FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

A Business Data Modeling Methodology for the Financial Industry

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Wirtschaftsinformatik

eingereicht von

Martina Resutikova
Matrikelnummer 1127575

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung: O.Univ.Prof. Dipl.-Ing. Dr.techn. A Min Tjoa
Mitwirkung: (Univ.Ass. Mag.rer.soc.oec. Elmar Kiesling PhD)

Wien, 31.01.2014

(Unterschrift Verfasserin) (Unterschrift Betreuung)

FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

A Business Data Modeling Methodology for the Financial Industry

MASTER'S THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Business Informatics

by

Martina Resutikova
Registration Number 1127575

to the Faculty of Informatics
at the Vienna University of Technology

Advisor: O.Univ.Prof. Dipl.-Ing. Dr.techn. A Min Tjoa
Assistance: (Univ.Ass. Mag.rer.soc.oec. Elmar Kiesling PhD)

Vienna, 31.01.2014

(Signature of Author) (Signature of Advisor)

Erklärung zur Verfassung der Arbeit

Martina Resutikova
Burggasse 123, 1070 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

(Ort, Datum) (Unterschrift Verfasserin)

Acknowledgments

This research project would not have been possible without the support of many people. The author wishes to express her gratitude to her supervisor, Associate Professor Christian Wolf and Assistant Elmar Kiesling who were abundantly helpful and offered invaluable assistance, support and guidance.

Deepest gratitude are also due to the company zeb/rolfes.schierenbeck.associates, especially to Diana Kartheuser for her valuable guidance, as well as to all participants of survey and case study research without whose knowledge and assistance this study would not have been successful.

The author would also like to convey thanks to the Informatics Department at INSA de Lyon as well as to Faculty of Informatics at the TU Wien for providing necessary background during author's studies.

Abstract

Data warehousing is successfully creating a layer between classical operational databases and reporting applications. In the financial industry, data warehousing architectures are crucial for creating market value and supporting business activities. Business data modeling, the initial phase of a DWH project, establishes a common language among stakeholders such as departments of a bank. However, we find that methodologies followed in the financial services industry today are insufficient in practice. In addition, these methodologies do not offer solutions for typical issues that occur during development of a data warehouse inside a bank. This leads us to the following research question: Can a structured methodology for the early development phase of financial DWH projects help prevent failure of these projects? In order to answer this question we will present a business data modeling methodology that establishes a basis for consistent data and clear product scope definition. Furthermore, it enhances top management prioritization and decreases interpretation flexibility caused by insufficient specifications and thereby helps to minimize the Business-IT gap.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

The amount of information to store and the pace of its retrieval from data sources have become crucial factors which developers take into account when they design or modernize data holding solutions. Data warehousing technology that was brought into practice more then 20 years ago (Inmon, 1992 and Kimball, 1996) represents a solution that addresses both aspects. Since a bank can be considered as an information processing engine, the implementation of a data warehouse is especially important in the banking sector. In order to be able to retrieve appropriate information, the bank needs to built its data warehouses properly. One way to assure this is to build data warehouses on exhaustive business requirements that reflect business need within a specific financial sector. The accuracy of the content of these requirements is therefore critical. Due to this fact, we need to identify factors that have influence on the quality of these requirements in order to help to reduce their impact on a final DWH product. We can structure these factors in form of an understandable methodology that contains rules of prevention and correction of their impact. This way, potential issues that require a solution in further project development phases can be captured in advance.

Various authors already described methodologies for data warehouse implementation. Especially Inmon (Inmon, 2005) defined principles for data-driven DWH-development. This guide is however rather theoretical and does not discuss the content of a DWH at all or only marginally. As a consequence, developers frequently face problems in applying such methodologies in practice (Huang et al., 2007 and Golfarelli et al., 2011). Moreover, a mutual and a straight-forward understanding of introduced terminology is a crucial element that needs to be acknowledged. In a bank, principally due to this element, projects have a potential to fail (Jukic and Nicholas, 2010).

Business requirements are key elements in the business data modeling. And it is the quality of business data modeling itself that creates the base for the success of the whole project. This thesis concentrates therefore on capturing crucial success factors in data warehousing as well as their importance within the business data modeling (BDM) phase. It offers prescriptive guidelines which aim to prevent these factors from occurring or at least decrease their impact on subsequent DWH development phases. We aim to find answer to the following research ques-

tion: **Can a structured methodology for the early development phase of financial DWH projects help prevent failure of these projects?**

This thesis introduces a structured methodology for business requirements collection/specification and business data modeling phases of a data warehouse project. This methodology is based on Inmon's model. The action research we follow in the context of the thesis helps us to formulate the methodology. The present thesis is structured in five chapters. The first chapter explains data warehousing in general and pinpoints problems that occur frequently during DWH implementation. Additionally, it describes the business data modeling phase of this process as well as identifies its deficiencies. The next chapter is the description of the research approach. It describes all techniques used step by step, together with the reasoning why a particular technique was used and how it was implemented. The third chapter discusses results of this research and the forth one introduces the methodology in detail. The last chapter sums up achieved elements and describes future work in improving this methodology as well as its limitations. A glossary is to be found at the end of the document.

The role of the author in the present master thesis is to conduct the research and guide participants through the process. She prepares necessary steps to be taken within respective research stages and processes collected information. Based on results of this processing, she formulates a BDM methodology and evaluates this one with help of research participants. She was however neither involved in the project development that is object of the case study nor was her opinion considered in survey results. Even though she was involved in a DWH project as an intern, especially in its data requirements collection and business data modeling phase, result of this master thesis is definitely not a pure description of actions taken in this project.

CHAPTER 2

State of the Art

This chapter describes fundamental data warehouse characteristics. It focuses on the reasoning why in the present master thesis we focus on Inmon's approach and describes its data modeling part. We close this chapter by pinpointing major defects of existing methodologies.

2.1 DWH Architecture

A list of reasons why operational and informational/analytical processing is separated (Inmon, 2005) helps us to understand incentives for analytical models implementation:

- The data serving operational needs is physically different data from that serving informational or analytic needs.
- The supporting technology for operational processing is fundamentally different from the technology used to support informational or analytical needs.
- The user community for operational data is different from the one served by informational or analytical needs.
- The processing characteristics for the operational environment and the informational environment are fundamentally different.

The beginning of data warehousing dates back to 1990, when Bill Inmon earned the moniker "Father of Data Warehousing" by coining the term in his seminal work entitled "Building a Data Warehouse" (Breslin, 2004). His book (Inmon, 2005) and the methodology of Ralph Kimball (Kimball and Ross, 2002) represent at the moment two main data warehousing models.

Main strengths of a data warehouse environment are the following:

- there is a single integrated source of data (the data warehouse) in data warehouse environment

- the granular data in the data warehouse is easily accessible
- the data warehouse forms a foundation for reusability and reconciliation of data

Data warehouse is "a subject-oriented, integrated, nonvolatile and time-variant collection of data in support of management's decisions" (Inmon, 2005).

Figure 2.1 illustrates the characteristic of subject-orientation. Classical operational systems are organized around functional applications of the company. For an insurance company, the applications may be for the processing of auto, life, health and casualty. The major subject areas of the insurance corporation might be customer, policy, and claim. For a manufacturer, the major subject areas might be product, order, vendor, bill of material, and raw foods. For a retailer, the major subject areas may be product, sale, vendor, and so forth. Each type of company has its own unique set of subjects (Inmon, 2005).

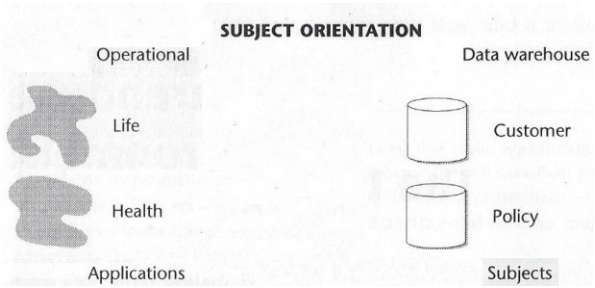

Figure 2.1 – Subject orientation of data[1]

[1] Source: Inmon, 2005

Of all aspects of a data warehouse, integration is the most important one (Inmon, 2005). Data is fed from multiple, disparate sources into the data warehouse. As the data is fed, it is converted, reformatted, re-sequenced, summarized, and so forth. The result is that the data - once it resides in the data warehouse - has a single physical corporate image (Inmon, 2005). This aspect is shown in Figure 2.2. Designers often do not consider the fact that the data they are operating on would ever have to be integrated with other data. Across multiple applications, there is therefore no application consistency in encoding, naming conventions, physical attributes, measurements of attributes, and so forth (Inmon, 2005). For example (cf. Figure 2.2), as far as encoding of gender is concerned, it matters little whether data in data warehouse is encoded as m/f or 1/0. What does mattes is that regardless of method or source application, data warehouse encoding is done consistently (Inmon, 2005).

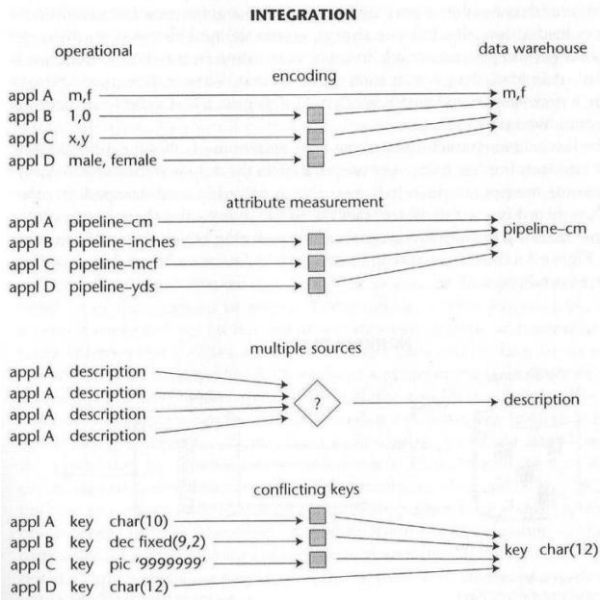

Figure 2.2 – Integration[2]

[2]Source: Inmon, 2005

The nonvolatility of data within a data warehouse is illustrated in Figure 2.3. Data in operational system is accessed one record at a time. In such an environment, data is updated on a regular basis. This is not the case in a data warehouse. Instead, when data in the data warehouse is loaded, it is loaded in a snapshot, static format. When subsequent changes occur, a new snapshot record is written. This way, historical data is established (Inmon, 2005).

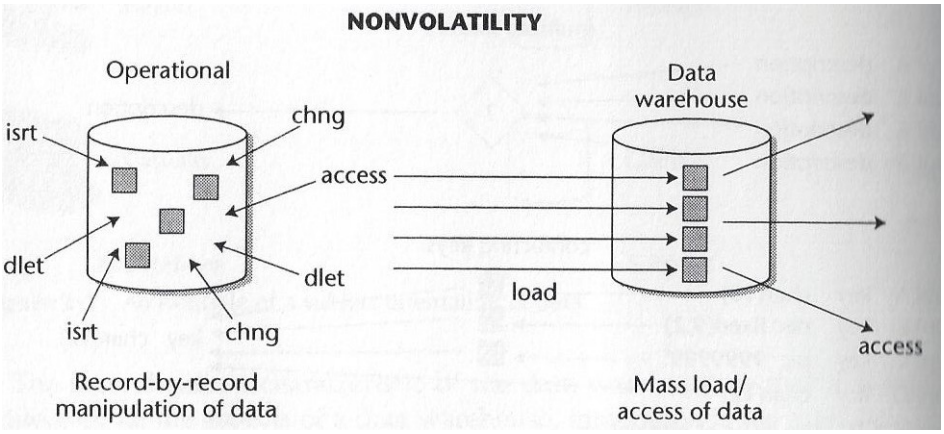

Figure 2.3 – Nonvolatility[3]

The issue of nonvolatility leads us to the final characteristic, to the time variancy illustrated in the Figure 2.4. Time variancy implies that every unit of data in the data warehouse is accurate as of some moment in time. There is therefore some form of time marking to show the moment in time during which a record is accurate.

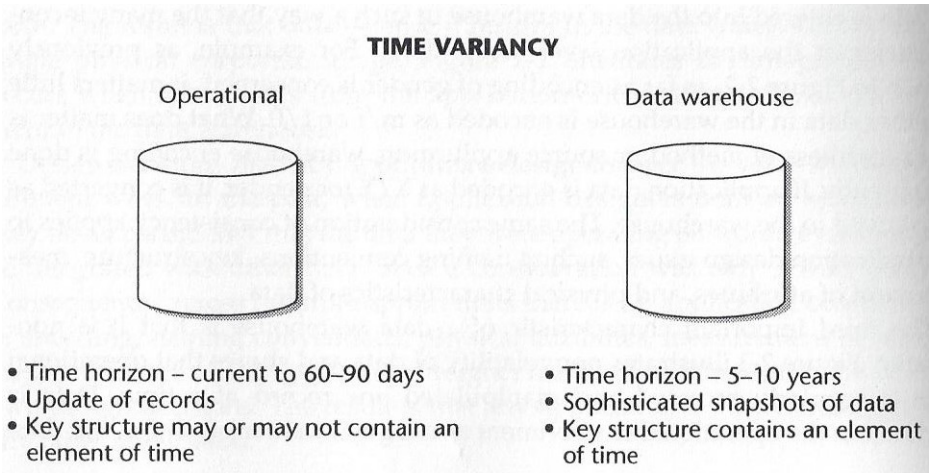

Figure 2.4 – Time variancy[4]

[3]Source: Inmon, 2005
[4]Source: Inmon, 2005

Understanding characteristics of data warehouse and advantages of such architecture in comparison to classical operational one opens discussion on possible implementation approaches.

2.2 DWH Architecture Approaches

Almost all approaches that focus on DWH development use either Inmon's or Kimball's model as a basis. Making good choices in DWH software tools and development approaches requires therefore understanding of the two main data warehousing models - Inmon's and Kimball's (Breslin, 2004).

In this section, we will describe their main characteristics and differences between them. At the end we will choose the one that fits the financial context best.

Bill Inmon advocates on one hand a top-down development approach that adapts traditional relational database tools to the development needs of an enterprise-wide data warehouse. From this enterprise-wide data store, individual departmental databases are developed to serve most decision support needs (Breslin, 2004).

Ralph Kimball, on the other hand, suggests a bottom-up approach that uses dimensional modeling, a data modeling approach unique to data warehousing. Rather than building an enterprise-wide database, Kimball suggests creating one database (or data mart) per major business process. These multiple data marts are expected to be highly interoperable. Enterprise-wide cohesion is accomplished by using another Kimball innovation, a data bus standard (Breslin, 2004). Kimball's dimensional modeling (cf. Figure 2.5) is different from the well known entity-attribute data models such as Entity Relationship Diagrams (ERDs). Kimball's tables are either fact or dimension tables. Fact tables contain metrics, while dimension tables contain attributes of these metrics. Facts are therefore central tables that are characterized by a certain number of dimension tables. Since the graphical representation of Kimball's model reminds a star (cf. Figure 2.5), this design is called star schema. Dimensional modeling violates normalization rules in order to increase query performance (Kimball and Ross, 2002).

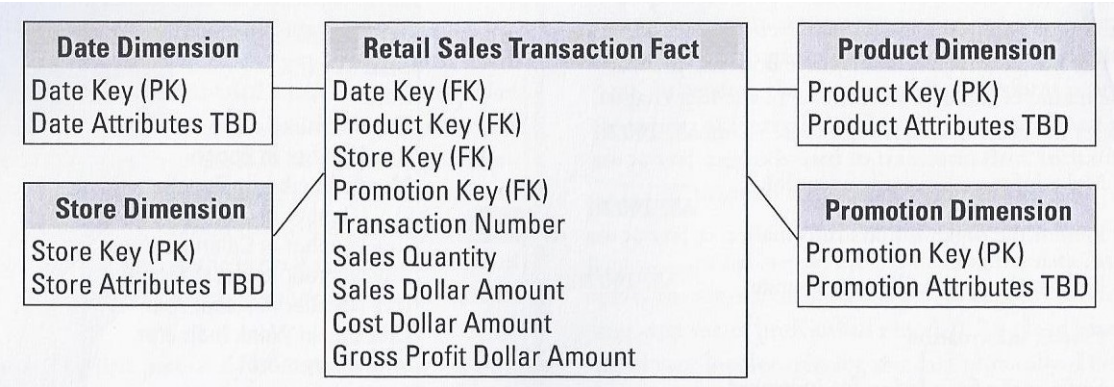

Figure 2.5 – Simplified example of Kimball's fact and dimension tables[5]

[5]Source: Kimball and Ross, 2002

The main criticism of the star schema is that it is overly restrictive because its structure forces the data warehouse designer to choose a single subject (Schuff et al., 2005). Indeed, the problem with a star join design is that it optimizes the access of data for one group of people at the expense of everyone else. Because the star join is shaped around user requirements and because user requirements vary from one type of user to the next, not surprisingly, different star joins are optimal for different types of users (Inmon, 2000). Moreover, star schema cannot effectively reflect complex business scenarios (Haughey, 2004).

Table 2.1 sums up eight characteristics favoring either Kimball's or Inmon's approach (Wells, 2003):

Nr.	Characteristic	Favors Kimball	Favors Inmon
1.	Nature of the organization's decision support requirements	Tactical	Strategic
2.	Data integration requirements	Individual business areas	Enterprise-wide integration
3.	Structure of data	Business metrics, performance measures, and scorecards	Non-metric data and for data that will be applied to meet multiple and varied information needs
4.	Scalability	Need to adapt to high volatile needs within a limited scope	Growing scope and changing requirements are critical
5.	Persistency of data	Source systems are relatively stable	High rate of change from source systems
6.	Staffing and skills requirements	Small teams of generalists	Larger team(s) of specialists
7.	Time to delivery	Need for the first data warehouse application is urgent	Organization's requirements allow longer start-up time
8.	Cost to deploy	Lower start-up costs, with each subsequent project costing about the same	Higher start-up cost, with lower subsequent project development costs

Table 2.1 – Specific characteristics favoring Inmon's or Kimball's model[6]

When bank's management decides to implement DWH architecture, it aims to build a base for strategic decision support (cf. first characteristic in Table 2.1). Additionally, within most banks, an enterprise-wide solution that builds a common language and understanding of data among bank's departments needs to be established. Kimball's idea of building each data mart on a single business process is in contradiction with this purpose (cf. second characteristic in the Table 2.1). Furthermore, data located inside a financial data warehouse is applied to meet multiple and varied information needs (cf. third characteristic in Table 2.1). Moreover bank is

[6]Source: Wells, 2003

rather composed of teams of specialists that are working within specific business functions or activities of business processes and often do not know in detail how other departments of the bank operate (cf. sixth characteristic in Table 2.1). Since the financial industry satisfies rather stable requirements (modifications occur principally due to changes in regulations), a bank can wait to see results over a longer time frame (cf. seventh characteristic in Table 2.1). In addition, if the bank decides to integrate another financial sector or department into the enterprise-wide data warehouse, the cost of such an implementation will be decreased in comparison to the cost of start-up project implementation of Inmon's approach (cf. eighth characteristic in Table 2.1). Even though source systems within the bank are relatively stable (cf. fifth characteristic in the Table 2.1), for our implementation purposes Inmon's model is more suitable. It satisfies financial industry needs in most aspects (six out of eight characteristics from the Table 2.1). Since posterior scope adjustments are critical according to Inmon's data warehouse implementation (cf. fourth characteristic in Table 2.1), our methodology will pay higher attention to this issue.

We will therefore concentrate in this thesis only on the methodology developed by Inmon and the one he describes in the fourth edition of his "Building the Data Warehouse" (Inmon, 2005).

2.3 Inmon's Methodology

Four levels of architecture represent a data warehouse according to Inmon (cf. Figure 2.6): operational, atomic, departmental and individual level. The first one, a so-called operational level holds application-oriented data that is used for high performance transaction-processing purposes. The second level, called the atomic level represents the data warehouse. It holds operational data that is not updated and forms therefore entities with corresponding historical data. The next level, the departmental level, stands for data marts, where uniquely derived data required by end users is present. The very last level is called the individual one and it is a place of heuristic analysis that determines the susceptibility of a system towards a particular behavior. At first sight, it may seem that this four-levels architecture creates high redundancy inside a data warehouse (cf. Figure 2.6).

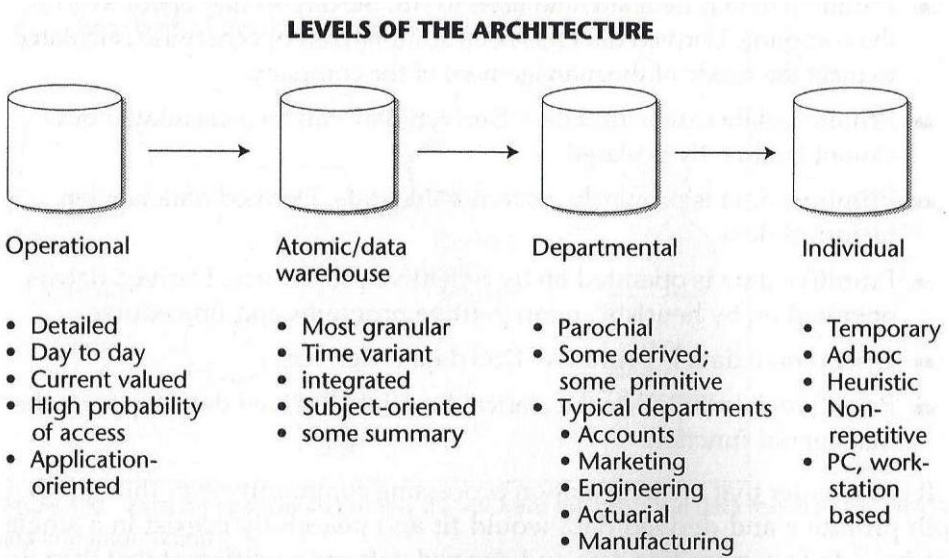

Figure 2.6 – Inmon's four-level architecture[7]

[7] Source: Inmon, 2005

Figure 2.7 shows that it is not the case. The four levels of architecture do not create redundant information. On the contrary, every simple transactional record acquires thanks to it a reporting and decisive power.

