An Information Requirements Collection and Analysis Model for Business Intelligence

Derk Pieter Busser

Faculty of Technology, Policy and Management

Delft University of Technology

A thesis submitted for the degree of

Master of Science

Systems Engineering Policy Analysis and Management

September 2011

1. Chair: Professor Dr. Yao-Hua Tan, TBM

2. First Supervisor: Dr. Ir. Jan van den Berg, TBM

3. Second Supervisor: Dr. Jos Vrancken, TBM

4. External Supervisor: Dr. Ir. Pieter Heij, De Bijenkorf

Signature from head of Msc committee:

Abstract

Incorrect and incomplete information requirements are the amongst the main reasons of business intelligence development project failure. Current methods for information requirements eliciting fail to gather complete sets of requirements with sufficient quality. Amongst the reasons for this failure are a lack of end-user involvement, communication problems and insufficient validation and iteration of requirements.

In this thesis, a new model has been designed by using a set of arguments and requirements that was derived from current methods and models. This model uses both a structured, goal driven, and unstructured, demand driven, method to elicit information requirements. The requirements methods that are incorporated into the model complement each other. Extensive prototyping and iteration is included in the model to validate the set of requirements. The process that is supported by the model is focussed on communication, transparency and stakeholder involvement.

The proposed model was validated by performing a case study at De Bijenkorf e-commerce department. The case study confirms the need for user involvement, communication and validation in the early phase of the business intelligence development lifecycle. The model has shown the ability to support stakeholders in the collection and analysis of a correct set of requirements of sufficient quality.

The model can be further developed by adding project management and documentation to get more transparency and to reduce complexity with regard to the process.

Acknowledgements

First I would like to thank my Master Thesis Commission. With a special thanks to first supervisor Jan van den Berg for supporting my Master and Bachelor thesis. Thanks to Jan I was introduced into the subject of Business Intelligence. Unlike many other subjects during my studies, business intelligence truly interests me. Therefor business intelligence is now the direction in which I will continue into the corporate world. Also, thanks to Jos Vrancken, second supervisor, who helped me getting more focussed and precise.

I would like to express my gratitude to my supervisor from de Bijenkorf, Pieter Heij. Thanks to Pieter I found this thesis subject. He supported me by contributing knowledge, hardware, software and all necessities to successfully complete my master thesis. Moreover, I can thank Pieter for my current job at de Bijenkorf. Furthermore, i would like to thank all employees of the e-commerce department for their participation during the case study and the numerous table soccer matches in between the hard work.

Last, but not least, I would like to thank Pietjeliefste, my family and friends for support.

Derk Busser, Amsterdam, The Netherlands, September 3, 2011

Contents

List of Figures

LIST OF FIGURES

List of Tables

LIST OF TABLES

1

Introduction

De Bijenkorf, a large Dutch retailer re-opened their online store in 2009. The web shop has produced a growth in turnover of three hundred percent annually since. To maintaining this steep growth, intelligent decisions need to be made based on insights derived from operational data. Currently, different data sources are used separately to gain insights into the operation. Most analysis is done on an ad-hoc basis, using tools such as Access, Excel, Google Analytics and Business Objects. The data that is used for these tools is dispersed through out the organization. Because of the current situation, at De Bijenkorf, the analysis of information for decision making is a difficult, time- and cost consuming process. Most of the employees within the organization are not trained or hired to perform this work. The manager of de Bijenkorf e-commerce department wants to develop a business intelligence system that provides actionable insights and intelligence to all employees that supports them in making sound decisions to achieve their goals.

1.1 Business Intelligence

Business Intelligence is a broad term that can be explained from different viewpoints. For example the strategic viewpoint has been defined by [Rouibah and Ould-ali, 2002] as:"Business Intelligence is a strategic approach for systematically targeting, tracking, communicating and transforming relevant weak signs into actionable information on which strategic decision-making is based.", others define Business Intelligence from a more practical viewpoint such as BI vendor Siebel:"Business Intelligence is a solution

suite that integrates data from multiple enterprise sources and transforms it into key insights that enable executives, managers, and front-line employees to take actions that lead to dramatic improvements in business performance". A similar viewpoint is taken by IBM:"Business Intelligence is a concept of applying a set of technologies to turn data into meaningful information. With Business Intelligence Applications, large amounts of data originating in many different formats (spreadsheets, relationship databases, web logs) can be consolidated and presented to key business analysts.., and armed with timely, intelligent information that is easily understood, and the business analyst is enabled to affect change and develop strategies to drive higher profit." and: "Business intelligence means using your data assets to make better business decisions. It is about access, analysis, and uncovering new opportunities." [Chang et al., 2006]. The information and intelligence that is gained through the use of BI systems is derived from structured, unstructured, internal and external data [Negash, 2004] that is stored in a large variety of data sources.

Business Intelligence is an umbrella term for a collection of methods, technologies and applications. Business Intelligence combines and enriches structured and unstructured data into information and insights. This information is accessible on a strategic, tactical and operational level. Business Intelligence has been developed over a period of 40 years, an overview can be seen in figure 1.1 by [Chang et al., 2006]. It can be seen that in the previous century the focus was on internal organizational data. BI systems and technologies focussed largely on strategic and tactic levels. In the last decade a focus shift has been made by including external data to complement the quality of BI systems. Due to technological growth and maturity, more work has been done in the use of BI systems by operational users and systems. An example of this are recommendation systems that receive intelligence by structuring data through semantics. [Moss and Atre, 2003] presents a list of more than twenty types of applications within business intelligence, including data mining, forecasting, business analysis and online analytical processing.

The development of business intelligence systems is a complex activity. In 57% of BI projects, the development time is longer than one year, the average length is 18 months [Gan, 2004]. The failure rate of BI projects, as found in case studies and literature, ranges between 70% and 80% [Goodwin, 2011]. Failure is here defined as all projects that are cancelled, over time, over budget and not according to the set of requirements

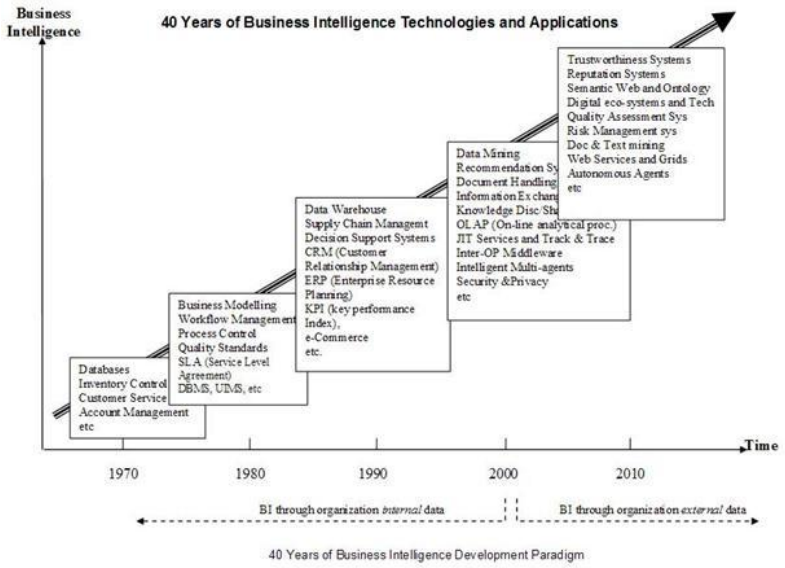

Figure 1.1: 40 years of Business Intelligence by [Chang et al., 2006]

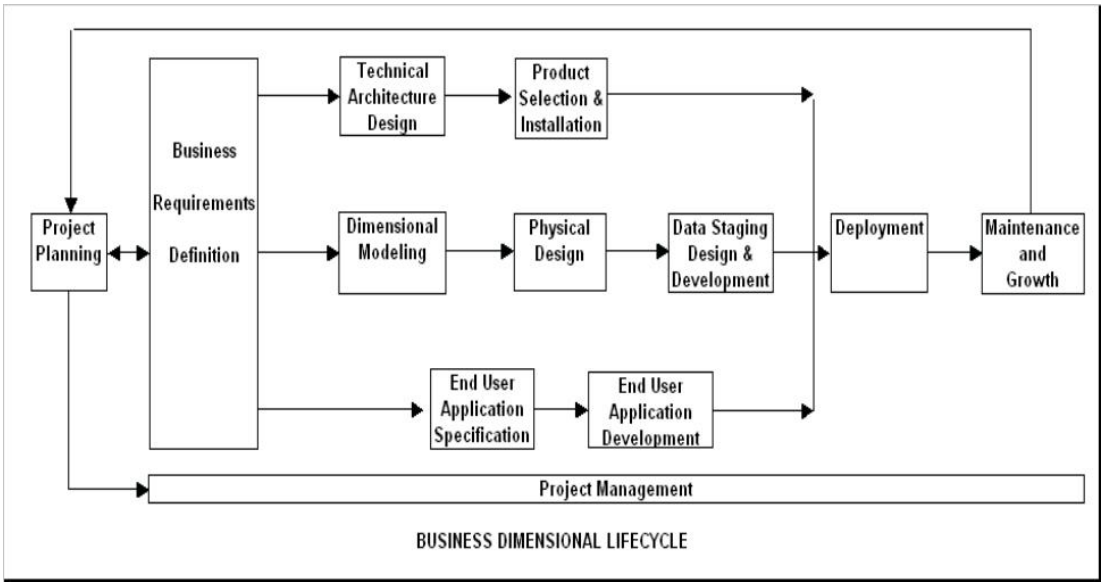

Figure 1.2: Business Dimensional Lifecycle by [Kimball et al., 1998]

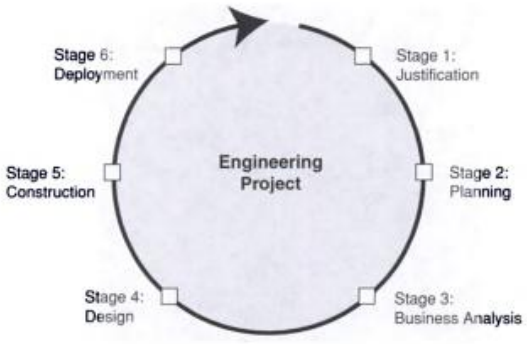

Figure 1.3: BI Development Lifecycle by [Moss and Atre, 2003]

defined. Although the definition of failure differs between studies, significant difficulties and complexities are found during development projects.

Business Intelligence development lifecycle models are similar to systems engineering and software development lifecycles. In figure 1.2 the lifecycle of business intelligence development by [Kimball et al., 1998] is visualized. Figure 1.3 by [Moss and Atre, 2003] gives a more generalized view of the business intelligence development lifecycle. During all phases of the development lifecycle, defects might arise that can cause projects to fail. In Information Technology the costs of changing systems due to defects grow exponentially when going deeper into the life cycle, this can bee seen in figure 1.4 by [Hood et al., 2008]. This was first presented by [Boehm, 1984] and is now considered common knowledge by the industry and academia. Even when business intelligence systems are operational, changes often need to be made, according to [Gan, 2004]: "75% of reports and custom applications will need to be changed within 6 months of deployment". Defects and problems that cause the need for these changes can occur in all phases of the development lifecycle.

[Goodwin, 2011] claims that in most cases, failure comes from"A combination of poor communication between IT and the business, the failure to ask the right questions or to think about the real needs of the business." Communication and ill-defined requirements are a common source of failure in IT projects. The engineers in IT projects view the project from a technological perspective and are not aligning the system with the end-users and the organization. According to [Ko and Abdullaev, 2007]:"there is always gap between the actual implementers and the users of the system." [Berry and

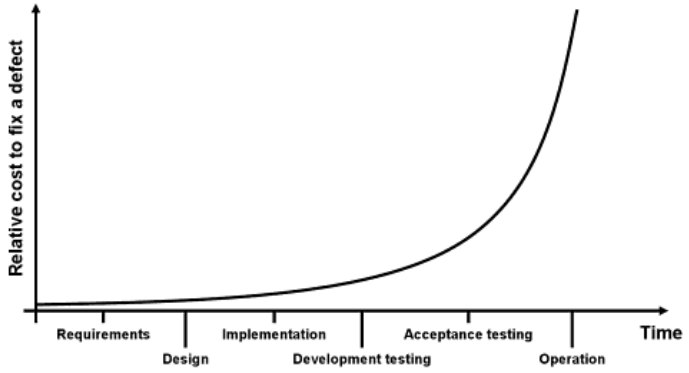

Figure 1.4: Costs of defects caused by wrong requirements by [Hood et al., 2008]

Linoff, 1999] also emphasizes this aspect of data warehouse development:"Defining the business problem is the trickiest part of successful data mining because it is exclusively a communication problem". Moreover a source of failure lies within the focus of the BI project according to [Ko and Abdullaev, 2007]:"Mostly BI projects considered to be the inner project which should meet the companies internal requirements. However it should reflect the needs of customers and deal with external issues like market situation or customers behavior."

Gathering information requirements is complex, [Davis, 1982] defines the three main obstacles for eliciting a correct and complete set of information requirements from organizations:

1. The constraints on humans as information processors and problem solvers

2. The variety and complexity of information requirements

3. The complex patterns of interaction among users and analysts in defining requirements.

Business Intelligence failure can arise in all phases of the development lifecycle, formulating a correct and complete set of requirements early in the process can reduce the chances of failure during further stages of development.

1.2 Research Goals

De Bijenkorf web shop has grown very fast in the last two years, a large amount of data is collected from the online operations. This data can be used to gain more intelligence and can be used to support decision making in the e-commerce department. In literature it has been seen that failure of Business Intelligence projects comes from setting the wrong requirements and a lack of communication between developers and end-users. Therefor, the design objective of this master thesis project is:

> ***Developing an Information Requirements Collection and Analysis Model for Business Intelligence***

Business Intelligence projects are developed over a long period. Depending on the size and scope, these projects can become complex. Complexity causes problems throughout the project and can lead to failure in all phases of the lifecycle. Information requirements are a common source for these problems and defects. Eliciting a complete and correct set of requirements can reduce the chance of failure during business intelligence projects. Currently, information requirements methods that have been developed for Business Intelligence fail to gather a set of qualitative requirements that represent the information needs that exist within organizations. The scope of this research within the lifecycle of business intelligence development is the information requirements phase and its position within the Business Intelligence lifecycle. A new information requirements collection and analysis model can improve the business intelligence development lifecycle.

During development, the needs of the end-users change because of changing perceptions and growing experience with business intelligence. The feedback and iteration that is needed to cope with changing needs involves information requirements and is thus also addressed during this research. A Business Intelligence system is intended to be flexible (not exhaustively) in that respect. The system requires a certain level of scalability [Labio et al., 1997b] to add new sources of data and to support new means of analysis on strategic, tactical and operation levels.

The design objective can be split up into multiple sub objectives and questions that support reaching the overall objective.

1. What are current methods for the gathering of information requirements for Business Intelligence systems?
   (a) What methods exist within Business Intelligence development frameworks?
   (b) What best practices are used in business with respect to information requirements gathering?
   (c) What methods exist in general information technology development?
   (d) What shortcomings can be found in current information requirements methods an models?
2. Design a model for information requirements analysis
   (a) What are the design requirements for an Information Requirements Collection and Analysis Model for Business Intelligence
   (b) Design the Model
   (c) How should the new method be applied in Business Intelligence Development Cycles
3. Validating the Method
   (a) What are the criteria for validating the information analysis model?
   (b) Can the model be theoretically tested by using the set of criteria?
   (c) Is the model valid according to the assessment of the case study using the validation criteria?

1.3 Methodology

A literature study has been done to answer the first two research questions. The research focus was on relevant literature in the field of Business intelligence. Research in conventional requirements analyses methods areas has also been reviewed to find possible new insights. Because Business Intelligence is multi-billion dollar industry, commercial sources have also been addressed.

The outcome of the literature research has been used as the basis for a set of requirements for a new information requirements collection and analysis model. The

development of this model was done by using design research as presented by Hevner et al. [2004]. In his article, Hevner presents seven guidelines for design research. The first three are important and explicitly used during this research. The first is the guideline to design an artifact, in this case a model. Secondly, the artifact has to be relevant with regard to a business problem. Finally, the designed artifact has to be evaluated rigorously.

The model that was developed has been tested and validated by doing a case study at de Bijenkorf E-Commerce department. The hardware, software and data needed was supplied by the Bijenkorf and access has been given to all data sources available. The development of the architecture and prototype has been properly documented because the goal is to develop a production version of the BI solution based on the resulting prototypes.

1.4 Structure of Thesis

The structure of the thesis is aligned with the objectives and sub-objectives that have been set out in this chapter. The first chapter is the Introduction chapter. The second chapter focusses on information requirements, both in the field of business intelligence and information technology development. The chapter thereafter presents a set of arguments that are elicited from literature research that stress the need for a redeveloped requirements collection and analysis model. These arguments are then transformed into a set of requirements for a new model. In the fourth chapter, the new requirements analysis model is presented and fully described. The fifth chapter covers the evaluation of the proposed model by theoretic validation and validation in practice at the e-commerce department at the Bijenkorf. Finally chapter six gives conclusions and recommendations regarding the research and implementation of the proposed model. An overview of the structure of the thesis can be seen in figure 1.5

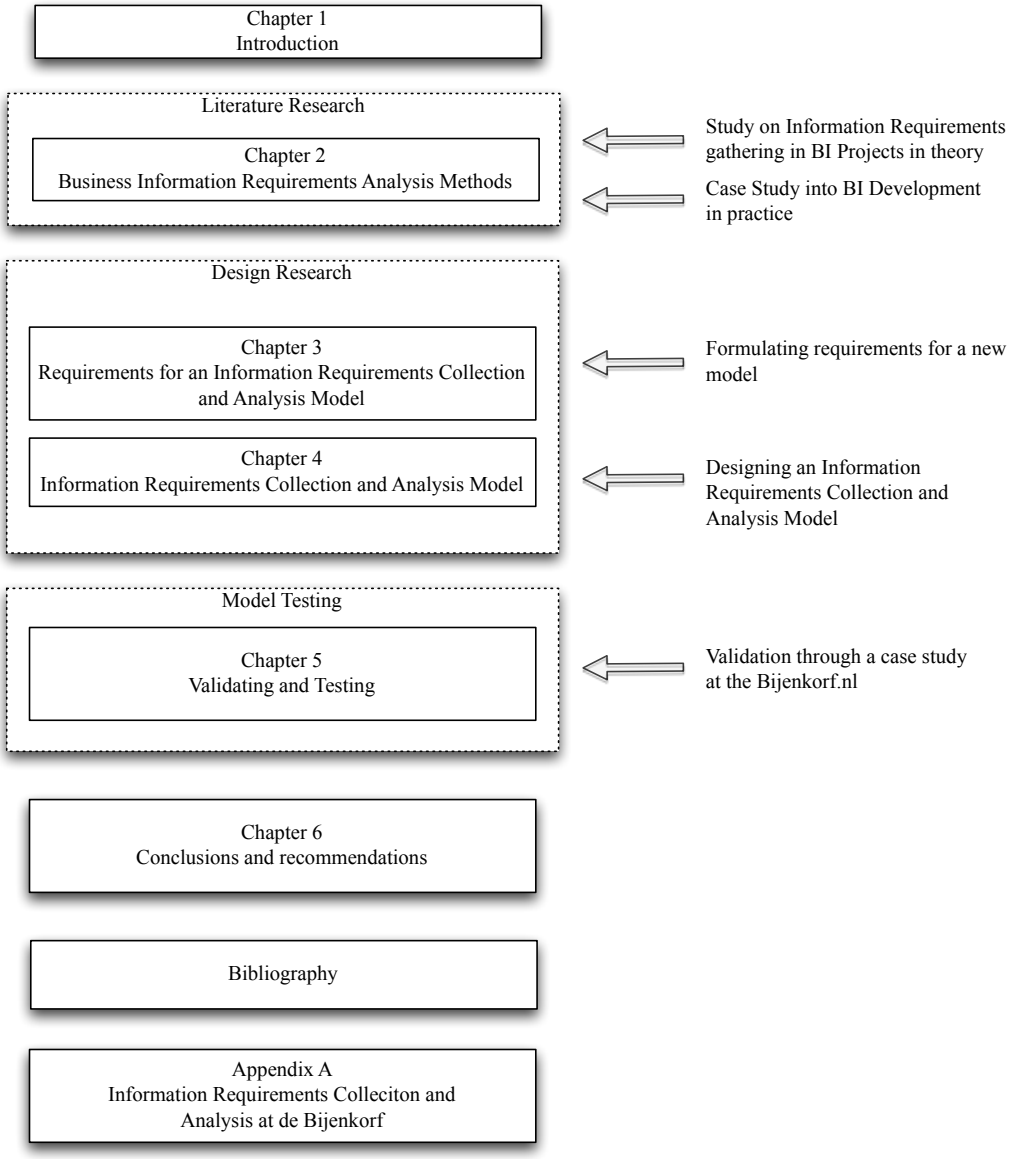

Figure 1.5: Structure of Report

2

Information Requirements Analysis Methods

2.1 Introduction

This chapter covers the current state of Business Intelligence Development and Information Requirements Analysis both in Academia and Business. The focus hereby is on methods that are used by developers. First, a general overview of business intelligence and development of such systems is explored. Following this, information requirements analysis methods are researched. Finally, best practices as found in literature and through case studies is elaborated upon. The research described in this chapter provides with a solid base of arguments for the development of an information requirements analysis method.

2.2 Business Intelligence

As said in the introduction, the methods, that are currently used, pose some problems in collecting correct information requirements. Setting wrong requirements for business intelligence, but also general information technology systems, can lead to project failure [Hood et al., 2008]. It is assumed that correct and complete requirements, in an early phase of development, significantly enhances the chance of project success.

In the first years of Business Intelligence development, two paradigms were leading, that of Ralph Kimball and Bill Inmon. The Inmon approach is characterized by being Subject-Oriented, Integrated, Non-volatile, Time-variant, and Top down, moreover

integration is done via the Enterprise Model. The data model is based on Entity and Relationships [Inmon, 1998]. Kimball preaches a Business-Process oriented approach that is bottom up and evolutionary. He uses a Dimensional Data model and integrates through conformed dimensions within that data model [Kimball et al., 1998]. A key difference is that in the Inmon paradigm, the data-warehouse is a complete artifact that can only be developed until the design is complete. The Kimball approach is more iterative and modular, the data-warehouse is a collection of smaller data marts that are connected through the data-warehouse bus. Because of this last characteristic Kimball is naturally used more in organizations as they start out with small data marts per department that are then used together in the Enterprise as the complete data-warehouse. An overview of the differences between the two approaches can be seen in table 2.1:

Criteria	Inmon	Kimball
Overall Approach	Top-down	Bottom Up
Architectural structure	Enterprise-wide DW	Data marts
Orientation	Subject	Business-process
Data Model	Entity Relationship Diagram	Dimensional, Star Schema
Primary Audience	IT Departments	End-users

Table 2.1: Inmon vs Kimball - Overview of differences between approaches

Because of the difference in the overall approach,Top-down vs Bottom up, the two researchers Kimball and Inmon use a different method for the elicitation of requirements, respectively a demand-driven and a supply driven method. These two methods reflect the difference in approach. Inmon mainly focusses on the data supply that is available, mostly ignoring the needs and requirements of the end-users. Kimball takes the end-users requirements as the main input for the data warehouse.

