relatively diverse data output options make the data management more flexible within the system, and it is considered rather important in a federated MDM mode (Ernst & Young, 2012).

6. RESEARCH DISCUSSION

Since the whole process of the research having been presented in previous chapters, this chapter aims to give a comprehensive discussion on the research. Generally speaking, the discussion would proceed in three phases. The first one is about the achievements of this research, while the second phase focuses on some uncertainties included in the research outcome, in the last phase, some out-of-research issues are pointed out for further study. Regarding the structure of the scientific research, those arguments in each phase are proposed in different perspectives. The arguments in the first phase refer to the functionality of the new designed database, which would improve the work efficiency of the sourcing department. The arguments in the second phase refer to the relationship between the implementation of the new database and those existing regulation in the company, some uncertainties should be noted in order to avoid potential conflicts. The arguments in the third phase refer to risk management of the whole project, concerning certain financial issues as well as some organizational behavior issues.

6.1. Achievements

In terms of the achievements of the research, they can properly satisfy most needs and expectations in the pre-research interviews (in Appendix 1). More specifically, the achievements could be concluded in the following five aspects.

- Data combination
- Additional evaluation factors
- The hierarchy of data structure
- Multi-functional data processing
- Data security

Firstly, data combination means that most essential sourcing related data dispersing in the three independent data systems are combined together in a new designed database. Thus, the diversity of the data in the new database is more than that in any single data systems in the company, and the “data overlapping” issue can be naturally solved with the application of the new combined data system. In the academic perspective, data combination is an essential approach to establish an effective database according to the federated MDM mode, which refers to a system with relatively centralized management and some flexibility in the lower-level operations (Ernst & Young, 2012). In such a condition, the efficiency of information retrieval on a large amount of data can be improved remarkably. Secondly, additional evaluation factors refer to those new added evaluation factors. As is shown in the pre-research interviews, there are some complaints about the lack of evaluation factors in the existing data systems. The category managers Mr. M claimed, “It is hardly to make comprehensive sourcing decisions without the consideration of many important evaluation factors. For example, I would like to know more detail information beyond the price of a component, so I prefer the cost-breakdown of the component rather than a price.” Many of the new added evaluation factors are proposed based on the needs of sourcing managers in Sandvik, while other factors are proposed according to some prior academic researches on the modern B2B cases. For example, the factor about “flexibility of technology” comes from the research by Erdem and Göcen in 2012. Thirdly, the hierarchy of data structure is established in the new database. It means that different types of data can be allocated and managed differently in the database, so that the whole system can operate more efficiently. According to the Analytical Hierarchy Process (AHP) theory (Saaty & Vargas, 2001), all sorts of data are classified into some groups and levels ensuring further data processing to be possible. For instance, some indirect information is extracted from the basic level for further processing. Otherwise, no further data processing can be achieved, since the obscure mixture of numeric and non-numeric data. The fourth achievement of the research is the ability of multi-functional data processing. Increased functionality, to some extent, improves the flexibility of the whole system, leading the database to a complete MDM federated mode (Ernst & Young, 2012). Since a certain number of measurable evaluation factors are introduced to the database, more complex data processing approaches can be done besides data storage and retrieval. As is mentioned in previous chapters, many mathematical calculations can proceed in the new system with the support of numeric data. For example, the item about “after sales services” can be quantified in the range from 0 to 3 according to some specific scoring standards. Then, the mathematical comparison of the service of different suppliers can be done through those numbers. Moreover, those numeric data can also be processed into a visualized outcome for users, so that a much more user-friendly user interface is eventually built up. Fifthly, data security is considered as an achievement, because a mature data supervising system is proposed in the research. The data supervising system is the centralized management of the federated mode (Ernst & Young, 2012), consisting of two parallel sectors. At the technical level, there is an automatic data supervisor module inside the system, which aims to detect any error occurs in the operation. When an error is targeted, a message will be sent to the terminals of the data supervising team, so that the IT experts in the team can repair the system. At the organizational level, the data supervising team is directly commanded by the chief manager of the sourcing department, without his or her permission, the team is not authorized to change any data in the system.