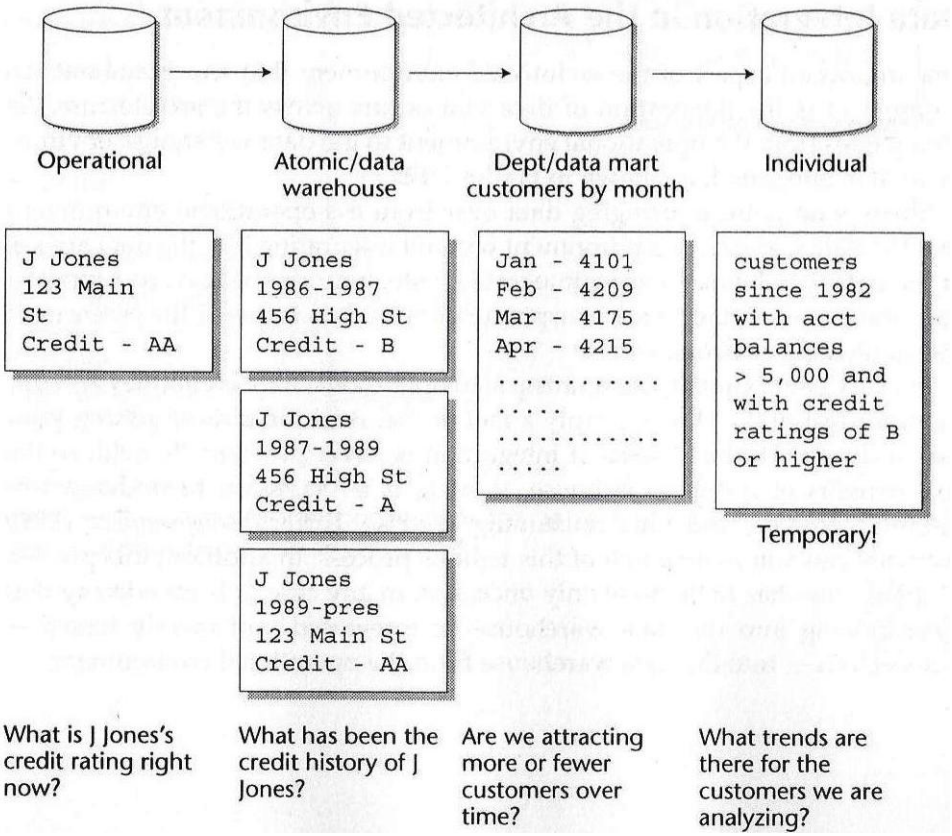

Figure 2.7 – Example of queries on different architecture levels[8]

[8]Source: Inmon, 2005

In order to interpret heterogeneous data among data applications (cf. Figure 2.8) and integrate it into the data warehouse in an appropriate way, Inmon proposes a three level data modeling approach: high-level modeling (called the entity relationship diagram, or ERD), mid-level modeling (called the data item set, or DIS), and low-level modeling (called the physical model) (Inmon, 2005).

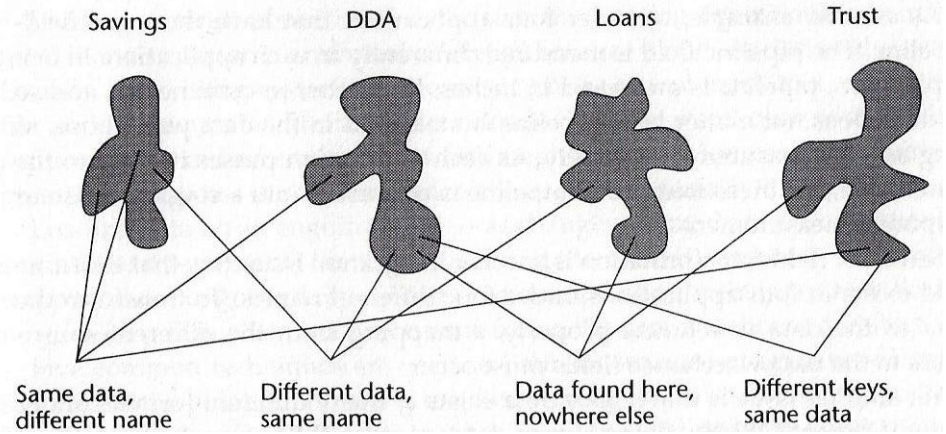

Figure 2.8 – Heterogeneous data across applications[9]

[9] Source: Inmon, 2005

The ERD corresponds to the actual business data model we focus on in the context of the present master thesis. ERDs feature entities and relationships. Entities, that are shown in ERD level are the highest level of abstraction. Entities, that belong to the scope of the model are determined by what is termed scope of integration (cf. Figure 2.9).

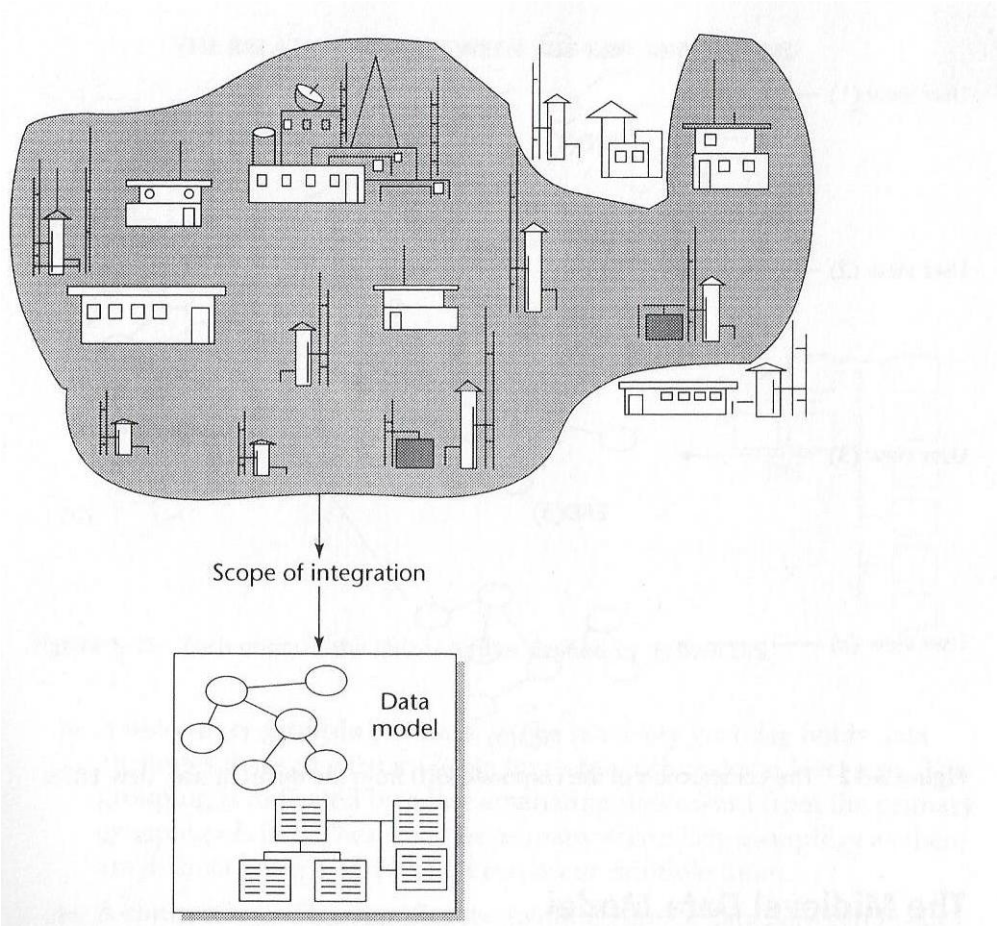

Figure 2.9 – Scope of integration[10]

[10]Source: Inmon, 2005

Scope of integration defines boundaries of the data model and must be defined before the modeling process commences (Inmon, 2005). The scope is agreed on by the modeler, the management, and the ultimate user of the system (Inmon, 2005). The corporate ERD is a composite of many individual ERDs that reflect different views of people across the corporation (cf. Figure 2.10). ERDs that represent known requirements of the decision support systems (DSS) community are created by means of user view sessions or Joint Application Design (JAD) sessions, which are interview sessions with the appropriate personnel in various departments (Inmon, 2005).

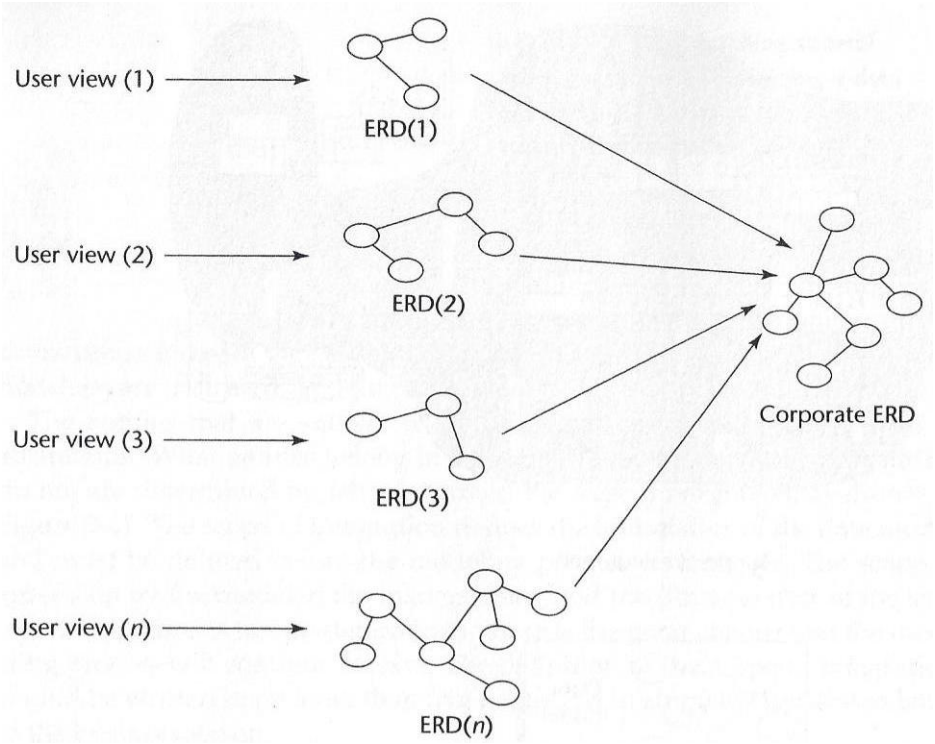

Figure 2.10 – Construction of corporate ERD from the different user view ERDs[11]

Since the present master thesis concentrates on business data modeling, subsequent development phases are introduced very briefly. The mid-level model analyzes entity attributes in more in detail. It identifies attributes of data in a data model and relationships among those attributes (Inmon, 2005). The physical model is created from the mid-level data model merely by extending the mid-level data model to include keys and physical characteristics of the model (Inmon, 2005). Inmon proposes an iterative implementation process where all development effort is being driven from the data model. According to Inmon, the reason why an iterative approach is more appropriate is the following: “Any architecture that must be implemented all at once, in a big bang, is doomed to failure in today’s world. There simply is too much risk and too long a

[11] Source: Inmon, 2005

period to wait until there is a payback. In addition, trying to freeze changes to consider any path that is revolutionary (rather than evolutionary) is unrealistic." (Inmon, 2005).

2.4 DWH Implementation Problems

Even though data warehouse architectures have been discussed for a long time by a lot of researchers, companies, and data storing solution providers and even though Inmon and Kimball are cited in virtually every paper discussing data warehousing, it is still difficult to apply available models in practice. Inmon's publications discuss the actual implementation and the data modeling in particular only marginally. He states, that the iterative implementation is derived from previously defined data model, but also in his book only a limited number of pages discusses the data modeling itself. He only introduces terminology that allows to define data modeling, but activities that one should follow or how one should implement them are discussed on a highly abstract level. We can consider his publication "Building the data warehouse" as a set of theoretical definitions. This is one of the reasons why developers face problems of applying data warehousing architectures in practice (Huang et al., 2007 and Golfarelli et al., 2011).

Inmon himself cites other implementation issues especially concerning methodologies (Inmon, 2005).

- Methodologies generally show a flat, linear flow of activities. They lack possibility of iterative approach execution.
- They often do not recognize the need to revisit one or more activities.
- Methodologies usually describe a prescribed set of activities to be done and do not consider the possibility of skipping or adding customized activities.
- They often do not consider size of system they help to build.
- They often do not include stopping checkpoints in case of failures.
- Methodologies are often sold as solutions, not tools. In this case, it tends to replace good judgment and common sense.

According to Eckerson, 2003, it is likewise important to include the following into methodologies:

- leadership, communication, planning and interpersonal relationships
- allocation of sufficient resources to create and maintain the data warehouse

We will aim to address these problems in the present master thesis.

CHAPTER 3

Research Process

Since estimating the impact of a business data model before its actual implementation is not straight-forward, we will use a qualitative research method in the context of the present master thesis. In order to answer our research question, we conducted action research (Berg, 2004).

We conducted the research in cooperation with consultants at zeb/rolfes.schierenbeck.associates (henceforth zeb/) and its clients. The process started with an initial survey that aimed to identify crucial factors that influence the success of DWH projects. Results of this survey were used to form the strategy of a case study that analyzed business data modeling part of a finished DWH project. The research process ended with a final survey. Its role was to evaluate the outcome of the present master thesis, the BDM methodology. Following section presents participants in the research process (cf. Section 3.1). Subsequent section (cf. Section 3.2) describes the research process in detail.

3.1 Presentation of Research Participants

Following table summarizes all participants and assigns them to specific research techniques they took part in (Table 3.1):

Participant	Initial survey	Survey Interview	Case study interview(s)	Final survey
Participant A	✓			✓
Participant B	✓			
Participant C	✓			
Participant D	✓		✓	✓
Participant E	✓	✓	✓	
Participant F	✓			
Participant G	✓			
Participant H	✓			

Participant	Initial survey	Survey Interview	Case study interview(s)	Final survey
Participant I	✓			
Participant J	✓		✓	
Participant K	✓			
Participant L		✓		
Participant M			✓	
Participant N			✓	
Participant O				✓
Participant P				✓
Participant R				✓

Table 3.1 – Research participants

Altogether, there were 17 participants engaged in the research process. We will present them more in detail in sections that discuss respective research technique they were involved in.

3.2 Action Research

Consulting pertinent literature is helpful for general and background information, but the literature will not provide much insight about specific conditions and situations (Berg, 2004). A solution to this problem offers action research. Action research is one of the few research approaches that embraces principles of participation, reflection, empowerment, and emancipation (Berg, 2004).

Action research is an appropriate research approach to be applied if we aim to achieve the following:

- A highly rigorous, yet reflective or interpretive, approach to empirical research
- The active engagement of individuals traditionally known as subjects, as participants and contributors in the research enterprise
- The integration of some practical outcomes related to the actual lives of participants in the research project
- A spiral of steps, each of which is composed of some type of planning, action and evaluation

Action research is intended to uncover or produce information and knowledge that will be directly useful to a group of people (Fals-Borda and Rahman, 1991 and Reason, 1994).

This approach endorses consensual, demographic, and participatory strategies to encourage people to examine reflectively their problems and particular issues (Berg, 2004).

The basic action research procedural routine involves four stages: identifying the research question (cf. Subsection 3.2.1), gathering the information to answer the question (cf. Subsection 3.2.2), analyzing and interpreting the information (cf. Subsection 3.2.3), and sharing the results with the participants (cf. Subsection 3.2.4).

3.2.1 Identifying the Research Question

The research process starts with brainstorming technique, literature analysis and discussions with zeb/ colleagues that work on financial DWH projects. The objective of these actions is to define a research question that once it is answered offers a solution applicable to financial data warehouse projects. The starting point is the project the author of the master thesis was involved in. The purpose of the project was to implement an enterprise data warehouse that would centralize all relevant business information. When she joined the project team, the project was at its beginning. After consultations with research participants, we researched potential topics and discussed their feasibility in the context of a master thesis with the tutor at zeb/.

Once we formulate this question, we need to research available literature in order to identify factors that influence the success of DWH projects. This literature concerns factors in general data warehousing and information systems fields as well as specific financial domain (Weir et al., 2003, Masiero, 2010, Wixom and Watson, 2001, Ojeda-Castro et al., 2011, Hwang et al., 2004 and Watson and Ariyachandra, 2005). This action leads to generation of a list of main factors, which influence the success of financial data warehousing projects (cf. Table 3.2).

Nr.	Factor
1.	Insufficient data consistency (validity, accuracy, usability and integrity of related data between applications and across an IT enterprise)
2.	Semantic data complexity (the understanding of the data takes a lot of time)
3.	Semantic data heterogeneity (the apparently identical piece of information has different meaning across data sources)
4.	Highly unstandardized data (no naming conventions applied, no proper data types used...)
5.	Insufficient data specification/definition
6.	Interpretation errors
7.	Data hard to understand (unclear or misleading data definitions)
8.	Incomplete data (the complete set of data needed was not acquired)
9.	Communication problems with the crucial actors
10.	Difficult relationship with clients
11.	Language gaps between designers and technical actors
12.	Insufficient concern from the client's side (causing their reduced availability)
13.	Missing support of top management
14.	Insufficient manpower allocation
15.	Problems in understanding roles and responsibilities
16.	Reduced availability of crucial actors

Table 3.3 specifies sources of factors from the Table 3.2:

3.2.2 Gathering the Information to Answer the Question

The survey method, if conceived in very general terms, is nothing but the general method of filling data matrices (Galtung, 1967). The survey is therefore the research technique we use in the

Nr.	Factor
17.	Difficulties in sharing the knowledge among requirements engineers
18.	Imperfections in building and maintaining mutual understanding
19.	Misinterpretations of the aim of the project
20.	Missing clear set of goals and priorities
21.	Unclear scope definition
22.	Complexity of the client's organizational structure
23.	Problems with the decentralized processes within the client's bank
24.	Absence of a clear methodology that could be followed
25.	Insufficiency of the followed methodology
26.	Problems with political awareness
27.	The skill set of crucial actors was either insufficient or not covering the actual needs
28.	Initial training in the topic was omitted
29.	The participants were looking for perfection
30.	Important design reality gaps
31.	Insufficient data quality

Table 3.2 – Factors influencing the success of financial DWH projects

Nr.	Weir et al., 2003	Masiero, 2010	Wixom and Watson, 2001	Ojeda-Castro et al., 2011	Hwang et al., 2004	Watson and Ariyachandra, 2005
1.					x	x
2.			x			
3.			x			
4.			x			
5.	x					
6.			x			
7.						x
8.						x
9.			x			
10.					x	
11.	x					
12.					x	
13.	x			x	x	
14.			x			
15.				x		
16.			x			
17.					x	

Nr.	Weir et al., 2003	Masiero, 2010	Wixom and Watson, 2001	Ojeda-Castro et al., 2011	Hwang et al., 2004	Watson and Ariyachandra, 2005
18.	x					
19.	x					
20.	x			x		
21.				x	x	
22.					x	
23.			x			
24.			x	x	x	
25.			x	x	x	
26.	x		x			
27.			x		x	
28.	x		x			
29.	x					
30.	x	x				x
31.	x		x	x		

Table 3.3 – Source of factors that influence the success of DWH projects

stage of gathering information. The initial survey covers in the context of the research process discovery function. We gather data from respondents in order to attempt to discover what might be responsible for some phenomenon or behavior (Judd et al., 1991). The role of this survey is to identify crucial factors influencing the success of financial data warehousing projects. This survey was sent to a convenience sample of 33 experts in financial DWH technologies. A senior manager that has been working at zeb/ for more than ten years and that is actively participating in financial data warehousing research provided the list of the experts. The majority of interviewed experts are managers and senior managers from zeb/. All of them had participated in more than one DWH project during their professional career. This list was extended by external actors.

Eleven participants answered to our survey. These eleven participants are all of them actively involved in DWH development of different projects at different banks/financial institutions. Their answers have therefore equal weight.

Participants involved in the initial survey are listed in the Table 3.4.

Participant	Description
Participant A	Senior Manager, IT unit, zeb/ employee since 1999
Participant B	Senior Manager, IT unit, zeb/ employee since 2002, actively participating in financial DWH research
Participant C	Senior Manager, IT unit, zeb/ employee since 2002
Participant D	Manager, Steering unit, zeb/ employee since 2011
Participant E	External consultant
Participant F	External consultant
Participant G	Senior Manager, IT unit, zeb/ employee since 2000
Participant H	Senior Manager, IT unit, zeb/ employee since 2002
Participant I	Manager, IT unit, zeb/ employee since 2004
Participant J	Manager, IT unit, zeb/ employee since 2006
Participant K	Ex-Manager, IT unit, zeb/ employee from 2007-2013
Participant L	Manager, IT unit, zeb/ employee since 2005

Table 3.4 – Participants involved in the initial survey

These participants have been involved in DWH projects in banks and financial institutions listed in Appendix C.

Figure 3.1 presents organizational chart of participants from zeb/ that are involved in the initial survey. As we can see, there are only two interdependencies among these participants. Participant A and participant C share the same partner and participant I belongs to the participant's B unit. Neither these participants nor the others base their answers on the same projects. We can therefore say that this sample of initial survey participants is heterogeneous.