2.3 Current Business Information Requirements Analysis Methods

In the last decade, new methods for the setting of requirements have emerged on top of the supply and demand driven methods [Taa et al., 2010]. This can be seen in table 2.2. When requirements analysis is done, the general perception in literature is that there are still big hurdles to overcome in order to set the right requirements. These problems

Researchers	Requirements Method
(Kimball, 1996)	Demand-driven
(Inmon,2002), (Winter and Strauch,2004)	Supply-driven, Data-driven
(Winter and Strauch,2004)	Demand-driven, Requirement-driven
(Niedrite et al. ,2007), (Giorgini et al. ,2008)	Goal-driven
(Mazon et al. ,2007)	Model-driven
(Romero et al., 2007),(Skoutas et al., 2007)	Ontology-driven

Table 2.2: Information Requirements Analysis - Overview of methods by [Taa et al., 2010]

are dependent on the method being used. The following sections cover the information requirements models currently used during business intelligence development.

2.3.1 Supply-driven information requirement analysis

The supply driven method as proposed by Inmon, amongst others, supports the view that the data is leading in the development of a data warehouse. In this method, data models that currently exist in the enterprise are analyzed and the data is mapped onto a data warehouse model. User requirements are ignored in the first phase and are only included after the data model is constructed.

The first, obvious problem with this method is that end-user requirements are not addressed [Winter and Strauch, 2004]. The lack of end-user involvement causes a low acceptance rate of the system amongst stakeholders. Large resources are being wasted, because the supply of information is not matched to the demand of the users. The resulting Business Intelligence system is over-engineered [Burmester and Goeken, 2006], difficult to manage and complex for users to operate [Britos et al., 2008]. The information requirements that are derived through supply driven methods are focussed on quantitative data in existing sources. Qualitative and unstructured data are not considered in this method.

For users that are uncommon with Business Intelligence and do not require complex information or intelligence, the supply driven method and the resulting information requirements are sufficient [Golfarelli, 2008]. He notices that the method itself has a couple of strengths. Costs and implementation time are low and mainly depend on technical skills of the developers and complexity of the data. Moreover, the dimensional

model is more stable because existing data structures and models are used as a basis for the new model.

In the introduction chapter, the main reasons for data warehouse failure have been acclaimed to a lack of end-user involvement. The goal of this research has been to develop an information requirements model that can be used to gather requirements that are correct, complete and fit the needs of end-users. The supply driven method is not suitable as the basis or input for such a model. Lessons learned into the pro's and con's of this method have been taken into account while formulating the requirements for a new information requirements analysis model.

2.3.2 Demand-driven information requirement analysis

Demand-driven information requirement analysis methods are based on the requirements and needs of the end-users. [Hansen, 1997] claims that end-users are critical for the success of the data warehouse. They are the only stakeholders that can specify the business goals of the business intelligence solution. Demand-driven requirements methods analyze needs from the perspective of users and organization. The methods take into account the goals that they have set and the decisions that they need to make. The corresponding data is sought after the user requirements are collected. This makes the demand-driven method a more hybrid model, because the supply side is also taken into account.

The requirements eliciting method by [Kimball et al., 1998] is demand-driven and is performed in the following sequence:

- Identify and prepare interview team
- Select interviewees
- Schedule interviews
- Prepare interview questionnaires
- Conduct user kick-off and prepare interviewees
- Conduct business user interviews
- Conduct IT data audit interviews

- Publish interview write-ups and incorporate feedback.
- Analyze interview findings
- Document findings and review
- Publish requirements deliverables
- Prioritize and revise project scope
- User acceptance/project review

The questions asked depend on the preparations that have been made by the interviewer. A template is given in the Data Warehouse toolkit by [Kimball et al., 1998]. The interview consists of 3 main parts (excluding introduction and wrap-up) that cover responsibilities (place and position within the organization), business objectives and issues (going into goals, decisions and kpi's) and analysis requirements (what analysis is done, what data is used, etc).

Validation is done by a user acceptance review and the availability of data is checked with IT data audit interviews. The end-users have been included in the process but only by developing a deep list of requirements, further there is no feeling with the data or the final functionality of the system.

A number of researchers propose methods using demand-driven information requirements analysis. Similar to the Kimball method, the techniques to obtain these requirements include interviews, guided sessions and business process analysis. In figure 2.1 the Information Requirements Analysis and data modeling method as proposed by [Winter and Strauch, 2003] is visualized. The model is based on both the demand of information and the available supply of data, making it a hybrid requirements analysis method. A common practice in the methods that have been studied is that the user needs and available data are mapped onto a data schema. This schema, or model, is then presented for evaluation to the end-users.

Depending on the type and size of the data model the complexity will be an obstacle for end-users to truly understand the model. End-users are mainly interested in the front-end applications and the user experience that comes with these applications. When analyzing literature, this final evaluation step lacks true meaning because of the complexity for common end-users.

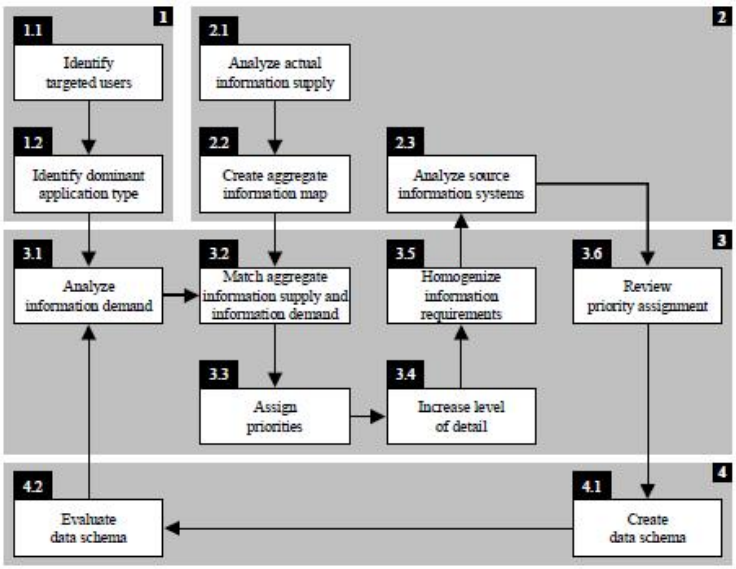

Figure 2.1: Proposed Demand-driven Information Requirements Analysis by [Winter and Strauch, 2003]

The demand-driven is characterized by its user-centric approach. Problems in this method are caused by the users and the communication with developers. [Valusek and Fryback, 1985], [Barone et al., 1999] and [Burmester and Goeken, 2006] mention a number of obstacles in the requirements elicitation process, caused by users. Communication between data warehousing architects and end-users is an obstacle. The stakeholders tend to use a different terminology, this is also confirmed by [Britos et al., 2008], [Grosser et al., 2005], [Felgaer et al., 2006] and [Cogliati et al., 2006]. The terminology used by technical stakeholders is not understood by end-users. The involvement of end-users decreases and the acceptance of the system will be lower because of this. The terminology problem is also emphasized by [Goodwin, 2011]: "They [the stakeholders] have not sat down and created a common lingua franca for BI. How can you meet expectations when you have one side speaking Latin and the other speaking Greek?"

The costs and time of development increase when end-users are involved. This is caused by growing complexity of the system and corresponding requirements. Developers have to cope with requirements from different perspectives that can be conflicting. Individual end-users also pose obstacles when deriving information requirements. The capability of end-users to define and formulate their requirements and informa-

tion needs is limited [Davis, 1982]. End-users can only give a subjective view of their business-processes, goals, decisions and analytic needs. Possible interesting insights and opportunities can be missed because they have been out of scope of current analysis. Moreover, end-users can be unaware of the possibilities of business intelligence [Golfarelli, 2008].

When information requirements are collected, the corresponding data model has to be traced back to existing data sources. During this data audit, information requirements that can not be met will be uncovered. The end-users either have to be disappointed, or other data sources have to be found that can provide the required data.

The demand driven method is a user-centric requirements analysis method. The engagement of end-users during the process promotes the acceptance of the resulting business intelligence system.

2.3.3 Goal-driven information requirement analysis

The goal-driven method is based on the organizational model [Giorgini et al., 2005, Mazon et al., 2007, Taa et al., 2010]. In this method, the organizational model is viewed as a combination of stakeholders, goals, decisions and information. Examples of organizational goals are"increasing sales","increasing traffic" and"keep within the budget". These goals can be strategic, tactical and operational. Goals can be achieved by making decisions, that are based on information. The basic representation of this view on the organizational model can be seen in figure 2.2. This figure depicts the stakeholders who want to achieve goals, the stakeholders can be organizations and the members of an organization. The goals are owned by stakeholders or organizations, they have an interest in reaching these goals. Stakeholders can reach goals by making decisions. These decisions are actions taken by stakeholders, and are based on information that is available within the organization. Information can be structured, unstructured, quantitative, qualitative, easy to access, etc.

The goal-driven method for information requirements analysis takes a top down approach to elicit information requirements. Input for the goal-driven requirements method is the organizational model. Documentation such as models, job descriptions and presentations can be used as a source for this model. Besides documentation, the input of end-users is used to completely model the organization. The involvement

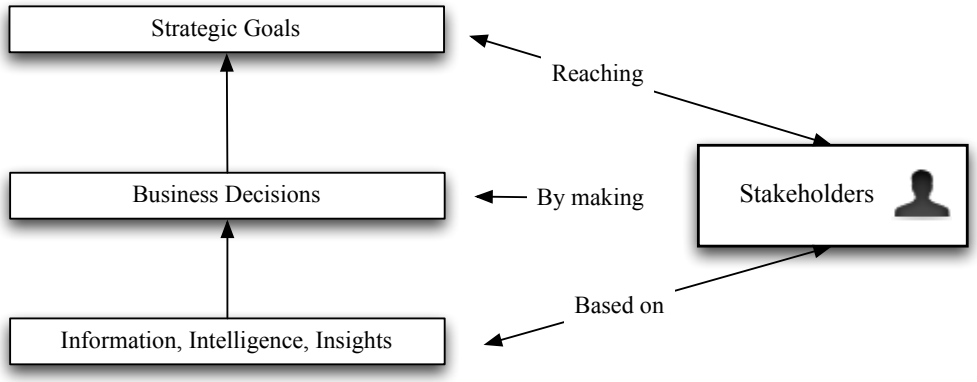

Figure 2.2: Goal-driven Requirements Analysis

of end-users has a similar effect on the modeling process as during demand driven requirements analysis. The method is a complex process with high costs for requirements analysis due to conflicting requirements and problems with eliciting knowledge.

The decisions that are a part of the organizational model have an unique character, most decisions will only be taken once. It is very costly to provide structural information and intelligence for all these ad-hoc decisions [Herrmann, 2004].

One of the most structured methods that uses goal-driven requirement analysis is the Goals, Decisions and Information (GDI) model by [Prakash and Gosain, 2003].This model corresponds to the representation of Goal Driven requirements analysis in 2.2. [Kumar et al., 2010] further developed the GDI model of [Prakash and Gosain, 2003] by adding stakeholders to the model. This enables the method to model complex organizations with a large number of stakeholders. A graphical representation of this enhanced AGDI model can be seen in 2.3.

The goal driven method works from the top, the goals, down to the information requirements (the dimensions and metrics in the data warehouse). The complex goals can be split up into simpler goals until a finite list of simple goals has been reached. Goals can be achieved by actions that are triggered by decisions, these decisions are also complex and can be split up into simpler goals until an exhaustive list of simple decisions is found. The decisions are based on information, the lowest level in the GDI model.

The addition of stakeholders in the GDI model is based on the assumption that goals owned by stakeholders from top to bottom in an organization. To achieve these goals, stakeholders within the organization can be delegated. Information is also owned by and available through stakeholders and can be interchanged between them [Kumar et al., 2010]. Stakeholders can be internal or external.

To assess whether goals are achieved, their performance has to be measured. These key performance indicators (KPI's) are not represented in the AGDI model. Measuring KPI's, and the success of goals that have been set, provides an indication as to which decisions need to be made, in what size or level and which decisions have been successful. For operational systems to work autonomously, the status of goals has to be constantly measured. Drilling down into the model, KPI's can provide insights into the quality of the decisions that have been made. The information that is needed to measure the KPI's will most likely come from internal or external data sources and are thus required in the business intelligence system. It is recommended that in the AGDI model KPI's are included in the analysis, making the model a AKGDI Model. Data quality is not directly addressed in goal driven requirements analysis, [Gosain and Singh, 2009] use the model to formulate the goal of information quality as one of the goals during data warehouse development.

The information requirements that are elicited by using the goal driven method are a representation of the organizational model. The maturity, completeness and quality of information requirements depends on the time spend by developers and stakeholders, their capabilities, experience and knowledge of the organization. The information requirements that will not be found using this method are those not directly related to the goals or decisions within the organizational model. Also, information requirements that fall outside of the cognitive capabilities of the end-users are not found.

The goal-driven method is a capable requirements analysis method that is mostly used. Figure 2.2 is a generalized interpretation of the goal-driven method, different variants of the method exist.

2.3.4 Model and Ontology-driven information requirement analysis

Two newer methods found in literature are model-driven and ontology driven. The model driven method is closely affiliated with the goal-driven method but takes the

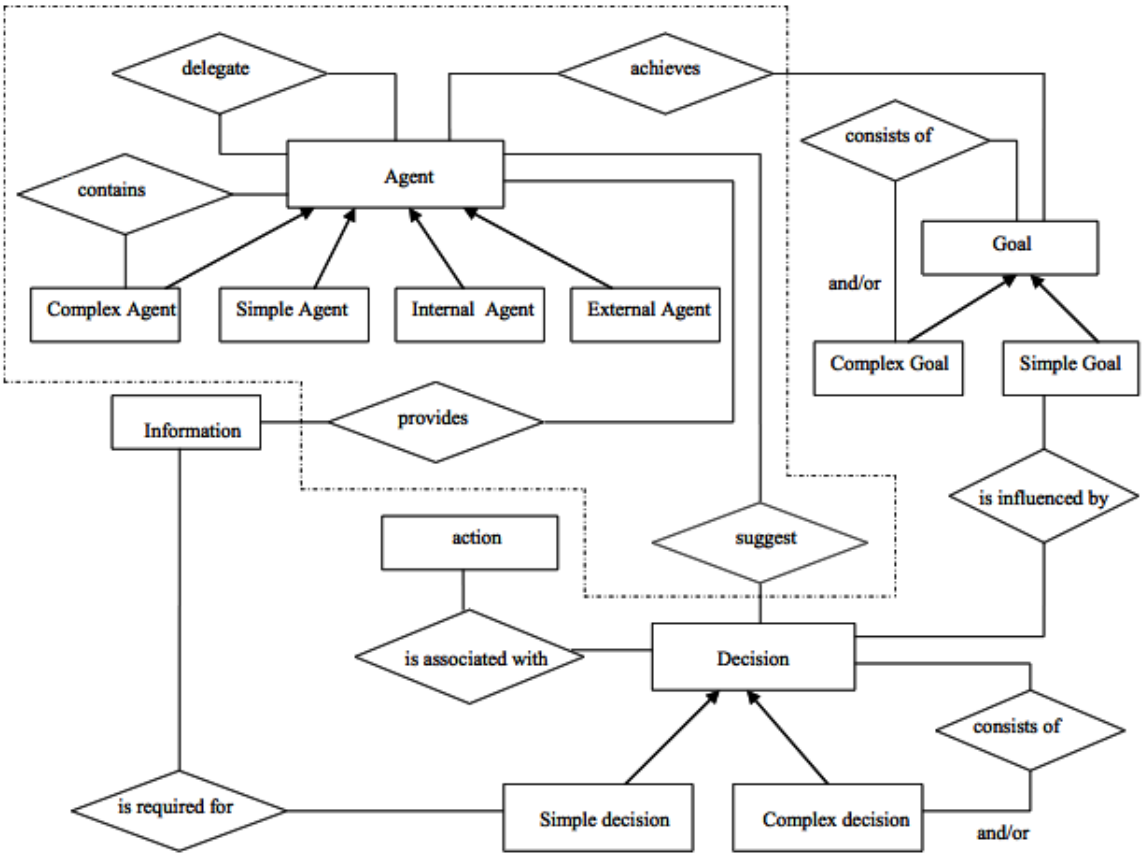

Figure 2.3: AGDI model by [Kumar et al., 2010]

	Supply-Driven	User-Driven	Goal-Driven
Basic approach	Bottom-up	Bottom-up	Top-Down
Users involvement	Low: DB Administrators	High: Business users	High: Top management
Constraints	Existence of a reconciled data level	Business users must have a good knowledge of the processes and organization of the company	Willingness of top management to participate in the design process
Strengths	The availability of data is ensured	Raise the acceptance of the system.	Maximize the probability of a correct identification of the relevant KPIs.
Risks	The multidimensional schemata do not fit business user requirements.	Quick obsolescence of the multidimensional schemata due to changes of the business users.	Difficulties in being supported by top management and in translating the business strategy into quantifiable KPIs.
Targeting organizational level	Operational and tactical	Depends on the level of the interviewed users, typically tactical	Strategic and tactical
Skills of project staff	DW designers	Moderators; DW designers	Moderators; Economist; DW designers
Risk of obsolescence	Low	High	Low
Number of source systems	Low	Moderate	High
Cost	Low	High	High

Figure 2.4: Comparisson of Requirements Analysis tools by [Golfarelli, 2008]

business model as a leading instrument in eliciting requirements. The motivation and the thoughts behind this are that the data-warehouse is closely aligned with the business model. The second method, the ontology driven requirements analysis is based on using metadata that exists in the current IT infrastructure, and is thus more similar to the supply driven method. Because the focus of this research is more on user-centric models to elicit information requirements, these methods will not be further elaborated upon.

2.3.5 Validation in Information Requirements Methods

A comparison of the three information requirements methods that have ben reviewed can be seen in figure 2.4.

The comparison has been made by [Golfarelli, 2008]. Most findings have been similar during literature research, although one point of view seems to be contradictory. In a supply driven method, all source systems will be used as opposed to custom requirements, by other methods, that might only need a small part of the available data for the data warehouse. Therefor, supply driven methods use a large number of

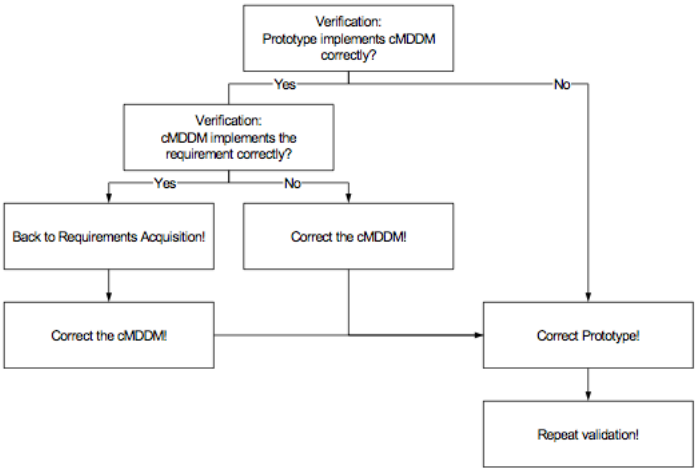

Figure 2.5: Validation by [Burmester and Goeken, 2006]

source systems, while the other methods require only the necessary systems.

A common risk factor in all methods is missing, the risk of unvalidated requirements. When reviewing the methods, most of them incorporate requirements validation. The methods do not elaborate upon how this validation should be done. [Burmester and Goeken, 2006] also stress this by saying that in literature there is a claim that there is a need for verification and validation of requirements but that authors, such as [Winter and Strauch, 2004] do not provide the tools to do so. He also cites [Bonifati et al., 2001] who claims that the development in business intelligence is a (long) trial and error process in which the requirements and data models are constantly improved. [Burmester and Goeken, 2006] then presents a method of validation by which users interact with early versions of the model, in this case prototypes of the cMDDM model, this can be seen in figure 2.5. The notion of prototyping a model could only be found in literature by one other author [Paim and de Castro, 2003]. He presents possible methods for validation of requirements, not a proposed method for validation and verification. The [Burmester and Goeken, 2006] prototyping method is much promising but has not been implemented in a case study and is not yet generalized for other methods of goal-driven requirements analysis let alone demand-driven requirements analysis.

Business intelligence systems should be flexible and capable to cope with changing requirements. Front end applications, that are becoming more important, are often

costly to develop and difficult to change and alter after development. The demand and goal driven methods both include the end-users in the early phase requirements analysis. During the succeeding phases, the end-users are not further involved. The validation of the requirements is done by presenting the list of requirement or the model that includes the requirements. The (mostly non-technical) stakeholders are not trained or capable to fully understand these models and lists, and are thus not capable of making a good assessment of the information requirements. In the final stages of business intelligence development, the end-users are presented with the front end applications that correspond to the set of information requirements. If problems with the system lead back to the information requirements phase, considerable time, costs and complexity will incur to change and alter the solution.

2.4 Best Practices in Information Requirements Analysis

In best practice, a number of cases can be found of BI consultants and engineers who use prototyping as a primary method for validating and verifying information requirements. Common methods and tools used in prototyping Business Intelligence solutions are wireframes, mockups and prototype setups. With the first two methods, the risk still remains that the information value of the proposed requirement is too low to base a decision upon. Developing a prototype setup of the system brings the issue of costs, the eventual system being a propagation of the prototype and not having a solid (technical) foundation and architecture. Moreover, the end-users can have misconceptions of the capabilities of the final system. Although these risks exist when using prototyping as a method for the verification of requirements, successful cases can be found in business.

2.4.1 Goal-oriented prototyping method at SNS Bank

One example is the development of a new dashboard for internet sales by the SNS Bank in The Netherlands, presented by [Wolfert and Dikmans, 2011]. In this case, the business intelligence development immediately included the end-users by defining the goals of the dashboard together. The experts then started drawing up mockups for the actual dashboard and a second round of user engagement started by voting for a group of promising ideas.