6.2. Uncertainties

In any scientific research, uncertainties that may happen during the implementation should always be taken into consideration, because they would be the inducement for some unexpected results in the future. Concerning this research, the following three issues could not be assured during the research, as most of them require coordination with the company's internal management.

- Corporation-level information sharing
- Data update regulation
- User's rights and priority

The first uncertain issue is whether all the necessary information can be shared at the corporation level. As is mentioned before, the new designed database would just be applied in the sourcing department of Sandvik Mining and Construction (China), but some work about information collection requires the support from other departments. As a result, the inter-department information connection tends to become an important uncertainty in the system. For example, the experienced sourcing engineer Mr. H (2013) complained that he was always annoyed with the drawing update issues among different departments in the company. In the current situation, a big number of production lines are transferred from Finland to China, since its relatively low production cost and shorter geographic distance to the Asia-Pacific market. However, only a certain percentage of R&D work is transferred at the same time. Thus, some products manufactured in China still necessarily require their specific drawings from Finland, and when a drawing has been modified in Finland, some sourcing components would be changed, but the sourcing department in China is always not informed in time. In addition to the first issue, how to ensure all the necessary data to be updated in time is regarded as the second uncertainty in the research. It means that some information may be changed actually, but which is only shared in small individual groups instead of being updated in the corporation-level data system. Take drawings as an example again, when an employee in the company does not update the latest drawings in time, he or she will not be punished in the current situation. In other words, nobody is really responsible for updating the latest drawings. This issue may result in the disorder or even chaos on the daily work among designing, sourcing, and manufacturing departments. Thus, a clear corporation-level data update regulation is highly recommended in order to improve the total efficiency of inter-department communication. The third uncertainty is about another concern of sourcing engineers – how to reasonably assign the user rights to employees at lower administrative level. In the pre-research interviews, a 6-year experienced sourcing engineer Mr. K (2013) complained, "When I want to search some suppliers' information in system P, I find that a lot of information is not open to sourcing engineers like me." Since some data in the system are confidential for the company, they are not totally open to all users. As a result, it is necessary for the management of the company to make a consensus on the user rights' arrangement at least in the department. In other words, the company needs to achieve a balance between potential risks and progress in work efficiency.

6.3. Other Issues

Despite many different aspects have been included in the research, there are still some fields not mentioned in the paper. To some extent, these unmentioned issues may affect the implementation of the new designed database in real business, so it is necessary to discuss these issues in this part. The discussion would proceed with two main concerns as following.

- Financial budget
- Resistance from the employees

The first issue is financial budget of this project, which is never mentioned to the public. In modern business, it is common that some details of the financial budget in a company are regarded as confidential information, especially when the specific project has not been proposed in the market. According to the sourcing manager Mr. G (2013), the project of building up an MDM-oriented database in the sourcing department of Shanghai subsidiary of Sandvik Mining and Construction (China) would be outsourced to some companies, so the budget was not open to the public. If some important information of the financial budget is leaked out before a project bid, the company may soon lose its bargaining power in the market. In this case, a general design of the database is proposed without consideration of financial budget, so some potential expenses should be calculated according to the design as the project budget. Reversely, the designed database could also be modified according to the specific financial situation of the company. The second issue is that some resistance may occur during the implantation of the new database. According to the previous study of organizational behavior and business case management, in a company, it is common that the resistant attitude of some employees may come out when a new change happens in their daily work. An unfamiliar data system can easily make some employees feel upset, because they always want to avoid uncertainties in work or life. In addition, some cultural factors may also affect the implementation of the database. In the pre-research interviews, the electrical sourcing engineer Ms. S (2013) claimed that she and some colleagues would like to share parts of their personal task within a small group rather than accomplish it independently. Asking for a little help from each other is not considered as workload personally, but as an approach to keep relatively close relationship with other people in daily work. So, if the new database encourages employees to work more independently, some resistance may come out among them.