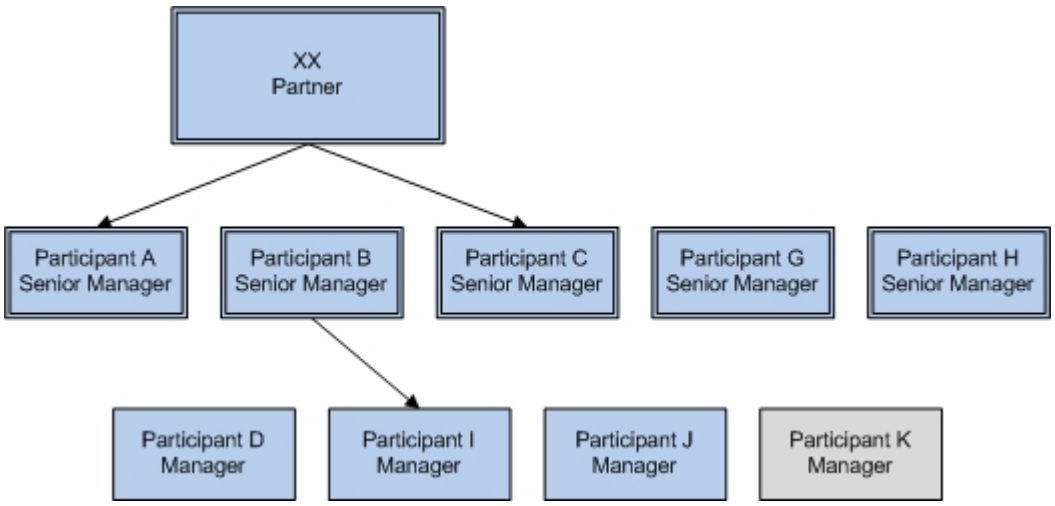

Figure 3.1 – Organization chart of zeb/ initial survey participants

The initial survey contains general questions on financial data warehouse projects. Particular importance is given to problems faced during their development. The goal of the survey is to identify crucial factors that influence the success of financial DWH projects. Results of the survey serve as the prerequisite for the following stage in the research process (cf. Subsection 3.2.3).

First version of the survey (cf. Appendix D) required filling in complex information on these factors (the project phase where the factor occurs, success of which phase it influences, or what corrective or preventive action can be implemented in order to reduce its impact). It focus particularly on problems faced in the development of financial DWH projects. This version of the survey aims to identify actions that either have been taken in order to correct issues in respective project's context or that might help to prevent these problems from occurring. These techniques can either be the result of actions that has been taken during the project development or actions that participants personally suggest based on their professional experience. In addition, participants were asked to identify the five most significant factors and order them according to the importance of the impact they have on the success of financial DWH projects.

The initial version was however too complex and required significant effort from participants. Consequently we simplified the survey. In this simplified version, participants had to choose five, in their opinion most influencing factors and mark the order of their importance (cf. Appendix E). Since we changed neither the formulation of factors nor the ordering task of the five most influencing factors, we were able to include results provided to the first version of the survey in the next research stage as well. The result of the initial survey is a list of factors

ordered according to the importance of the influence they have on the success of financial data warehouse projects.

3.2.3 Analyzing and Interpreting the Information

Semi-structured interviews and case study help us to conduct this stage of the research process. Semi-structured or open-ended interview is a formal unstructured research technique that leads to verbal responses (Galtung, 1967). The objective of these interviews is to get additional information from survey participants concerning factors they identified as the most influencing and to collect information within the case study.

Some participants of the initial survey took part in an additional interview session (cf. Table 3.1; a list of the survey participants can be found in Table 3.4). Due to availability issues, only one participant was interviewed. This interview concentrated on six, in participant's E opinion, most important factors. The set of survey interview questions can be found in Appendix F. In addition, participant L was interviewed. This one does not provide answers to the survey, but prefers to participate uniquely in the interview part. This interview discusses crucial factors identified after analysis of the initial survey. The set of survey interview questions can be found in Appendix G.

The purpose of a survey interview is to get additional information on crucial factors. Important in these interviews is to set a mutual understanding of these factors right from the beginning. Therefore, a definition is proposed for every discussed factor, so that this one can be argued. These factors are afterwards examined with the help of participants. The objective of this exchange is to reveal preventive actions that can be introduced in order to reduce the impact of respective factors to the minimum. A template of such an interview protocol can be found in Appendix B.

The next step in this stage of the research process is the case study. Case studies are preferred strategies when the investigator has little control of events, and when the focus is on a contemporary phenomenon within some real-life context (Yin, 1994). The essence of a case study is that it tries to illuminate a decision or set of decisions: why they were taken, how they were implemented and with what result (Schramm, 1971). The purpose of the case study is on one hand to verify the correctness of results gained in the previous research stage (cf. Subsection 3.2.2). It covers therefore the demonstration research function (Judd et al., 1991). On the other hand, its role is to analyze the business data modeling phase of a data warehouse project at a big Austrian bank (henceforth BANK) focusing on corrective and preventive actions (taken or recommended) that reduce the impact of crucial factors that we identified during the initial survey (discovery research function according to Judd et al., 1991). We aim to discover how project members achieved completeness and correctness of the available data, how was the product scope defined and what methodology was followed during the project development. A "how to" conduct a financial data warehouse project based on experience collected during the "Business Data Model Project" at BANK is derived. The project started in 2009 and was completed in 2013. The Business Data Model was released in 2011. Necessary information to assess this project was provided by people that were actively involved in the project development. Particularly the business requirements, data collection, and business data modeling were analyzed.

The objective of analysis of subsequent project phases is to discover strengths and weaknesses, which the members of respective teams identified in the released BDM.

Table 3.5 summarizes participants of the case study with the description of the role they had within the studied project:

Reference	Position
Participant D	Member of the business requirements and data requirements collection/definition phase
Participant E	External consultant, member of the business data modeling team
Participant J	Participant on a preceding DWH project at the same bank
Participant M	Member of the business data modeling team
Participant N	Project manager of the mapping and implementation phase (the one that follows the business data modeling)

Table 3.5 – Participants of the case study

The outcome of the case study is an analysis of the BDM phase of a DWH project at the BANK with the focus on the crucial factors and solutions implemented in order to reduce the impact of these factors.

Two preliminary introduction meetings were conducted before the case study started in order to get an introduction to the studied project:

- *Initial Meeting* – General project Overview given by Participant J
- *Initial Documentation explanations Meeting* – Provided by Participant D

All conducted interviews are documented, recorded and protocoled in a separate document. The outcome is summarized in the Chapter 4.

Figure 3.2 summarizes the time plan of the case study.

Nr.	Activity	Start	End	Duration
1	Preparation	10.06.2013	14.06.2013	1w
2	1st session of interviews	17.06.2013	28.06.2013	2w
3	Break	01.07.2013	05.07.2013	1w
4	2nd session of interviews	08.07.2013	19.07.2013	2w
5	Break	22.07.2013	26.07.2013	1w
6	3nd session of interviews	29.07.2013	09.08.2013	2w
7	Finalization	12.08.2013	16.08.2013	1w

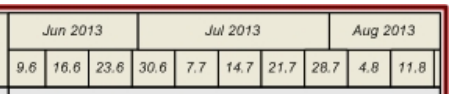

Figure 3.2 – Case study time schedule

Preparation: One week period of time when interviews with case study participants are scheduled.

Session of interviews: Two weeks period of time when interviews are conducted. There are two sessions in total (initially three of them were planned, but due to time constraints, the last one did not take place).

Break: One week period when all collected information/documentation is verified, protocoled and analyzed. The break week serves for the following interview scheduling as well.

Finalization: One week period, where the case study is finalized and its outcomes are derived.

In the following section, we describe the procedure of the first and second interview session, respectively.

For every case study interview, the author prepared a set of questions as well as a set of backup questions. A template of a document including case study interview questions is provided in Appendix A. Each interview starts with setting of the interview session objectives. An exchange of clear target questions and responses follows. At the end of all interviews, there is a time devoted for an open discussion. The interviewer initiates this discussion by asking participants if they feel some elements were omitted or not stated clearly. All interviews are recorded and protocoled.

The first session of interviews aims to present participants and to get general overview of the project altogether with identifying its main strengths and weaknesses. Study of the project setup was part of backup questions. Participants, interview dates as well as links to corresponding question protocols are summarized in Figure 3.6.

Reference	**Interview Date**	**Case study Interview questions**
Participant D	27th of June 2013	Appendix H
Participant E	2nd of July 2013	Appendix J
Participant J	26th of June 2013	Appendix H
Participant M	1st of July 2013	Appendix I

Table 3.6 – First session of case study research

Questions are almost identical for all four participants. There are slight changes in the set of target questions for the Participant M. Since he was responsible for tasks defined within the business data modeling phase, he was able to provide additional information on the working model of this project phase. Apart from that, it is mainly in this project phase, where elements that were initially classified as out of scope might need to be reintroduced. Due to this fact, the participant M is challenged to suggest a solution that could minimize the obligation to change the product scope, especially during the business data modeling phase. Questions that differ from general interview questions are highlighted directly in the document in yellow (see Appendix I).

In case of the Participant E, an important part of questions differed from the ones that we asked other participants. On account of the fact that he is a consultant from an external company that develops the DWH solution that was in the end implemented (henceforth SOLUTION), he is able to answer specific questions concerning the training that was introduced in order to teach participants to use this solution. In addition, his company uses the studied project as a reference for the business modeling. In the context of a presentation that he gave at an international conference, among others, lessons learned were presented. We discussed them with the Participant E because they represent an important source of preventive actions that we will consider implementing in our final methodology. He as well provides answers concerning the workflow of the project before and after the business data modeling part. In order to differentiate aforementioned questions, they are highlighted directly in the document in yellow (cf. Appendix J).

There is a slight time delay in the first interview session. Last two interviews are conducted during the break weak (cf. Figure 3.2). Reasons of this delay were time constraints and availability conflicts between the interviewer and interviewees.

The second session of interviews aims to examine models and methodologies followed during project development (e.g. scrum, waterfall model) altogether with data quality issues and Business–IT gap.

Figure 3.7 summarizes participants of this second session:

Reference	Interview Date	Case study Interview questions
Participant D	16th of July 2013 2013	Appendix L
Participant E	18th of July 2013 2013	Appendix O
Participant J	17th of July 2013 2013	Appendix N
Participant M	17th of July 2013 2013	Appendix M
Participant N	10th of July 2013 2013	Appendix K

Table 3.7 – Second session of case study research

In this second session, interview questions differ from one participant to another. Participant N can identify main imperfections of the business data model released in 2011, because he was the project manager of project phases following the business data modeling. As project manager, he can also answer questions concerning project management.

Participant D can provide detailed information concerning the business requirements and data collection phase.

Both, Participant M and Participant E can provide additional information on the data specifications issues within the business data modeling as well as exhaustively describe the implemented model and explain actions taken step by step.

Concerning Participant J, his role is to provide background of the project and describe data quality issues more in detail (but from the perspective of the previous project). Even when he participated in a DWH project that preceded the one that is the object of our case study, his contribution is very valuable. In both projects, the one we are studying and the one that took place just before this one, data quality management was introduced only at the end. Therefore all the other participants are not able to provide satisfactory answer to data quality questions, since this one was still ongoing at the time when we conducted the interviews (June/July 2013).

3.2.4 Sharing the Results with the Participants

The final stage of the research process aims to share results with participants. Since we were already facing availability issues and time conflicts between research participants in previous research phases, we needed to select a research technique that is not too time consuming and that's complexity would not discourage the participants. The research technique we select in this research stage is therefore a survey. By implementing this specific research technique, we can not only present results to the participants, but also conduct results evaluation. A survey is a successful research technique because it provides theoretically relevant data that are amenable to statistical treatment (Galtung, 1967).

Participants involved in the final survey are listed in the Table 3.8.

Participant	**Description**
Participant A	Senior Manager, IT unit, zeb/ employee since 1999
Participant B	Senior Manager, IT unit, zeb/ employee since 2002, actively participating in financial DWH research
Participant C	Senior Manager, IT unit, zeb/ employee since 2002
Participant D	Manager, Steering unit, zeb/ employee since 2011
Participant E	External consultant
Participant J	Manager, IT unit, zeb/ employee since 2006
Participant L	Manager, IT unit, zeb/ employee since 2005
Participant M	Manager, IT unit, zeb/ employee since 2007
Participant O	Partner, IT unit, zeb/ employee since 2002
Participant P	Senior Manager, IT unit, zeb/ employee since 2006
Participant R	Senior Manager, IT unit, zeb/ employee since 2009

Table 3.8 – Final survey participants

Using the Likert scale, the participants were asked to provide their opinion on five statements. Their answers allows us to assess the success of our methodology. Since we aimed to reduce the impact of the crucial factors we identified in the present thesis, we asked participants to which extent they think the methodology we presented in the final survey can influence the impact, which these factors have on the success of DWH projects (cf. Appendix P).

The whole research process is summarized on the timeline in Figure 3.3.

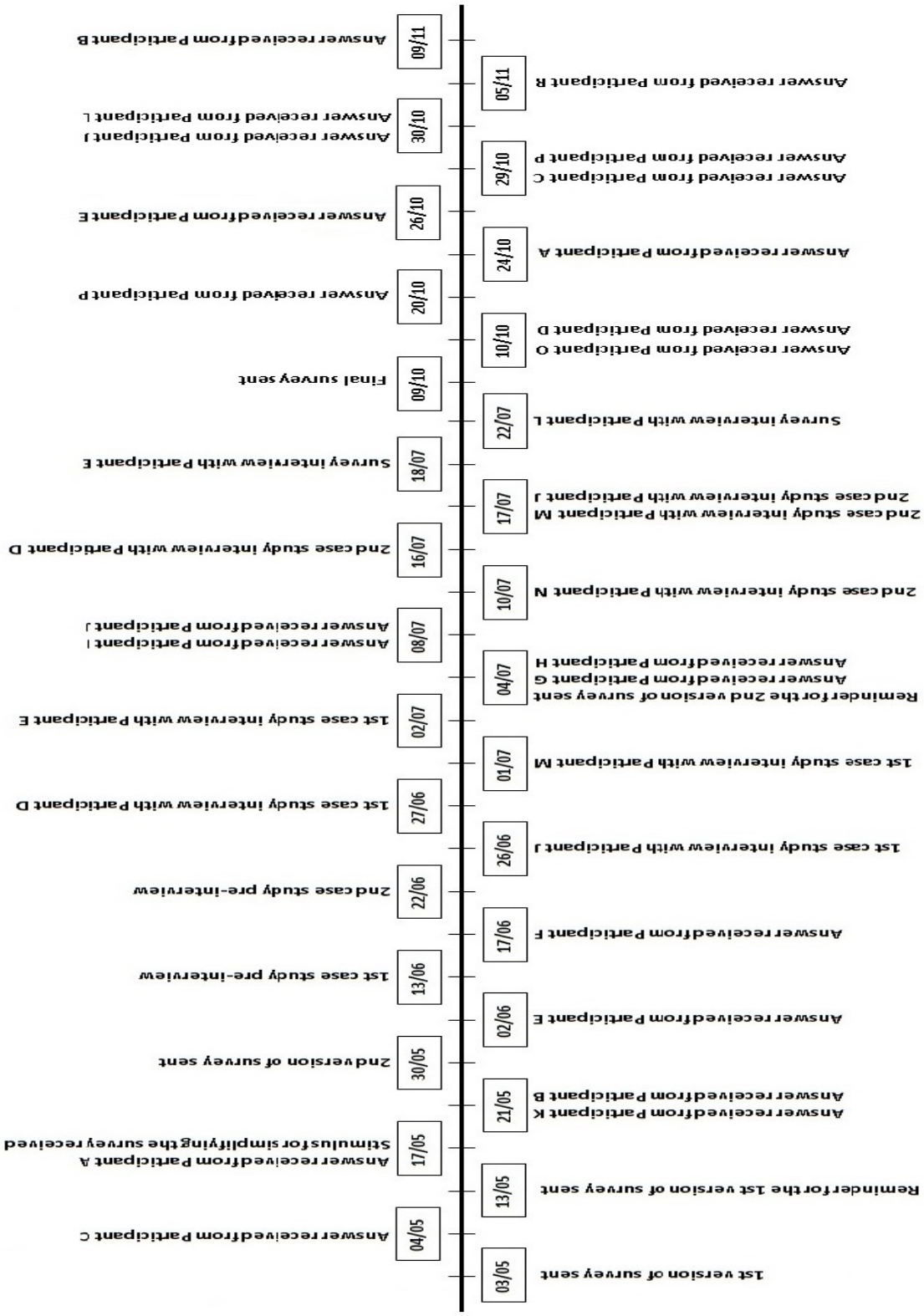

Figure 3.3 – Research process timeline

CHAPTER 4

Results

In this chapter, we concentrate on the description of collected, processed and analyzed results coming from research techniques introduced in Chapter 3. Information that is described in this chapter results from conducted research (e.g. introduced figures were either provided by research participants themselves or they have been created based on topics discussed within conducted interviews).

Crucial factors identified during the second stage of the research (cf. Subsection 3.2.2) are being discussed in Section 4.1. Subsections of this section offer a detailed view on the crucial factors. These subsections present in addition to the information gained during the initial survey results of supplementary interviews with initial survey participants. In section 4.2 an overview of the explored project is given. Furthermore the crucial factors are analyzed regarding their possible impact on the project success (third stage of the research described in Subsection 3.2.3).

4.1 Crucial Factors Influencing the success of DWH projects

Objective of the analysis performed during the initial survey (cf. Appendix E) was to identify the factors, which developers usually face in financial DWH projects. The survey represents the second stage of the research process, entitled "Gathering the information to answer the question" (cf. Subsection 3.2.2).

After completion of the initial survey, answers from participants are processed. All participants provided their answers based on different projects executed in different banks. The same weight is therefore assigned to all answers.

Analyzed factors have to be ordered according to the participant answers to the initial survey. The formula applied to assign a score to a factor is:

$$score_{factor} = \sum_{i=1}^{n_{factor}} 1/order_i$$

n_{factor} corresponds to the number of times the factor appears in participants ordering.

The score of a factor is the sum of multiplicative inverses of all order values that participants assigned to the factor.

The following example helps us to understand the scoring task.

Label	Factor
A	Insufficient manpower allocation
B	Problems in understanding roles and responsibilities
C	Reduced availability of crucial actors
D	Difficulties in sharing the knowledge among requirements engineers
E	Imperfections in building and maintaining mutual understanding
F	Misinterpretations of the aim of the project
G	Missing clear set of goals and priorities

Table 4.1 – Factors to be ordered - illustrative example

In order to score the 7 factors that are listed in Table 4.1, we posses 2 answer sheets (cf. Figure 4.1) based on which the scoring proceeds.

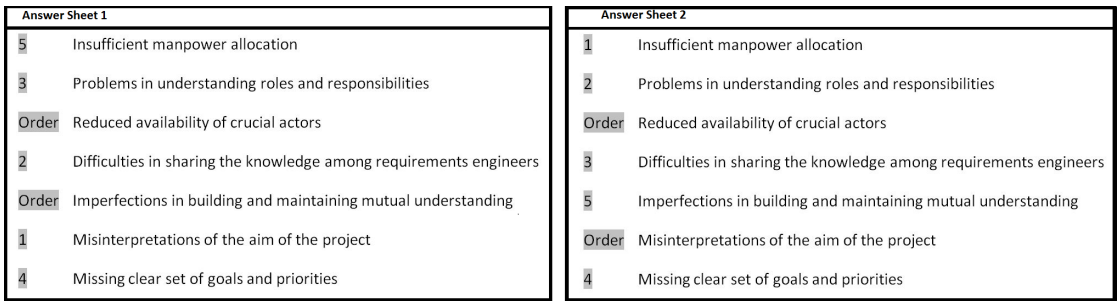

Figure 4.1 – Example of answer sheets

Scores are calculated as follows:

$$n_A = 2$$

$$score_A = \sum_{i=1}^{2} 1/order_i = 1/5 + 1/1 = 1.2$$

$$n_B = 2$$

$$score_B = \sum_{i=1}^{2} 1/order_i = 1/3 + 1/2 = 0.83$$

$$n_C = 0$$

$$score_C = \sum_{i=1}^{0} 1/order_i = 0$$

$$n_D = 2$$

$$score_D = \sum_{i=1}^{2} 1/order_i = 1/2 + 1/3 = 0.83$$

$$n_E = 2$$

$$score_E = \sum_{i=1}^{1} 1/order_i = 1/5 = 0.2$$

$$n_F = 1$$

$$score_F = \sum_{i=1}^{1} 1/order_i = 1/1 = 1$$

$$n_G = 2$$

$$score_G = \sum_{i=1}^{2} 1/order_i = 1/4 + 1/4 = 0.5$$

This scoring leads to following ordering (cf. Table 4.2):

Order	**Factor**
1.	Insufficient manpower allocation
2.	Misinterpretations of the aim of the project
3.	Problems in understanding roles and responsibilities
3.	Difficulties in sharing the knowledge among requirements engineers
5.	Missing clear set of goals and priorities
6.	Imperfections in building and maintaining mutual understanding
7.	Reduced availability of crucial actors

Table 4.2 – Result of the scoring example

Table 4.3 lists the ordering of factors after score allocation.

Nr.	Original number	Order	Factor
1.	13.	1	Missing support of top management
2.	21.	2	Unclear scope definition
3.	20.	3	Missing clear set of goals and priorities
4.	5.	4	Insufficient data specification/definition
5.	31.	5	Insufficient data quality
6.	1.	6	Insufficient data consistency
7.	27.	7	The skill set of crucial actors was either insufficient or not covering the actual needs
8.	6.	8	Interpretation errors
9.	11.	8	Language gaps between designers and technical actors
10.	7.	10	Data hard to understand
11.	19.	10	Misinterpretations of the aim of the project
12.	22.	10	Complexity of the client's organizational structure
13.	9.	13	Communication problems with the crucial actors
14.	3.	14	Semantic data heterogeneity
15.	15.	14	Problems in understanding roles and responsibilities
16.	8.	16	Incomplete data
17.	18.	16	Imperfections in building and maintaining mutual understanding
18.	23.	16	Problems with the decentralized processes within the client's bank
19.	25.	16	Insufficiency of the followed methodology
20.	4.	20	Highly unstandardized data
21.	30.	21	Important design reality gaps
22.	2.	22	Semantic data complexity
23.	10.	22	Difficult relationship with clients
24.	12.	22	Insufficient concern from the client's side
25.	14.	22	Insufficient manpower allocation
26.	16.	22	Reduced availability of crucial actors
27.	17.	22	Difficulties in sharing the knowledge among requirements engineers
28.	24.	22	Absence of a clear methodology that could be followed
29.	26.	22	Problems with political awareness
30.	28.	22	Initial training in the topic was omitted
31.	29.	22	The participants were looking for perfection

Table 4.3 – Factors ordered according to initial survey results

In Table 4.3 the major success factors are highlighted in green. The remaining ones can be considered as too general and will not be discussed further.