The team then used these mockups to create concept dashboards (without data), these concepts where then validated with the end-users. The first operational business intelligence solution was built as a prototype. This prototype was then extensively used and tested by the end-users, which resulted in the a set of requirements that has been used as the basis for the final business intelligence system. The key lessons learned in this process have been that the end-users involvement has provided a significantly big acceptance rate within the organization. Moreover, it gave the stakeholders a business intelligence system upon which they can base their decisions and reach their goals.

Problems that have been perceived by the developers were the availability and quality of data from datasources, it turned out to be difficult to tie different sources together. Moreover the quality of the data was essential to keep the end-users convinced of the quality of the system.

The example presented here was a non-academic, pure commercial, best practice based project. Looking into the development process, these methods and tools do surface. The early phase requirements analysis was done by looking into the goals of the end-users, and can be seen as a goal-driven method, the validation was also user-centric and based upon rapid prototyping. Although this case has proven successful, the risk that is caused by using wireframes and mockups still remains, also the availability of data has shown to be a problem.

2.4.2 Prototyping with new Technology

Complications that are caused by a lack of data during validation make that the strategy for verification and validation of requirements is not full proof. With growth and emergence of technology, the development of business intelligence systems has become less complex. The Composite Software company [Software, 2011] provides software that enables 'virtual' data warehousing as a means of prototyping the intended data warehouse, this can be seen schematically in figure 2.6. In the other parts of the business intelligence stack, similar new tools are available to speed development time. These tools can assist in rapidly developing prototypes that can validate requirements during the requirements analysis phase of a project. Example tools are RapidMiner, Pentaho, Jaspersoft, Spotfire and Tableau. The advantages of using these tools for validation of information requirements is that in a relatively easy time period and at low costs, requirements can be tested by end-users using sample data from source

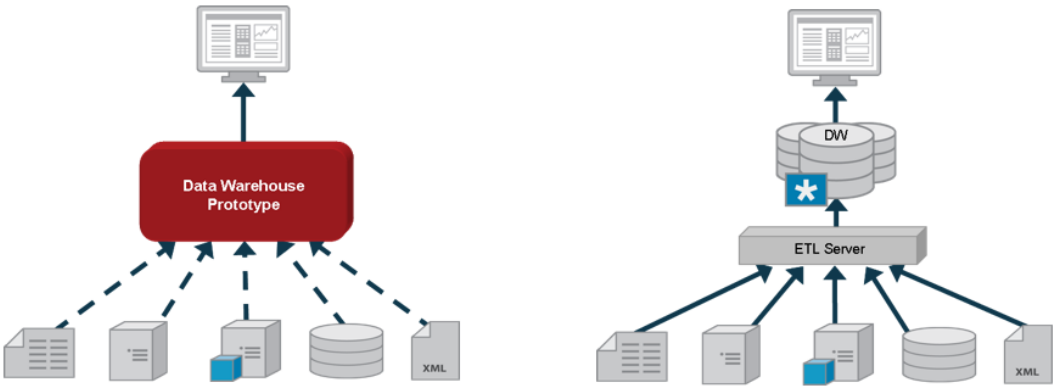

Figure 2.6: Datawarehouse prototyping by the Composite Software Company [Software, 2011]

systems. The focus in these applications is front-end, for example, dashboards, OLAP and reporting but also ETL and Data warehousing.

Prototyping helps the end-users understand the information requirements and needs for a business intelligence system. Moreover, prototyping can be done rapidly at low costs.

2.5 Requirements Engineering in General IT Projects

To get inspiration for requirements for a new requirements model for business intelligence projects this section will briefly explore requirement engineering methods in other (conventional) IT development projects.

Requirements engineering has emerged since the 1970s. In figure 2.7 an overview of the history of requirements analysis by [Hood et al., 2008] can be seen. Naming has gone through a change from customer requirements to user requirements, stakeholder requirements and back to customer requirements. The aim of requirements analysis has always been on the stakeholders that are affected by the system. This can be as an end-user, organization, maintenance engineers, etc. A consideration in this perspective is that because of the stakeholders involved and the heterogeneous character of stakeholders the requirements that are elicited eventually all need to be considered and leveraged to create a final set of system requirements.

To elicit the requirements for a system, [Davis, 1982] has proposed four strategies:

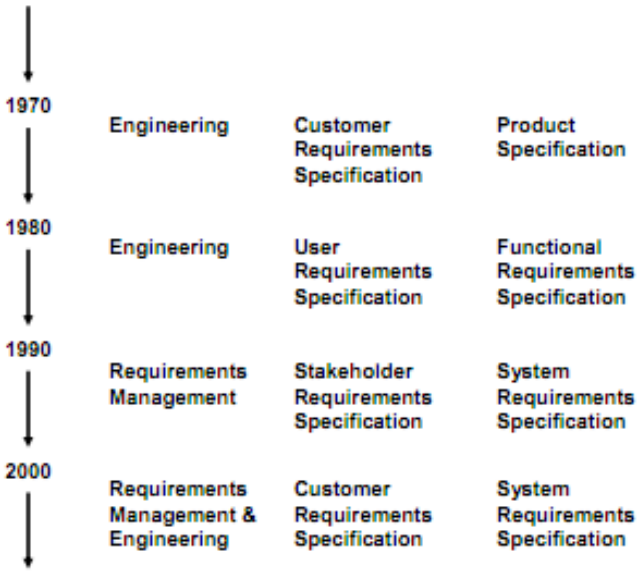

Figure 2.7: History of Requirements Analysis by [Hood et al., 2008]

- Asking
- Deriving from other systems
- Synthesis from characteristics of the utilizing system
- Discovering from experimentation with an evolving information system

In practice, asking is mostly used, Davis claims that because of their subjective view on things this may not yield to a complete set of requirements and thus other strategies need to be used. Asking covers all interaction with stakeholders that result into (functional and non-functional) requirements. Stakeholders first need to be identified and can be internal or external to the organization. The interactions include interviews, questionnaires, guided sessions, brain storming, etc. Asking in these projects is similar to the demand-driven methods in BI requirements eliciting, thus similar pitfalls exist.

When gathering requirements by deriving them from other systems, these systems can include other IT systems but also organizational systems or conventional physical systems. An example of this can be the re-development of a currently used system. The system as-is can be the basis for the requirements, short-comings or missing elements can be identified by end-users to complete the set of requirements. Moreover,

systems of other companies, organizations or vendors can be leading in developing a set of requirements. In the case of Business Intelligence projects, the data warehousing solutions are mostly custom-made and do not provide a solid basis for replication or requirements setting. Existing front-end applications, ad-hoc analysis and other BI related applications do offer some basis for requirements. The uniqueness of Business Intelligence solutions limits this option however. Model- and ontology based requirements analysis methods fit in this category, although most likely forms of asking are required to complete the requirements. This also holds for Goal-oriented analysis, if you see the set of Goals and Decisions as an existing system.

In the third strategy, Davis presumes the availability of an existing system. Requirements that need to be gathered are based on this existing system. This strategy also encompasses requirements eliciting from business processes, goals and decisions, thus goal-oriented requirements analysis is very similar.

The final strategy is close to the third strategy but is characterized by its evolutionary approach by experimentation, or prototyping. Requirements are tested by prototyping and rapid development. In Business Intelligence projects this is already much seen in best practice cases. The main reason to use this last strategy is when:"There is no well-defined model of information requirements", which is very likely in these BI projects due to their unique characteristics. Although in practice this method is often used in BI projects, there are no validated methods for requirements eliciting available in theory that explicitly use prototyping as a tool.

In figure 2.8 the methods found in literature are mapped to the four strategies by [Davis, 1982]. None of the methods cover all strategies.

2.6 Conclusions

The methods and models that have been found in literature all present a different specific approach to elicit information requirements. The methods all have strengths and weaknesses. None of the methods found is capable of creating a near complete and correct set of requirements that meet the needs of end-users. In practice and general information systems development, requirements are tested and validated extensively after early phase requirements have been formulated.

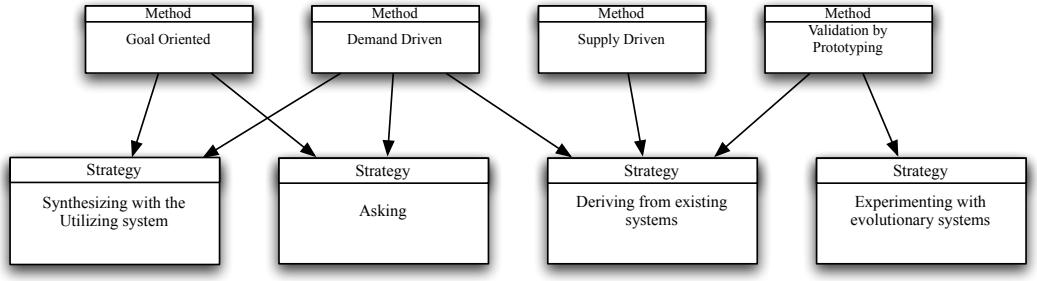

Figure 2.8: Strategies by [Davis, 1982] covered by Information Requirements Analysis Methods

[Davis, 1982] emphasizes that to elicit a complete and correct set of requirements different strategies have to be considered. The four strategies for requirements elicitation by [Davis, 1982] are asking, deriving from other systems, synthesis from the utilizing system and discovering by experimentation.

During the information requirements phases in business intelligence development, there is a need for a model that combines different approaches to requirements elicitation methods. The goal-driven method provides a solid basis of information requirements, to complete this set, the demand-driven method can be used to further elicit requirements together with end-users. Using validation and verification of the information requirements by experimentation and prototyping, the set of requirements can be further developed.

3

Requirements for an Information Requirements Collection and Analysis Model

In the previous chapter, an overview has been given into current requirements analysis for business intelligence projects. In conventional information requirements, more research has been done into requirements eliciting. The previous chapter described the four requirements gathering strategies by [Davis, 1982]. When using only one business intelligence information requirements method, not all four strategies are covered.

In this chapter, information requirements methods have been combined and further developed. The goal has been to design an information requirements collection and analysis model that covers all requirements gathering strategies and to resolve issues and problems that exist in current analysis methods.

3.1 Requirements

Requirements that correspond to the four strategies of [Davis, 1982] and underlying methods, together with the requirements for the overall process, found in the previous chapter, together form the requirements categories for an information requirements model. The shortcomings in the current requirements methods and models are mapped onto these categories. Besides shortcomings, the essential elements for information requirements eliciting found in current methods have been included in the set of require-

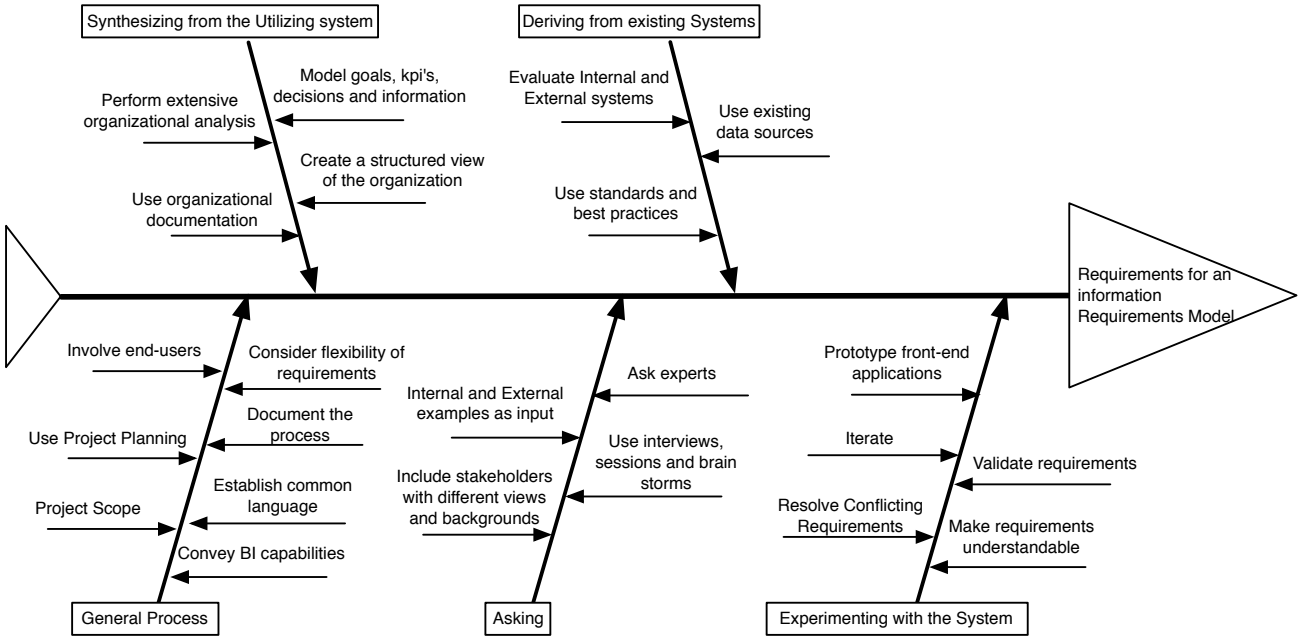

Figure 3.1: Requirements for an Information Requirements Model

ments. A visualization of the requirements that have been collected is visualized in figure 3.1.

The following sections cover the requirements for each strategy and the general process.

3.1.1 General Process

The requirements that correspond to the general information requirements process can be seen in figure 3.2. The end-users need to be involved, this creates engagement and commitment to the success of the process and system. To include end-users optimally, communication barriers such as language problems and miscommunication have to be minimized. This can be done by introducing business intelligence, its capabilities, limits and terminology.

Based on the definition of the business problem, expectations of end-users and capabilities of business intelligence and developers, the project scope should be defined. Together with the stakeholders, project management, planning and documentation of

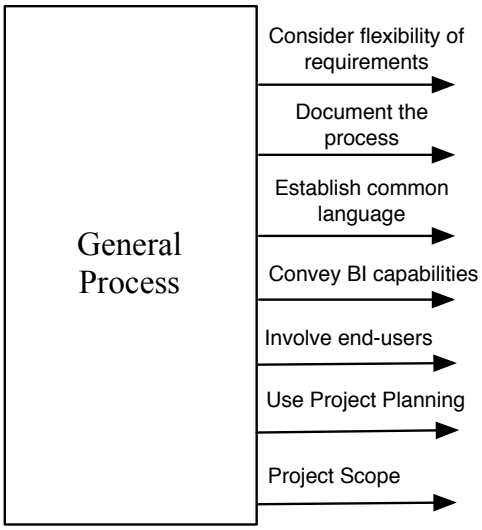

Figure 3.2: Requirements for the general process

requirements have to be set up and maintained. This helps in creating a transparent process. The documentation has to be maintained because during the process, information requirements will change. To cope with these changes, requirements for iteration and validation have been set according to the strategy for experimenting with the system.

3.1.2 Asking

Asking is done in both demand- and goal-driven requirements methods. Requirements to support asking can be seen in figure 3.3. To elicit knowledge from the end-users, different approaches can be used. Interviews, facilitated sessions and brainstorms are techniques that support gathering information from stakeholders. Doing this with stakeholders that have different views and backgrounds helps getting a more complete set of information requirements. This approach also produces requirements that can be conflicting, by experimenting with the system, these can be resolved. Besides using internal stakeholders, experts and external stakeholders should be consulted, opening the views of the stakeholders in the organization.

Asking can be supported by using input from internal and external sources. Examples are business intelligence systems that are already in place, dashboard and reports

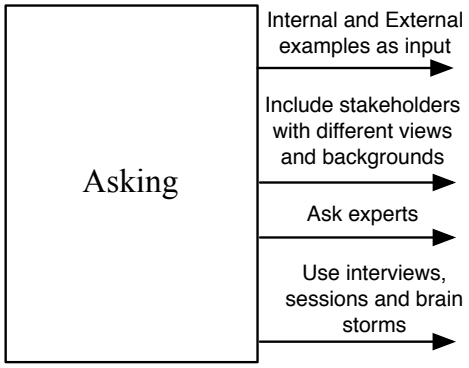

Figure 3.3: Requirements to support Asking

available throughout the organization and commercial of the shelf systems. Organizations that support research in Business Intelligence development have lists of common information requirements, reports and dashboards available that can be input for information requirements analysis.

By asking the for needs and requirements for a business intelligence system, the end-users feel involved in the process. This promotes commitment to the process and acceptance of the system.

3.1.3 Deriving from existing systems

The information requirements do not have to be engineered from scratch. The requirements to support deriving information requirements from existing systems can be seen in figure 3.4. Existing internal and external information systems can be used as a basis for requirements. In business intelligence, diverse data sources are combined to create insights and intelligence. Synergies that combined existing information systems promise can be the source of additional information requirements. External systems that are already developed can be commercial of the shelf software, open source systems and examples of reports, dashboards and other front-end applications. These include best practices by developers and standards developed in the industry.

Information requirements that are formulated need to be mapped onto existing systems. A data audit provides with the insights into the available data and the structure

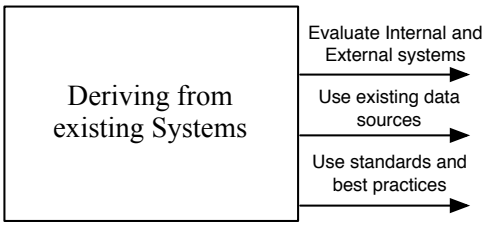

Figure 3.4: Requirements to support Deriving from existing systems

of this data. The data should not be leading in eliciting requirements, viewing the problem as a technical problem, but data should not be ignored.

3.1.4 Synthesizing of the Utilizing system

During the demand- and goal-driven methods, the main goal is to elicit information requirements for the utilizing systeml. The requirements supporting synthesizing of the utilizing model can be seen in figure 3.5. The system should reflect the requirements and needs that exist within an organization. The goal-driven information requirements method provides with means to analyze organizations. The organization is seen as a set of stakeholders, their goals and the supporting decisions and underlying information. The utilizing system should be modeled to give a clear and structured overview of the needs and requirements within that organization. The knowledge of this organization can be derived from the stakeholders by asking. Another source of knowledge is organizational documentation, examples are organization charts or job descriptions.

Complementary to the project scoping that is required by the process, the organizational model should give an overview of the stakeholders that are part of the utilizing system, and that should be included in the process.

3.1.5 Experimenting with the system

Eliciting requirements directly is difficult, requirements change during the project. Initial requirements are changed and altered due to, data availability, data connections, new insights, altered perceptions and growing end-user experience with business intelligence. This point stresses the need for constant iteration and validation of the set of requirements, they are evolving alongside the project and the system. Because of this

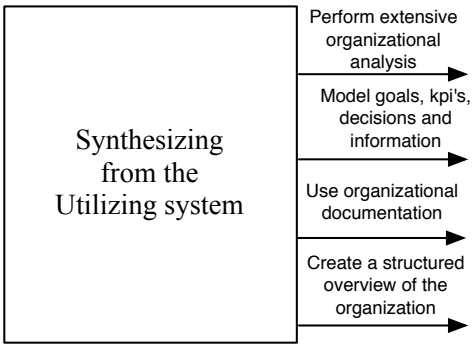

Figure 3.5: Requirements to support Synthesizing of the Utilizing system

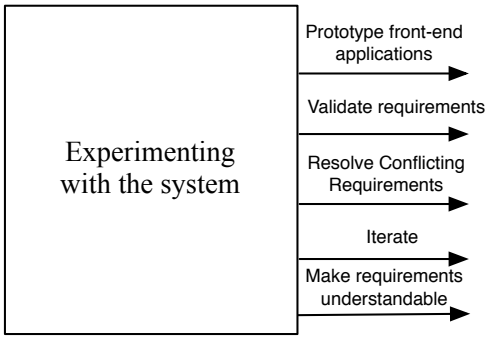

Figure 3.6: Requirements to support Experimenting

flexibility, the evolutionary approach found in multiple best practice cases and literature, is a very suitable method. The system grows, starting with an incomplete set of initial requirements and iterates stepwise until the system is more or less complete. As already mentioned in the introduction, the intended new method is not supposed to mitigate in the flexibility of requirements during operation of the BI system, but merely in the development cycle. If the set of requirements are not validated with the end-user, they will not reflect the actual information needs of the end-users. In complex organizations, requirements can be conflicting, and thus need to be mitigated. The end-users will be interacting with the front end applications of the BI system. BI projects are very costly. The development of data warehouses and front-end applications cost a lot of time and money. In the last couple of decades a large number of tools has become

available that make rapid prototyping and rapid development of BI solutions possible, at low cost. Using prototyping as a means of validation gives the end-users a feeling of what the system is going to be like. Prototypes are more understandable to end-users than lists or models with information requirements.

3.2 Conclusions

The requirements for an information requirements model that elicits a set of correct and complete requirements are based on existing methods. To gather a complete and correct set of requirements, a combination of methods and strategies has to be adopted. The characteristics of the demand- and goal-driven methods overlap, a combination of the two methods however, complement each other. The requirements for a new information requirements model that have been derived from reviewing the methods are supplemented by requirements for validation and the general process. The four strategies of requirements gathering for information systems by [Davis, 1982] are used to map and structure the requirements for an information requirements model. The next chapter presents the proposed model.

3. REQUIREMENTS FOR AN INFORMATION REQUIREMENTS COLLECTION AND ANALYSIS MODEL

4

Information Requirements Collection and Analysis Model

4.1 Introduction

In the previous chapters, the current status of information requirements in business intelligence has been evaluated. A comparison has been made between different, much used, tools from academia and business. Based on this research, the need for a new requirements analysis method for business intelligence projects has become clear. Based on observations during the research, positive and negative, a set of requirements has been developed. This set of requirements provides a solid basis for a new, consolidated, information requirements collection and analysis model for business intelligence.

In figure 4.1 a graphical representation of the proposed new model can be seen. 4 main phases have been designed, the introduction phase, goal-driven phase, demand-driven phase and a phase within the consolidation and validation. Clear iteration exists between the validation phase and the requirements gathering phases. The iteration can also be limited to one of the requirements phases or only to the consolidation and verification phase. It should be noted that this model does not have to be followed strictly and that during the business development life cycle it is possible to fall back onto the requirements analysis stage.