7. CONCLUSION

In the modern world, international business has been significantly influenced by the accelerating globalization as well as advanced information technology. The former factor ensures a company to create considerable competitiveness via taking advantage of business resources all over the world. The latter one can provide more and more technical support for the operation of a business, even some strategic decisions. Sandvik, as a Sweden-rooted global company, is planning to expand its business in China and other Asia-Pacific countries based on its consistent localization strategy. This thesis aims to design an MDM-oriented supplier database for the sourcing department of Shanghai subsidiary of Sandvik Mining Corporation (China), improving the general efficiency and accuracy of sourcing data analysis. During the period of 4-month internship in the company, the designer gathered a lot of research materials by using existing material, questionnaire survey, qualitative interviews, observation, and in general action science. Through the systematic study on all these research materials, the real issues of the existing data systems were approached and finally tackled.

The whole process of designing an MDM-oriented supplier database comprises of four main phases. The first phase is theoretical analysis, which focuses on some academic concepts, such like master data, master data management, database, as well as relationships between sourcing decisions and master data. The second phase is to develop the general structure of the specific database. Before designing, it is necessary to make sure the sourcing procedures in Sandvik (China), its current challenges and requirements. In terms of the framework of the MDM-oriented database, it can be divided into six modules, including data collection, data categorization, data modeling, data consolidation, data export, and data supervising. Each functional module stands for a specific data processing sector, and they are connected with each other, forming an entire data processing system. The third phase is about the empirical study of the research. The designing process is analyzed in this part, and some practical experience is summarized. In addition, the outcome of the design is generally proposed in this part, referring to some practical functionality as well as multi-optional output approaches. The fourth phase aims to make a comprehensive discussion of the research. As a practical research, although some achievements are made in the design, but at the same time, there are still some uncertainties about the outside environment, especially some regulations and inter-departments coordination. Moreover, two additional issues, financial budget and risk management, are also discussed in this part, presenting some advices for further research in the future.

In terms of the project research, how to develop an MDM-oriented supplier database is a rather complex project, and this thesis has proposed a general structure for that. In order to accomplish the ultimate project, there is still a lot of work to be done. More specifically, the project requires necessary supports from both technical and organizational aspects. On one hand, a professional team should be established for the project. It must include certain professionals from different backgrounds, especially some experienced IT technicians, programmers, coders, and sourcing managers. All these team members can use their specific professional skills to solve different kinds of issues in detail, so that to carry out the whole project in reality. On the other hand, as mentioned before, certain organizational supports are also necessary for the implementation of the entire project. Actually, organizational supports refer to all other non-technical supports within the company, including financial support, management support, operating support, regulation support, etc. All these supports are important for the project. Without them, the project would be frustrated by series of practical difficulties during its implementation. Additionally, it is recommended to prepare a testing period before the wide adoption of the new database. In the perspective of risk management, leaving a time span for testing is necessary, because many potential issues would be found and tackled immediately without causing negative effects on the real business.

In conclusion, the project of my master's thesis in Tampere University of Technology offered me a great opportunity to obtain a lot of knowledge and experience through the research of "designing a supplier database to support sourcing analysis". During the research, I worked in Sandvik Mining and Construction (China) as a master data analyst for 4 months, where I gradually aggregated my knowledge about master data, master data management, sourcing data systems, supplier evaluation processes, etc. Different from the study in a university or from a literature, I could consistently learn from practice, and I believed the "learn-practice-learn" loop was an efficient approach for my study in Sandvik. Moreover, I was able to gather a number of research materials within the company, where I could get some useful literatures about the project, and easily interviewed those colleagues around me. Although the research is generally regarded as a successful one, it is not perfect. Reviewing the whole research process, it is clear that the communication between the designer and potential users was terminated after the stage of pre-research information gathering. As a result, the MDM-oriented database was designed without necessary information sharing with its potential users. In other words, those users were not involved in the designing process, so there was not consistent feedback from them while designing the database. If I could start over the research again, I would like to keep close contact with those potential users not just in the stage of information gathering, but also in the stage of designing. Thus, a continuous information sharing loop would be established between the designer and potential users, and the outcome of the design could be more practical and acceptable for the users.