Figure 4.2 illustrates dependencies between these factors. Key success factors have light orange background. Factors they influence directly have light blue background and factors they influence indirectly have violet background. Factors having grey background have been by survey participants identified as least important (cf. factors having order 22 in Table 4.3)

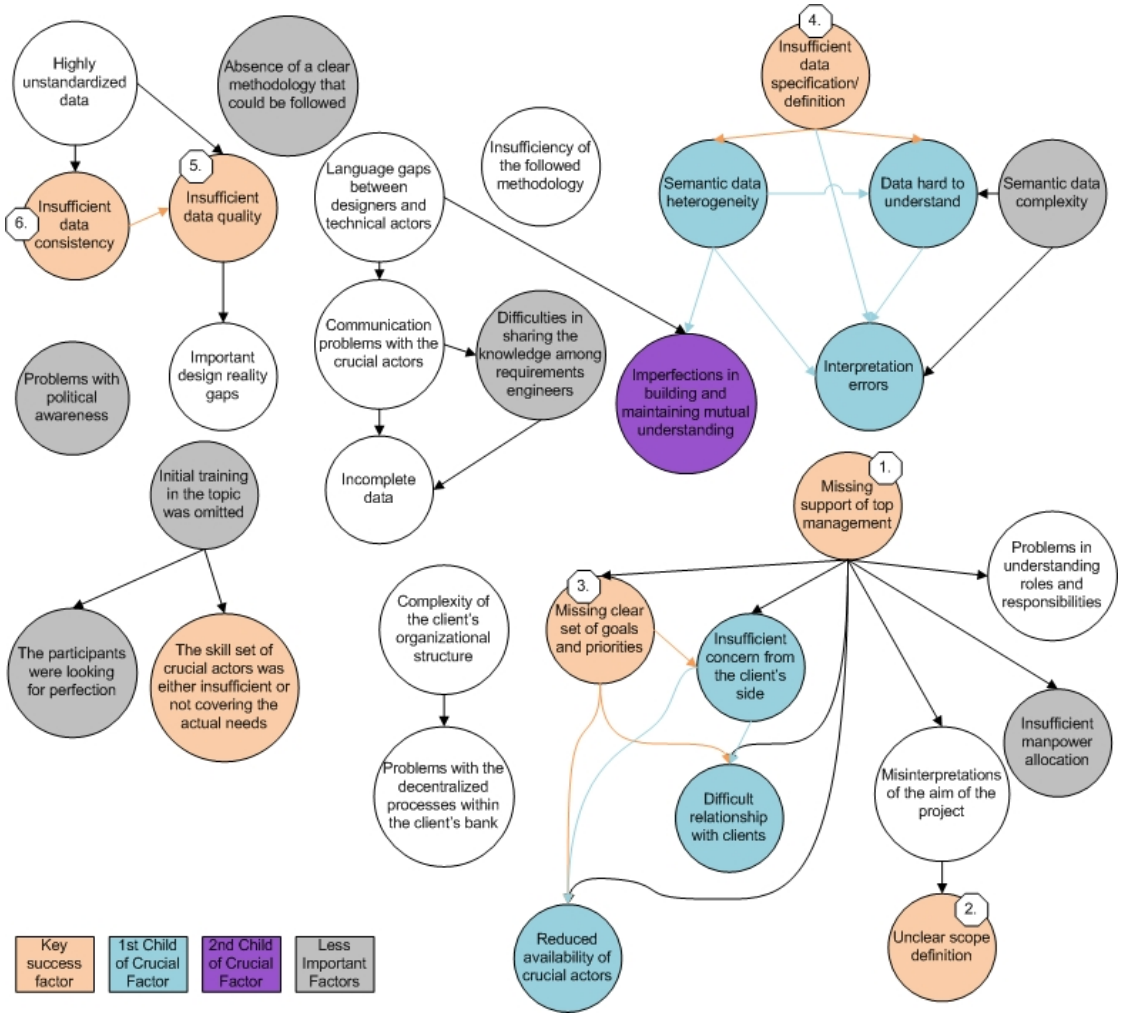

Figure 4.2 – Direct dependencies between factors

From the six highlighted key success factors, furthermore those success factors which encompass a significant number of sub-factors will be excluded as their analysis would violate time resource constraints. This action affects the factor number 5, “Insufficient data quality”, and factor number 1, “Missing support of top management”.

As a result, we get the final four key success factors (cf. Table 4.4).

Nr.	Order	Crucial factor
1.	6	Insufficient data consistency
2.	3	Missing clear set of goals and priorities
3.	2	Unclear scope definition
4.	4	Insufficient data specification/definition

Table 4.4 – Crucial factors

Within the following sections we will discuss these four factors. Result we will present are mainly the result of the third research stage, entitled "Analyzing and interpreting the information" (cf Subsection 3.2.3).

4.1.1 Data Consistency

As data consistency is part of the data quality issue, we discuss in this subsection the data consistency topic in the context of data quality itself.

Data quality should start to be acknowledged as early as possible. The definition of "as early as possible" is however ambiguous. What usually happens is that project members claim at the beginning of the project that this issue is very important. Nevertheless, a lot of project managers still do not consider it in an appropriate manner. In many cases data quality management does not start until the very end of the project. When the data quality rules are implemented at a later stage, it takes a lot more effort to analyze an existence and analyze the reason of a defect. If a bank already owns a certain product, has implemented an ETL process and has a DWH in place, or its data is being used by stakeholders and their employees identify inaccurate data quality at this stage, the corrupted data should be changed in the core-system. Consequently, the process of entering the data needs to be changed accordingly. The later the stage they observe such a defect, the more costly its correction. Moreover, not correcting this defect can have (according to its severity) tremendous impact on the final outcome. Not taking care of data quality and consistency could lead to lack of acceptance on user side.

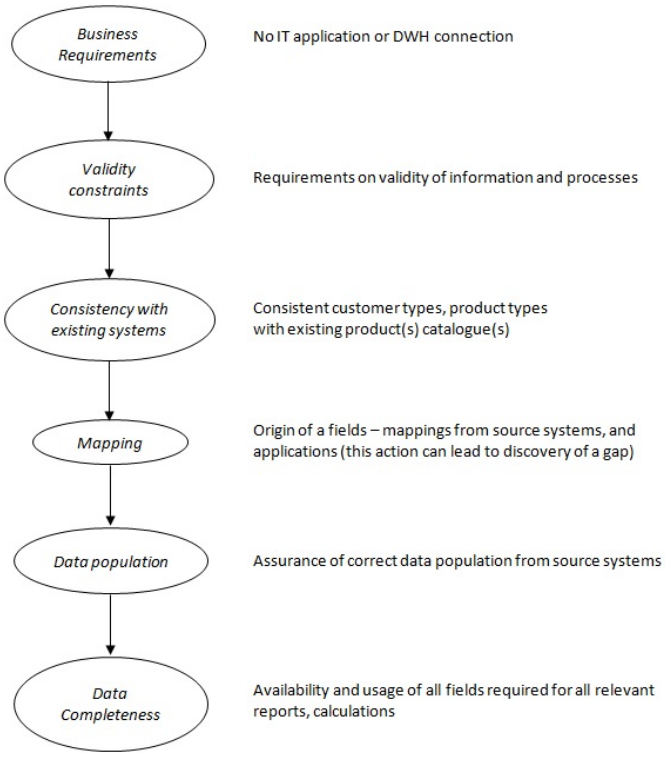

Figure 4.3 – Data supply chain in a bank

In order to understand the complexity of implementing consistency checks at later project development stage, we present a general supply chain in a bank (cf. Figure 4.3). Within a bank, procedures of data processing where data quality has to be assured start by defining business requirements. They represent basic business functions, which the bank needs to cover. They are however independent from all applications, source systems or a data warehouse. After the business requirements are defined, validity constraints requirements have to be specified. Validity checks verify for example required periodicity of verification of the information that runs within the bank in order to satisfy business requirements. Next step is to verify the consistency with existing systems in the bank. This verification contains for example the activity of checking if target entities have real support in existing systems. Once these checks have been executed, the mapping of required data can start. This action helps to identify potential gaps between required data and data located in current source systems of the bank. If a gap is discovered an appropriate strategy has to be implemented in order to cover this gap. When mapping checks are finished, last check in the supply chain follows. All data required for creating necessary reports has to be considered within the DWH product.

In order to prevent costly correction of data at a later stage, data quality rules should be defined from the beginning of the project and data quality workflow should be controlled. At this point, data quality requirements should be defined together with business requirements. Data quality checks are to be applied at different check points. Starting point is within source

systems that should be checked against data quality as well. The easiest way to make this is to implement an adequate tool. This tool allows in addition an automatic or semi-automatic execution of consistency rules. Having such a tool in place right at the beginning could help us to a certain extent. It is however preferable, to implement an in-database tool. Such a tool represents a considerable advantage for the bank.

In order to achieve required data quality, data semantics have to be taken into account as well. These topic will be discussed when we will define factors from the category "Issues Met while Building a Common Language".

4.1.2 Missing Clear Set of Goals and Priorities on the Top Management Level

This item is a subset of the category "Top Management Issues". In order to better understand this topic, we will position it in larger context by explaining the importance of top management itself.

Top management support is understood as the action of task setting, with a clear message behind them and with allocating clear time and human resources. Moreover, these actions need to include the emphasis and regular verification of the prioritization of generated tasks.

This factor is critical and needs a particular attention even before the project set-up. In a bank, there are usually several ongoing initiatives/projects/programs at the same time. Clear prioritization messages from the top management are therefore very important. For example if is an employee already involved in too many ongoing projects/tasks, he will automatically enter in an availability conflict when a requirement of execution of an additional task arrives.

The priority has to be set in a comprehensible matter (not just indicate projects that are "very important") otherwise resources will not be available.

Frequently top management does not take all the responsibility and leaves project prioritization "open". Such a message causes conflicted understanding of the project prioritization. Involved parties interpret this message according to their internal needs. If top management escalates prioritization task to their Board-1/Board-2 employees, very often resource availability conflicts appear. If the priority is not set at the top level, every department will follow their own interests and will not take into consideration needs of other departments. Due to conflicts on the board level, at some point, projects might not go forward anymore.

The following example illustrates the problem. There are two projects: one of implementing a risk data warehouse and another one of changing the core-banking system. Top management says that these two projects are both very important to the bank. They don't say for example: "We need to have the DWH implemented by 2014, but the core banking system is more crucial, so we will force the core banking part and schedule the data warehouse afterwards." Instead, they emit a misleading signal by saying that both projects are very important. Then if the risk board wants to have a data warehouse and financial board neglects its importance, a conflict on board level can occur and the risk board might not give any support anymore.

Setting a clear set of goals and objectives is mainly a top-down issue. When the top management sets priorities, their members communicate them to second line managers. Consequently, they need to release respective resources. In order for this to be effective, they have to emit a clear message: "*This* is important to us, because... So please provide *this* by doing *that*."

Depending on objectives one wants to achieve, project structure and the staffing has to be shaped accordingly. If the staff does not know what is to be achieved, they might go in the wrong direction. Top management's crucial role is to provide a clear picture on the target to be achieved. They have to assure the execution of management topics by defining smart targets, i.e. specific and measurable. They additionally need to manage bank's resources.

In order to prevent misleading prioritization within the organization, one central unit should be put in place. This unit takes care of projects that are scheduled and sets their priority. Furthermore, this unit should align all its activities with the board and then assign resources, budget and time. Its task is to analyze and identify conflicts (time/budget/resources/priority). In case a conflict is identified, they propose an adequate solution. They suggest techniques that have to be implemented in order to correct this conflict and make the project work according to original schedule. Members of this unit might need to reopen tasks or potentially even re-prioritize projects. In all cases, they have to provide the support necessary to execute these actions in the context of respective projects/departments. Additionally, if they observe that projects and involved departments have different priorities than the ones the top management allocated, then a board member message is needed to emphasize the priorities. This happens also if employees do not continue to contribute to a necessary extent. The central unit does not necessarily have to look for these conflicts continuously, but reminders have to be set regularly.

All project managers should assure that their projects follow the outlined schedule and use allocated resources. Implementation of a tool that simplifies these tasks should create the following added value:

- Workflow support
- Awareness – board knows about ongoing projects and can create corresponding reports
- Project leaders know what the status is and how much budget they have left
- Project management has an overview of waiting task queue for a given project (indicates on one hand why a specific task cannot start or on the other hand, when it can start)
- Stabilized communication between different stakeholders
- All communication flow is documented and archived
- Information and activities are centralized and easily accessed

4.1.3 Unclear Scope Definition

This factor belongs to the category of project management issues. We will start the description of this factor by positioning it into a wider context. This factor is one of the most crucial and not easy to be addressed.

Proper tools and templates need to support management. They are not needed for the project management itself, but are needed for the Project Management Offices (PMO).

The scope of the product needs to have clear boundaries which have an unambiguous understanding among all project participants. It has to be defined so that the need of reshaping

the scope is minimized. The scope additionally includes roles and responsibilities and allocates enough decision power to respective project members so that if the top management is not present, they still can make decisions if they need to.

The bank usually has certain expectations, but it is important to differentiate mandatory (can be achieved within a given time frame, budget and resources) from ”nice to have” expectations. The scope has to be agreed on by the key persons: business leaders and respective business experts. The following points have to be clarified:

- Clarify the scope - what are the major expectations, what are the respective goals
- Make sure that it is possible within a given time and budget to fulfill these expectations; if not, try to reduce the scope together with bank's representatives
- Sometimes the bank does not provide a clear scope (in this case, the project is a demonstration of “learning by doing”)

Before a DWH implementation, scoping questions have to be addressed as precisely as possible. Inmon himself emphasizes the importance of this task. In his implementation, growing scope and changing requirements are critical (cf. Table 2.1). The scope has to cover all business needs within bank business areas. In order to define what business areas the bank covers, it is recommended to use examples from other projects. Likewise, past experience helps to estimate the time needed for the completion of business requirements definition of one business area. In addition, interdependencies between business areas can be analyzed in order to identify supplementary elements that should be included in the product scope. It is very important to define the scope as soon as possible so that we avoid scope changes during the project development. Therefore, the scope has to be as realistic as possible.

The proposed DWH architecture has to fit in the IT landscape of the bank. In order to get an overview of this appropriateness, “as is” analysis is usually conducted. This analysis aims to discover and estimate what IT applications are present at the bank and how they look like. Elements that are to be included in the business data model are checked against the IT architecture. This activity helps to verify if data that is needed has already been covered in current systems or if there is a gap. In case a lot of gaps are discovered, a more precise gap analysis and evaluation process has to be executed in order to get as narrow scope as possible (and as precise as possible).

The scope should be realistic and manageable without any (or as little as possible) budget overrun. Deadlines and project plan has to be effective and efficient (one has to assure that project members are doing the right things and that they are doing these things right). Apart from that it has to be possible to deliver result within this scope in the required quality.

4.1.4 Insufficient Data Specification/Definition

Interpretation errors are the consequence of insufficient data specifications/definitions. Insufficiently defined data leads to flexibility of interpretation that increases the probability of interpretation errors.

This element can be understood as the semantical as well as conceptual part of data quality. In the context of the present master thesis it is however seen more as a component of a common language building within the bank and/or among business and technical actors. The reason why this topic is interconnected with the data quality is because establishing a common language could serve as a basis for validating data in terms of completeness, correctness and accuracy.

This factor represents the main topic of business data modeling. Initial motivation for establishing such a concept is an existence of possibly several subsidiaries and several involved business lines within the head office of the bank. This concept gets even more important in case several different operational systems have to be considered as possible data sources. Respective business areas have to be therefore consolidated.

Imagine a case where the data collected or specified in the context of either business requirements collection or business data modeling is not precise enough. It leads to increased flexibility of interpretation and therefore to possible implementation errors. If one needs to integrate all data from fragmented heterogeneous environments, a common language has to be established in order to overcome this ambiguous situation. Additionally processes within a bank have to be business-driven but supported by IT (IT-driven approaches are very often not accepted by business). IT is not a self purpose but it is here to serve and support business.

Achieving good quality of data/requirements specification can be realized by business data modeling conducted via workshops where all departments as well as IT are present. By strongly connecting business and IT and ensuring a strong cooperation, acceptance by both sides can be significantly increased. In case of the BANK, this helped to bring together Risk and Controlling world.

The described crucial factors have to be taken into account in development of financial data warehouses. The following chapter will discuss these factors from a perspective of the project in the analyzed bank (result of the third stage of our research process described in the Subsection 3.2.3).

4.2 Business Data Modeling in Practice at a Big Bank in Austria

This chapter examines the business data modeling part of a project that took place at the BANK. It puts it in the context specific to this project and analyzes its strengths and weaknesses. This chapter identifies on one hand issues inherited from previous project phases such as business requirements definition and data requirements collection. On the other hand, it tries to discover defects of the business data modeling, that were either observed in this phase itself or in following project development phases.

This section sums up results of the case study conducted in order to satisfy demonstrative and discovery research function (Judd et al., 1991) within the third stage of the research process, “Analyzing and interpreting the information” (cf. Subsection 3.2.3). The purpose of the present case study is to verify that the same factors as the ones listed in Table 4.4 occurred in this project (demonstrative research function according to Judd et al., 1991). In addition, we aim to discover preventive and corrective actions that can be implemented so that the impact of these factors can be reduced (discovery research function according to Judd et al., 1991).

The aim of the case study is to derive a "how to" conduct business data modeling in the context of a financial data warehouse project based on experience collected during the case study of the Business Data Model Project at the BANK.

4.2.1 Project Overview

The studied project was initiated by a need of unifying the risk and controlling world of the BANK.

Its main purpose was to harmonize in the Head Office, business requirements and definitions from different business departments – mainly Risk Management and Controlling. The purpose of the project was to define "what" data has to be integrated in the future Enterprise Data Ware-House (EDWH).

The final "product" of the project is one common Business Data Model enabling consistent reporting across departments out of one single data source.

The Business Data Model project assures the following requirements:

- Ensure availability of proper descriptions for all characteristics (attributes) of business entities
- Deliver characteristics of business entities that are agreed with Risk and Finance
- Maintain business entities and their attributes documentation and ensure smooth inclusion of new business requirements
- Decide on offered data source (in case of different source options)
- Define business entity model
- Define clear general rules for characteristics of business entities modeling
- Prepare and conduct training for business entities and for the design of their characteristics
- Validate existing data quality concept
- Set up data governance and data quality by defining business (functional) data quality requirements

4.2.2 General Project Background

Project success is interpreted differently by different project members. Even though in general the project was considered successful, it required a lot of changes in scope, budget, and time. The project was too complex, scope was a moving target and too many underestimations accompanied its evolution. The project needed to be rescheduled, because its development took more time than expected. Additional capacity was then included in order to keep the new schedule. These adjustments are however partially explained by the enlarged product scope in comparison to the initial one. The budget was overrun approximately by additional 40% from the initial value. This overrun is justified by the scope extension and is internally considered valid. Even

if the project is supposed to be in production in January 2014, it might take another half a year until this is the case. This project is however an important reference for zeb/ as well as for the external consulting company that provided the solution for the implemented DWH. This company gave a presentation on the business modeling strategy introduced in the studied project at a conference in Dublin in 2012 and another one is planned for this year (2013) in Dallas. This presentation is entitled *"Breaking up business silos by business data modeling"*.

4.2.3 Scope Evolution

This project was initiated because the BANK wanted to replace its current head office solution with a new one. This new solution is based on the new Enterprise Data Warehouse (EDWH), which stores all relevant risk and finance data for the BANK. The development of the old DWH (GDWH) was driven by Basel II and the data model was therefore heavily risk-driven. The new EDWH is driven by internal business needs and regulatory demands in several areas. This business data model allowed harmonizing requirements from all main steering functions including risk, controlling, accounting and regulatory. In order to build the EDWH, a vendor solution was implemented.

The scope in the EDWH project was defined by the bank. According to participants, its definition was unambiguous for them throughout the project development. It included primarily Risk and Controlling requirements. The BANK implemented one analytical application for controlling and one for risk management purposes. The EDWH contains therefore data which these two applications need for their execution.

Figure 4.4 – Initial assumption of the scope of the EDWH project

However, the scope was a moving target. At the beginning there was the initiative to implement a head office Controlling solution (cf. Figure 4.4). Business requirements needed for its correct execution were therefore collected. Everything was handled within the context of this application. Controlling analytical application was supposed to be fed directly from source systems (Loan, Securities, Derivatives, Payments…). Risk analytical application was supposed to be fed by GDWH. The BANK afterwards decided to replace the GDWH by a new DWH

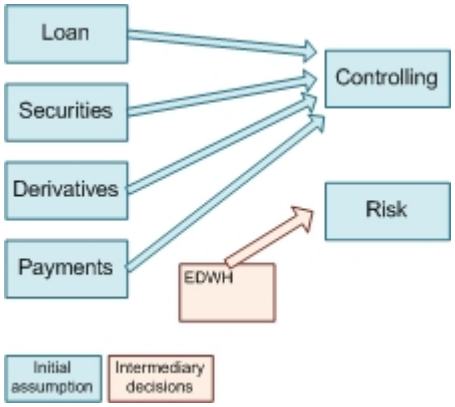

Figure 4.5 – Intermediary scope of the EDWH project

solution that was supposed to feed the "Risk" application (cf. Figure 4.5). This new DWH is called EDWH.

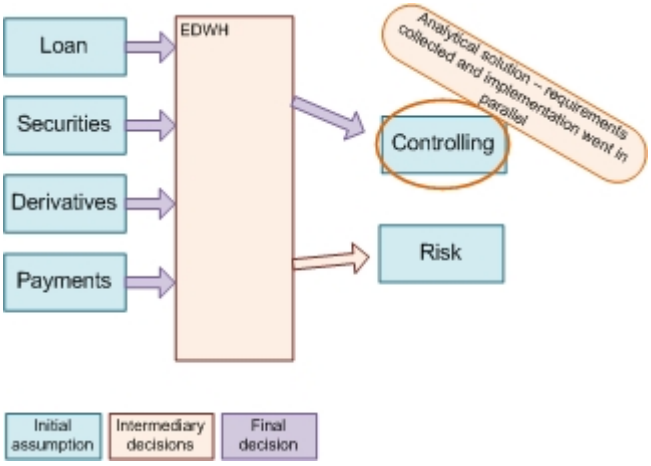

Figure 4.6 – Final scope of the EDWH project

The final decision taken (while the Controlling application was already being in development), was to include in EDWH also requirements for the Controlling application and make the EDWH an intermediary between source systems and analytical applications (cf. Figure 4.6). EDWH was therefore fed by data sources Loan, Securities, Derivatives, Payments, and so forth. It was build to satisfy business requirements of analytical applications that execute calculations needed for bank steering (Controlling and Risk). Based on these business requirements, corresponding data requirements were derived. Upon these requirements, the business data model was built.