4. INFORMATION REQUIREMENTS COLLECTION AND ANALYSIS MODEL

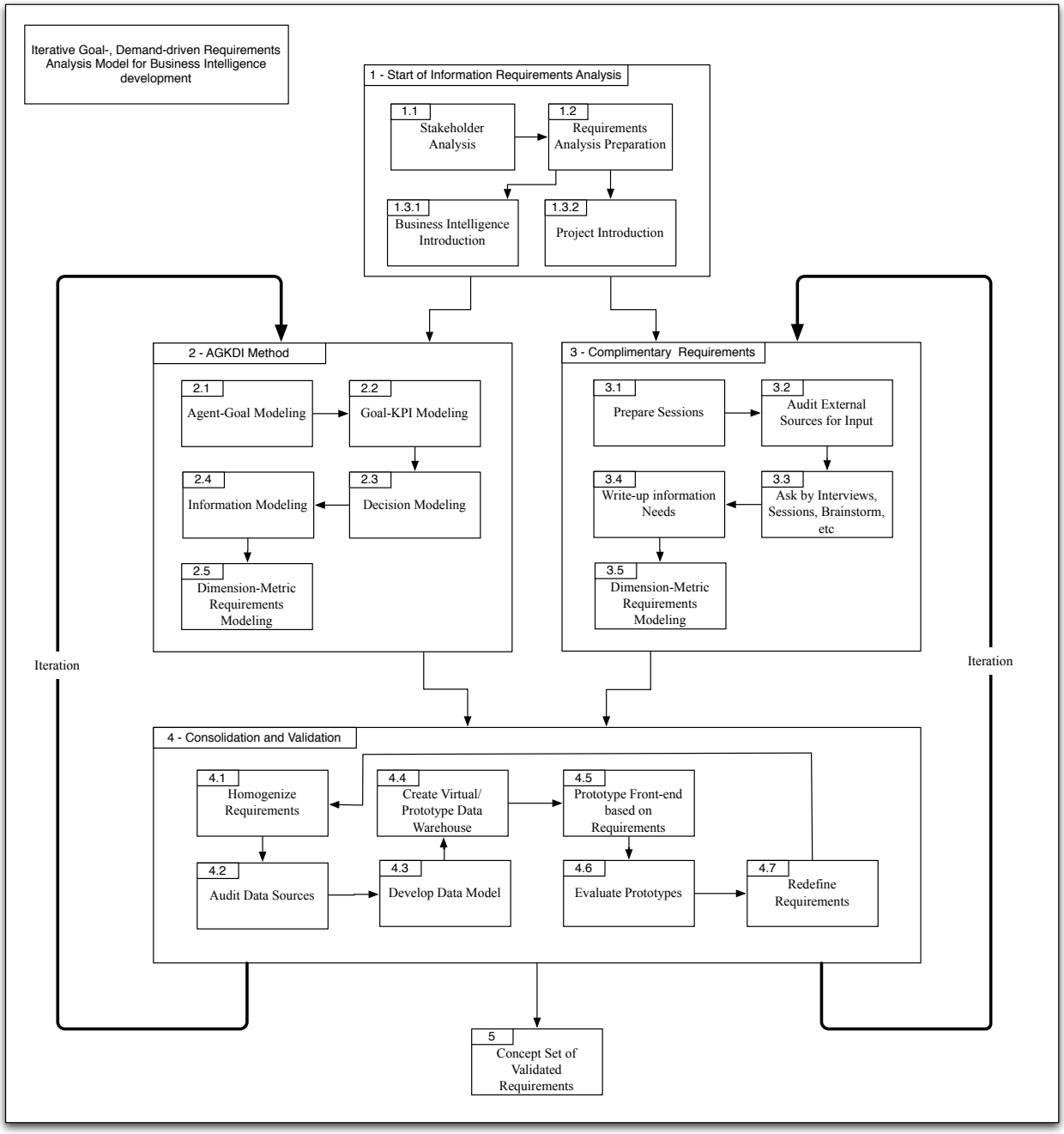

Figure 4.1: Diagram of Proposed Information Requirements Collection and Analysis Model

4.2 From a set of requirements to an information requirements collection and analysis model

The goal of the information requirements model is to cover the requirements that were set in the previous chapter. In figure 4.2, the four strategies and the requirements for the general process are mapped to the high level information requirements analysis model.

An explicit choice has been made to include both a goal- and demand-driven method in the model. The goal-driven method takes the organization as a basis and provides with a structured view of the utilizing system. The resulting set of information requirements is complemented by performing a demand driven requirements method. The demand driven process approaches information requirements both from internal and external existing systems. In contrast with goal driven methods, the demand driven method is less structured and rigid. The demand driven method requires the stakeholders, both developers and end-users, to be creative and to have an outward view.

The model is characterized by a focus on communication, end-users, validation, iteration, project management and project planning. These are derived from the requirements that support the general process of information requirements eliciting. Although they may seem trivial, the current requirements analysis methods do not adequately describe these elements in their models.

4.3 The place of the model in the business intelligence development lifecycle

The requirements analysis phase is, and should be one of the first elements within the business intelligence development lifecycle. This can also be seen in the business dimensional lifecycle by Kimball in figure 1.2. The new model includes various feedback loops, dimensional modeling and other elements that are also depicted in the lifecycle model by Kimball. The information requirements analysis is an integral part of business intelligence development. The overall model should facilitate iteration but also aspects that impact the entire lifecycle, such as project management. The interaction between the different phases in the business dimensional lifecycle lies outside of the scope of this

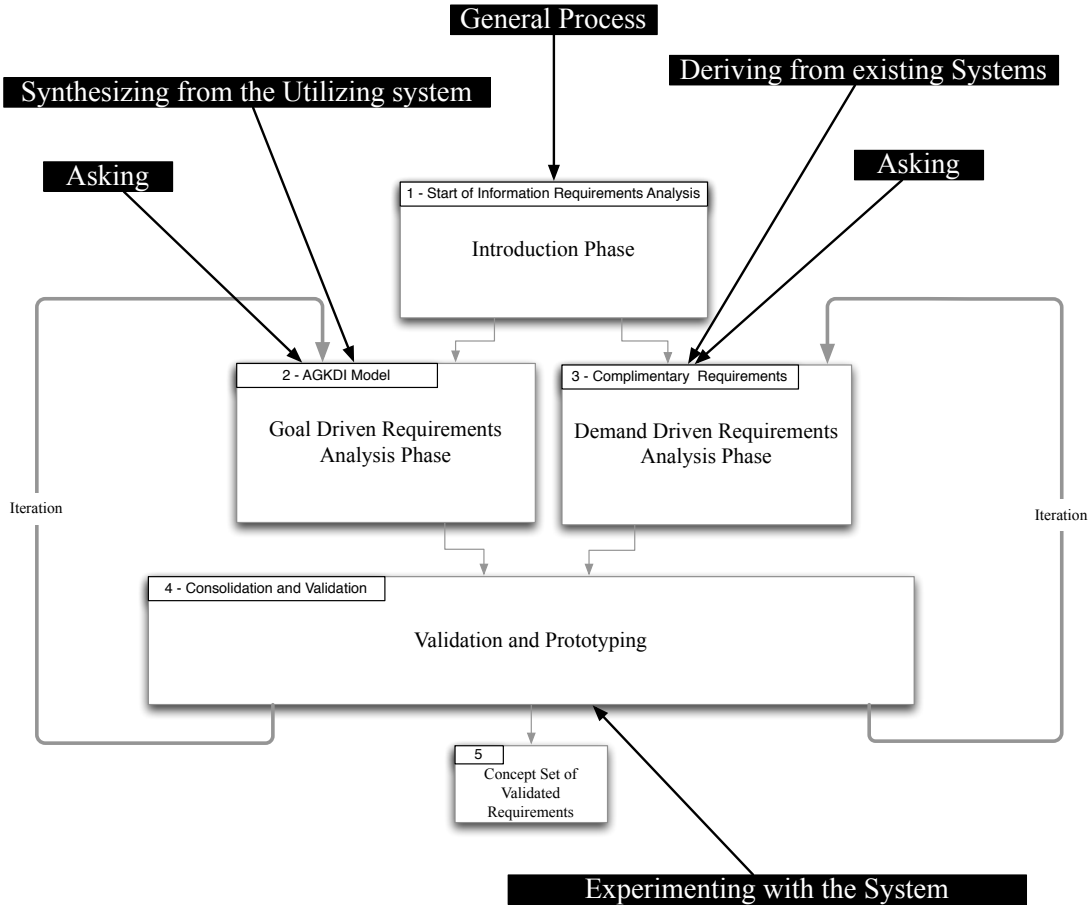

Figure 4.2: Set of Requirements mapped onto the Information Requirements Model

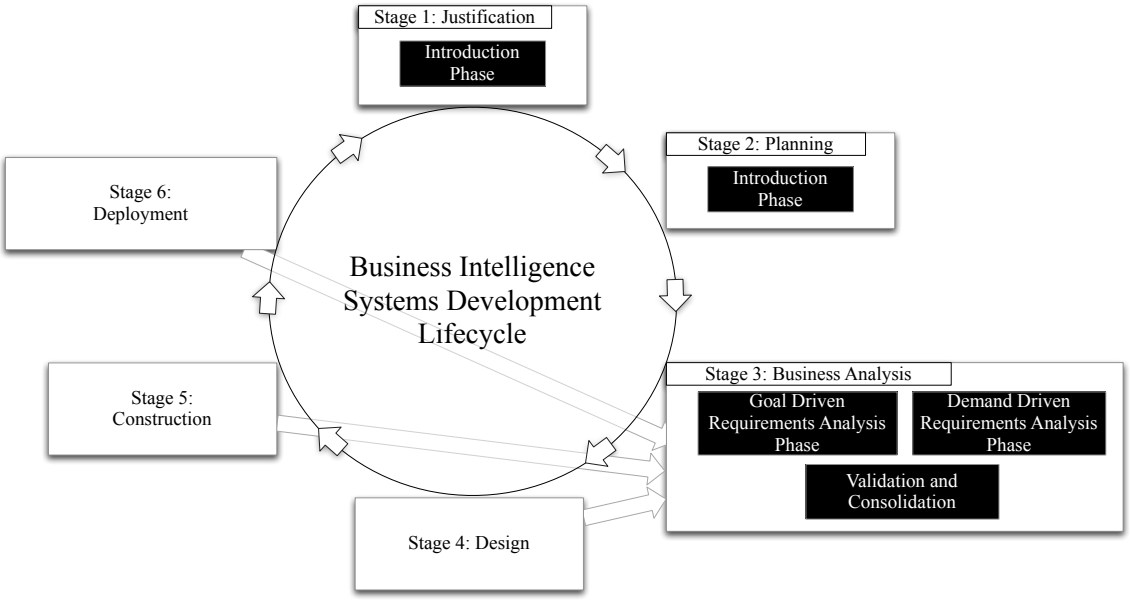

Figure 4.3: Set of Requirements mapped onto the Information Requirements Model

research but the success of business intelligence development is not only dependent on the quality of the information requirements.

In figure 4.3, the information requirements analysis model is mapped onto the business intelligence development lifecycle by [Moss and Atre, 2003]. The cycle is iterative. To emphasize on the position of requirements analysis, the feedback steps from the deeper stages of the model are also depicted. However, the focus of the information requirements model is on the first half of the cycle.

4.4 Working with the Information Requirements Analysis Model

In the following sections, all elements within the model will be elaborated upon and guidelines will be given as to how the model and including tools should be used in a development project.

4.4.1 Phase 1 Start of Requirements Analysis

In this model, requirements analysis is part of the start of a business intelligence project. It is presumed that the project team, project definition and scope have been mostly set during preliminary sessions between the project sponsor and developers. The first phase of requirements analysis is the start. This can be seen as both an exploratory phase for both developers and end-users and as an introduction into business intelligence and the scope of the project for both parties.

4.4.1.1 Phase 1.1.1 Project Introduction

The project introduction is a phase in where developers can lay out the proceedings of the project. The scope of the project will be discussed by both parties, this can be combined with the business intelligence introduction. The following phases of the requirements analysis can be conveyed. This way the end-users have an idea of the project proceedings. Moreover, they can start collecting material, documents, examples and can start thinking about the requirements that they would like to have in the intended system.

4.4.1.2 Phase 1.1.2 Business Intelligence Introduction

This phase focusses on giving a broad introduction of business intelligence, its capabilities and limits. This main goal of this phase is to include end-users in the project from the start. Also, the introduction should get the end-user more comfortable with the terminology that is used during the project by the development team. This helps to enhance the communication between the stakeholders. Finally, the insight into the capabilities and limits of Business Intelligence (as far as this can be shown during an introduction) should open the creative minds of the stakeholders, enabling them to get the most out of the project.

4.4.1.3 Phase 1.2 Requirements Analysis Preparation

This phase encompasses all preparations that need to be made in order to successfully go through the analysis method. Preparations concern both organizational planning and gaining more basic knowledge into the project. All stakeholders need to be informed of the project. Together with the two parties, a planning has to be made for first and

second round sessions with stakeholders. Depending on the size and complexity of the project, the number of meetings can be decided.If necessary, more meetings can be planned during the project. This choice depends on the number of iteration loops that is needed. Besides planning, preparations have to be made for the contents of the analysis cycles. Presentations on business intelligence and the project need to be made. Through research in internal and external documentation a basic knowledge of the project should be acquired.

4.4.1.4 Phase 1.3 Organization Analysis

The organization analysis is a multistep stage that answers the question as to why and what has to be developed for whom. First, the problem definition has to be clear. A common problem definition is that information is spread out through an organization in different systems and within the knowledge of the people within that organization, a business intelligence platform can provide a more structural basis for sharing, querying and analyzing information. This will probably answer the what on a high level as well. The biggest element in the organizational analysis is a stakeholder analysis. The stakeholders need to be identified and mapped onto a model.lThis gives insight into how the organization is structured and what the landscape of end-users will look like. The stakeholders that have been identified can then be included in the further proceedings of the method. The stakeholder analysis will also be an important input into phase 2.1, the Agent-Goal modeling phase.

4.4.2 Choosing between Phase 2 or 3

Time limits of stakeholders possibly require the information requirements phases to be done simultaneously. Although this is possible, it is preferred that the second phase, which is very structured is done first. The requirements that remain after this phase has been worked through will probably surface when going through phase 3. The goal of the two phases is the same, to elicit requirements from the end-users for the BI project, the methods are very different.

The AGKDI Model is structured and rigid, the third phase is relatively unstructured. The goal is to get a complete set of requirements and with the contrast between the two phases, all types of stakeholders will get a chance to convey their requirements. This also strokes with the fact that, although formal organizational charts exist and

should provide a basis for the functioning of organizations, informal structures and connections between stakeholders exist and will surface during demand driven modeling.

4.4.3 Phase 2 AGKDI Model

The AGKDI model is a model that is based on the GDI, goal-driven requirements analysis method by [Prakash and Gosain, 2003] and Kumar et al. [2009]. It has been enhanced by adding stakeholders and key performance indicators. A visualization of the relations within this model can be seen in figure 4.4. The output of the model is a set of visualizations and a list of dimensions and metrics. The method has a top down approach. It starts with the stakeholders and their goals, drilling down through key performance indicators, decisions and the supporting information. The stakeholders that are involved in this method should also be approached top down, so first the executives down tol operators and the lowest level employees intended to work with the system.

During the sessions with the stakeholders, sources can be used as an input for analysis such as performance documents, work-flow descriptions, process models, etc. The tools that can be used to find the elements of the models together with the stakeholders can be, for example, interviews or brainstorms.

4.4.3.1 Phase 2.1 Agent-Goal Modeling

This phase is a continuation of the organizational analysis. The organization is taken as an input. The top stakeholder is the organization that has a certain goal, for example to maximize profits in a retail organization. This complex goal consists of multiple simple goals, in this case it is assumed that the maximization of profits is split up in getting more income and making less costs. The goals can be delegated to other stakeholders in the organization. For example, increasing income can be delegated to sales persons and making less costs to buyers or sorcerers. In a BI project, the highest goal, owned by the highest stakeholder is most likely the manager or executive that heads the organization under scope of the project. He is the starting point. The agent-goal modeling phase ends whenl the lowest stakeholder (intended to use the system) that has been delegated the most simple task has been modelled. Depending on the organization, a large set of stakeholders and goals will be found. For the stakeholder(s) with whom the session

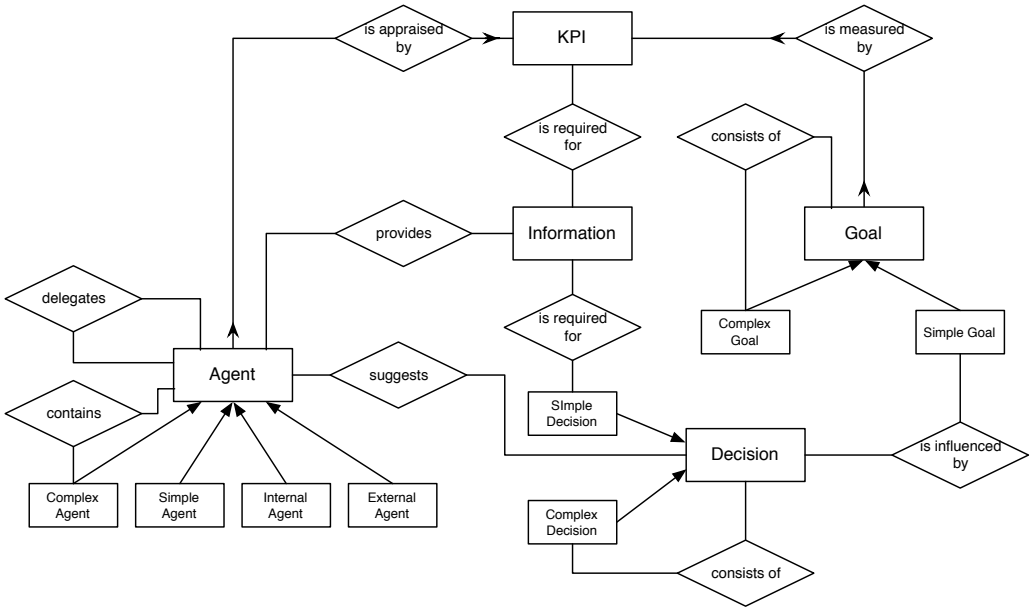

Figure 4.4: The AKGDI Model Visualized

is being held, the goals that are delegated to them, or that they delegate to another stakeholder need to be further analyzed during the next phases.

4.4.3.2 Phase 2.2 Goal-KPI Modeling

The goals that have been found under phase 2.1 belong to stakeholders. They want to achieve these goals in order to, when consolidated back to the top complex goal, reach the main goal of the organization. Within organizations, stakeholders are appraised according to the goals that they have, or have not reached. To measure if the set goals has been reached, either quantitative or qualitative, Key Performance Indicators are formulated with certain thresholds relating to the goal success. In the example of a retail organization, the maximization of profits could have the goal of a certain amount of profit to be reached. When this threshold is passed, the goal of the organization has been achieved. All goals can have one or more Key Performance Indicators, which is in turn calculated by using information that is available to the stakeholders.

4.4.3.3 Phase 2.3 Decision Modeling

To reach the goals that have been set during the previous phase of the model, decisions need to be made. These decisions result in actions that in turn influence the goals. Decisions can, similar to goals, be broken up from complex decisions into a finite number of simple decisions. These decisions are suggested and made by the stakeholders concerned with the set of goals. The stakeholders base these decisions on information that is available to them. In retail organizations this information can vary, for example the expected demand, the temperature, number of visitors, supply, etc. The decisions that they make can be, for example, to stock more products, to change the set-up of a store, etc.

4.4.3.4 Phase 2.4 Information Modeling

Using the decisions and KPI's found during the execution of the method, information needs can be collected. These information needs can be both internal and external. There has to be a likelihood that information can be obtained from data that is available. The information needs can be formulated on a high but exact level, for example, the historic profits, or historic sales, future expected sales, etc.

4.4.3.5 Phase 2.5 Dimension-Metric Modeling

The final phase of the AGKDI model translates the information needs in Dimensions and Metrics that can be found in existing data sources. Dimensions are discrete values that give perspective to data. Examples of dimensions are Time, Location, Customers, etc. Metrics are the numbers that correspond to the data and is characterized by a set of dimensions. These dimensions and metrics come together in facts. An example of a fact is a sale, which has dimensions time, location, customer and payment method and the metrics amount, number of articles, etc.

4.4.3.6 Drawing up the AGKDI Model

To gather all the elements of the AGKDI model will take, depending on the size of the organization, a fair share of time. During the modeling sessions, post-its and large sheets can help in structuring the model. The resulting model will have a significant size, therefor it is recommended to document the model using electronic modeling tools

such as Microsoft Visio or Omnigraffle. A number of Goal-modeling tools have been developed for the Tropos research project, these can also be used, but the problem is that these are not adepted to the AGKDI model, but merely to a Goal-Decision model. The Dimensions and Metrics can be combined into a dimensional model, this can be the input for the development of a data warehouse that is part of the Business Intelligence solution. Phase four of the model deals with requirements that conflict with each other due to different owners, originating methods or validation cycles.

4.4.4 Phase 3 Complementary Demand Driven Requirements

As said, the second phase is very structured, this phase is more unstructured and has the goal to find all remaining information requirements. This can be done by asking the stakeholders found in the previous phase about information needs that they have that are not directly captured in that phase because they do not directly relate to immediate goals that have been set. Examples from external cases can provide a source for new ideas and insights. Existing systems can also be a source of inspiration for information requirements.

4.4.4.1 Phase 3.1 Prepare Sessions

During the preparation of the requirements gathering sessions an overall strategy for the session should be developed. What are methods to elicit the information needs that are currently not found through the AGKDI method. Example methods are brainstorms, interviews, examples from internal and external projects, etc. Moreover, this phase is also open for stakeholders and experts other than used in the previous phase. This supports the goal of gathering a complete set of requirements.

4.4.4.2 Phase 3.2 Audit internal and external sources for input

One of the possible approaches, that is recommended to use during this phase, is the use of external sources as input for information requirements. This input can be, for example, dashboards and reports that other similar organizations use, lists of requirements, existing dimensional models, etc. External input can also come from external stakeholders, such as experts and other end-users. This phase is recommended as a means of inspiration and to trigger discussion in the following section.

4.4.4.3 Phase 3.3 Asking

Asking is one of the main strategies to elicit additional information requirements from stakeholders. The asking phase can encompass interviews, facilitated sessions, brainstorms, debates, discussions and other forms of direct communication with stakeholders. All these methods can be interchanged and used together. The main goal is to get those requirements that could not be found using the more structured method.

4.4.4.4 Phase 3.4 Write-up information needs

Although this method does not have a clear structure, the management of the information requirements should be done in an orderly manner. Although this phase is explicitly mentioned, it should be clear that requirements need to be documented very precisely and should stay clear of ambiguity. Moreover, the ownership, source and change of requirements need to be documented.

4.4.4.5 Phase 3.5 Dimension-Metrics Modeling

This phase is exactly the same as phase 2.5, but applies to the information requirements as found in the previous phases.