REFERENCES

1. Accenture Research. (2011). PHILIPS sense and simplicity, Master Data Management Strategy Whitepaper

2. Akamp, M. & Müller, M. (2012). Supplier management in developing countries. Journal of Cleaner Production. Pp. 1-9.

3. Briggs, P. (1994). Case study: vendor assessment for partners in supply. European Journal of Purchasing and Supply Management, 1 (1), Pp. 49-59.

4. Cebi, F. & Bayraktar, D. (2003), An integrated approach for supplier selection. Logistics Information Management, 16 (6), Pp. 395-400.

5. Chen, W.H. (2011). Short list selection of intermediary in China: Marken Case Study. Tampere University of Technology. Faculty of Business and Technology Management. International Master's Programme in Business and Technology. Papers 4, Pp. 23-42.

6. Chen, X., Lin, Y, Liu, M., Gilson, M. K. (2001), The Binding Database: data management and interface design, Bioinformatics, (18), Pp. 130-139.

7. Choy, K. L., Lee, W. B., Henry, C. W., Lau, L. C. (2005), A knowledge-based supplier intelligence retrieval system for outsource manufacturing. Knowledge-based Systems, 18, Pp. 1-17.

8. Choy, K. L., Lee, W. B., Lo, V. (2002). An Intelligent Supplier Management Tool for Benchmarking Suppliers in Outsource Manufacturing. Expert System with Applications, 22, Pp. 213-224.

9. Choy, K. L., Lee, W. B., Lo, V. (2003). Design of a Case Based Intelligent Supplier Relationship Management System – the Integration of Supplier Rating System and Product Coding System. Expert System with Applications, 25, Pp. 87-100.

10. Chu, T. C. & Varma, R. (2011). Evaluating suppliers via a multiple levels multiple criteria decision making method under fuzzy environment. Computers & Industrial Engineering. Issue 2. Pp. 653-660.

11. Clark, S. (2009), Guide to sourcing database developers: How to source the right database developers for your business, Computer.

12. Dickson, G. W. (1966). An analysis if vendor selection systems and decisions, Journal Purchasing, 2(1), Pp. 28-41.

13. Dreibelbis, A., Hechler, E., Milman, I., Oberhofer, M., van Run, P., Wolfson, D. (2008). Enterprise master data management: An SOA approach to managing core information, Upper Saddle River, NJ: IBM Press.

14. Erdem, A. S. & Göcen, E. (2012). Development of a decision support system for supplier evaluation and order allocation, Expert Systems with Applications, 39, Pp. 4927-4937.

15. Ernst & Young. (2012). Recommendation of Sourcing MDM Modes

16. Fagerfjäll, R. (2012). The Sandvik Journey: the first 150 years. Production: Bokförlaget Max Ström. Sandvik AB. Pp. 60-61.

17. Ghodsypour, S. H. & O'Brien, C. (1998). A decision support system for supplier selection using an integrated analytic hierarchy process and linear programming. International Journal of Production Economics. Volumes 56-57. Pp. 199-212.

18. Greeff, G., Ghoshal, R., Mackay, S. (2004), Chapter 10: Production data collection and performance analysis, Practical E-Manufacturing and Supply Chain Management, Pp. 271-307.

19. Gummesson E. (1993). Qualitative Methods in Management Research (2nd Ed.). Printed in the United States of America.

20. Ha, S. H. & Krishnan, R. (2008). A hybrid approach to supplier selection for the maintenance of a competitive supply chain. Expert System with Applications. Issue 34. Pp. 1303-1311.

21. Halpin, T., Evans, K., Hallock, P., Maclean, B. (2003). Chapter 2: Database Modeling, Database Modeling: With Microsoft Visio for Enterprise Architects.