4.2.4 Development Process

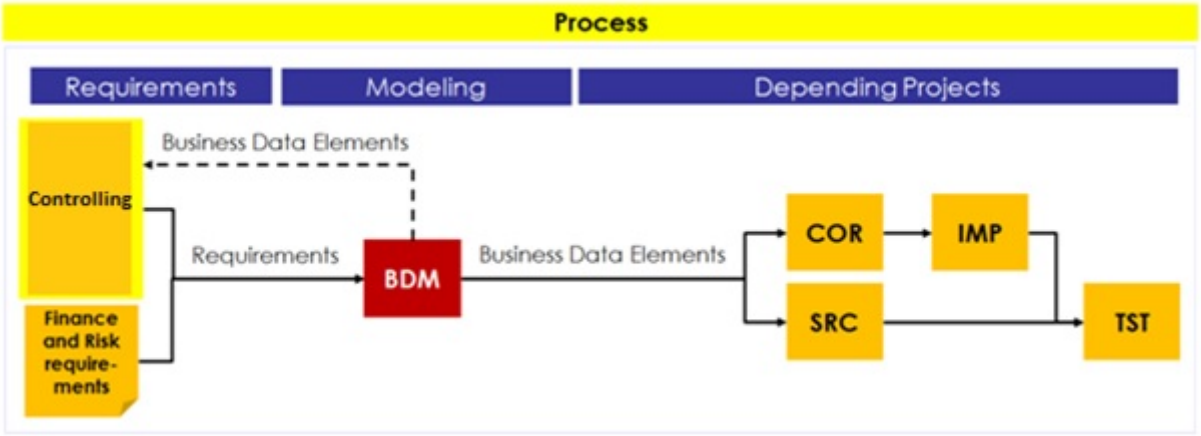

Figure 4.7 – EDWH project development process

Figure 4.7 illustrates the project development process. The very first project phase was requirements collection/definition. These requirements were collected from Controlling and Risk department. After the completion of these requirements, the requirements team handed them to the Business Data Model (BDM) team. This team created a data model based on data requirements collected in the previous stream in form of ERD model of business entities. A set of characteristics (attributes) builds up every business entity. A complete BDM was handed to EDWH Core requirements and Analytics (COR) and Implementation (IMP) as well as Sourcing (SRC) team. The COR team was collaborating with the IMP and interconnected with SRC team. The responsibility of members of these teams was to deliver physical implementation of previously defined BDM together with analytical and source information. The final phase was testing (TST).

In the following section, we describe the first two phases in more detail.

4.2.5 Business Data Modeling

Requirements collection task was based on Controlling and Risk requirements. Since the analytical application for controlling was being developed at the same time as the EDWH, new requirements coming from this analytical application had to be reintroduced into the EDWH. This caused the appearance of moving targets in the scope (cf. Figure 4.6).

The BANK delivered the list of contact persons that could provide business expertise on specific subject areas, more precisely by risk project lead and by controlling project lead. They both assigned contact persons having required expertise in derivatives, loan, off balance, or securities. They identified the employees that were experts in specific subject area as well as the ones that possessed general knowledge related to multiple subject areas. Employees that possessed knowledge about multiple subject areas, were distributed among all subject areas. Contact persons were controllers and risk department employees from the BANK. No explanation of their

role in the process was necessary. They had their system and they knew how it worked. The requirements and data collection was implemented within workshop sessions. At the beginning, these sessions aimed to explain to the BANK controllers and risk people what role and responsibilities they had within business data modeling and what results they had to deliver. They were later on asked to deliver necessary information and to verify this delivery after the requirements team processed the information it contained. Most participants were self motivated. They organized necessary meetings by themselves and voluntarily identified contact persons that could solve specific business issues. There was a certain time allocated to their tasks and they knew how important these tasks were for their managers. Both managers (risk and controlling) were present at almost all workshops. The objective of this first phase was to gather risk and controlling business requirements as well as collect all data necessary for both risk and controlling analytical applications to work properly.

All information assembled in the previous phase was handed over to BDM team. The business data modeling part was divided into "small interconnected projects". Every one of these projects represents a business data model of one subject area (one business concept in the BANK) such as Customer, Limit or Service. 22 business concepts were included in the Business Data Model. Figure 4.8 visualizes the subject areas decomposition. This figure illustrates the initial project plan that consisted of 22 deliverables and foresaw two phases for validation/review of the chosen working approach. The modeling approach is called a Kick start or Jump start, which means quick modeling. There were two weeks devoted for modeling of every subject area. This approach allowed quick results. It was however too time and effort challenging and required increased daily working hours for all participants of the modeling team. This is the reason why this approach was slightly changed after the second review phase. Three modeling weeks were then allocated to every remaining subject area.

The EDWH project analysis is closed by an overview of the outcome of the BDM from the point of view of streams that had to base their results on the final BDM release and provide lessons learned.

4.2.6 Lessons Learned

This section offers an analysis of advantages and disadvantages discovered in the studied project as well as proposed improvements.

Parallel Business Data Modeling

Despite the initial skepticism concerning the parallel business data modeling of subject areas (especially due to their interdependencies), this fear was proven unfounded. A model where several weeks are devoted to one subject area was very effective and it sped up the modeling process.

Management Commitment

An element that contributed significantly and helped to produce satisfactory results was the overall management commitment and awareness. Since the scope and time frame were chal-

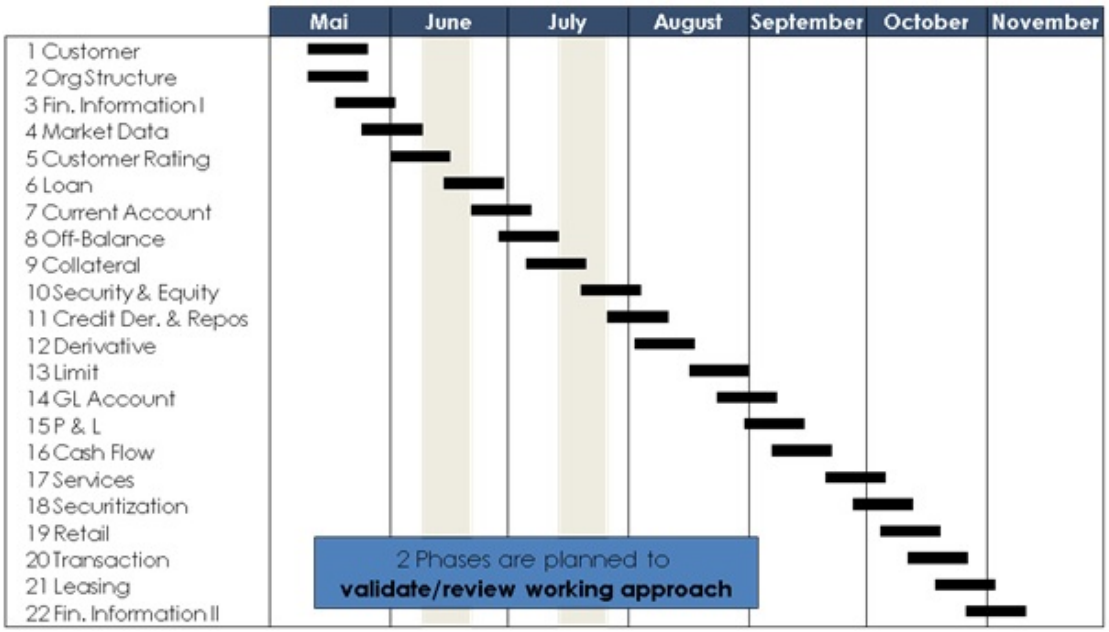

Figure 4.8 – Initial timeplan of Business Data Modeling

lenging, the support from overall project managers was required. The overall project manager has decades of experience. He took his role of reminding his managers and co-workers to release available resources and deliver required results in a high level of effectiveness. He had a good PMO support as well. The management commitment was so high, that managers even participated in workshops, which made the process more efficient and apart from that it showed to all other participant the importance of the project.

Data Specifications/Definitions

Insufficiency in definitions was considered highly problematic. The requirements stream already delivered unsatisfactory results. This was the case especially due to weak verification and validation process. Even though the commitment of management was really high, they did underestimate the importance of verification process. Management underestimated the importance of the verification process. This led to weak quality of partial system delivery.

Skill Set of Project Members

Even though in our case (besides the previously mentioned issue), managers were in general highly experienced, it was their surroundings that was problematic. A lot of participants did not have sufficient skill set needed for business data modeling. Business side was often lacking technical skills and IT side did not have sufficient knowledge considering business problematic.

Communication Model

Contrary, a positive point of the business data modeling was the communication and working model from management to its working groups. There were two managers present, one for IT and another one for business. Vertical communication and task escalation produced very good results. However the horizontal communication between business and IT was not easy (cf. Figure 4.9). It took some effort to establish a working bridge between these two worlds.

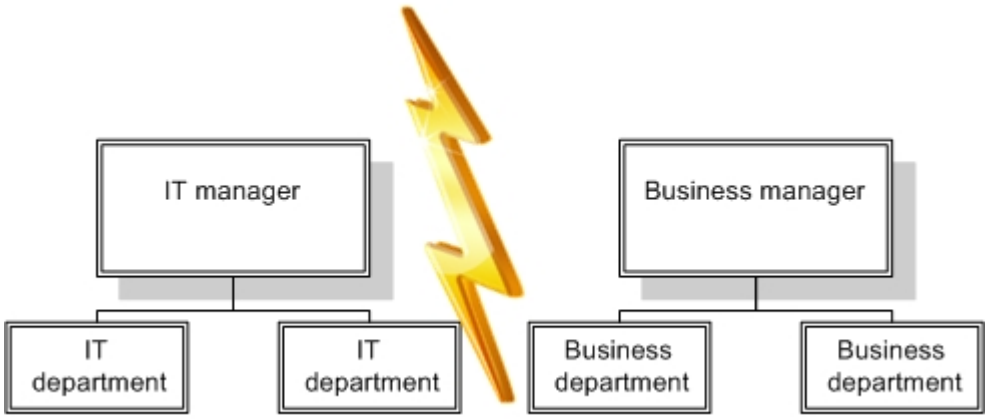

Figure 4.9 – Business–IT communication gap in the bank's organization

Working Model

The general working model and the implementation of a waterfall-like model was not appropriate. Particularly the handover between phases caused a lot of problems.

The source of this problem was partially the fact that different people were staffed in every project development phase.

- Team 1: requirements team
  - Participant D involved
- Team 2: BDM modeling team
  - Participant D and M involved
- Other teams: sourcing, mapping, implementation and testing team
  - Participant N involved in streams that followed the BDM phase

Team separation caused incoherence issues and since many definition imperfections were confronted in the project, the risk of wrong interpretations was faced.

If the project would start all over again, people who participated on its development would recommend making subsequent streams (especially the ones conducted by IT) actively participate in workshops right from the beginning.

Participation Model

The BDM was too ambitious in terms of business-driven approach. IT side was at the beginning completely left out of the process. At the beginning of the modeling phase, subsequent project streams (implementation, mapping and sourcing conducted by IT) were not involved at all. It was already during the project setup, where need of their involvement was identified. During the project development (approximately in the middle of the business data modeling), two people from subsequent project streams started to participate in workshops. Their participation unfortunately weakened and existed in listener mode only. Moreover, even though it was a known fact, that they participated only passively, no one forced them to change that. This caused decreased level of quality and the resulting sourcing problems. It was not until the last third of the business data modeling, where IT started to actively contribute in the process. Even if the time needed to explain and find consensus during workshops was then increased, the handout time from business to IT was decreased. Once the handover was not that difficult anymore, the process became smoother. This action was successful but not to full extent. The IT side was still facing a lot of interpretation problems at the mapping phase. Apart from that, the release version of the BDM was 41. For additional one and a half years, four to six people were fully working on the business data model (this was partially justified with the need of including sourcing and analytical information into the BDM). The final release represented the version number 261. This large difference indicates the imperfection of the first BDM release (version 41).

Business-IT Gap

Another issue that was faced in the BDM part of the project was the gap between business and IT. In the end a solution was found. Processes and guidelines were implemented, but not right from the beginning. Additional solution was a BDM training implementation in form of a presentation of ERD (Entity Relationship Diagram) features. The role of this training was to explain the model to business side and therefore make it up to date. This was done only after some models were created, because it was only then that modelers realized that this kind of representation is not completely understandable for business. On average, it took half a year for business to really understand how modeling worked (in case participants were interested in this technique, the assimilation time was significantly reduced). In addition, the document entitled “EDWH Foundation” was created. This document was produced in order to increase the understanding of implemented features as well as minimize flexibility of interpretation. It describes architecture and mappings as well as presents roles of all implementation levels.

Business Lead in the Business Data Modeling Process

A very accurate statement of the COO and CIO concerning BDM concept and methodology: “IT does not need a warehouse, you (business) need to need it.” when talking to their business-

side management level colleagues. He addresses with this statement the business. He states that if business needs a data warehouse, IT can provide it, but business needs to offer necessary support.

Business requirements and the adequacy of their content as well as their completeness is one of the most important success factors of business data modeling. This factor is often labeled mission critical. It is the key to a successful BDM. The project must be understood as well as its basic business context. This is afterwards necessary for a successful data requirements definition.

Data Quality Management

Implementation of the data quality management could increase the data quality and prevent some problems. It decreases imprecise definitions and integrity violations. Lack of such an element was faced during handover between project phases. In fact, data quality indicators were just being defined while the case study was conducted (June 2013). Even though the quality management is considered very important element to be implemented from the beginning of the project, it was not until the last third of DWH implementation that this kind of management was launched.

In this chapter we described results of the empirical action research process. Results of the initial survey that allowed identifying crucial factors in financial data warehousing were presented. Results from the case study that focused on a finished DWH project (especially on the business data modeling part) were summarized. In the context of this case study, main strengths and weaknesses of the financial business data modeling were gathered. Based on crucial factors and lessons learned from the studied project, we take Inmon's business data modeling methodology and adapt it to the financial industry as well as include specific features that are important in this field. We create a methodology that is usable in practice and that takes into consideration both aspects of business data modeling, human and technical. This methodology includes managerial techniques and modeling standards in addition to the fact that it reflects current needs in the financial data warehousing. This methodology is described in detail in the following chapter.

CHAPTER 5

A Methodology for Financial Business Data Modeling

Previous chapters introduced data warehousing in general and compared the development methodologies of Inmon and Kimball. For our purposes, Inmon's model appeared more appropriate. One chapter in this thesis described the research process. We first presented techniques that this process used, their objectives and order in which we implemented them. We afterwards described results they led us to. Based on information gained throughout the whole research process as well as analysis of literature discussing Inmon's development strategies, we formulate a practical business data modeling methodology for the financial industry. Due to restricted (time) resources, it discusses a limited number of elements.

The objective of our methodology is to decrease the impact of factors summarized in the Table 5.1).

Category	**Factor**
Top Management Issues	Missing clear set of goals and priorities
Data Quality Issues	Insufficient data consistency
Project Management Issues	Unclear scope definition
Issues Met while Building a Common Language	Language gaps between designers and technical actors
Issues Met while Building a Common Language	Insufficient data specification/definition

Table 5.1 – Final set of factors that are covered in the BDM methodology

We did include in the originally defined four crucial factors the factor "Language gaps between designers and technical actors" because it was ranked high in the results of the initial survey (ordered eight in the Table 4.3) and because during the case study it appeared to be a critical element.

There exist a lot of issues that do not appear to be directly connected to business data modeling on the first sight. One of them is the support of the top management (cf. category "Top Management Issues" in the Table 5.1). This one is however a crucial success element in corporations, including banks, where several projects run in parallel. In addition these projects involve to a smaller or grater extent the same resources (human, material, and so forth). In order to assure smooth project development, all project participants must have a clear picture of the role and responsibility they have within different projects. They additionally need to be aware of the priority their tasks have, and of the time that is allocated to their execution.

All the other elements have a clearer connection to business data modeling. Without having a clear product scope (cf. category "Project Management Issues" in the Table 5.1), the project management might need to reintroduce elements they considered not relevant at the beginning. The later these elements discovered, the more costly their reintegration. As we can see in the Figure 5.1, this cost raises exponentially with time.

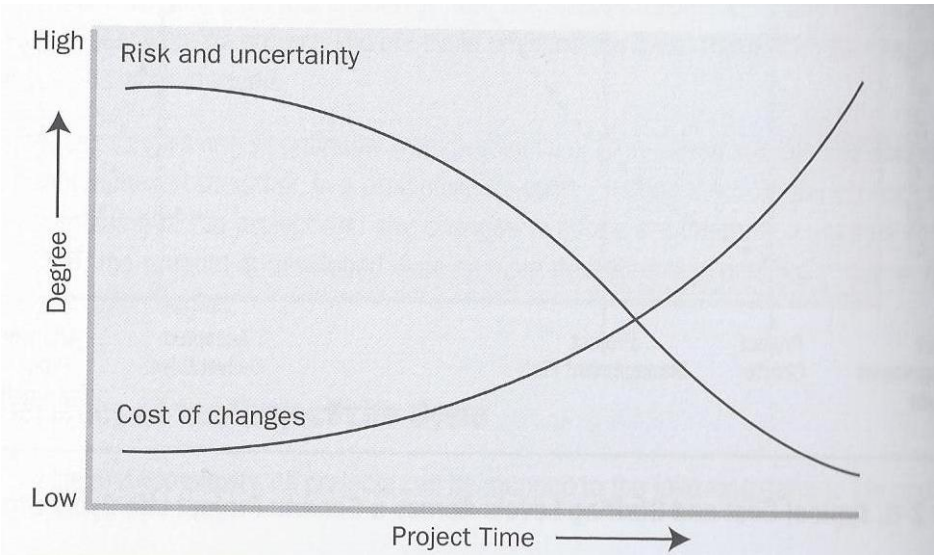

Figure 5.1 – Cost of changes with increasing project time[1]

Data consistency (cf. category "Data Quality Issues" in the Table 5.1) or the data quality itself behaves in similar way as the scoping issue. The sooner in the bank's supply chain (cf. Figure 4.3) the project management introduces data quality rules, the fewer defects they will face and less their correction will cost. Moreover, the identification of the source of these defects gets more complex and challenging with time.

Last two factors we will introduce (cf. category "Issues Met while Building a Common Language" in the Table 5.1), represent the added value of the business data modeling itself. When the bank fills or at least decreases language gaps between technical and business actors and assures unambiguous data definitions for the whole organization, it will be making a big step towards establishing a commonly understandable language within this organization.

[1] Source: PMBOK, 2013

Before we start to introduce our methodology, we need to discuss general assumptions and scope delimitation:

- The methodology follows Inmon's principles and respects zeb/ best practices, methods and processes. Information concerning zeb/ as well as which part of the methodology follows their instructions is not revealed to third parts due to confidentiality issues.
- This methodology represents just a very small part of the whole DWH project life cycle. It discusses uniquely its conceptual design part. This is the reason why we concentrate neither on any software development methodologies nor on any Agile BI techniques. Even though some agile elements (such as for example regular team meetings) are mentioned, they are not implicitly linked to any agile methodologies per se.
- The size of the organization is not taken into account. We develop the methodology in the context of an implementation of a financial data warehouse at a large banking corporation. Our methodology is built for DWH projects at big banks that consists of several departments. In special cases, when one would need for example to implement a data warehouse within a small enterprise, it would not be reasonable to follow our methodology. A lot of issues we aim to solve will with high probability never occur in small structures.
- Within top management issues, we concentrate on the prioritization aspect. We do not analyze techniques of goal setting in the context of this thesis.
- Specific (project) prioritization techniques have not been analyzed in the thesis due to the fact that these techniques differ from one financial institution to another. No general prioritization/decision making guide can be therefore proposed. Moreover, choosing an appropriate prioritization technique requires additional and rather complex research.
- We assume that the central unit that assures compliance with the project prioritization has already been set up (this unit is discussed in the top management part of our methodology (cf. Section 5.1)).
- Details of project scheduling techniques are not covered in the thesis.
- Project planning as well as time and content of deliverables release is out of scope.
- Budget allocation is out of scope as well.
- Our methodology discuss the BDM phase of a financial DWH project development. It creates a basis for a smooth continuation of subsequent project phases. However it does not discuss these phases at all. This restriction applies to the whole thesis.
- Corrections of errors in source systems (found during activities within data consistency assurance) as well as techniques that can be implemented for this purpose are considered out of scope. This would be the responsibility of specific data owner, a data quality role that we do not specify here.

- We do not discuss tooling in our methodology, even though automation of processes as well as usage of appropriate project management and data quality tools facilitate tasks necessary to be executed within these two domains.
- We concentrate in the current thesis only on data consistency issues. Data quality requirements are therefore out of scope in the business requirements definition, data requirements collection as well as in the business data modeling phases.
- Data quality part of our methodology discusses neither data quality indicators nor tools or methods. These elements require complex research, that is out of scope of the present master thesis. Such a research would lead to a whole new master thesis topic.

Our methodology is modeled in Business Process Model and Notation (BPMN), one of the most widely used languages for business process modeling. We do not use complex BPMN elements, so that the model remains readable also for anyone not familiar with this language. In addition, BPMN notation is supposed to be understandable for technical developers and business analysts alike (ISIS, Summer Term 2012).