4.4.5 Phase 4 Consolidation and Validation

When the previous two phases have been completed (as far as possible) the next step is to consolidate and validate the set of information requirements.

4.4.5.1 Phase 4.1 Homogenize Requirements

Because the requirements that have been found come from two separately performed methods they need to be combined and made homogeneous. It is possible that requirements can be conflicting, together with the stakeholders and owners of the system it should become clear how these conflicts can be resolved. There is no standard approach to solve these conflicts, therfor it is a process that requires the engagement of the stakeholders involved.

4.4.5.2 Phase 4.2 Audit Data Sources

The information requirements and corresponding dimensions and metrics need to be validated from the perspective of the available data. This can be done through a data audit. The data audit requires the participation of the owners of the data, such as IT developers, database operators and business analysts. A data audit can be an interview or for example a review of currently existing data models such as Enterprise-relationship models, etc. When external data is needed, the owners of that data need to be contacted. It is imaginable that data is not always available from third parties, or that the data provider asks a fee. The data audit has the goal of identifying these issues and to uncover if all data is available. The data audit also gives the project team a feeling of the possibilities that come with the available data, this can be used in a feedback loop as inspiration for new requirements.

4.4.5.3 Phase 4.3 Develop Data Model

The information requirements and the available data and corresponding data schemas and models now need to be combined into a data model. Multiple types of data models are available, common methods are dimensional models, star and snowflake schemas. Depending on the preference of the project team and the stakeholders a choice can be made. If all data is available, the dimension-metric models from phases 2.5 and 3.5 can be used as the main input. During the development of the data model, the source systems that have been found in the previous phase have to be mapped to the data model.

4.4.5.4 Phase 4.4 Create Virtual/Prototype Data Warehouse

A large part of data warehouse development is concerned with Extraction, Transformation and Loading or ETL. This encompasses the extracting of all data from the source systems, transforming it so that it fits into the data model and loading the data into the system. ETL is a timely and costly operation and until the requirements have been set in stone, the ETL and corresponding data warehousing system has to be flexible. This flexibility is required with respect to the time and costs for the development and the costs of the hardware, such as database servers, connections, etc. New technology makes this possible through open source software and cheap hardware or cloud hosting

solutions such as Amazon S3. Other technological advancements such as virtual data warehousing gives the flexibility of having cheap prototypes during data warehousing development. Another option would be to develop the data warehouse and to load it with sample data, this can save time and costs. Even though the prototyping phase can be a bit rough on the edges, the quality of the data that it holds has to be as high as possible under the circumstances.

4.4.5.5 Phase 4.5 Prototype Front-end based on Requirements

To validate the requirements prototypes can now be made in this phase to create a feeling for the final system that the end-users will be using. Prototypes can be very static, for example just a draw up of the applications (in that case there is not yet a need for a virtual data warehouse of a prototype data warehouse) in wireframes or using a graphical user interface. More hands-on prototypes can be developed using other technological analytics tools. Depending on the type of front-end application, reports, dashboards, OLAP cubes a choice in technology can be made. Unless the previous phase has been skipped, the prototype data warehouse should be the provider of the data that is the basis for the prototypes. Similar to the data warehouse prototyping tools, tools for the prototyping of BI front-end applications are widely available. Examples of these tools are Pentaho BI Suite, Jaspersoft BI Suite, Vanilla BI, Rapidminer, Weka, Tableau, Access, Excel, etc.

4.4.5.6 Phase 4.6 Evaluate Prototypes

The prototype applications should be presented to the end-users. They must be well informed of the fact that these are prototypes and not the final developed application. As said the data quality should be ensured, also in the prototype applications, otherwise the end-users will not trust the system and its development. The goal of this phase is to evaluate the set of requirements by the end-users. The evaluation can be done in a number of ways. An issue list can be kept, sessions can facilitate evaluation, this depends on the situation. Initially, a feedback loop should be made between this phase en the previous phase. If needed the feedback loop can be extended to go back to phase 2 or phase 3.

4.4.5.7 Phase 4.7 Redefine Requirements

The evaluation and validation of prototypes should lead to more refined requirements. After a number of loops between phase 4.6 and 4.5 and phase 4 and 2 or 3 this should lead to a final list of requirements that is the outcome of the model, or phase 5.

4.5 Conclusions

The key aspects of the proposed new requirements analysis model includes the involvement and engagement of end-users and a combination of a structured method, the AGKDI model, with unstructured analysis methods to elicit a set of information requirements for business intelligence development projects.

The model is developed according to the requirements that were set in the previous chapter. The next step was to validate the model by means of a case study at de Bijenkorf e-commerce department.

5

Validation and Testing

The information requirements analysis model that has been presented in the previous chapter has been validated and tested by means of a case study at de Bijenkorf e-commerce department. The goal of validating the model is to evaluate the capability of the model to be used in a requirements analysis phase in a business environment.

[Hevner et al., 2004] presents the case study as an observational method to evaluate the artifact that has been designed, in this case the information requirements collection and analysis model. The case study is an observation method used to induct knowledge or to test methods and models. Alongside a number of criteria, one or multiple case studies can be evaluated [Yin, 2002].

This chapter elaborates on the set up and execution of this case study.

5.1 Validation criteria

The goal of this case study is to evaluate the ability of the model to support the gathering of a complete and qualitative set of information requirements in an business environment. During a case study, the researchers have executed the information requirements model in a business environment, de Bijenkorf. To measure the success of the model, the case study has been assessed by evaluating a set of criteria. In the second and third chapter of this research, a set of arguments and requirements has been formulated. With these, a new model has been designed. These arguments and requirements are the basis for the criteria by which the case study is evaluated.

The first criterium is user involvement and communication. The utilizing system that is to be developed has to fit the needs of the organization and stakeholders. The end-user involvement includes communication between developers and end-users and the perceived barriers to communication. A parameter indicating the end-user involvement, is the number of touch points between developers, stakeholders and prototypes. These can be meetings, emails, phone calls or quick conversations. The usage frequency of the prototypes is also used as a qualitative metric.

The completeness and quality of the resulting set of requirements is the second criterium. Although difficult to measure, this has been the key argument for designing an information requirements collection and analysis model. Completeness and quality of the requirements can not be quantified and thus need to be based upon the perceptions of both the developers and end-users. An argument for using both a goal- and demand-driven method has been that the methods individually are not capable of finding a complete set of requirements. Combined, the two methods are supposed to complement each other. When evaluating the completeness of the set of requirements, the matter in which the two requirements methods complement each other is thus a parameter. The quality of requirements is measured subjectively by looking at the evaluation of the end-users with regard to the prototypes.

The final criterium is the flexibility of requirements during the process. Flexibility and quality of requirements have been the main arguments for including validation, iteration and prototyping into the model. If all information requirements have only been prototyped once and no iteration was necessary, based on the end-users evaluation, the model does not require the validation and consolidation phase. The parameter for this criterium is the number of iteration cycles, adjustment and changes to the information requirements.

The criteria will be measured alongside the phases of the model and the overall process during the case study at de Bijenkorf.

The hypothesis, set for the validation of the information requirements model by means of a case study, is that the end users feel, and are involved in the process. This will result in a complete and correct set of requirements that has been developed through a number of prototyping cycles. The information requirements collection and analysis model contributed to a correct set of requirements of sufficient quality.

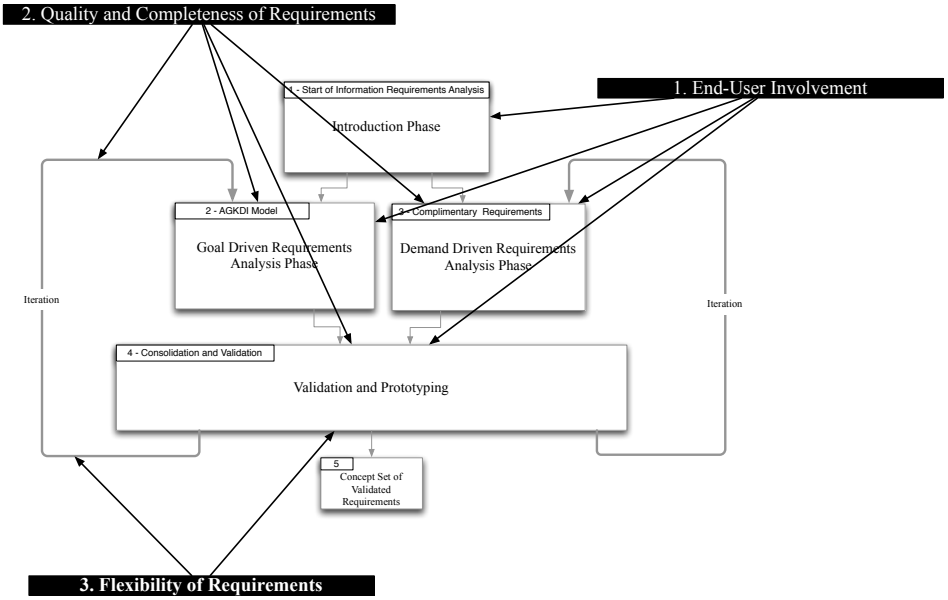

Figure 5.1: Theoretic Validation of the Model

Besides evaluating the criteria that have been set, the lessons learned during the execution of the model in the case study will be discussed.

5.2 Validation by theory

Before the model is validated by means of a case study, validation is briefly done by evaluating the model and underlying theory with the use of the criteria. In figure 5.1 the three criteria are mapped onto the model.

The first criterium, end-user involvement was included in all phases of the model. The model prescribes developers to include end-users from start to finish of the project. Involvement is used as input for the requirements and as evaluation of the set of requirements.

The quality and completeness of requirements is difficult to assess when looking at the model. The two complementary requirements gathering methods are intended to provide with a (more) complete set of requirements. By including end-users and validating the requirements by prototypes, quality is addressed in the model. Quality is hereby seen as subjective to the end-users.

The flexibility of information requirements is addressed in the model by the use of validation and iteration cycles. When requirements change during the business intelligence development lifecycle, the model can be used again to cope with these changes. The model is designed to address changing requirements at low costs by using prototyping as a validation method.

The three criteria that have been set for validation are all addressed in the model. The quality, completeness and flexibility of requirements are difficult to assess by theoretically testing the model. However, the model is designed to create a correct set of requirements and to cope with changing requirements and needs.

5.3 Case study at de Bijenkorf

This section describes, on a high level, the set up and execution of a case study at de Bijenkorf and the results of the case study with regard to the validation criteria that have been set. First, the business intelligence challenge at de Bijenkorf will be described, then the set up and execution of the process will be briefly elaborated upon. Then the set of validation criteria will be evaluated and finally the lessons that have been learned during the case study will be presented. The full report of this case study can be found in Appendix A.

5.3.1 Business Intelligence Challenge at de Bijenkorf

The organization de Bijenkorf is a traditional luxury department store and has 13 physical stores that are spread across the Netherlands, the biggest stores are those in Amsterdam, The Hague and Rotterdam. The customers of de Bijenkorf represent an overall image of the Dutch society including all ages, genders and races. The online store of de Bijenkorf was relaunched in march 2009 after the first online store was unsuccessful.

A new strategy was formed in which only de Bijenkorf core-activities and management of the web shop are kept in-house. All other (mainly technology and operational) activities are outsourced to specialist companies in the Netherlands. The online product portfolio contains more than 25.000 products, divided in 8 categories.

To get information about the online customers, their behavior, the products that they are interested in and the overall performance of the web shop, analyses is done

using the data that is collected through de Bijenkorf web shop. These analysis are currently done ad-hoc with conventional tools such as Microsoft Access, Excel and Google Analytics. Because of the outsourcing strategy of de Bijenkorf, the datasources for these analysis are distributed amongst different external databases. These databases contain for example clickstream-, order- and customer data.

Examples of analytical queries that exist within the e-commerce department are: "how can we segregate our customers?", "how can we sell more cosmetics?", "what is the retention rate of customers that reach the web shop through Search Engines?", "what are the current trends in our internal search functionality?."

The current analysis methods of de Bijenkorf.nl are not capable of answering these questions. Because of the ad-hoc nature of these techniques, it is difficult to provide a structured view of business performance. The business users at de Bijenkorf would like to be able to access analysis at any given moment. Business Intelligence methods can facilitate these goals with reports and dashboards.

The employees of de Bijenkorf e-commerce department are mostly skilled in social sciences and alpha studies. The manager and the members of the business development team are the only end-users that have technical skills and a deeper understanding of information systems and business intelligence.

5.3.2 Information requirements analysis at de Bijenkorf

During a period of six months, the research for the information requirements analysis model and execution of the case study to validate this model, has been done at de Bijenkorf. As seen in the previous chapter, the business problem is real and there is a clear need for a business intelligence system that provides with intelligence and insights.

The full description of contents of the case study can be found in Appendix A. The remainder of this subsection is a summarized description of the case study.

5.3.2.1 Introduction phase

The kick-off of the project, for the developer or researcher, was an ad-hoc analysis assignment at the e-commerce department. The goal of this assignment was to get an introduction into the industry, the operation of de Bijenkorf.nl, the organization and the available data. The project was then introduced to the developers by the

stakeholders. By using the ad-hoc analysis as a showcase for business intelligence and the available data, the end-users got a feeling of the possibilities of the project.

Together with the project sponsor and the e-commerce manager, the developers agreed upon the business problem definition and the scope of the development process. Decisions were made with regard to timing, the intended end-users, priorities and technology. Because of the research that was done in parallel, clear project planning and management was not established.

Part of the preparations for the remainder of the project was an organizational/stakeholder analysis.

5.3.2.2 Goal Driven Analysis

The goal driven analysis was done by using the AKGDI method. To prepare for the modeling during this phase relevant documentation was reviewed. Employees of de Bijenkorf have performance based contracts that include a set of goals to be met in their job descriptions. Together with the organizational model this has been used as the main source of input. The preparations further entailed planning the sessions together with the stakeholders. Because of the scope of the research, modeling phase was limited to the e-commerce management and online-marketing team.

In the first session, together with the e-commerce manager and the online marketing manager, the entire AKGDI (within the scope of the research) was run through. During the session, large sheets of paper and post-its were used to document the model. After the first session the model was digitized and presented to the stakeholders. To get deeper knowledge of the decisions the are made by the agents deeper in the organization, another session was held with the online marketing manager and the off-site marketeer. The model was refined by the efforts of this meeting. This step is relevant because all stakeholders that are going to work with the system should be included in the requirements analysis process.

The resulting model, that can be seen in figure 5.2, was then transformed into a list of information requirements. The corresponding dimensions and metrics together form the dimensional data model.

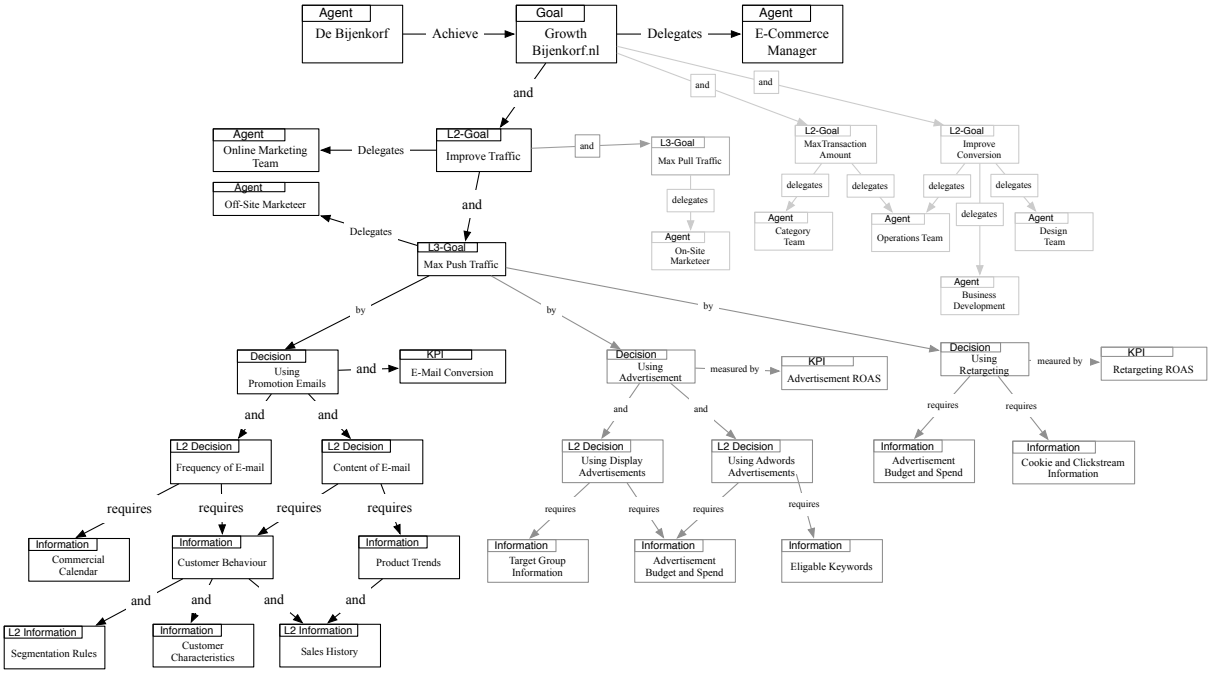

Figure 5.2: AKGDI Model at de Bijenkorf

5.3.2.3 Demand Driven Analysis

For the demand driven method, the first step was making preparations, this included planning sessions with the stakeholders and auditing internal and external sources as input for these sessions.

From the organization, the currently existing information systems, reports and dashboards were used as inspiration and input. For strategical aspects, the main input was the weekly report. This report was put together by the off-site marketeer and distributed to the CEO, e-commerce manager and the e-commerce team. The report contains performance indicators such as the weekly profits and revenue indices. Per team a key performance indicator is included in the report, for example the conversion for the design and marketing teams. Other input came from dashboards and BI systems such as business objects, google analytics and lightship.

To broaden the scope of the analysis, external sources were used. For de Bijenkorf these were, for example, reports and dashboards built by other developers and those available in commercially available systems. The business intelligence solution at the

e-commerce department of Selfridges, the owner of de Bijenkorf was also used as input.

Together with the e-commerce manager and the members of the assortment- and online marketing team, a list of information requirements was formulated over the course of a couple of weeks. This was done by having multiple meetings, interviews, sessions and brainstorms. These information requirements were then transformed into a dimensional data model by formulating the dimensions and metrics that correspond to the requirements.

5.3.2.4 Validation through Prototyping

The requirements that were derived from both methods were homogenized by using the dimensional data model. The information requirements were mapped onto this model unless the required dimensions and metrics were already in place.

The first prototype was a copy of the original weekly report, that was manually updated in Excel every week. By using ETL processes, this was automated. By connecting data in a prototype data warehouse, new insights and information requirements could be produced. After development, all stakeholders were asked to comment on the prototype in any way possible. The comments were processed and the information requirements, dimensional model, ETL processes and prototypes were adjusted. After which the stakeholders were again asked for their input. The weekly report had six major revamps and numerous small adjustments in between the iteration cycles. In a similar manner the other reports and dashboards were prototyped and iterated.

Any comments, adjustments and modifications to the information requirements, based on the experience with the prototypes, were made as soon as possible. The adjustments made to the information requirements during the process were documented immediately. The corresponding prototypes, data model and ETL processes were adjusted to start a new iteration cycle. During the validation phase, the stakeholders were involved and a large number of requests were made for new information and to change requirements. An example of a prototype can be seen in figure 5.3

5.3.3 Information requirements model evaluated through validation criteria

Prior to the case study, three validation criteria have been set to evaluate the information requirements model. These criteria are: the involvement of stakeholders and the

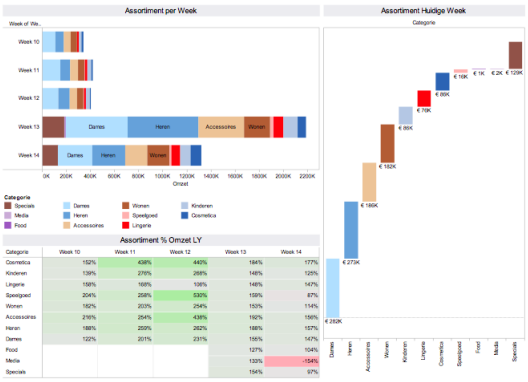

Figure 5.3: Example of Tableau Dashboard Prototype

communication between them and the developers; the completeness and quality of the set of information requirements and finally the matter of iteration during the process.

The overall process and execution of the four phases within the model, during the case study, have been assessed according to these criteria. The resulting evaluation is described in the following sections.

5.3.3.1 End-user involvement

One of the main arguments for developing an information requirements analysis model was the lack of end-user involvement in the requirements analysis process during business intelligence development. The proposed information requirements analysis model has the goal of including the end-users in all the phases of the model.

During the case study at de Bijenkorf, the end-users were involved in all phases of the model. In the introduction phase, the project sponsor, e-commerce manager, and online marketing manager were the most involved stakeholders. Their interest was to convey their needs and requirements with regard to the resulting business intelligence system. By performing an ad-hoc analysis assignment prior to the start of the information requirements gathering, the end-users got a sense of the possibilities and capabilities of business intelligence. This basis provided with a good basis to overcome any problems with terminology and understanding with regard to business intelligence.

During the two requirements analysis phases, the end-users were involved into the process. Their input was used to complete the requirements analysis phases. During the requirements phases, the input and involvement of the end-users was limited. The

members of the management had a more complete view on the organization and the possibilities of the available data.

The actual decision makers got highly involved in the process during the validation phase. The prototypes that were used during this phase gave them a feeling of what the system would look like and what the possibilities were. This triggered the end-users to become involved into the process and to convey their needs. During the validation phase and the corresponding meetings and communication, a large number of requirements were added to the system. It can be deducted that the (hands-on) experience with prototypes of the utilizing system triggers decision makers and end-users to get more involved in the process.

Some prototypes were developed during more than ten iteration cycles, indicating the number of large modifications and adjustments posed by the end-users. Overall, the stakeholders were highly involved in the process. Depending on the type of stakeholder, the starting point of involvement differs.

5.3.3.2 Completeness and quality of the set of Information Requirements

To reach the goal of a more complete and qualitative set of requirements, two methods, a goal driven and demand driven requirements analysis method, have been combined. The assumption has been that the two models complement each other during information requirements eliciting. The completeness and quality of the set of information requirements as found during the case study is subjective to the stakeholders. Because only one model has been used to elicit the requirements, conclusions can not be directly drawn with regard to the completeness and quality of all requirements found. However, by looking at the differences in the requirements that were found during the individual requirements analysis phases, conclusions can be drawn with regard to the ability of the two methods to complement each other.

The stakeholders of de Bijenkorf e-commerce department were both involved in the goal driven phase and the demand driven phase.