22. Kokemuller, J. & Weisbecker, A. (2009). Master Data Management: Products and Research, Fraunhofer IAO, Germany.

23. Large, R. (2006), Interpersonal communication and successful supplier management: analysis of external communication behavior of purchasers. Zeitschrift für Betriebswirtschaft 76 (10), Pp. 1005-1034.

24. Larry E. Daniel & Lars E. Daniel. (2012). Chapter 39: Database, Digital Forensics for Legal Professionals: Understanding Digital Evidence from the Warrant to the Courtroom.

25. Lightstone, S., Teorey, T., Nadeau, T. (2007). Chapter 1: Introduction to physical

database design, Physical Database Design: The Database Professional's Guide to Exploiting Indexes, Views, Storage and More, Pp. 1-13.

26. Loser, C., Legner, C., Gizanis, D. (2004). Master Data Management for Collaborative Service Processes, Institute of Information Management, University of St. Gallen, Switzerland.

27. Loshin, D. (2009). Master Data Management, Burlington, MA: Morgan Kaufmann. Printed in the United States.

28. Makkonen, H. & Olkkonen, R. (2013). The conceptual locus and functionality of key supplier management: A multi-dyadic qualitative study. Industrial Marketing Management. Pp. 2-13.

29. Möller, K. (1994). Interorganizational marketing exchange: Metatheoretical analysis of current research approaches. Research Traditions in Marketing. Pp. 347-372.

30. Oracle. (2009). Master Data Management, an Oracle White Paper, Pp. 1-41.

31. Otto, B. (2012). How to design the master data architecture: Findings from a case study at Bosch, International Journal of Information Management, Pp. 337-346.

32. Ott, R. L. & Longnecker, M. (2010). An Introduction to Statistical Methods and Data Analysis, Belmont, CA, USA.

33. Pardo, C., Missirilian, O., Portier, P., & Salle, R. (2011). Barriers to the "key supplierization" of the firm. Industrial Marketing Management. 40 (6). Pp. 853-861.

34. Phillips, J. (1999). Chapter Five—Data Collection Techniques, Accountability in Human Resource Management, Pp. 110-146.

35. Razmi, Z., Rafieri, H., Hashemi, M. (2009). Designing a decision support system to evaluate and select suppliers using fuzzy analytic network process, Computers & Industrial Engineering, 57, Pp. 1282-1290.

36. Rosenberg, J. M. (1987). Dictionary of Computers, Information Processing & Telecommunications, 2nd ed., New York, John Wiley & Sons.

37. Saaty, T. L. (1980). The analytic hierarchy process, New York. NY: McGrawHill.

38. Saaty, T. L. & Vargas, L. G. (2001). Models, methods, concepts & applications of the analytic hierarchy process, Kluwer Academic Publishers.

39. Sandvik Annual Report, (2012). Sandvik AB.

40. Sandvik World, (2013). Sandvik AB.

41. Silberschatz, A., Korth, H. F., Sudarshan, S. (2006). Database System Concepts, 5th Ed. New York: McGraw-Hill.

42. Smith, H. A. & McKeen, J. D. (2008). Developments in Practice XXX: Master Data Management: Salvation or snake oil? Communication of AIS, 23, Pp. 63-72.

43. Stevenson, W. J. (2008). Operation Management (9th ed.). New York: McGrawHill.

44. Verma, R. & Pullman, M. E. (1998). An analysis of the supplier selection process, Omega International Journal of Management Science, 26(6), Pp. 739-750.

45. Wang, G., Huang, S. H., Dismukes, J. P. (2004). Product-driven supply chain selection using integrated multi-criteria decision-making methodology, International Journal of Production Economics, 91 (1), Pp. 1-15.

46. Wise, R. & Morrison, D. (2000). Beyond the exchange: the future of B2B. Harvard Business Review. November-December. Pp. 86-96.