Figure 5.2 – BPMN elements

The following elements (cf. Figure 5.2) are used in the model:

- Start event - represents starting stimulus that triggers a business process
- End event - represents end of a business process
- Task - represents a task that needs to be executed within the business process
- Repetitive task - represents a task that repeats until the condition needed for its execution is valid
- Sub process - encapsulates business sub process
- Parallel gateway - tasks that come from such a gateway are executed in parallel
- Exclusive gateway - only one of tasks that come from such a gateway is executed

5.1 Top Management

The main goal of BDM implementation is to build a common language inside an organization. This kind of language is crucial when building a DWH fed by many sources and used by several lines of business. A comprehensive methodology of business data modeling that focuses on the content of business requirements promotes communication between all involved parties and

ensures active participation of important stakeholders. This kind of methodology could moreover make the transition from content into data models/structures smoother since it requires less transformation and avoids some errors. It prevents misinterpretations and supports harmonized and integrated data content.

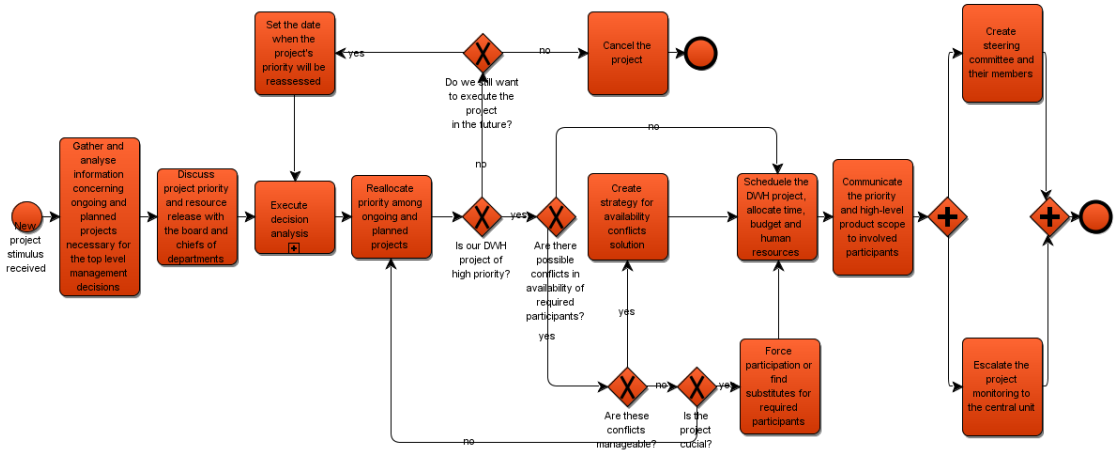

Figure 5.3 – Methodology part that concentrates on the top management prioritization

Business data modeling at a bank starts with the prioritization activity by top management (factor "Missing clear set of goals and priorities" from the Table 5.1) when they identify a need to a pilot project that creates a BDM or when new business requirements that enlarge an already existing BDM arrive (cf. the prioritization process in Figure 5.3). Top management sets priority to this new project but it respects the priority of other planned and ongoing projects. Top management members open the prioritization task after discussions with board members as well as heads of involved departments. If the outcome of the project is needed very soon or if the importance of its implementation is high for the bank at the moment, then top management approves the project, allocates corresponding time and budget and releases necessary resources. If there are conflicts in crucial actor's availability and the project is of essential importance, either top management forces their availability (they re-prioritize tasks they are involved in) or they find satisfactory substitutes. In case this project is not crucial, they re-prioritize and postpone or eventually completely cancel the project. Finally, they communicate final prioritization as well as project time allocation, and available budget to concerned human resources together with high-level scope. High-level scope has a form of project objectives, goals and expectations. Top management verifies the proper execution of regulations it has delivered in a so called steering committee that meets regularly. This body includes leaders of the bank and their project managers involved in the topic. If the bank executes the project in cooperation with a consulting company, project managers of external consultancies complete the steering committee.

The Figure 5.4 illustrates the organizational structure of management bodies.

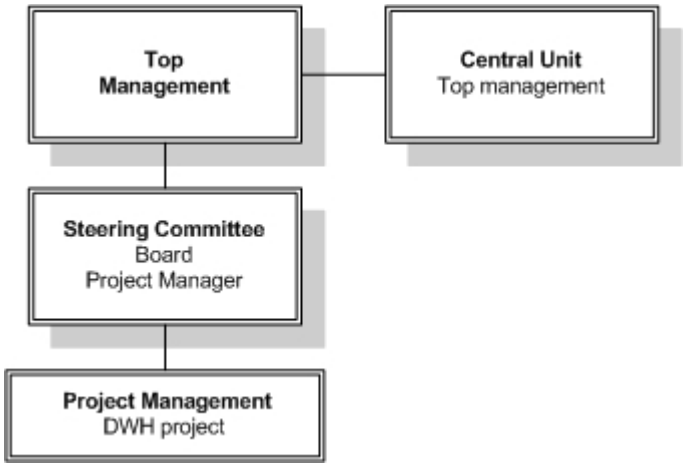

Figure 5.4 – Management organization chart

In order to assure the project prioritization compliance, top management escalates this monitoring task to a central unit that acts as an operative working body (cf. Figure 5.4). It is composed out of top management members (or their representatives). Their principal responsibility is the project prioritization surveillance. This unit verifies in regular meetings that the priority, time, budget and resources constraints that the top management set to our project (and to other ongoing projects in the bank) are respected by all involved parts. Likewise, if any violation is noticed by the project management or project members of the DWH project, this issue is escalated back to the unit. Their members implement specific techniques to find an appropriate solution to respective problems. The same applies if the unit itself discovers a conflict. Furthermore, they align all their activities with the top management. This alignment is done by quarterly reports to the top management within steering committee meetings. Only selected members of the unit participate on the steering committee. This assures the consolidation between these two bodies. The central unit might need to reopen specific tasks within our project or maybe even require re-prioritization of bank's projects/resources if they observe significant conflicts. Moreover, they have to provide necessary support to execute proposed actions. They have the right to require assistance from the top management to gain necessary weight to their actions.

Once the top management defines the priority of the DWH project and communicates it to concerned parts together with the project objectives, roles and responsibilities description, the project management needs to define clear scope and their boundaries as well as schedule project's progress and release of deliverables.

5.2 Scope Management

The project's success is directly influenced by active stakeholder involvement in the discovery and decomposition of needs into requirements and by the care taken in determining, documenting and managing requirements of the product, service, or result of the project (PMBOK, 2013).

Additionally, business side (and not IT) must have the lead in business requirements definition, because they are data consumers. IT has to follow business in order to deliver quality data that would satisfy business requirements. IT is involved in the scoping process, but it has only supporting role. This action starts to fill the Business-IT gap.

In order to be able to clearly define the product scope (cf. factor "Unclear scope definition" from the Table 5.1), that is unlikely to change during further project development, it has to be based on a clear set of business requirements that are derived from the high-level scope. Once this has been done, the project management can define a highly reliable and stable product scope. This is the reason why a considerable amount of time has to be dedicated right at the beginning of the project to the business requirements definition.

Every business requirement consists of following:

1. Person responsible for this requirement definition within decisive body (explained in the following paragraph and the glossary)
2. Person responsible for this business requirement fulfillment within the bank
3. List of contact persons and tasks they execute in order to fulfill this business requirement
4. Business value and objective
5. Project objectives that this business requirement satisfies
6. Systems and applications that participate in the execution of this business requirement (this information is afterwards completed in the context of data requirements collection)
7. Detailed description of this business requirement
8. Decomposition into business functions and business processes
9. Impact on other requirements
10. Dependencies with other requirements
11. Available documentation and contact persons (store and link physical documentation and/or note people responsible for it)
12. Functional and nonfunctional requirements
13. Requirement assumptions and constraints
14. Test cases (all cases that can happen during this requirement fulfillment; they will be tested in the test phase of the project)
15. Acceptance criteria (criteria that need to be fulfilled in order for this requirement to be accepted in the final release of the project)
16. Stability of validity of this business requirement (represents the possibility of requirement to be changed, replaced or deleted in the future based on current knowledge)

17. Priority
18. Relevance to high-level product scope (helps to identify if this requirement is potentially out of scope)
19. Status (in or out of scope or to be decided)
20. Approval status (current status of completeness and correctness of this business requirement)

Before we start to explain the scope definition process, we must clarify some terminology. The meaning and interpretation of some elements we introduce can be misleading. The so called *business requirements definition team* (BR team) is a team that processes business requirements. It is an active working body. The BR team (cf. Figure 5.5) consists of heads of departments or application managers that act actively in business areas of the bank, external consultants (if the project is executed in cooperation with a consulting company) and project managers or their substitutes.

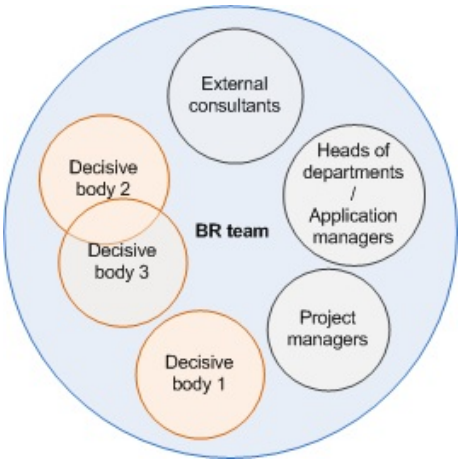

Figure 5.5 – BR team composition

Within BR team, project management elects a so called *decisive body*. Project management identifies one decisive body per business area. One person can at the same time be a member of multiple decisive bodies. It is required, that at least one IT professional that administrates source systems at the bank and one business person are part of every decisive body (cf. Figure 5.6).

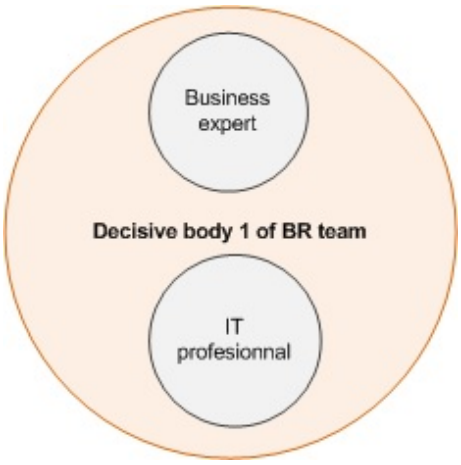

Figure 5.6 – Decisive body composition

The BR team needs additional help from *workshop participants* so that it can define business

requirements. Within the bank, workshop participants are responsible for the adequate execution of business functions and business processes. Employees that execute tasks that participate in added value production of these business processes and functions as well as the ones that directly execute these functions should be workshop participants. Project management can likewise require the participation of additional IT professionals. It is the responsibility of project managers and heads of departments to nominate these employees. One person may participate in more than one business areas workshop (cf. Figure 5.7).

Figure 5.7 – Workshop participants composition

Figure 5.8 illustrates involvement of personnel for one business area.

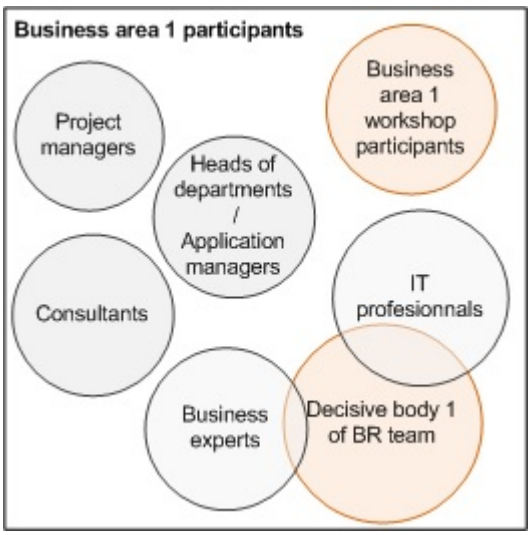

Figure 5.8 – Example of involvement in a business requirements collection for a business area

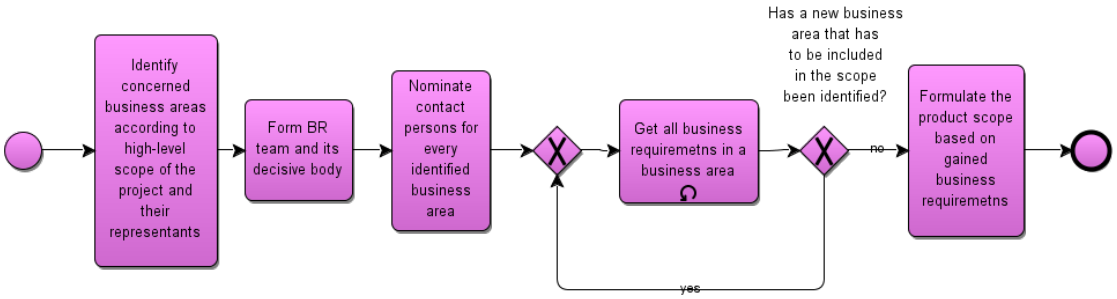

Figure 5.9 – Encapsulated process of scope definition

The whole process of business requirements collection starts with identification of business areas involved in the processes and other activities that are to be covered by the project (cf. Figure 5.9 and the detail to the sub process "Get all business requirements in a business area" in Figure 5.10).

Top management indirectly provides the list of these business areas. They are part of the high-level scope delivered earlier. Once the project management identifies these areas and their representatives, they create BR team. This team with help of project management nominates employees that are responsible or participating in processes and applications that are needed to be covered so that the project's goals set by the top management are achieved. It is important to define a decisive working body within this team. The decisive body prepares and conducts all workshops, takes necessary decisions and has the responsibility for the final result.

For better understanding of difference between *business area*, *business process*, and *business function* a short definition follows. The definition of *business area* differs throughout different projects and banks. A business area represents either an analytical application or it can be a department of the bank. Business area is the biggest operational subset of the high-level scope. Within one business area, several business processes and business functions satisfy business needs of the bank. *Business process* is a set of tasks and activities, that satisfy a specific business requirement. Under *business functions* we understand typical bank functions such as RWA calculation or collateral allocation that have to be implemented so that the bank is able to support its basic functionalities as well as meet the reporting and regulatory requirements.

Once the project management finished the initial setup, project members follow the further described formula for all business areas (cf. Figure 5.10). There is also a possibility of merging two or more business areas, if they involve the same people or if their inter-connectivity is high.

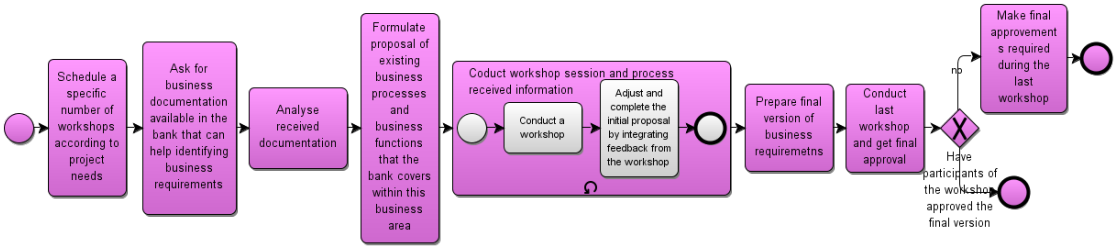

Figure 5.10 – "Get all business requirements in a business area" sub process from Figure 5.9

Expert judgment and facilitated workshops techniques with knowledgeable and experienced parties helps to solve scoping issues (PMBOK, 2013). The decisive body schedules workshops with all required participants. Facilitated workshops take place once a week. The total number of these workshops is project dependent. At least two weeks before the first workshop takes place, the decisive body demands workshop participants to send documentation that helps to understand business requirements within this business area. They analyze this documentation so that they can generate a proposal of business applications, processes and functions that cover specific business need necessary for the satisfaction of project objectives. This proposal serves as guideline for the first workshop. The purpose of the first workshop is to verify the proposal, make corrections and add missing business requirements (current and future). The decisive body of the requirements team processes adjustments and corrections requested during the first workshop. They then discuss this version with remaining BR team members and then with workshop participants during the following workshop session. With their help, the decisive body completes business requirements and corrects errors. Time between two workshop sessions serves for updates and corrections. Decisive body presents the final version of business requirements of a specific business area during the last workshop session. If workshop participants approve this version during the workshop session, the business area requirements definition phase is considered complete and is closed. If not, the decisive body makes required changes and requests final approval. This approval does not require physical presence of workshop participants.

Business requirements are presented and collected in a so called traceability matrix (cf. Figure 5.11). Requirements traceability matrix is a grid that links product requirements from their origin to deliverables that satisfy them. The implementation of a requirements traceability matrix helps to ensure that each requirement adds business value by linking it to the business and project objectives (PMBOK, 2013).

Requirements Traceability Matrix

Project Name:	
Cost Center:	
Project Description:	

ID	Associate ID	Requirements Description	Business Needs, Opportunities, Goals, Objectives	Project Objectives	Test Cases
001	1.0				
	1.1				
	1.2				
	1.2.1				
002	2.0				
	2.1				
	2.1.1				
003	3.0				
	3.1				
	3.2				
004	4.0				
005	5.0				

Figure 5.11 – Simplified traceability matrix example[2]

The scope has to be approved by all core stakeholders at the beginning of the project. After requirements collection phase, we can define the product scope based on the collected requirements.

The document that describes the scope is called product scope statement and contains the following elements (PMBOK, 2013):

- product scope description
- assumptions
- constraints

The scope is agreed on by the modeler, the management, and the ultimate user of the system (Inmon, 2005).

Since business requirements were defined with participation of department leads and their subordinates as well as project managers, and since they also approved these requirements in sense of scoping and relevance issues, the scope that is defined based on these requirements has a high potential to be realistic and stable.

The correctness and completeness of product scope is re-verified before the business data modeling phase itself starts (cf. Section 5.4).

[2]Source: Inmon, 2005

5.3 Data Requirements Collection

When the project management defines a detailed product scope, it can start to gather requirements on actual data that the future data warehouse will hold. These requirements are derived from previously defined business requirements. Every business requirement requires a set of data that allows its proper execution.

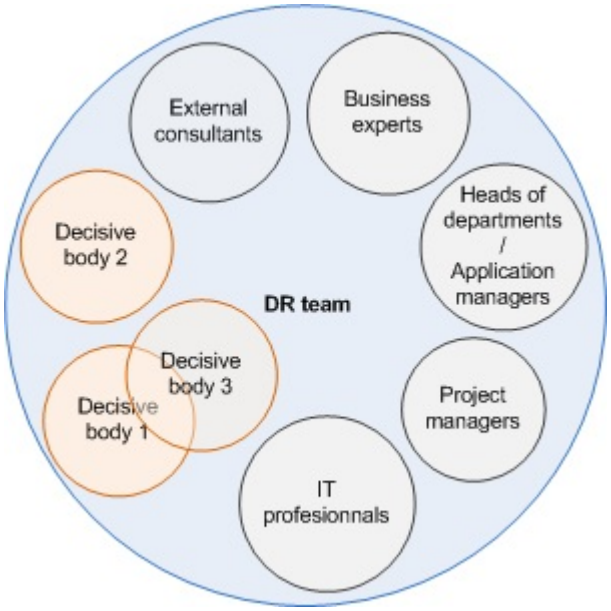

Figure 5.12 – Data requirements team composition - an illustrative example

Before the data requirements collection starts, project management needs to define subject areas, in which the business data model will be decomposed. This decomposition is done based on experience of project participants from previous similar projects and on structure of source systems as well as organization of subject areas in the bank. Typical subject areas are for example customer, bank's products, limit, collateral, and so forth.

The next step that needs to be executed by the project management is to form a *data requirements collection team* (DR team). This team includes several decisive bodies, one per business area. The DR team is an active working body, similar to the BR team (cf. Figure 5.5). However, the DR team includes business experts, that are specialists in a given business area or one of its subsets and IT professionals that provide data, which these experts consume (cf. Figure 5.12). Their presence in DR team help us therefore to continue to close the Business-IT gap. One decisive body has the responsibility to collect complete data requirements for all business requirements in respective business area. All business requirements identified in the business requirements definition phase that were kept in scope of this project have to be covered. The project management involves in the DR team as many participants from BR team as possible, because they are already familiar with bank's activities. By involving the same people, the data consistency is increased, since the understanding of business requirements will be passed on to data requirements.

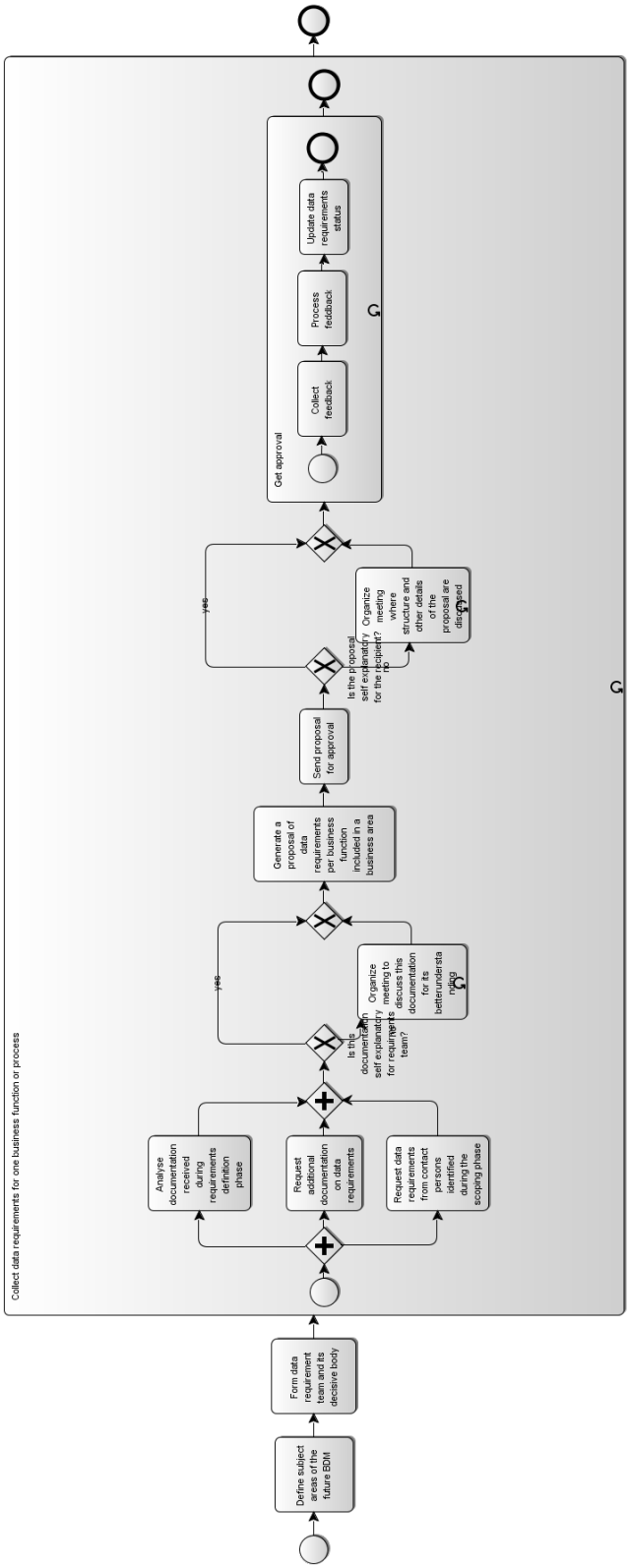

Figure 5.13 – Definition of data requirements

Data requirements collection is organized per business areas and repeats as many times as many business areas are involved in the scope of the future data warehouse product (cf. Figure 5.13 and for repetition the sub process "Collect data requirements for one business function or process" in this figure). First, the DR team analyzes documentation received during the business requirements definition phase and at the same time, the decisive body requests additional documentation that allows its members to derive data requirements. In parallel they request data requirements from contact persons identified during the scoping phase, in case they exist in an easily interpretable form. Based on received documentation or requirements list, the decisive body creates the proposal for the list of data requirements ordered by business requirement. They send it to respective contact persons. In case either available documentation is not self explanatory for the decisive body or if the structure of the list of data requirements needs to be explained to contact persons a meeting that allows to discuss these topics is organized. If needed, more than one meeting can be scheduled. Until the proposal is acceptable for contact persons or the deadline set by project management is not reached, the proposal is updated based on the feedback that contact persons provided.