To get a more complete and qualitative set of requirements, a goal-driven and demand-driven requirements method have been combined to be complementary during information requirements analysis. Both methods were performed completely (within the scope of the project). The sets of requirements were very similar, especially the

requirements that are directly related to goals, decisions and stakeholders within the organization showed a significant overlap.

During the demand driven method, requirements were found that are not directly used in decisions by stakeholders, but do relate to the goals that exist within the organization. At de Bijenkorf e-commerce department, there are no employees directly responsible for customer relationship management. During the demand driven analysis the management of the e-commerce department conveyed their need for insights into to the behavior of customers. Because these requirements are not linked to any decisions made by stakeholders, they can be seen as indicators that show the performance of the organization with regard to the relationship with the customer.

It has been observed that the two methods do complement each other, but a significant part of the requirements found by both methods overlap. The conclusion can be drawn that the methods complement each other and that a more complete set of requirements can be found than when only one method is used during requirements analysis.

With regard to the quality of the information requirements, no conclusions can be drawn besides commenting on the perception of the quality as seen from the developers and stakeholders. The quality of the requirements should be measured quantitatively by looking at the improvement in the outcome of the goals that have been reached by making decisions upon the information and intelligence from the developed business intelligence system.

5.3.3.3 Iteration in the process

The final criterium is the need for iteration and validation during the information requirements analysis process. If the requirements would have been set in one or two iterations, there is no direct need to validate the requirements during the early requirements analysis phase.

As seen in the description of the case study, a number of iteration cycles have been done during the requirements analysis at de Bijenkorf. Especially the non-management end-users and decision makers were very active during the validation phase. The end-users were triggered to get more engaged by the prototypes. They used the validation step as a source for new information requirements. The experience with the system

and business intelligence that the end-users gain during the process contributes to the number of iteration cycles.

The weekly report that was prototyped for de Bijenkorf was redeveloped more than six times, numerous small adjustments and modifications were made in between the iterations. The other prototypes were also iterated numerous times and a large number of adjustments and modifications have been made.

The validation phase has proved to be an essential element of the model during the case study at de Bijenkorf. The iteration did not only validate and evaluate the information requirements that were found using the requirements analysis methods, but also uncovered a number of information requirements that were missing.

5.3.4 Situation before and after modeling at de Bijenkorf

The goal of the requirements modeling at de Bijenkorf was to collect and analyze a set of information requirements for a business intelligence system. After modeling at de Bijenkorf was finished, the project owner decided to propagate the prototype system into the production system. This section will briefly cover the implications of the project on the e-commerce department.

Dashboards and reports for the assortment team and the online marketing team were developed. The two teams were involved during the development lifecycle. The analytics were made available through the intranet at de Bijenkorf.

Before the current system, the online marketing team spent on average four hours weekly creating the week report. The week report can now be printed at all times directly from the server. The reports can be dynamically built using filters and selections. Similarly, time is saved by having all marketing related data available through a single reporting system.

For the assortment team, dynamic dashboards are available, giving them combined information about sales, returns, traffic, data quality and search engine optimization data through the entire product hierarchy. The dashboards are very graphic and include product images to create context.

A number of clear improvements in business operations for de Bijenkorf e-commerce have been found after the case study. Time spent on analysis was drastically lowered. Automating analysis reduced the time for analysis from hours to minutes for some employees. Connecting data sources opened up a range of new possible insights that were

previously unavailable to the teams. By shifting all analytical requests to a single system and a dedicated analyst, employees can focus on their core activities. The clear set of information requirements helps end-users by presenting actionable insights. Previously, end-users had to collect all available data and search for insights and information themselves.

5.3.5 Lessons learned at de Bijenkorf

A number of key lessons have been learned. First of all, data quality is essential for acceptance and engagement of the stakeholders. If data is incomplete or incorrect, the end-users will not be committed to the system and will not be encouraged to help further developing the system. Communication is important, the end-users need to be engaged at all times, this stimulates the process and makes the stakeholders committed to the project and system.

When looking at the model, a number of changes and modifications can be recommended. During the case study at de Bijenkorf, the order of the process was the introduction, then the unstructured method, the structured method and finally the validation step. The two analysis phases should switch. The goal-oriented requirement analysis method AKGDI gives a structured view of the operations and business of an organization and includes the essential and common information requirements. To complete the information requirements the demand-driven method should be used. This method also covers the information requirements that are not immediately related to any decision, goal or KPI that is documented within an organization. Moreover, it gives creative and less structured employees a platform to think of information requirements that they will use.

Another learning is that iteration during validation does not immediately imply going back to the requirements model and going through the entire model again. By keeping short communication lines and small barriers to make modifications and adjustments, the validation phase was very successful.

Because research and the case study were done in parallel, good structure and process management was lacking. When projects become large, complex and include a large number of stakeholders, good project- and requirements management is essential.

The requirements that were derived from the demand driven method did not directly relate to decisions made by stakeholders, they could be related to the goals of the

organization. In the model, two different documentations are used, a model for the goal driven method, and a list for the demand driven method. It could be possible to create a single medium to organize and document the requirements. A possible solution can be a CASE software tool to document the information requirements analysis model.

5.4 Conclusions

The case study has been used to validate the model alongside three criteria. The model, as used at de Bijenkorf, past the validation by the means of these criteria. The completeness and quality of the requirements that are a result of the model can not be validated.

The problem with a single project case study, is that it is difficult to generalize the findings. The business intelligence development cycle at de Bijenkorf is fairly simple due to the size of the organization and the short period of operation. To completely validate the model, more case studies need to be done in various environments. Also, the observations that have been done are subjective, because the researchers acted as developers during the case study.

To improve the case study and the conclusions that can be drawn from the case study, quantitative metrics can be used to assess the quality of the model.

For de Bijenkorf, the model proved to be a successful method for requirements analysis. Currently, the set of requirements is used as input for the further development of the business intelligence solution. The prototype that has been developed, is used as a basis for the business intelligence platform.

The model can be improved by taking into consideration the lessons that were learned during the case study. This includes providing quality data during the validation process; following the phases as specified by the model; elaborating upon project planning, management and documentation.

6

Conclusions and Recommendations

6.1 Conclusions

Business Intelligence is an integral element of business processes for many companies. Information systems and management information systems feed employees with intelligence and insights that they use to make decisions. System development usually starts with requirements collection and analysis. One of the causes of failure in projects is setting the wrong requirements.

In this thesis, three information requirements analysis methods have been analyzed. Key lessons learned from this research are that the communication between developers and stakeholders is essential and has to be established at project start and has to continue until project finish; there is a lack of validation of requirements and the current methods fail to produce a complete set of requirements of sufficient quality.

A set of requirements has been formed that argue a need for the development of a new model. An information requirements collection and analysis model was developed with communication and validation as key aspects.

The model, as designed during this research, was validated by means of a case study at the e-commerce department of the Dutch retailer de Bijenkorf. The validation was done by evaluating three criteria: the end-user involvement; completeness and quality of the resulting information requirements; the need for iteration during requirements analysis.

The stakeholders of de Bijenkorf were highly involved during the requirements analysis phase. The validation phase within the model was used as another source of input for requirements, especially by end-users and decision makers.

The two requirements analysis phases complemented each other, this resulted in a more complete set of information requirements. The goal driven requirements method forms a basis set of requirements that is then completed by requirements found using the demand driven method. The first method is structured and the information requirements are a reflection of the organization. The second, demand driven method, utilizes the creativity of end-users and external sources.

The validation phase within the model was intensively used at de Bijenkorf. This resulted in multiple iteration cycles to refine information requirements. During the prototyping sessions, a large number of new information requirements were formulated. This can be acclaimed to the fact that the end-users got more experienced with the system and business intelligence.

Other lessons learned with regard to the model are that project management is essential during the whole project; the sequence of the model should be followed; the requirements need to be properly documented.

It is difficult to generalize the findings of the case study because of two reasons. First, the case study has only been performed in one project, more testing and validation is needed. Second, the situation at de Bijenkorf was fairly simple. The requirements analysis phase is one element in the business intelligence lifecycle. Problems can arise further down the cycle where the process will become more complex.

6.2 Recommendations

The information requirements collection and analysis model supports the gathering of a correct set of requirements with sufficient quality. The model can be utilized through out the business intelligence development lifecycle. By using a structured and less structured requirements eliciting techniques, the methods used in the model complete each other.

Extensive validation gives the end-users a front seat during development. The advancements in technology causes reduced time and costs for prototyping and evolutionary development.

The information requirements collection and analysis model focusses on end-users and stakeholders with limited knowledge into business intelligence or technology. By using the model, the information requirements can be set together with the developers and stakeholders. This creates commitment to the system by the end-users. The resulting business intelligence system will fit the needs of the end-users more closely. The BI system will give the end-users actionable information and insights upon which they can base their decisions.

Having a well designed business intelligence system will save time and costs for organizations. The end-users can focus on making decisions in stead of waisting time doing ad-hoc analysis.

In academia, a fair amount of research is being done, but a large part of the knowledge lies within corporations who are not willing to share this because of competition. More knowledge should be made available by these corporations to stimulate research in the business intelligence area by academics.

Gathering information requirements differs very much from eliciting conventional requirements for IT systems. This research has combined the existing research and corresponding methods into a complete model. The model should be tested more extensively during various types of business intelligence development cycles.

The goal-oriented model was not completely developed for the Bijenkorf case. Future efforts should try to prove that the model is capable of modeling various types of organizations, goals, decisions and information requirements. Also, documenting these modeling activities should be supported by a set of standard rules, and preferably with a graphic user interface or other software tool. There is a need for such a tool to support the modeling activities from the AKGDI model including the development of the data model and mapping of the source systems.

The position of information requirements in the Business Intelligence lifecycle has to be further researched. The changes in the requirements analysis method can lead to problems further down the road of development.

Bibliography

Daniel J. Abadi, Samuel R. Madden, and Nabil Hachem. Column-stores vs. row-stores: how different are they really?, 2008.

Suhail Ansari, Ron Kohavi, Llew Mason, and Zijian Zheng. Integrating e-commerce and data mining: Architecture and challenges, 2001.

D. Barone, E. Yu, J. Won, L. Jiang, and J. Mylopoulos. Enterprise modeling for business intelligence. *The Practice of Enterprise Modeling*, pages 31–45, 2010.

G. Barone, M. Holsten, S. Ratchev, and F. Weber. User requirements and ce needs. 1999. 16

Robert Baumgartner, Georg Gottlob, and Marcus Herzog. Scalable web data extraction for online market intelligence. *Proc. VLDB Endow.*, 2(2):1512–1523, 2009.

J. Becker, C. Janiesch, D. Pfeiffer, and S. Seidel. Evolutionary method engineering: towards a method for the analysis and conception of management information systems. pages 3922–3922, 2006.

M. Berry and G. Linoff. *Mastering data mining: The art and science of customer relationship management.* John Wiley & Sons, Inc., 1999. 4

B.W. Boehm. Software engineering economics. *IEEE Transactions on Software Engineering*, 1984. 4

A. Bonifati, F. Cattaneo, S. Ceri, A. Fuggetta, S. Paraboschi, et al. Designing data marts for data warehouses. *ACM Transactions on Software Engineering and Methodology*, 10(4):452–483, 2001. 22

BIBLIOGRAPHY

P. Britos, O. Dieste, and R. GarcÌa-MartÌnez. Requirements elicitation in data mining for business intelligence projects. *Advances in Information Systems Research, Education and Practice*, pages 139–150, 2008. 13, 16

L. Burmester and M. Goeken. Method for user oriented modelling of data warehouse systems. pages 366–374. Citeseer, 2006. xi, 13, 16, 22

D.P. Busser. A new requirements analysis method for business intelligence development projects. Master Thesis, 2011.

E. Chang, T.S. Dillon, T. Dillon, and F.K. Hussain. *Trust and reputation for service-oriented environments: Technologies for building business intelligence and consumer confidence*. Wiley, 2006. xi, 2, 3

Wingyan Chung, Hsinchun Chen, and Jr. Jay F. Nunamaker. A visual framework for knowledge discovery on the web: An empirical study of business intelligence exploration. *J. Manage. Inf. Syst.*, 21(4):57–84, 2005.

Ju Chunhua. Effectiveness of olap-based sales analysis in retail enterprises. volume 3, pages 240–244, 2008.

M. Cogliati, P. Britos, and R. García-Martínez. Patterns in temporal series of meteorological variables using som & tdidt. *Artificial Intelligence in Theory and Practice*, pages 305–314, 2006. 16

Cooley, Mobasher, and Srivastava. Web mining: Information and pattern discovery on the world wide web, 1997.

G. B. Davis. Strategies for information requirements determination. *IBM Systems Journal*, 21(1), 1982. xi, 5, 17, 25, 27, 28, 29, 35

Stephen G. Eick. Visualizing online activity. *Commun. ACM*, 44(8):45–50, 2001.

P. Felgaer, P. Britos, and R.Ó.N. GarcÍA-MartÍNez. Prediction in health domain using bayesian networks optimization based on induction learning techniques. *International Journal of Modern Physics C*, 17(3):447–456, 2006. 16

I. Gam and C. Salinesi. A requirement-driven approach for designing data warehouses. 2006.

LLC Survey of DM Review readers. The Gantry Group, 2004. 2, 4

Sonja Gensler, Marnik G. Dekimpe, and Bernd Skiera. Evaluating channel performance in multi-channel environments. *Journal of Retailing and Consumer Services*, 14(1): 17–23, 2007.

Paolo Giorgini, Stefano Rizzi, and Maddalena Garzetti. Goal-oriented requirement analysis for data warehouse design, 2005. 17

Frank G. Goethals. Data warehousing and business intelligence for e-commerce: by alan r. simon & steven l. shaffer. *SIGMOD Rec.*, 31(2):71–72, 2002.

M. Golfarelli. From user requirements to conceptual design in data warehouse designña survey. 2008. xi, 13, 17, 21

Matteo Golfarelli, Stefano Rizzi, and Iuris Cella. Beyond data warehousing: what's next in business intelligence?, 2004.

Bill Goodwin. Poor communication to blame for business intelligence failure, says gartner, 2011. 2, 4, 16

A. Gosain and J. Singh. Towards data warehouse business quality through requirements elicitation. pages 464–469. IEEE, 2008.

A. Gosain and J. Singh. Achieving data warehouse quality using gdi approach. *InternationalJournalofInformationStudies*, 1(2):110–117, 2009. 19

H. Grosser, P. Britos, and R. García-Martínez. Detecting fraud in mobile telephony using neural networks. *Innovations in Applied Artificial Intelligence*, pages 613–615, 2005. 16

Y. Guo, S. Tang, Y. Tong, and D. Yang. Triple-driven data modeling methodology in data warehousing: a case study. pages 59–66. ACM, 2006.

Himanshu Gupta. Selection of views to materialize in a data warehouse, 1997.

Li Haigang. Applications of data warehousing and data mining in the retail industry. volume 2, pages 1047–1050 Vol. 2, 2005.

BIBLIOGRAPHY

W. R. Hansen. Vorgehensmodell zur entwicklung einer data warehouse-lösung. *Das Data Warehouse-Konzept*, pages 311–328, 1997. 14

RVL Hartley and A. Synopsísl. Transmission of information1. 1928.

C. Herrmann. Exploring the structural dimension of data warehouse organizations: Results of a survey and implications. pages 350–358. Citeseer, 2004. 18

A.R. Hevner, S.T. March, J. Park, and S. Ram. Design science in information systems research. *Mis Quarterly*, pages 75–105, 2004. 8, 53

E. Hochsztain, S. Mill #225, B. Pardo, J. M. Pe #241, and E. Menasalvas. A framework to integrate business goals in web usage mining, 2003.

C. Hood, S. Wiedemann, S. Fichtinger, and U. Pautz. *Requirements Management*. Springer, 2008. xi, 4, 5, 11, 25, 26

Xiaohua Hu and Nick Cercone. A data warehouse/online analytic processing framework for web usage mining and business intelligence reporting. *Int. J. Intell. Syst.*, 19(7):585–606, 2004.

B. Inmon. Wherefore warehouse? *Byte*, 23(1):1, 1998. 12

Yong Soo Kim, Bong-Jin Yum, Junehwa Song, and Su Myeon Kim. Development of a recommender system based on navigational and behavioral patterns of customers in e-commerce sites. *Expert Systems with Applications*, 28(2):381–393, 2005.

R. Kimball, L. Reeves, W. Thornthwaite, M. Ross, and W. Thornwaite. *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses with CD Rom*. John Wiley & Sons, Inc., New York, NY, 1998. xi, 3, 4, 12, 14, 15

M. Klesse and R. Winter. Organizational forms of data warehousing: An explorative analysis. 2007.

Il Ko and Sarvar Abdullaev. A study on the aspects of successful business intelligence system development. In Yong Shi, Geert van Albada, Jack Dongarra, and Peter Sloot, editors, *Computational Science – ICCS 2007*, volume 4490 of *Lecture Notes in Computer Science*, pages 729–732. Springer Berlin / Heidelberg, 2007. 4, 5

Ron Kohavi. Mining e-commerce data: the good, the bad, and the ugly, 2001.

Ron Kohavi and Foster Provost. Applications of data mining to electronic commerce. *Data Mining and Knowledge Discovery*, 5(1):5–10, 2001.

Ron Kohavi, Llew Mason, Rajesh Parekh, and Zijian Zheng. Lessons and challenges from mining retail e-commerce data. *Mach. Learn.*, 57(1-2):83–113, 2004.

M. Kumar, A. Gosain, and Y. Singh. Agent oriented requirements engineering for a data warehouse. *ACM SIGSOFT Software Engineering Notes*, 34(5):1–4, 2009. 44

M. Kumar, A. Gosain, and Y. Singh. Stakeholders driven requirements engineering approach for data warehouse development. 2010. xi, 18, 19, 20

Wilburt J. Labio, Yue Zhuge, Janet L. Wiener, Himanshu Gupta, H #233, ctor Garc #237, a Molina, and Jennifer Widom. The whips prototype for data warehouse creation and maintenance. *SIGMOD Rec.*, 26(2):557–559, 1997a.

W.J. Labio, D. Quass, and B. Adelberg. Physical database design for data warehouses. In *Data Engineering, 1997. Proceedings. 13th International Conference on*, pages 277–288. IEEE, 1997b. 6

Beate List, Robert Bruckner, Karl Machaczek, and Josef Schiefer. A comparison of data warehouse development methodologies case study of the process warehouse. In Abdelkader Hameurlain, Rosine Cicchetti, and Roland Traunmüller, editors, *Database and Expert Systems Applications*, volume 2453 of *Lecture Notes in Computer Science*, pages 203–215. Springer Berlin / Heidelberg, 2002.

J.N. Mazon, J. Pardillo, and J. Trujillo. A model-driven goal-oriented requirement engineering approach for data warehouses. pages 255–264. Springer-Verlag, 2007. 17

Andreas Meier, Nicolas Werro, Martin Albrecht, and Miltiadis Sarakinos. Using a fuzzy classification query language for customer relationship management, 2005.

L.T. Moss and S. Atre. *Business intelligence roadmap: the complete project lifecycle for decision-support applications*. Addison-Wesley Professional, 2003. xi, 2, 4, 41

S. Negash. Business intelligence. *The Communications of the Association for Information Systems*, 13(1):54, 2004. 2

BIBLIOGRAPHY

F.R.S. Paim and J.F.B. de Castro. Dwarf: An approach for requirements definition and management of data warehouse systems. pages 75–84. IEEE, 2003. 22

Shan L. Pan and Jae-Nam Lee. Using e-crm for a unified view of the customer. *Commun. ACM*, 46(4):95–99, 2003.

N. Prakash and A. Gosain. Requirements driven data warehouse development. In *CAiSE Short Paper Proceedings*, pages 13–17. Citeseer, 2003. 18, 44, 90

Stefano Rizzi, Alberto Abell #243, Jens Lechtenb #246, rger, and Juan Trujillo. Research in data warehouse modeling and design: dead or alive?, 2006.

K. Rouibah and S. Ould-ali. Puzzle: a concept and prototype for linking business intelligence to business strategy. *The Journal of Strategic Information Systems*, 11 (2):133–152, 2002. 1

Chris Rygielski, Jyun-Cheng Wang, and David C. Yen. Data mining techniques for customer relationship management. *Technology in Society*, 24(4):483–502, 2002.

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Analysis of recommendation algorithms for e-commerce, 2000.

Arun Sen and Atish P. Sinha. A comparison of data warehousing methodologies. *Commun. ACM*, 48(3):79–84, 2005.

C.E. Shannon. Prediction and entropy of printed english. *Bell System Technical Journal*, 30(1):50–64, 1951.

Dominik Slezak, Jakub Wroblewski, Victoria Eastwood, and Piotr Synak. Brighthouse: an analytic data warehouse for ad-hoc queries. *Proc. VLDB Endow.*, 1(2):1337–1345, 2008.

Composite Software. Data warehouse prototyping. Website, 2011. URL `http://www.compositesw.com/solutions/data-warehouse/data-warehouse-prototyping/`. xi, 24, 25

E. Soler, V. Stefanov, J.N. Mazon, J. Trujillo, E. Fernndez-Medina, and M. Piattini. Towards comprehensive requirement analysis for data warehouses: Considering security requirements. pages 104–111. IEEE, 2008.

I.Y. Song and K. LeVan-Shultz. Data warehouse design for e-commerce environments. *Advances in Conceptual Modeling*, pages 374–387, 2010.

Srini S. Srinivasan, Rolph Anderson, and Kishore Ponnavolu. Customer loyalty in e-commerce: an exploration of its antecedents and consequences. *Journal of Retailing*, 78(1):41–50, 2002a.

Srini S. Srinivasan, Rolph Anderson, and Kishore Ponnavolu. Customer loyalty in e-commerce: an exploration of its antecedents and consequences. *Journal of Retailing*, 78(1):41–50, 2002b.

Jaideep Srivastava, Robert Cooley, Mukund Deshpande, and Pang-Ning Tan. Web usage mining: discovery and applications of usage patterns from web data. *SIGKDD Explor. Newsl.*, 1(2):12–23, 2000.

V. Srivastava. Requirement elicitation for data warehouse using gdi diagram and tropos methodology, 2008.

A. Taa, M.S. Abdullah, and N.M. Norwawi. Rameps: a goal-ontology approach to analyse the requirements for data warehouse systems. *WSEAS Transactions on Information Science and Applications*, 7(2):295–309, 2010. 12, 13, 17

Tableau. Fast analytics and rapid-fire business intelligence from tableau software. Website, 2011. 98

Christian Thomsen and Torben Bach Pedersen. A survey of open source tools for business intelligence. In A. Min Tjoa and Juan Trujillo, editors, *Data Warehousing and Knowledge Discovery*, volume 3589 of *Lecture Notes in Computer Science*, pages 74–84. Springer Berlin / Heidelberg, 2005.