47. Wolter, R. & Haselden, K. (2006). The What, Why and How of Master Data Management, Microsoft Corporation.

48. Zhang, Y. F. (2012). A multi-criteria approach for supplier election: a Sandvik case study. A seminar report of Academic Writing 2. Tampere University of Technology, Finland.

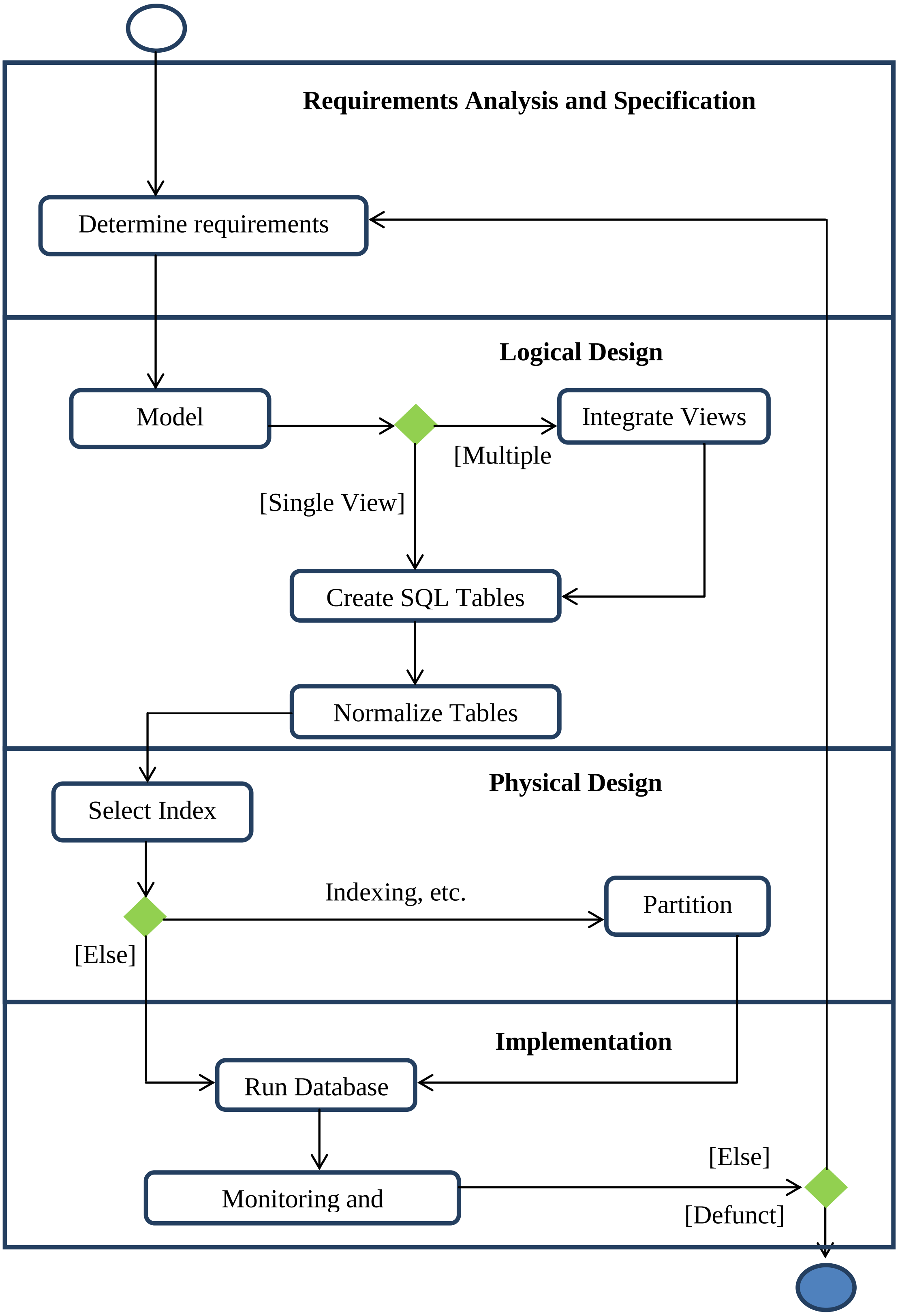
49. Zeydan, M., Colpan, C., Cobanoglu, C. (2011). A combined methodology for supplier selection and performance evaluation. Expert Systems with Applications. Volume 38. Issue 3. Pp. 2741-2751.