The role of data requirements is to complete previously defined business requirements. Every data requirement contains following information (some information was already included in business requirements):

1. Person responsible for the data requirements collection within the decisive body of the DR team
2. Person responsible for the data within the bank (resp. the responsible of the data source where the data comes from)
3. Source system the data comes from (this information allows the decisive body to understand business requirements much better, because it provides understanding of its original logic, context and relationships/dependencies; data coming from source systems requires expert analysis; workshops within the requirements collection phase need therefore the presence of data deliverers that possess sufficient knowledge on data characteristics)
4. Data type and format of this data requirement
5. Frequency of validity verification of the data requirement (if applicable)
6. Additional information this data requirement needs to be associated with in order to be usable for its intended purpose
7. The purpose this data requirement is to be used for
8. Clear and unambiguous business definition of the data
9. Link to business requirement(s) that requires this data
10. Subject area it is to be modeled in (this information is a proposal for the next phase (business data modeling) and can be in this phase associated to more than one subject area, since the clear differentiation between subject areas is to be delimited in the business data modeling phase)

5.4 Scope Re-verification

As we mentioned in the scoping section (cf. Section 5.2), we will re-check the correctness of the product scope before the business data modeling starts.

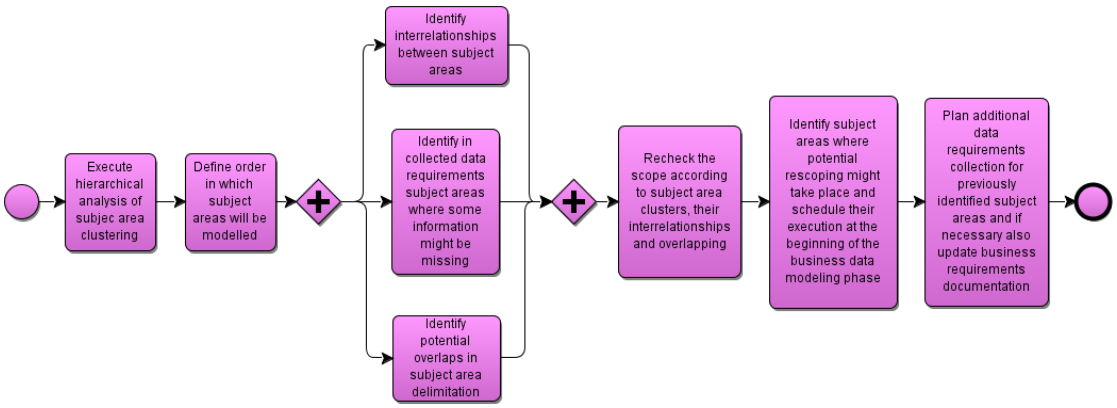

Figure 5.14 – Scope analysis process at the beginning of the business data modeling phase

The scope re-verification phase (factor "Unclear scope definition" from the Table 5.1) starts with hierarchical analysis of subject areas defined in the previous phase. The project management schedules these subject areas according to inter-dependencies between them (cf. Figure 5.14). Project management orders subject area analysis, where interrelationships between subject areas are verified together with potential overlaps. At the same time, DR team analyzes collected data requirements and identifies subject areas, where some information might miss. According to results of this analysis, the scope is rechecked and changes are documented. Project management schedules subject areas where changes might take place, at the beginning of the business data modeling phase and schedules additional data requirements collection within this subject area if necessary. If needed, business requirements documentation is updated.

5.5 Common Language

The project needs to start to build common language during the design phase. The modeling team needs to agree on common wording. Their members have to develop a shared understanding of business objects within the business data model, since the data that specifies them is defined from their different point of views. We consider the point of view of consumer (business) and supplier (IT) of the data. Homogeneous entities among all involved parties (business and IT) as well as departments of the bank appear during workshop sessions and following review phases.

If the project manages to successfully implement an integrated data model that creates consistent data structure of the bank as the whole, we will decrease potential for interpretation errors in subsequent project development phases. Especially when this model offers clear unambiguous description of all its business entities.

Difficulties that can complicate the task of building a common language such as motivation or communication problems are on one hand decreased by the previous top management prioritization and on the other hand by creating teams that integrate business analysts among different mindsets (business experts and IT professionals). Business analysts understand to a certain extent the point of view of both. They enhance the communication between business experts and IT professionals by making a bridge between these two worlds. Active participation of both sides, IT and business is likewise a crucial aspect that fills the Business-IT gap. IT needs to understand the terminology used by business side so that they can overtake the implementation task and conduct it with as few interpretation errors as possible. Business analysts should be the ones that will take over the communication task and its guidance during workshop sessions. In additionally they can provide translations and extended descriptions. When a business analyst guides workshop session, participants have tendency to take an active part, since discussed issues are presented in a language understandable for them. Active participation and increased motivation can be additionally enhanced by management level presence during workshop sessions.

In this phase, we call face-to-face workshops Joint Application Design (JAD) sessions (according to Inmon, 2005). They are implemented in order to understand and realistically structure business and data requirements from all involved departments in form of a widely understandable business data model. JAD sessions represent an open discussion of wording, definitions and homogeneity issues that are illustrated with carefully designed examples.

JAD sessions take place regularly in order to get an impact on the knowledge sharing among crucial actors.

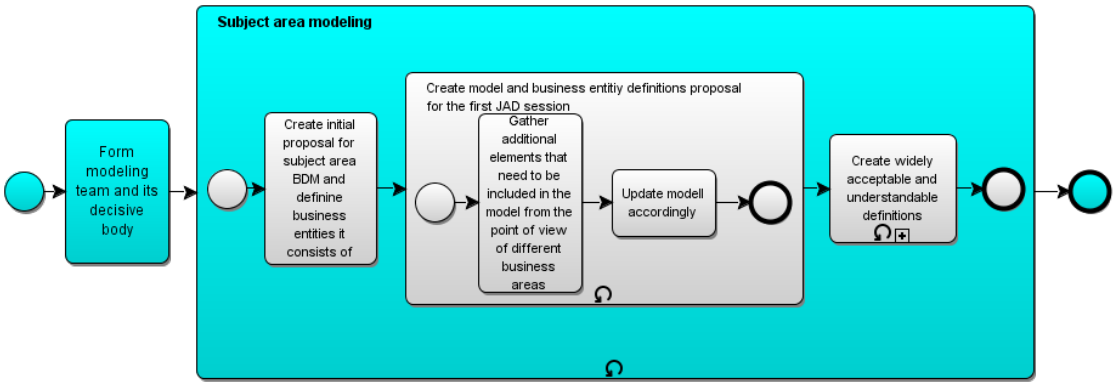

Figure 5.15 – Business data modeling

The business data modeling (cf. Figure 5.15), the crucial phase for building the common language, starts when the project management forms the modeling team and its decisive body (one per subject area) (cf. Figure 5.16). The project management involves in the modeling team as many members of DR and BR teams as possible, so that the consistency and data interpretation has a fluid continuity. In order to fill or at least decrease the gap between business and IT, business experts, IT professionals as well as business analysts (bridge between business and IT) are members of this team. At least one business analysts is a member of every decisive body. The Business-IT gap is therefore under control.

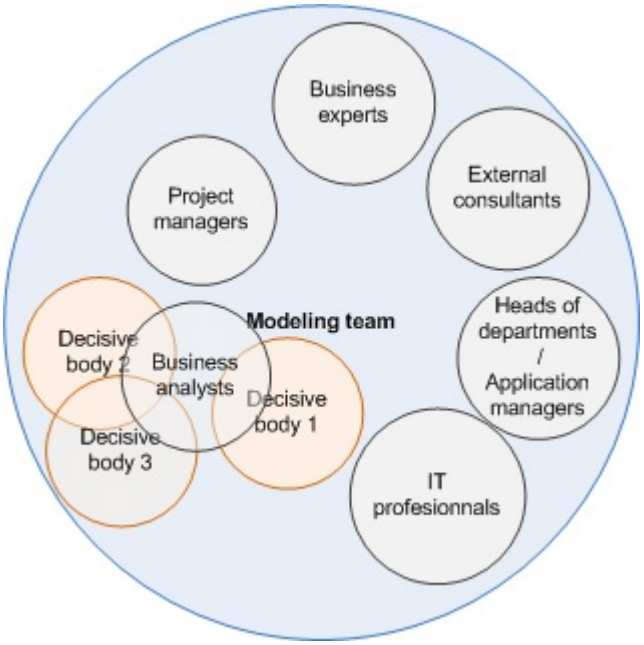

Figure 5.16 – Modeling team - an illustrative example

The sub process "Subject area modeling" (cf. Figure 5.15) repeats as many times as many subject areas the project management defined .

Before the first JAD session, the decisive body creates a proposal for the business data model and its content with as precise definitions as possible. Business data model is composed out of business entities at the highest level of abstraction. This model is iteratively extended of business entities required by all involved business areas (cf. sub process entitled "Create model and business entity definitions proposal for the first JAD session" in the Figure 5.15).

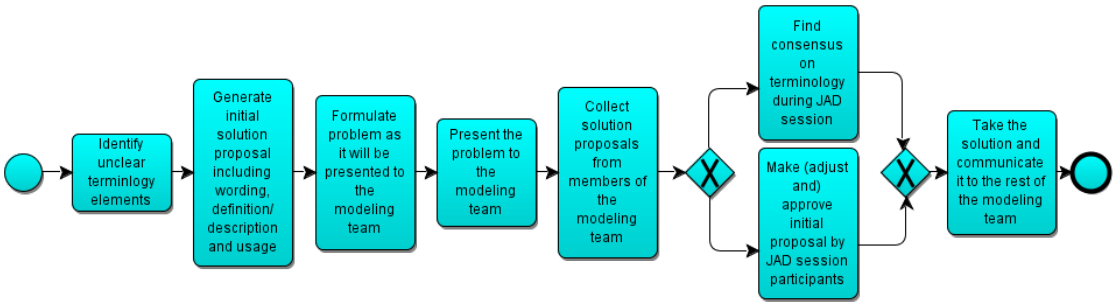

Figure 5.17 – "Create model and business entity definitions proposal for the first JAD session" sub-process from Figure 5.15

Afterwards (before every JAD session), the decisive body identifies terminology elements that can potentially create misleading interpretation from JAD participants' view (cf. Figure 5.17). The decisive body prepares a proposal definitions for conflicted terminology and discusses them with the rest of the modeling team. Team meetings, where members of the modeling team discuss these open issues take place regularly. During these sessions, the decisive body, with help of the rest of the modeling team, generates a proposal solution or at least shows a direction to follow in problem solving. Since this solution is a consensus between all involved parties (IT, business and involved business areas of the bank), the probability of creating an acceptable common language increases. Particularly when they prepare precise definitions and specifications that minimize interpretation flexibility as well as prepare a good starting point to fill the Business-IT gap. The proposal as well as all other open issues are then presented to JAD participants. The role of JAD participants is to correct and update this proposal so that it satisfies all involved parties as well as propose mutual solution to open issues.

At the beginning of every JAD session, objectives and expectations are announced and it is the role of the decisive body to assure, that they are satisfied and followed properly.

During JAD sessions, the decisive body either finds consensus for conflicted terminology elements or presents the proposal they created and make adjustments to this one so that it is accepted by JAD participants in the end.

In order to decrease on one hand the amount of missing data and on the other hand data insufficiency in terms of breath and depth for satisfaction of a specific business requirement as well as data correctness and reliability, the decisive body collects and includes data constraints in the business data modeling. If we want to consider data specification satisfactory, we need to include the information on the reevaluation frequency, where applicable. This information is necessary to assure data reliability and regulatory correctness.

Discussion of appropriate language, symbols and units usage as well as generation of clear and understandable definitions is discussed during JAD sessions with business experts, IT professionals and business analysts. The decisive body assures the quality of these elements when it base its decisions on participants expertise and testimony.

As we already mentioned, goal of JAD sessions is to build a common understanding of a subject area among different business departments of the bank. The modeling team assures this kind of understanding, when it creates an entity-relationship diagram that represents main business entities needed for the representation of a given business area. Data requirements (either

old ones that directly provide from existing source systems of the bank or new ones that have been identified during modeling workshops) characterize these business entities. Every BDM in addition aggregates elements that enhance reporting functionality of the future DWH.

5.6 Data Consistency

Since data is often considered one of bank's main assets, its quality highly influences the quality of bank services. Moreover, processing reliable information brings the bank the advantage of better and more precise day-to-day decision-making as well as long term strategy formulation. This assures a better position on the market. This section concentrates therefore on data quality issues, especially on the data consistency. Methodology part that focuses on data consistency assurance (factor "Insufficient data consistency" in the Table 5.1) includes data consistency itself, data validity, accuracy, usability, and integrity within the bank, among its applications, and across an IT infrastructure.

The data consistency assurance process is modeled in the Figure 5.18.

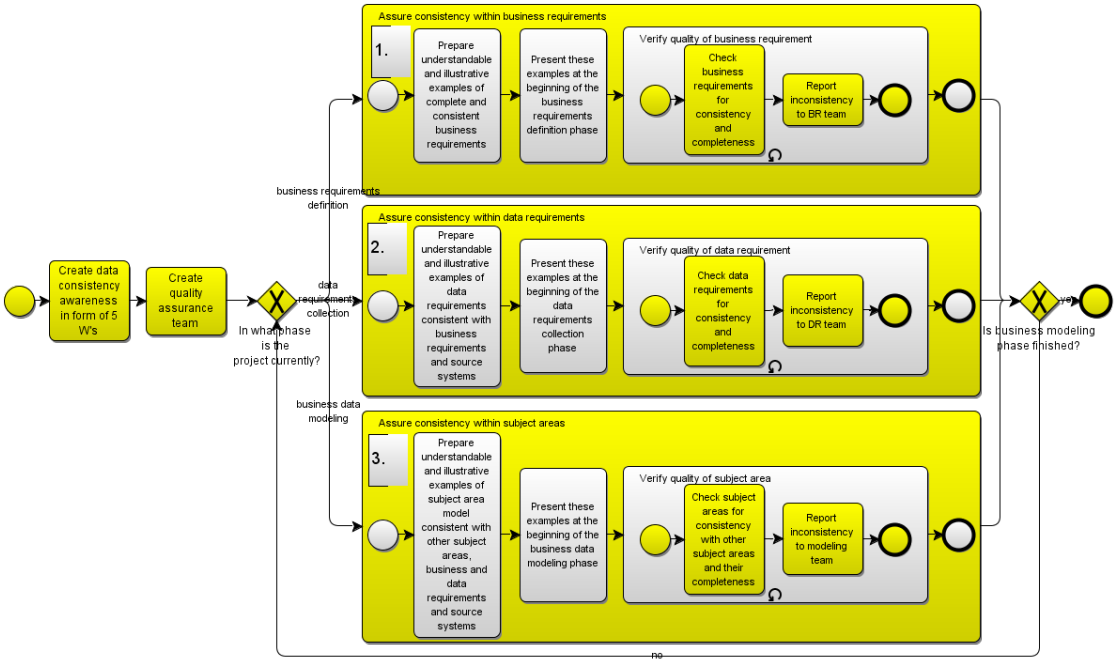

Figure 5.18 – Business process for data consistency assurance

Data quality management starts when the project management spreads the data quality awareness among data consumers and providers (cf. very first task in the Figure 5.18). They communicate this awareness in form of 5 W's (definition of What, Who, Why, When and Where). This means that in order for this process to be effective, they need to present what everyone needs to do and why these actions are important. Employees need to be aware of negative effects of incorrect data entry, mistreatment, inadequate storage and documentation of in-

formation. Awareness of the importance of the data quality starts at the beginning of the project, because already the scope definition that is composed out of business requirements definition needs to follow consistency principles. With data consistency awareness among all project team members, the data quality increases and the quality of the final product is therefore enhanced. This awareness is spread to all departments that participate on the current project.

The 5 W's strategy on data consistency awareness is then the following. Everyone who works or participates in any way on business activities that fulfill project objectives has to process all information with respect to its consistency with other data and tasks they are actively working on. All this is done so that in the end, they increase the final data quality within the future data warehouse. The importance of these actions is explained to everyone involved. Every employee on each level is responsible for data quality assurance within activities he was assigned to.

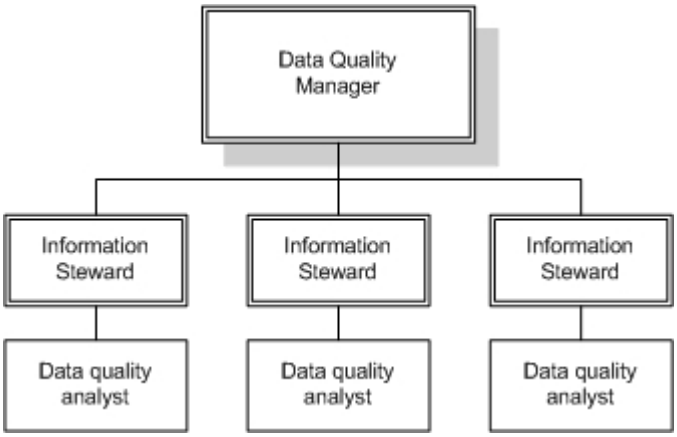

Figure 5.19 – Data quality assurance team - organizational chart

In order to assure and verify that this awareness is effective, project management creates a data consistency assurance team (cf. Figure 5.19) with a data quality manger that leads it. Main responsibility of the DQ manager is to establish and adapt data consistency methods and standards. He additionally gathers and tracks data consistency issues as well as assures that these issues are solved. Within a business area, there is at least one information steward that ensures optimal data consistency. He understands business information and can therefore assure correct and precise specifications. He defines data consistency targets within his business area and monitors them. At least one data quality analyst supports each information steward. His responsibility is mainly to report, correct, and measure data consistency management. This team verifies that data quality in source systems corresponds to the quality required in the data warehouse, verifies data completeness and the level of perfection of their specification. It means that they assure common semantics of required data and their unambiguous understanding. In addition, they verify the origin of data together with the way and purpose for which this data will be used. The role of members of the data consistency assurance team is to make sure that a specific level of data quality among all subject areas inside the business data model is reached. In addition, they assure the quality of interdependent entities and relationships between them. This team remains unchanged throughout the whole process, from the beginning of the project,

until the end of BDM phase.

The consistency assurance team prepares before each phase (business requirements definition, data requirements collection and business data modeling phase) examples of how consistent results look like and presents them at the beginning of each of these project phases. Members of the data consistency assurance team afterwards, verify the consistency just before a sub-element (business requirement, data requirement, or subject area) is finished. With every new closed sub-element, they check the consistency between all previously closed sub-elements. This is especially important in the business data modeling phase, where data should not be redundant and where subject areas are interconnected. Therefore relationships between them have to be modeled in a consistent way. The consistency assurance team cooperates all the time with respective teams (BR team, DR team, or modeling team).

Sections that discuss business requirements definition (cf. Section 5.2), data requirements collection (cf. Section 5.3) and business data modeling (cf. Section 5.5) provide among others additional information on data consistency issues.

Our business data modeling methodology covers topics that can lead to data quality issues. It aims to minimize their impact by providing recommendations and optimizing processes of business requirements and data collection as well as their structurization into a business data model. Members of the data consistency assurance team verify that the design of the BDM excludes redundant storage. In case they identify at the first sight identical data coming from multiple data sources, a more profound analysis and workshops or JAD sessions with business experts and IT professionals will be scheduled. This way they will identify the data source that suits our business requirement best. In case they discover an inconsistency in semantics of business elements in different business departments of the bank, they discuss these issues during workshop sessions. They aim here either to solve different wording problem or they identify and specify separate business terminology elements. For consistency purposes, data formats are consolidated as well. In order to assure specific data quality level, they associate data to specific business requirement so that they assure that it is being used in the way intended for its purpose. If the data is used in intended way and for originally intended purposes, the correctness of reports this data is involved in have increased reliability. In case of incoherence within data sources, the data consistency assurance team have to correct concerned data directly in source systems so that this one does not propagate into other systems that use it.

Figure 5.20 illustrates our final methodology.