Nectaria Tryfona, Frank Busborg, and Jens G. Borch Christiansen. starer: a conceptual model for data warehouse design, 1999.

J.R. Valusek and D.G. Fryback. Information requirements determination: obstacles within, among and between participants. In *Proceedings of the twenty-first annual conference on Computer personnel research*, pages 103–111. ACM, 1985. 16

BIBLIOGRAPHY

Michael R. Wade and Saggi Nevo. Development and validation of a perceptual instrument to measure e-commerce performance. *Int. J. Electron. Commerce*, 10(2): 123–146, 2006.

David W. Wallace, Joan L. Giese, and Jean L. Johnson. Customer retailer loyalty in the context of multiple channel strategies. *Journal of Retailing*, 80(4):249–263, 2004.

Hugh J. Watson and Barbara H. Wixom. The current state of business intelligence. *Computer*, 40(9):96–99, 2007.

Robert Winter and Bernhard Strauch. A method for demand-driven information requirements analysis in data warehousing projects, 2003. xi, 15, 16

Robert Winter and Bernhard Strauch. Information requirements engineering for data warehouse systems, 2004. 13, 22

R. Wolfert and S. Dikmans. Data inzicht = doel in zicht. Presentation, 2011. 23

R.K. Yin. Case study research: Design and methods, applied social research methods series, vol 5. 2002. 53

Yan Zhu, Christof Bornhövd, Doris Sautner, and Alejandro Buchmann. Materializing web data for olap and dss. In Hongjun Lu and Aoying Zhou, editors, *Web-Age Information Management*, volume 1846 of *Lecture Notes in Computer Science*, pages 201–215. Springer Berlin / Heidelberg, 2000.

Appendix A

Appendix: Information Requirements at de Bijenkorf.nl

This chapter describes the validation of the information requirements model that has been done at de Bijenkorf. The entire research was done at de Bijenkorf in the first half of 2011, the case study included a number of phases and iterations, all described here.

A.1 Bijenkorf.nl Business Intelligence Goals

In section 1, the general background of the challenge posed by de Bijenkorf has been presented. To summarize, the online retail channel of this large Dutch Department store has grown since the start in 2009 with an average sales index of around 300%. To continue this growth, decisions need to be made on a strategic and tactical level to, amongst others, gain more traffic, retain customers, convert more, and sell more often and with a higher average spend. Online retailing differs from conventional retailing due to the large amounts of data that is generated during operations. In the stores the amount and type of sales provide most information. Customer relationship data can be collected with loyalty programs.

Because of online metrics, new insights are available to decision makers (on all levels). For example channels through which users buy their products, bottlenecks in the"store", or the types of customers and their purchasing habits. Currently, the employees of the e-commerce department of de Bijenkorf, use a range of systems to access

the data. They base their decisions upon ad-hoc analysis. Because of the distributed character of these systems, the data that is needed for these insights (information needs) is difficult to obtain. The data is not connected and heavy transformation needed before use. To overcome these hurdles, de Bijenkorf has set the goal of developing a Business Intelligence environment that enables all employees to achieve their business goals by making decisions based on insights found in the collected (internal and external) data. Most of the employees of the e-commerce department have a non-technical background. The system has to be easy to use, without too much analytical and technical hurdles. A strong visual focus is needed in the system.

The employees own and influence the strategic and tactical goals and decisions that need to be made. Decisions for operational goals and systems should be automated in the future. The business intelligence system should enable this with regard to, for example, personalized mailings, offerings etc. The operational level has not been part of this research. The business intelligence solution that will be (partly) designed has to be able to cope with future developments in this area.

deBijenkorf.nl has been used as a case to validate the model that has been developed during this research. The research into information requirements analysis has been done in parallel to the case study at de Bijenkorf. The final requirements analysis model has not been followed strictly in the order that has been presented in the previous chapter. When used in other cases, the order should be kept by using good project management, this helps in structuring the process and aligning all stakeholders in this process.

This chapter is a report of the validation of the newly developed model by means of the case study at de Bijenkorf. The focus hereby is merely on the process and the findings during this process that can help developing the model than on the requirements that have been found. The reporting will be done according to the steps of the model as all elements of the model have been fulfilled during the case study.

A.2 Start of Requirements Analysis

The introduction and starting phase of the project at de Bijenkorf was the start of this research. Prior to starting the Business Intelligence development process, an analysis assignment has been done. The goal was to map the customer base of de Bijenkorf web shop, answering questions like “what is the churn-rate of our customers over time?”,

"what type of customers do we have?", "where do they come from?", etc. This analysis was done by using rapid analytics tools such as Access, Excel, ThinkCell and Tableau. The result of this assignment was not only having insights into the customer base of de Bijenkorf but it also served as an introductory phase for all stakeholders.

As an analyst, and executor of the information requirements analysis method, the assignment meant a deeper knowledge of the business, the organization and the possible information needs. A deep insight into the available data is gained through this exercise. For the problem owners, or stakeholders, this exercise gives a feeling of what analytics and Business Intelligence is and how it can help the organization.

To start the project, a briefing was held to define the goals and scope of the project. During this meeting, the possibilities and capabilities of BI and the developers were conveyed as well as the needs of the project owners, de Bijenkorf. Both parties agreed upon the course of the project, the timing and the contents of the project.

The remainder of this section reports the organizational modeling that has been done in the preliminary phase of the case study at de Bijenkorf.

A.2.1 de Bijenkorf.nl Organization and Stakeholders

de Bijenkorf e-commerce department is a fast growing organization, both in size and in revenue, in 2009 the webshop was managed by 6 employees, in a loose structured informal setting. Currently, two years later, more then 25 employees work at de Bijenkorf, grouped in 5 teams lead by a department manager. In figure A.1 the organizational chart can be seen.

A.2.1.1 Management

The E-Commerce manager deals with all strategic goals of de Bijenkorf.nl, he is responsible for the Balance Sheet and Profit & Loss, thus he needs to make a number of well informed decisions. His overall goal can be formulated as: "The Continuation of Growth of de Bijenkorf Online Retailing activities". this goal is delegated to him by the board of de Bijenkorf corporation. He heads the management team, that consists of all managers of the 5 teams in the e-commerce department.

A. APPENDIX: INFORMATION REQUIREMENTS AT DE BIJENKORF.NL

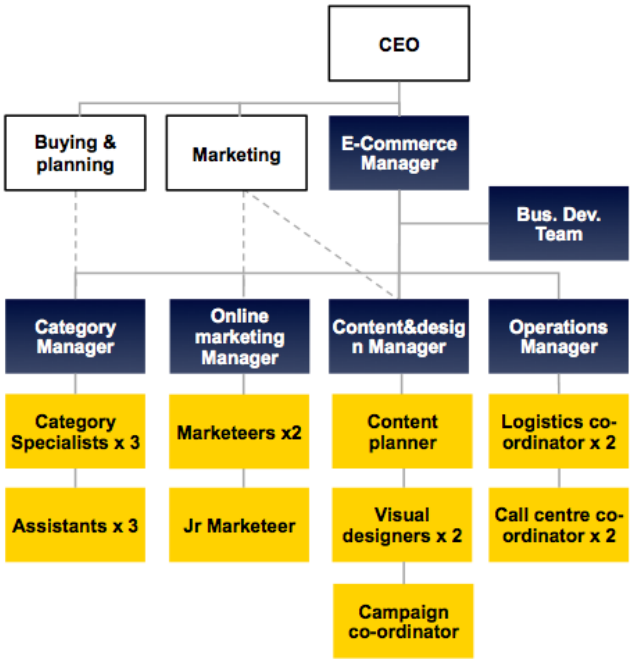

Figure A.1: Formal Chart of de Bijenkorf E-Commerce department

A.2.1.2 Business Development

This team is charged with the development of technology for the web shop. Developing a Business Intelligence solution is one of these tasks. Other tasks are for example, testing, new features, the it-outsourcing strategy, etc. Because this team has only recently been formed, the information needs will not be covered in this research, but they will be kept in mind when developing the BI prototype.

A.2.1.3 Category Team

The category team is tasked with selecting, sourcing and managing all products in the webshop. The team is subdivided by category, de Bijenkorf sells products in 8 main categories, Gents, Ladies, Children, Accessories, Cosmetics, Home, Gifts and Multimedia.

A.2.1.4 Content & Design Team

The communications team consists of content specialists and designers. The content specialists focus on online content such as event pages, blogs, information pages, etc. The designers are firstly responsible for all images and figures on the website, these change very frequently (at least once a week) according to the change in theme, products or in case of events. Moreover, they are also responsible for the photography of products and the placement of these photos.

A.2.1.5 Online Marketing Team

Online marketing has two main objectives, or goals, firstly to generate as much traffic as possible, this can be seen as"off-site" marketing. This is done through on- and offline advertisement, affiliate marketing, e-mail marketing and retargeting. Secondly their objective is to convert as much of this traffic as possible, thus"on-site" marketing.

A.2.1.6 Operations Team

Operations is responsible for the after-sales and customer care. All products that are sold are fulfilled through an outsource partner that has a very large warehouse, they pick, pack and send off all products. Because there are almost 25000 different products being sold, a large number of problems can arise from the moment an order comes in and the customer receives his order. Also, about 20 percent of all orders are returned, handling this process is also a task of the operations team.

A.2.2 Scope of information requirements

The stakeholders at de Bijenkorf web shop have a combined complex goal, to maintain current growth in sales through the web shop, which can be split up in a large number of sub-goals that become smaller and less complex. Because of limits in time during the research, only the information requirements for the Management, Online Marketing and Category team has been evaluated using the model.

A.3 Information Requirements Analysis

The scope of the project has been set during the organizational analysis, due to time constraints only the information requirements of the e-commerce manager, category team and online marketeers will be elicited during the case study. Although the final model was not yet finished when the case study commenced, the main outline was agreed upon. First, a structured method for requirements analysis with a focus on the Goals and Decisions that lie in the organization. These requirements were then complemented by using a less structured more creative demand-driven process with interviews, sessions and a large number of examples from internal and external sources.

A.3.1 Unstructured Demand Driven Requirements Analysis

The ad-hoc analysis assignment that was performed to get an insight into the business of de Bijenkorf webshop was also the basis for the preliminary requirements analysis. Moreover, during this assignment the analysis had a highly iterative prototyping character that was deemed useful by both analyst and stakeholders. Although the model was not finished, all elements of phase three of the model have been touched upon.

A.3.1.1 Preparations

The preparations for the analysis was done by the developer of the Business Intelligence solution. A number of decisions need to be made in order to roughly outline the requirements analysis process. The stakeholders that participate in the process need to be selected, this was already done during the organizational analysis. Then a planning was made to agree upon the moments for interviews and collective sessions. In this case, the decision was made to use the existing weekly dashboard as a starting point and also looking at dashboards that are used by the e-commerce industry.

For the questions and issues that would be raised during the interviews and sessions different sources were used that describe a more demand-driven requirements analysis method, such as the Data warehouse toolkit by Kimble.

A.3.1.2 Auditing external and internal sources for input

As said, the weekly dashboard was used as a starting point for the analysis. This weekly dashboard was maintained by one of the e-commerce employees, the off-site marketeer.

The original week report was made by using Excel and a number of Macro's. The main source for this analysis was Google Analytics, this is a free metrics tool by Google that measures all clickstream activity online but also measures sales that are done through e-commerce applications. Although Google Analytics is a great source for mission critical data, it is not very accurate. On average, about 5 percent of all orders will not be tagged by Analytics due to technical limitations and privacy settings. Suboptimal data as a source for the weekly report was conceived the main issue with that analysis. Also, the marketeer would spend a weekly average of 4 hours on copying and pasting the data from the source to the report. Finally, because Excel is used, the report was prone to errors due to the lack of data quality checks and human error, e.g. deleting cells and formulas etc. The information that was conveyed through the report was a collection of KPI's such as the conversion rate, number of sales, amount of sales, number of return items and value, etc. These KPI's and the corresponding graphs and figures where the first information needs that should be used in the business intelligence development process.

Google Analytics and the weekly Excel report were not the only reports or dashboard systems being used by the e-commerce department. Company-wide two management information systems are used to convey financial data, Business Objects, a commercial BI solution and Lightship, a custom built front end application based on an Oracle database. The operations team uses Crystal Reports to gain intelligence on the operational aspects of the business. For advertisement spending, Google Adwords is used and the dashboard that comes with that system. Affiliate marketing sales can be viewed through the custom built Zanox dashboard. The e-commerce webshop system, provides with a custom built dashboard giving data into orders and customers. All these reports and underlying data sources can be seen as input for information requirements analysis. Moreover, they are part of the motivation for developing a business intelligence system.

To get more inspiration for the information needs, external sources were used. Firstly, academia provides a lot of research into the information needs that exist at firms and industries. Also, the internet is a major source, there are a lot of examples of dashboards, reports and best practices in the industry. There are also databases available that contain lists of KPI's and requirements for information systems ordered by industry, type of application, etc. A selection of interesting online sources was made to

present to the stakeholders of de Bijenkorf as an input for the next phase, the interviews and sessions.

A third channel through which input for the interviews and sessions is collected is by looking at what other companies in the industry use in their operations. de Bijenkorf is a subsidiary of Selfridges Co. Ltd., a large department store in the United Kingdom, which also has an online division. As part of the case study, the business intelligence system at the e-commerce department of Selfridges was reviewed and used as an input for the sessions at de Bijenkorf. Selfridges uses a commercial business intelligence system by IBM, Coremetrics that is targeted at measuring online activity, similar to Google Analytics but more advanced. Although the input of the visit at Selfridges was very useful this was also a critical moment for de Bijenkorf with respect to the choice between a COTS and a custom made business intelligence solution. The reason for this is that the system that was used by Selfridges was highly inflexible. Every adjustment required consultants and programmers to write custom code. Also quantitative restrictions were set on the number of analysis.

A.3.1.3 Interviews and Sessions

With the preparations and examples ready, the next step were a number of interviews and sessions with the stakeholders at de Bijenkorf. The order of interviews was top-down, so first the E-commerce manager and then down to the other stakeholders. During these sessions, a list of information requirements and KPI's has been made.

During the interviews it could be noticed that when the stakeholders where talking about information needs, they mostly referred to key performance indicators. KPI's are only a part of the complete set of information requirements. From the perspective of this research, KPI's measure the success of goals. Information needs contribute to reaching these goals by supporting stakeholders during their decision making processes.

A range of different reports was already being used at de Bijenkorf, a large part of the information needs could be copied directly without any adjustments or need for immediate prototyping. With the goal of connecting the underlying data sources of these reports with each other in a business intelligence solution a range of new possible information became available. By using examples as described in the previous chapter, new possibilities could be translated into new information requirements.

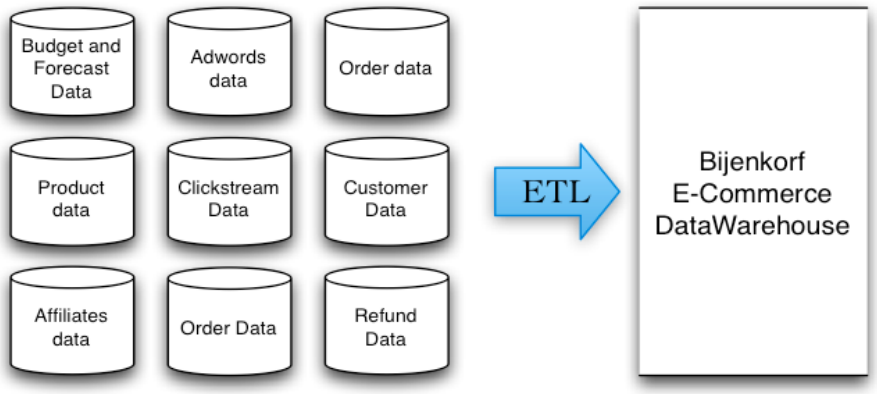

Figure A.2: Possible Datasources for de Bijenkorf DW

Connecting data sources enabled insights and intelligence that were impossible to access previously. During the interviews and sessions, these possibilities led to a large number of information requirements. To support this, all available (internal) data sources were depicted in an image that can be seen in figure A.2. These sources come from the systems that are used by the e-commerce department. The new insights that can be found by connecting data from different sources have also been found were also available in numerous external examples.

The result of this phase was a list of ideas for new insights and information that should be made available through a business intelligence solution. According to the model, during the interviews and sessions, and the entire process, the information requirements should be documented. In this case study this was partially done. The reason for this is that a clear project planning and management element was lacking. Another reason for this is because unlike provisioned by the model, during the case study, the information requirements were all prototyped and presented to the end-users in a short time period.

A.3.1.4 Write-up and Dimensional Modeling

The write-up of the information requirements has been a bit lacking during the case study because project management was not utilized. The information requirements were documented using shared documents. It is recommended, as said in the model, that this is done properly and that ownership and history of information requirements

is documented as well. During de Bijenkorf case, the prototypes are the primary documentation form of the requirements. They provide a historic view of the requirements.

As for the dimensional modeling, because during the case study rapid prototyping was used, the dimensional modeling proceeded this phase. The dimensional modeling uses the information requirements as an input. The information requirements determine the dimensions and metrics in the dimensional model, but also the grain of the model and analysis. The grain directly corresponds with the depth of the dimensions and the corresponding metrics. Using fact tables and aggregates, performance can be optimized for different types of grain in the dimensional model.

In figure A.3 one of the dimensional models shows the fact table for customer orders and corresponding dimension tables. A decision that has been made here is not to include a time dimension, so analysis can be done on a daily level and not deeper. This decision has been made after the development of prototypes and corresponding to the goal of the business intelligence solution.

During the case study, the model was developed by using modeling software provided by the database manufacturer Oracle MySQL. The reason for this was that the model can be easily implemented as a database model afterwards. Also because of the graphical interface it is easier to communicate to stakeholders than a list of dimensions and metrics. A large number of tools and methods exist to translate information requirements into dimensions and metrics.

The dimensional model is an input for the prototyping phase.

A.3.2 Structured Requirements Analysis using the AKGDI Model

The second phase of the requirements analysis model was the execution of a goal-oriented model, the AKGDI or Agent-KPI-Goal-Decision-Information model. In contrast with the more demand-driven phase three that was done first at de Bijenkorf, this model is a structured approach to gathering requirements within an organization.

Because of constraints in time and for the purpose of this report the AKGDI model was drawn up for a part of the e-commerce department only. The organization and stakeholder analysis serves as an input for the first element of the AKGDI model, the Agent-Goal modeling. For the first meeting, the e-commerce manager and the online marketing manager were invited to participate. With a whiteboard, post-its, a large sheet and markers the modeling was facilitated.

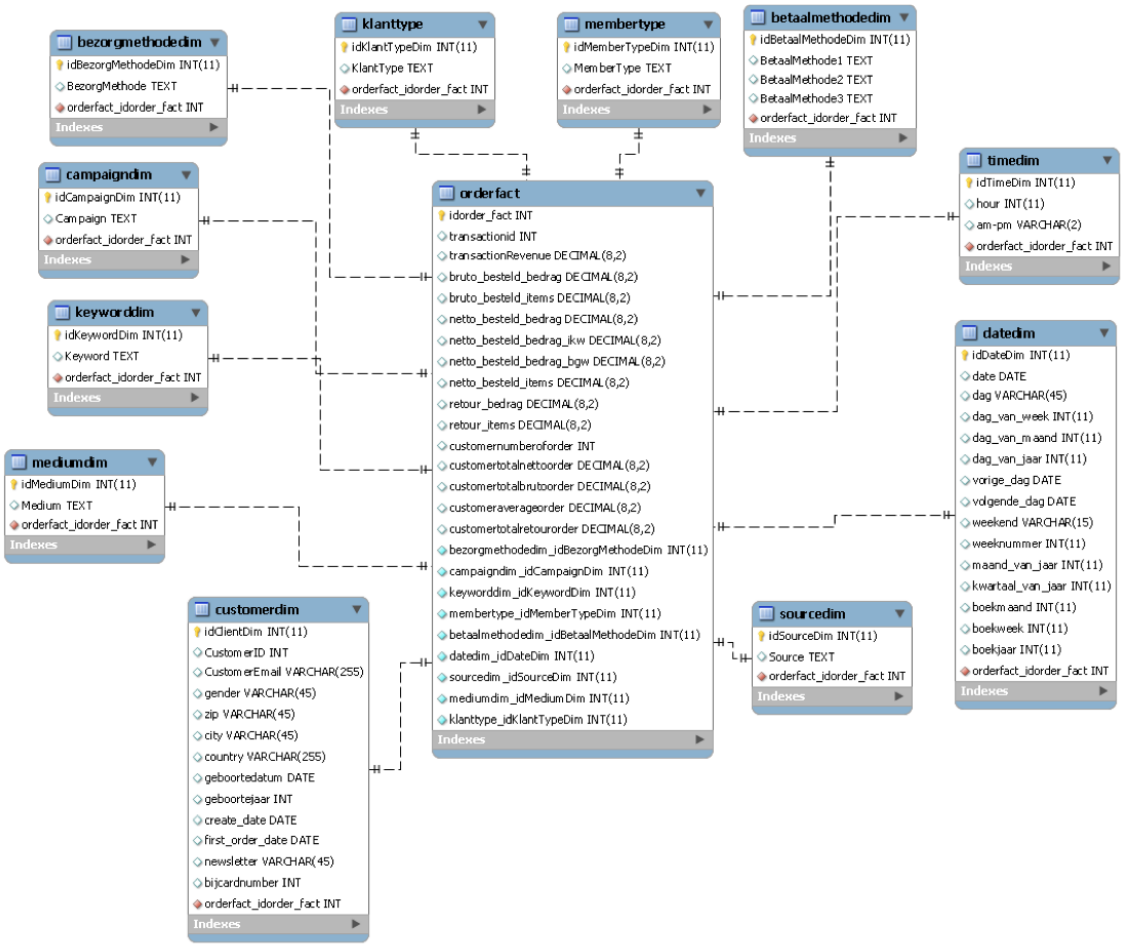

Figure A.3: Dimensional model for fact table Orderfact

Before the modeling was started, the model was explained to the stakeholders. An example of a finished model by [Prakash and Gosain, 2003] was presented. This model did not yet include the addition of KPI's so this was further explained to the stakeholders. The model was presented as an element in the larger requirements model.