50. Zynapse, (2010), The Business Benefits of Material Master Data Management, A Zynapse White Paper, A Zycus Inc. Division.

http://www.sandvik.com/Global/About%20Sandvik/The%20Sandvik%20World/The%20Sandvik%20World%202013.pdf

http://www.sandvik.com/Global/Investor%20relations/Annual%20reports/Annual%20Report%202012.pdf

APPENDIX 1: *Database Life Cycle*



(Source from: Lightstone et al, 2007, Physical Database Design)

APPENDIX 2:

A summary of pre-research interviews in the sourcing department of Sandvik Mining and Construction (China) in Shanghai

1. Top 5 disadvantages of the current data systems

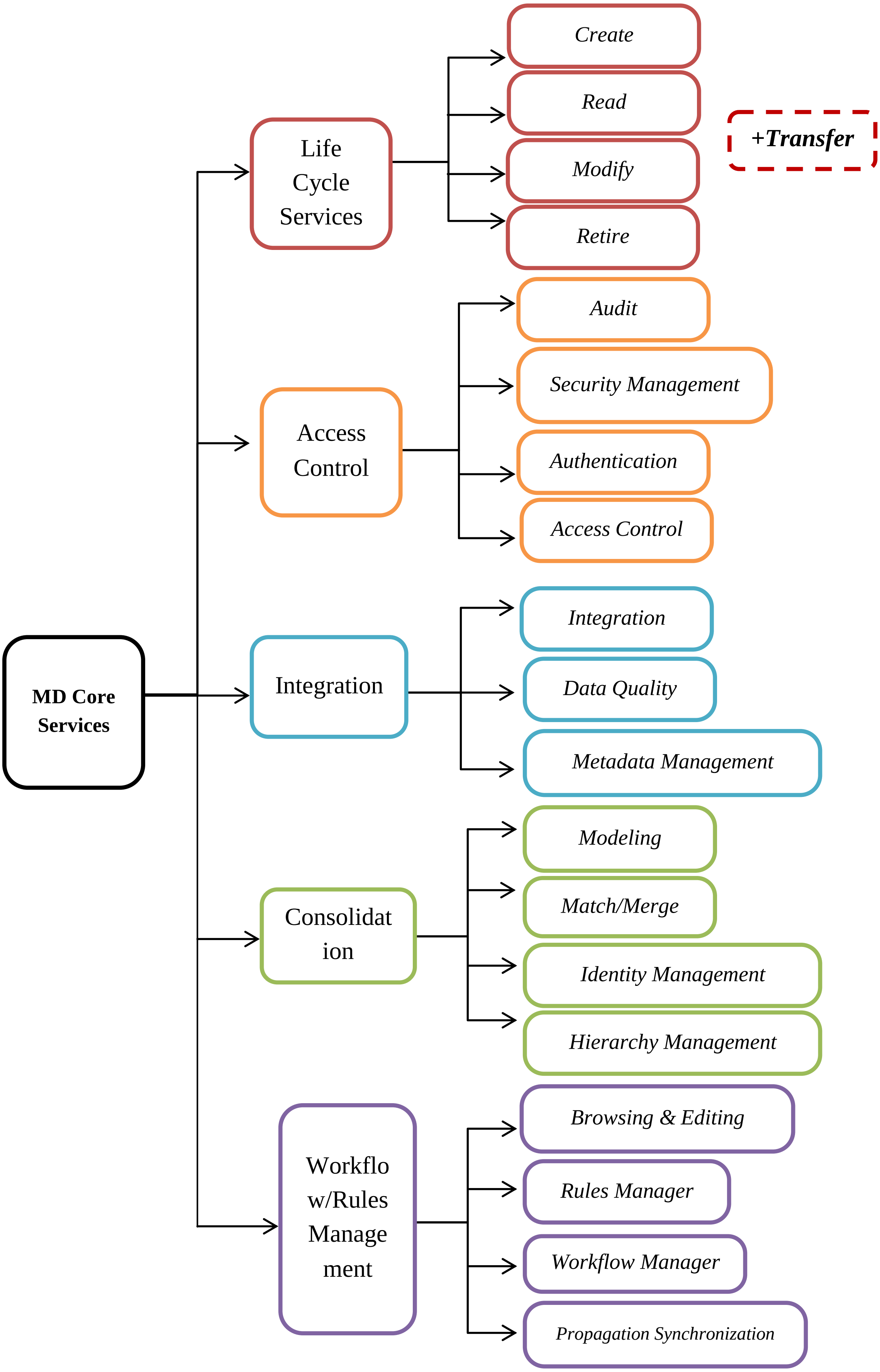
Disadvantages	**Notes**
1. Overlapped information	3 independent data systems shared a lot of similar functions and data.
2. Not updated data	Sourcing engineers suffered from those old drawings available in the databases.
3. Detail information unavailable	Several managers complained the lack of certain sourcing information in detail.
4. Unable to access	Some necessary information was not directly open to lower level employees.
5. Relatively complex to use	People were confused about different operation standards for each data system.