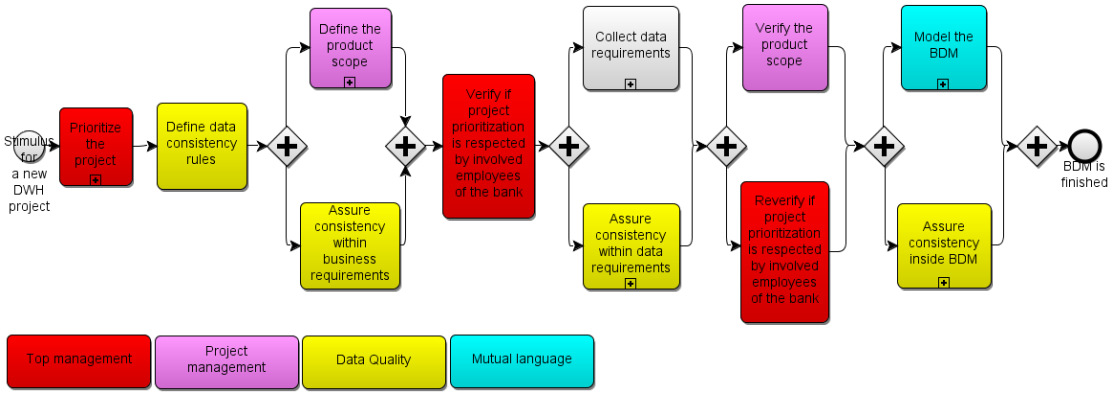

Figure 5.20 – Final business data methodology

The process starts with a stimulus for a DWH project within a financial institution. The top management of this institution executes the first task. They prioritize the project among other ongoing, scheduled and planned projects within the institution (cf. Figure 5.3). When the top management approves the project, the process continues with the definition of data consistency rules (cf. first two tasks in the Figure 5.18). The task of scope definition (cf. Figure 5.9) and consistency assurance within business requirements (cf. sequence number one in the Figure 5.18) are executed in parallel. When these two tasks are finished, the central unit verifies if the project prioritization is respected. Collection of data requirements (cf. Figure 5.13) and consistency assurance within these requirements (cf. sequence number two in the Figure 5.18) are as well executed in parallel. The project management afterwards re-verifies the product scope (cf. Figure 5.14) at the same time as the central unit verifies once again if the project prioritization is respected. When these two tasks are completed, the modeling team creates the business data model (cf. Figure 5.15) at the same time as the assurance team assures the data consistency inside it (cf. sequence number three in the Figure 5.18).

This chapter closes the business data modeling part as well as the main part of the master thesis. The following chapter concludes results of the phase validation (cf. Subsection 3.2.4) as well as discusses limitations and future work to be conducted in order to enrich currently defined methodology.

CHAPTER 6

Conclusions and Future Work

6.1 Evaluation

This section summarizes results of the final research stage (cf. Subsection 3.2.4).

The final survey we shared with a homogeneous group of eleven respondents provides evaluation of our prescriptive methodology. It allows to asset the ability of the methodology to reduce the impact of crucial influence factors (cf. Table 5.1) by evaluating following statements with help of the Likert scale.

1. The process illustrated on the Figure 5.3 can help to reduce the impact of the factor "Missing clear set of goals and priorities" in financial DWH projects.

2. The process illustrated on the Figure 5.18 can help to reduce the impact of the factor "Insufficient data consistency" in financial DWH projects.

3. The process illustrated on the Figure 5.9, Figure 5.10 and Figure 5.14 can help to reduce the impact of the factor "Unclear scope definition" in financial DWH projects.

4. The process illustrated on the Figure 5.15 and Figure 5.17 can help to reduce the impact of the factor "Language gaps between designers and technical actors" and "Insufficient data specification/definition leading to interpretation errors" in financial DWH projects.

5. The methodology illustrated in Figure 5.20 can help to reduce the impact of crucial factors (cf. Table 5.1) identified during the second stage of the research process (cf. Subsection 3.2.2).

The Likert scale is graded according to the Table 6.1:

Likert Scale	Grade
Strongly disagree	1
Disagree	2
Neither agree nor disagree	3
Agree	4
Strongly agree	5

Table 6.1 – Grading of the Likert scale

Table 6.2 summarized final grades of the statements:

Statement	Grade
1	4.1
2	4
3	4.3
4	4.3
5	3.9

Table 6.2 – Phase validation results

These results suggest that the methodology we introduced in this master thesis can reduce the impact of crucial factors in financial data warehousing.

However, participants observed a number of limitations and suggested following improvements:

1. Additional support of goals setting by top management and involved business departments could be included in the prioritization process. Only communication of goals is not sufficient.

2. Goals and priorities are different dimensions. They should not be consolidated into one statement. Moreover, the process described could cover the priority issue, but does not cover the goal issue.

3. Communication of project priority/importance for the enterprise resp. organization as well as enforcement of participation (if necessary) are ongoing senior management tasks.

4. Top management process does indeed help to set the goals and priorities among the current project landscape. Goals and priorities within the BDM are not touched.

5. Advices to overcome common types of conflict would be even more helpful. Furthermore, there should be a loop for prioritizing the content/information/goals. Right now there's only a yes/no decision on the project and a prioritization loop against further projects, but no project internal loop which also influences the later priority against other projects.

6. Harmonization and definition of quality requirements for the business requirements should be included in the methodology.

7. Depending on the three phases of DWH project as described in the data consistency connected process, the core to reduce the impact of data consistency factor in every sub process is to describe, present and agree on the requirements of DQ based on illustrative examples. Measures seem to be high-level and generic. They only aim to provide examples and assure a verification step. That is usually the case, but does often not solve the issue.

8. Data Consistency process is basically correct, but rather superficial. Acceptance of separate quality assurance team might be doubtful in practice, often perceived as “administrative burden”.

9. The process of establishing data consistency assures explicit validity of project results on an organizational level. Another interesting question would be how to find the most efficient amount of effort to spend on these processes (economic perspective).

10. The process of scope definition mainly focuses on the product scope which is of course the most important scope to be defined. Scope definition of related subject areas (like customer or market data) are therefore only touched indirectly as “needed to be covered for a complete product representation”.

11. These processes help to reduce the risk of unclear scope. There are however two measures which help additionally. First, a formalized sign-off of the requirement (real signature on paper) from senior executives. Second, in many cases misunderstandings are caused because business cannot imagine what they are going to get when the project is finished. The deliverable of a DWH project is too abstract to them. Therefore examples, non-working prototypes, pilots etc. should be used to reduce this risk.

12. The process of common language establishing only partly cover factor "Language gaps between designers and technical actors" and "Insufficient data specification/definition leading to interpretation errors", since they focus on the definition phase. The second issue is the common understanding of the definition; this would require a sort of knowledge management within the total project from design to implementation, to ensure the transfer of the know-how between the different project stages.

13. These processes will help to clarify identified problems and help to raise the awareness for misunderstandings in the team. There is no impact on additional identification of misunderstandings (e.g. in synonyms or homonyms).

14. Of course it is useful to do all the tasks listed in the final methodology. But it is only one of many similar helpful processes we can choose. In every (customer)-situation one has to evaluate what process is useful. E.g. it may not be necessary to prioritize the project if there is enough budget and the right people on board and so forth.

15. These factors, which seem to summarize the processes defined before, cover only the BDM of a DWH. The BDM is only a small part of the DWH project. This methodology therefore helps to impact the success of a BDM project but it does not impact the success of a total DWH project. There are many other factors, which would have to be considered, if a total financial DWH is in question.

16. The final process is straight-forward without possibilities of looping back in the process. Experience from practice: loops back to previous steps occur during the whole process.

Remarks from final survey participants can be partially explained by the scope of the thesis. We covered only the priority issue in the thesis (cf. point number two and four). Data quality requirements are not included in the business requirement definition (cf. point number six) in our processes. The thesis focuses uniquely on product scope (cf. point number ten).

Other remarks of final survey participants are explained in the respective chapters of the thesis. The prioritization is monitored by the organizational unit (consisting of top management members). This supports the statement that this task is an ongoing senior management task (cf. point number three). Internal loop within the project for prioritization purposes (cf. point number five) is partially solved in the final process, where the respect of prioritization is re-verified after every project phase (cf. Top management tasks in Figure 5.20). Even if the data consistency supporting process is rather superficial (cf. point number seven and eight), additional activities taken into account are described in the thesis chapters (cf. Section 5.6). Even if the DQ team is considered an "administrative burden" (cf. point number eight), it is not completely independent structure. It contains field expert resp. other project members. The know-how hand-over (cf. point number twelve) is assured by involvement (to a maximal extent) of the same project members in different project development phases.

However some issues should be introduced into the methodology in the future. The process should be enhanced by additional prioritization within the project itself (cf. point number five). Economic side of data consistency should be included (cf. point number nine). The effort invested in this process should not have a tremendous impact on the budget of the project. Additional implicit harmonization of business requirements (cf. point number six) can help to create even more reliable product scope and smoothens the whole development of the project development. If the understanding is unified, than interpretation conflicts are reduced. Formal sign-offs and generation of prototypes and other example types (cf. point number eleven) allows to make the scope more reliable in the future. The methodology should in the future support the whole DWH project development (cf. point number 15) and it should provide a parametrization of issues to be implemented and the level of detail of these activities (cf. point number 14). The process should be likewise supported by more repetition and re-entrance loops (cf. point number 16).

6.2 Limitations and Future Work

Limitations and possibilities for future research fall into the following categories: factors dependency, human factor, and data quality.

First, the methodology could involve prevention and correction tactics, that decrease the impact of more than five factors. In addition, before a project starts, an analysis of factors that have a potential to occur in a project can be executed. The methodology can be afterwards tailored for consideration of these specific factors. This means that we can choose during the project development what actions we want to take and what factors we want to minimize in the context of specific project circumstances.

Second, we did not consider the human factor in our methodology or only marginally. Even if we manage to create a "perfect" business data modeling methodology, the project can still fail due to human factors. Reasons for this failure can be various. Either the staffing is insufficient or involved project members have insufficient knowledge to be able to execute allocated tasks. This problem can be partially solved by an appropriate training. Another topic that would need to be taken into consideration is motivational techniques and techniques that allow to increase the level of engagement of project members. This is partially related to staffing and selection of project members, but can also additionally increase motivation and engagement level in a carefully selected team that produces high quality results. Another aspect of human factor is the communication that is the better suitable for the task execution. This communication can be seen on the one hand as communication of task that project members need to execute or the communication within project teams or on the other hand the communication model such as workshops, interviews or other, that suits the best and is the most efficient in a given context.

Finally, data quality should be in the future extended to its other aspects, not only consistency. Necessary extensions would be for example compliance with regulatory standards or analysis of return on investment to data management implementation from market or other point of view. Proper data quality management should involve all parts of the bank and should for instance start in the front office, where employees need to create correct, complete and reliable data about customers, products, services, etc. Tactics that assure data quality at this level are not being developed in the current work. Once front office data has satisfactory quality, they must be correctly inserted into source systems and source system have to provide satisfactory transactional and application data. Our methodology does not concentrate on the quality of these activities, since we are not developing the whole data quality life cycle, but we only want to assure data consistency at the business and data requirements and business data model level.

The goal is to sustainably improve and ensure overall data quality of all systems and therefore of generated business reports.

The methodology has been defined in the context of the financial environment. It is therefore specific for this industry. The reuse of the methodology or its part in an industry of similar complexity might be possible. It however needs to be verified using appropriate research techniques.

6.3 Conclusions

In the present master thesis, we introduced a business data modeling methodology, that helps to minimize the impact of critical issues in the financial data warehousing. This methodology aims to answer the following research question: Can a structured methodology for the early development phase of financial DWH projects help prevent failure of these projects?

Even though the business data modeling is the base for the further correctness of the financial DWH implementation process, a methodology that covers only the BDM can not prevent the failure of the whole DWH project. If however the BDM is defined exhaustively, it simplifies a lot the tasks of members of subsequent project phases.

CHAPTER 7

Glossary

BANK
Definition:
Big bank in Austria, which's BDM project was the object of case study within the thesis. The name of the bank is not revealed due to confidentiality issues.

BDM
Meaning:
Business Data Model

BR
Meaning:
Business Requirement

BR Team
Meaning:
Business requirements definition team
Definition:
The team of project members that's role is to define business requirements of the bank as a whole (within the high-level scope of the project). BR team consists of heads of departments or application managers that act actively in business areas of the bank, external consultants and project managers or their substitutes.

Business analyst
Definition:
Business analyst understands to a certain extent the point of view of business and at the same time IT side. He enhances the communication between these two mindsets by building a bridge in-between.

Business entity
Definition:
Business entity represents a business concept such as Customer, Physical Person, Legal Person, Loan, Fixed Deposit or Current Account.

Crucial factor
Definition:
One of the five factors we aim to minimize the impact in the thesis (cf. Figure 7.1).

Decisive body
Definition:
Part of the BR (or modeling) team that's role is to animate workshops (or JAD) sessions. They have the responsibility of delivering business requirements within one business area (or business data model of a subject area) at a satisfying level of quality.

DQ
Meaning:
Data Quality

DSS
Meaning:
Decision support system

DWH
Meaning:
Data WareHouse

EDWH
Meaning:
Enterprise Data WareHouse

ERD
Meaning:
Entity Relationship Diagram

Factor
Definition:
The 31 factors that we face during development of financial DWH projects (cf. Figure 7.1).

GDWH
Meaning:
Group Data WareHouse
Definition:

Old DWH solution at the BANK, that provided data for risk analytical purposes.

High-level scope of the project
Definition:
Set of project objectives, goals and expectations. High-level scope is written by top management and is delivered to project management when the project's existence is approved.

JAD
Meaning:
Joint Application Design
Definition:
A form of workshop; they are interview sessions with the appropriate personnel in various departments.

Modeling team
Definition:
The team of project members that's role is to model the BDM as a whole.

PMO
Definition:

Project Management Office
Meaning:

Roles and responsibilities of this organizational structure lie in execution and support of tasks connected with project management.

SOLUTION
Definition:
Commercial DWH solution that was implemented at the BANK for EDWH.

Subject Area
Definition:
Subject area represents an area of highly interconnected business entities such as Customer, Bank's Product or Limit.

5 W's
Definition:
Definition of Who, What, When, Where and Why within business processes.

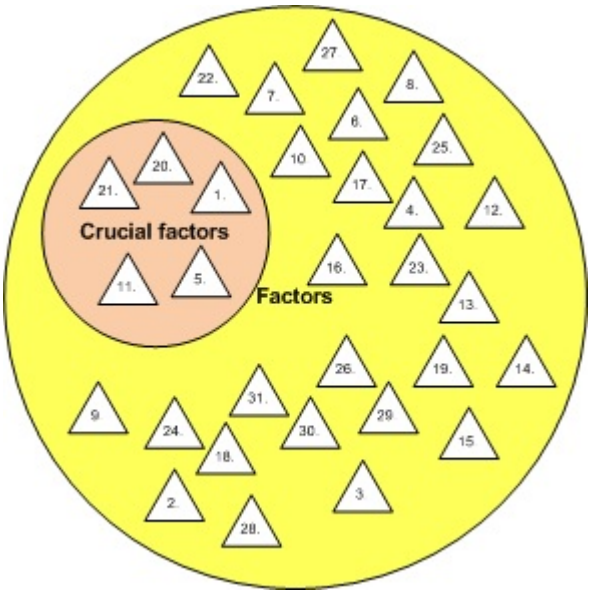

Figure 7.1 – Graphical representation of factors and crucial factors

Bibliography

Anderson, J. (1990), *Cognitive Psychology and Its Implications, 3rd edition*, W.H. Freeman and Company. (Not cited.)

Ashcraft, M. (1989), *Human Memory and Cognition*, Scott, Foresman and Company. (Not cited.)

Berg, B. L. (2004), *Qualitative research methods for the social sciences*, Pearson Education, Inc. (Cited on pages 17 and 18.)

Breslin, M. (2004), 'Data warehousing battle of the giants: Comparing the basics of the kimball and inmon models', *Business Intelligence Journal* pp. 6–20. (Cited on pages 3 and 7.)

Carneiro, L. and Brayner, A. (2002), 'X-meta: A methodology for data warehouse design with metadata management', *Proc. of the 4th Intl. Workshop on Design and Management of Data Warehouses (DMDW'02* . (Not cited.)

Collins, A. and Quillian, M. (1972), 'How to make a language user', *Organization of Memory* pp. 309–351. (Not cited.)

Eckerson, W. (2003), 'Smart companies in the 21st century: The secrets of creating business intelligence solutions', *TDWI web site* . (Cited on page 15.)

Fals-Borda, O. and Rahman, M. A. (1991), *Action and Knowledge: Breaking the Monopoly With Participatory Action Research*, Intermediate Technology / Apex. (Cited on page 18.)

Galtung, J. (1967), *Theory and methods of social research*. (Cited on pages 19, 24, and 28.)

Golfarelli, M., Rizzi, S. and Turricchia, E. (2011), 'Modern software engineering methodologies meet data warehouse design: 4wd', *Lecture Notes in Computer Science* **6862**, 66–79. (Cited on pages 1 and 15.)

Haughey, T. (2004), 'Is dimensional modeling one of the great con jobs in data management history?'. (Cited on page 8.)

Huang, S., Chou, T. and Seng, J. (2007), 'Data warehouse enhancement: A semantic cube model approach', *Information Sciences* **117**, 2238–2254. (Cited on pages 1 and 15.)

Hwang, H., Ku, C., Yen, D. and Cheng, C. (2004), 'Critical factors influencing the adoption of data warehouse technology: a study of the banking industry in taiwan', *Rep. Prog. Phys.* **37**, 1–21. (Cited on pages 19, 20, and 21.)

Inmon, B. (1992), *Building the Data Warehouse*, Wiley Publishing, Inc. (Cited on page 1.)

Inmon, B. (2000), 'The problem with dimensional modeling', *DM Review Magazine* . (Cited on page 8.)

Inmon, B. (2005), *Building the Data Warehouse – Fourth edition*, Wiley Publishing, Inc. (Cited on pages 1, 3, 4, 5, 6, 9, 10, 11, 12, 13, 14, 15, 63, and 68.)

ISIS (Summer Term 2012), *Course material written by team of 188.924 Workflow Modeling and Process Management*, TU Wien. (Cited on page 54.)

Jarvenpaa, S. L. and Machesky, J. J. (1986), 'End user learning behavior in data analysis and modeling tools', *Association for Information Systems* pp. 152–167. (Not cited.)

Judd, C. M., Smith, E. R. and Kidder, L. H. (1991), *Research methods in social relations: Sixth edition*. (Cited on pages 21, 24, and 41.)

Juhn, S. H. and Naumann, J. (1997), 'The effectiveness of data representation characteristics on user validation', *Association for Information Systems* pp. 212–226. (Not cited.)

Jukic, N. and Nicholas, J. (2010), 'A framework for collecting and defining requirements for data warehousing projects', *Journal of Computing and Information Technology – CIT* **18**(4), 377–384. (Cited on page 1.)

Kimball, R. (1996), *The Data Warehouse Toolkit*, Wiley. (Cited on page 1.)

Kimball, R. and Ross, M. (2002), *The Data Warehouse Toolkit: The complete guide to dimensional modeling - Second Edition*, John Wiley and Sons,Inc. (Cited on pages 3 and 7.)

Leitheiser, R. and March, S. (1996), 'The influence of database structure representation on database system learning and use', *Journal of Management Information Systems* **12**(4), 187–213. (Not cited.)

Masiero, S. (2010), 'Design-intentionality gaps: Explaining expectation failure of information systems in developing countries'. (Cited on pages 19, 20, and 21.)

McGee, W. C. (1976), 'On user criteria for data model evaluation', *ACM Transactions on Database Systems* **1**(4), 380–387. (Not cited.)

Nordbotten, J. and Crosby, M. (1999), 'The effect of graphic style on data model interpretation', *Information Systems Journal* **9**(2), 139–156. (Not cited.)

Ojeda-Castro, A., Ramaswamy, M., Rivera-Collazo, A. and Jumah, A. (2011), 'Critical factors for successful implementation of data warehouses', *Issues in Information Systems* **12**(1), 88–96. (Cited on pages 19, 20, and 21.)

PMBOK (2013), *A Guide to the Project Management Body of Knowledge (PMBOK®Guide) Fifth Edition*, Project Mnanagement Institute, Inc. (Cited on pages 52, 56, 62, and 63.)

Reason, P. (1994), 'Co-operative inquiry, participatory action research and action inquiry: Three approaches to participative inquiry', *Handbook of Qualitative Research* pp. 324–339. (Cited on page 18.)

Schramm, W. (1971), 'Notes on case studies of instructional media projects.', *Working paper, the Academy for educational Development* . (Cited on page 24.)

Schuff, D., Corral, K. and Turetken, O. (2005), 'Comparing the effect of alternative data warehouse schemas on end user comprehension level'. (Cited on page 8.)

Tan, J. K. H. and Benbasat, I. (1993), 'The effectiveness of graphical presentation for information extraction: A cumulative experimental approach', *Decision Sciences* **24**(1), 167–191. (Not cited.)

Todd, P. and Benbasat, I. (1992), 'The use of information in decision making: an experimental investigation of the impact of computer based decision aids', *Management Information Systems Quarterly* **16**(3), 373–393. (Not cited.)

Watson, H. and Ariyachandra, T. (2005), 'Data warehouse architectures: Factors in the selection decision and the success of the architectures'. (Cited on pages 19, 20, and 21.)

Weir, R., Peng, T. and Kerridge, J. (2003), 'Best practice for implementing a data warehouse: A review for strategic alignment', *Proceedings of the 5th Intl. Workshop (DMDW'2003), H.-J. Lenz, P. Vassiliadis, M. Jeusfeld and M. Staudt (eds.), Berlin, Germany* . (Cited on pages 19, 20, and 21.)

Wells, D. (2003), 'Choosing the right data warehousing approach', *TDWI FlashPoint* . (Cited on page 8.)

Wixom, B. H. and Watson, H. J. (2001), 'An empirical investigation of the factors affecting data warehousing success', *Management Information Systems Quarterly* **25**(1), 17–41. (Cited on pages 19, 20, and 21.)

Yin, R. K. (1994), *Case Study Research: Design and Methods*, SAGE Publications, Inc. (Cited on page 24.)

Yin, R. K. (2009), *Case Study Research: Design and Methods*, SAGE Publications, Inc. (Not cited.)

Zeid, A. (2006), 'Your bi competency center: A blueprint for successful deployment', *BUSINESS INTELLIGENCE JOURNAL* **11**(3), 14–20. (Not cited.)