2. Top 5 expectations for a new database

Expectations	**Notes**
1. A multi-functional database	It should be used for both data retrieval and data analysis.
2. Data update in time	The latest updated data are really valuable for nowadays' business.
3. More data should be included	Managers require more specific data to detect the reality of a case.
4. Open to all	Lower level employees also need to use the database for daily work.
5. Easy to operate	A new database should never be too complex to use in real life.

APPENDIX 3:

Framework of master data core services (Modified from Loshin, 2009)



APPENDIX 4:

The list of sourcing related factors in the new database

(1) Order No.	(24) Item Name	(47) Unit Price
(2) Order Type	(25) Item Code	(48) Direct Material Cost
(3) Purchase Officer	(26) Local Item No.	(49) Direct Processing Cost
(4) Purchase Line	(27) Item Description (Loc.)	(50) Direct Packing Cost
(5) Stock Room	(28) Item Description (Eng.)	(51) Transportation Cost
(6) Issue Date	(29) Sub-category	(51) Inventory Cost
(7) Due Date	(30) Sub-category Code.	(52) Overhead
(8) Product Code	(31) Technical Category	(53) Profit
(9) Unit Weight	(32) Supplier Name	(54) Lead Time
(10) Total Weight	(33) Supplier Type	(55) Drawings
(11) Intake Last 12 Months	(34) Tele. No.	(56) Quality Failure Rate
(12) Hits Last 12 Months	(35) Business Area	(57) Defect Fix Time
(13) Quality Outstanding	(36) Payment Term	(58) Expected Delivery Date
(14) Value Outstanding	(37) Supplier Code in System	(59) Earlier Delivery (days)
(15) Total Quantity Received	(38) Local Supplier ID	(60) Late Delivery (days)
(16) Total Weight Received	(39) Country	(61) Flexibility of Capacity
(17) Total Value Received	(40) Address	(62) Flexibility of Tech.
(18) GRN No. *	(41) Region No.	(63) Flexibility of Transport
(19) Advice Note No.	(42) Selected Currency	(64) After Sales Service
(20) Invoice No.	(43) Turnover	(65) Allocated Capacity
(21) Promise Date	(44) Owning Team	(66) Quality Standard
(22) Receipt Date	(45) Latest Price	(67) Perfect Order Fulfillment
(23) Unit	(46) Purchase Year	(68) R&D Involvement

(GRN No.* = Goods Received Note Number)