A.3.2.1 Agent-Goal Modeling

After the explanation, an agent-goal model was developed. To create context, the goal of de Bijenkorf E-commerce department, maximizing the growth of de Bijenkorf.nl webshop, which is seen as a subgoal of the entire organization. This goal has been delegated to the E-commerce manager. The main goal can be split up in a number of sub goals, in the strategy of the e-commerce department this is done in three main subgoals, improving traffic, maximizing the transaction amount of customers and improving the conversion of the webshop. The decision was made only to look at the first subgoal, improving webshop traffic. This goal is delegated to the online marketing manager by the e-commerce manager. Improving traffic can be further decomposed into two subgoals, maximizing push traffic and maximizing pull traffic. Push traffic is creating traffic by external means such as advertisement, search engines, etc, pull traffic is generated by giving the customers a reason to return to the webshop. The first subgoal is chosen to further develop in the model. This goal is delegated to the off-site marketeer by the marketing manager. An overview of the agent-goal modeling performed can be seen in figure A.4

A.3.2.2 KPI-Goal Modeling

The next step was to find the Key Performance Indicators for the goals that have been defined. A goal can have multiple KPI's. Key Performance Indicators can be composed out of a number of information subsets. For example, one of the main KPI's used at de Bijenkorf.nl to measure the growth of the webshop is the year-over-year growth in turnover index. Corresponding to these sub-goals, three sub-kpi's exist, the year-over-year growth index of the average transaction size, the conversion rate and traffic. Because goals tend to be reached by making numerous decisions and the outcomes of these decisions differ, KPI's that measure the effect of certain decisions that have been made should also be taken into account. The next step is thus to model the decisions that need to be made to reach the goals that have been found.

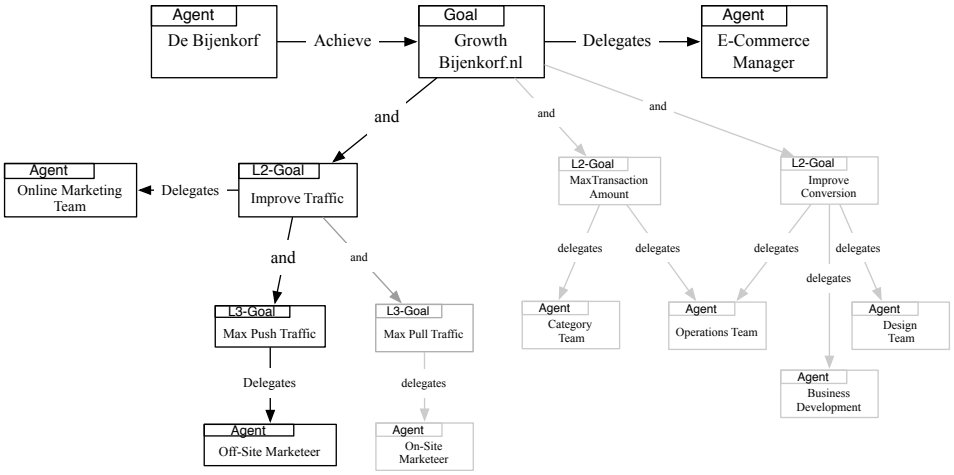

Figure A.4: Partial Goal-Agent model for de Bijenkorf.nl

A.3.2.3 Goal-Decision Modeling

The deepest goals, or most simple goals can be reached when agents or stakeholders make sound decisions. The success of the goals can be measured by using key performance indicators. For the off site marketeer, three decisions are made on a daily basis to maximize the pull traffic to deBijenkorf.nl: the decisions to use retargeting advertisement; display advertisement and to send out promotional e-mails. These decisions in turn have sub-decisions that need to be made. The decision to use promotional mail will be the scope of further modeling. Sub decisions that have to be made while sending promotional mails are: the frequency and moment of sending mail, the content and recipients of the mail. Sending an email can be marked as an event or dimension in the KPI's that correspond to all currently selected goals, especially conversion and traffic, but also the average spending per transaction. The goals and decisions can be seen in figure A.5

A.3.2.4 Decision-Information Modeling

The next step has been to find the information that is required to make decisions that lead to reaching the set of goals. This was done during the first session with the e-commerce manager and the online marketing manager. The information needs that

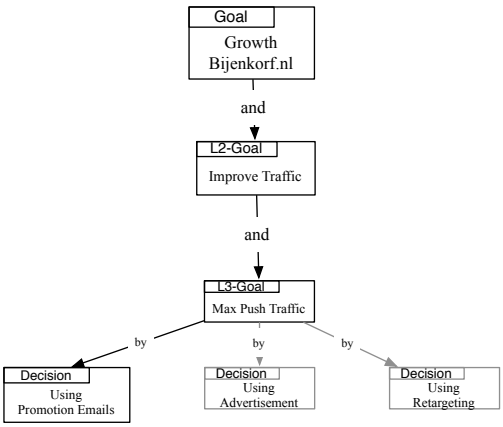

Figure A.5: Goal-decision modeling for de Bijenkorf.nl

support making the decision as to the frequency of promotional mails and the contents of those mails can be found in figure A.6.

The frequency of mailing depends on the commercial calendar, that is set through an interdepartmental group at de Bijenkorf. For each customer, the frequency of a mailing depends on their behavior. This customer behavior insights are also required for deciding upon the content of the email. Customer behavior is composed out of three pieces of information: the generalized segmentation rules or dimensions that are agreed upon together with the customer relationship management department; the characteristics of customers and particularly the sales history of customers. The payment history also contributes to the information requirement of product trends. Other information that contributes to insights into product trends are mostly unstructured sources of data, for example weblogs, product meetings, conferences, etc. Other sources can be, search volumes on product keywords in search engines, the prices for advertisements, product sales in other departments or companies, etc. Most of these sources can not be easily implemented into a Business Intelligence solution.

The e-commerce manager and the online marketing manager have a good knowledge of the decisions that need to be made in order to maximize pull traffic to the web-site. But the decision to use promotional emails and the corresponding goals are delegated to the off-site marketeer. Therefor, after this session, another session together with the off-site marketeer was held to complement and supplement the modeling that had been

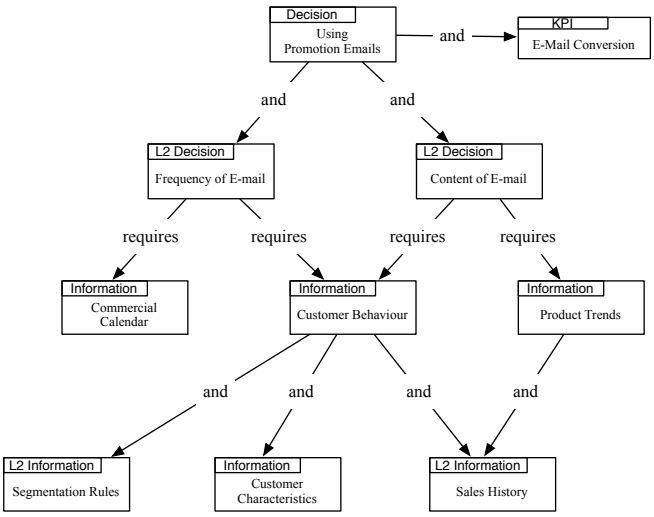

Figure A.6: Decision and Information model for promotion emails

made.

The proceedings of this session were similar to the session that was done with the managers, the difference being that the examples given came from the session already done at de Bijenkorf.

The total AKGDI model so can be seen in figure A.7. The complete model of de Bijenkorf.nl would be a much bigger model with multiple layers.

A.3.2.5 Dimensional Data Model

The final step in the method has been to design a dimensional model using the information requirements. During this phase, a large part of the dimensional model was already developed. The product history could be generated because of the detailed ordering data and the available product data. The transaction data also includes a campaign and medium dimension. These dimensions contain detailed information into the promotional emails that have been sent in the past. With the use of these dimensions the effect of emails can be measured and the transactional history can be filtered according to this data. To be able to prototype and develop the customer behavior information requirements the customer dimension needed to be adjusted. The segmentation rules result into a type of user, this can be either a dimension that is related to the trans-

A. APPENDIX: INFORMATION REQUIREMENTS AT DE BIJENKORF.NL

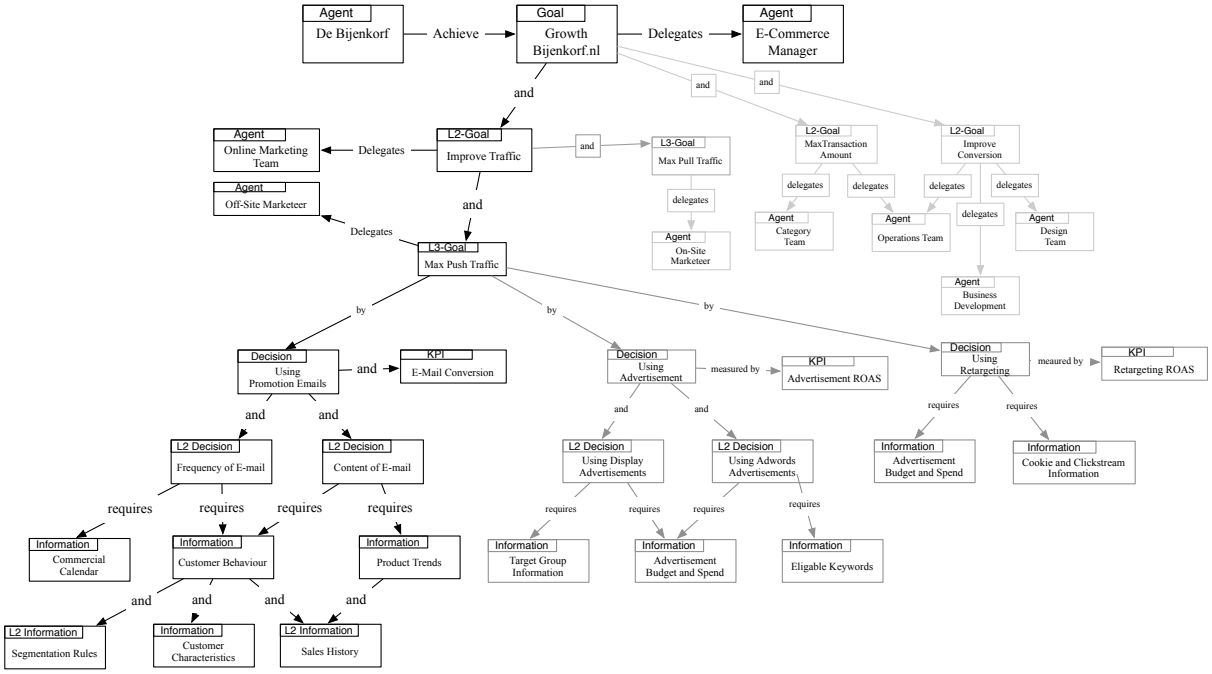

Figure A.7: AKGDI Model at de Bijenkorf

action fact table or to the customer dimension, or both. The first will result into more processing because the type relates to the cumulative expenditure of the customer and needs to be calculated for all transactions, the second needs less processing because it can be calculated per customer, the shortcoming is that the historic course of customer types can not be directly analyzed. Because de Bijenkorf does not immediately need such analysis, the customer type is modeled as an attribute in the customer dimension.

The resulting modified dimensional model that is used to provide the information requirements as found by using the AKGDI model can be globally seen in figure A.8.

A.3.2.6 Lessons learned at de Bijenkorf

Although the AKGDI model has been used after most of the information requirements had been found through other methods, the model helped structure the requirements analysis as a whole. It helps the stakeholders to take a step back and to observe their daily activities from a structured perspective. This helps them to find the information requirements that they take for granted and uncovers new requirements that they did

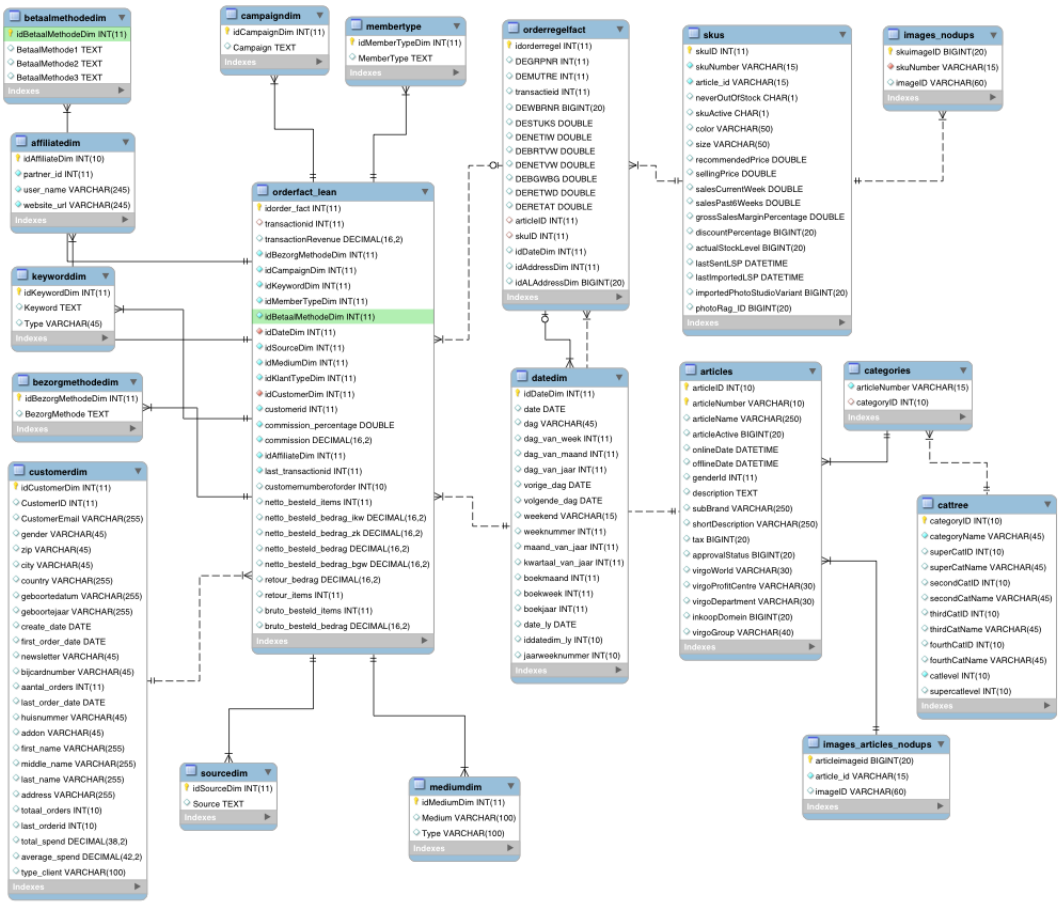

Figure A.8: Part of the Dimensional Model after AKGDI modeling

not yet know about. It is recommended to use this model first, as it can be presumed that a large part of the information requirements that exist within an organization can be found through this method. The model takes a deep look into the processes and operations of organizations, because of this, going through multiple cycles of the model can take a lot of time and will therefor be costly. Moreover, a fair amount of analytical reasoning skills is required from both the developers and stakeholders to be able to construct the models. The following section will elaborate on how the initial requirements that were found using this model were validated by prototyping.

A.3.3 Validation through prototyping

During the unstructured information requirements eliciting process at de Bijenkorf, the prototyping immediately followed sessions and interviews. This does not mean that the process of prototyping differs from phase four of the model as described in previous chapter. The only difference is that in this view, phase three has a internal validation loop that uses prototyping as a means of validation.

This section describes the prototyping process from the selection of technology to the definition of finalized information requirements and everything in between.

A.3.3.1 Technology

For prototyping, different approaches can be taken. All prototypes can first be drawn with the use of wireframes. Mockups can be made using with graphical software such as photoshop or visio. During the case study at de Bijenkorf, a test environment was set up. The environment required four technical elements and tools:

- A (virtual) data warehouse or data mart
- An Extraction Transformation and Loading tool (ETL) and corresponding process
- A front-end prototyping tool
- A test-environment, hardware or virtualized

The input of the prototyping phase is the dimension-metric model that has been produced in the previous phase. The model can be implemented as a database model

for a data warehouse. At de Bijenkorf a MySQL database was used to prototype the data warehouse in. MySQL is an open source database with high performance, good support and also great compatibility with Business Intelligence tools in the market. Different database engines can be used, in this case the MyISAM engine was used, which is more or less the default non-transactional MySQL engine available. Other considerations have been column-oriented engines such as InfoBright and InfiniDB. For non-transactional purposes, these engines can provide much faster performance. Their support, stability and ease of use are still not optimal. These decisions are not final, and the setup of the prototype is very flexible, the data warehouse can be migrated to other engines and even other frameworks. MySQL community server comes with a number of tools such as a Workbench and a Query Browser enabling fast development of the data warehouse and optimal querying capability.

The data warehouse is filled with (sample) data through the use of an ETL tool named Pentaho Data Integration or PDI. This is also open-source software developed as part of the Pentaho BI stack, a collection of BI tools. PDI is java based and has a GUI (graphical user interface) development tool and a command line running environment that can be scheduled. Using the GUI tool a workflow can be built that extracts data from any data source, then the data can be transformed so that it fits the data model and finally the data can be loaded into the data warehouse environment. An example of a workflow in PDI can be seen in figure A.9. The processes can then be scheduled using a batch script and a cronjob. A common choice in actual business intelligence development is to code the entire ETL process in custom software to precisely fit the news of the application. Setting up the ETL processes with a GUI tool is already a cost- and timely activity, taking around 70% of the prototyping and development time, hard coding it would make the entire process exponentially more expensive.

Because the main goal of the business intelligence solution for de Bijenkorf e-commerce department was to create dashboards and reports for the team, a front-end development application was chosen. Similar to the argument for the ETL process, developing a custom made visual dashboard or report layer for Business Intelligence takes a lot of time and is expensive, this also holds for adjustments that need to be made after initial development. OLAP, online analytics processing is more flexible but is less graphical and clear to end-users. Therefor during de Bijenkorf case study a graphical

A. APPENDIX: INFORMATION REQUIREMENTS AT DE BIJENKORF.NL

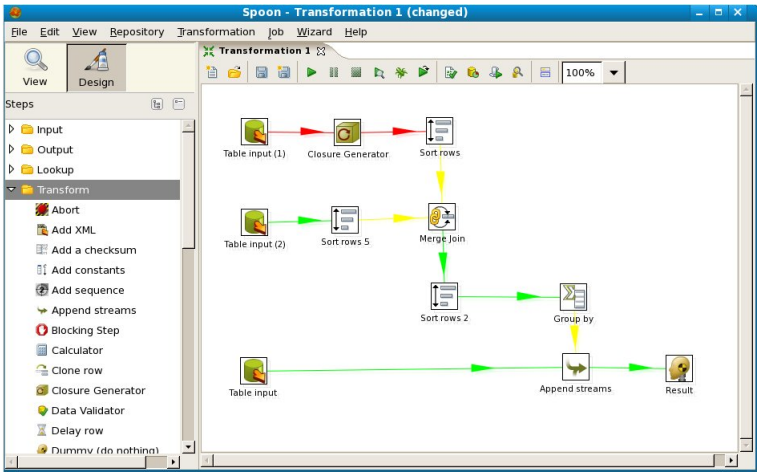

Figure A.9: Example of an ETL process developed with PDI

rapid analytics software package was used, Tableau. Tableau provides with"Fast analytics and Rapid-fire Business Intelligence" [Tableau, 2011] through combining OLAP capabilities with a large focus on graphical representation of data. An example of one of the dashboards that was developed using tableau can be seen in figure A.10. Tableau makes it easy to use any data source, in this case our prototype data warehouse as an input for dashboards and reports that can be published as image, pdf or on a web server. Although other solutions exist, such as RapidMiner, Spotfire, Microstrategy, Excel, Pentaho, JasperSoft, ThinkCell, etc, Tableau was the first choice because of the relatively low costs, and ease of use. During the first ad-hoc analysis assignment at de Bijenkorf, Tableau proved to be a very good tool with dashboards and reports that can be adjusted quick and are easy to understand by the end-users. This can be seen in figure A.10. Another argument for Tableau is the use of an internal indexing technology that speeds up the analysis process.

The final technical decision needed to be made in the development process was the choice for the computing environment. Firstly if the setup would be a hardware solution or hosted on a virtual environment. A virtual, cloud, environment could provide scalability if necessary, but because of the lack of virtualization options within de Bijenkorf the choice was made to set up a test BI hardware environment. External cloud services would require a high level of security because of the sensitivity of the data that is used. The hardware solution involved buying a low-cost (under 700 euro)

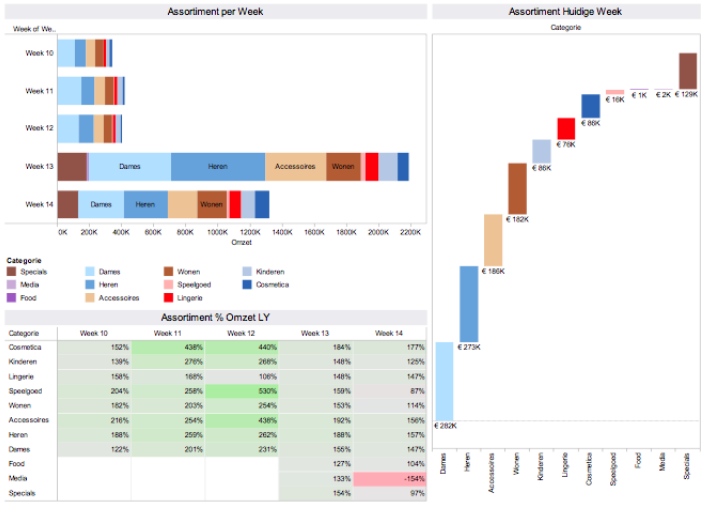

Figure A.10: Example of Tableau Dashboard Prototype

machine with most money spend on maximizing CPU and memory power. All the tools have been installed on this machine. Also taken in consideration during this decision was the expected size of the data warehouse and type of backups necessary. An overview of the prototyping environment can be seen in figure A.11

A.3.3.2 Data Audit and Data Model Implementation

During the case study, the first sessions and interviews provided a large number of information requirements. These were then transformed into a preliminary data model, a part of this model can be seen in figure A.3. Before implementing the Data Model, a data audit was needed to map all data sources and corresponding data to the required dimensions and metrics of the data model. There are two options for this, auditing by looking or by asking. Documentation and sample data from data sources is used as an input. When asking, the stakeholders and experts that work with that data are asked for their knowledge of this data source and availability of data. At de Bijenkorf, a combination of the two has been used. The reason for this is that some data sources are poorly documented and expert knowledge is necessary. Alongside auditing if the data is available, the accessibility of the data sources should be investigated. The questions that will highly impact the prototype with this respect are: how will the data be available and how often can it be accessed.

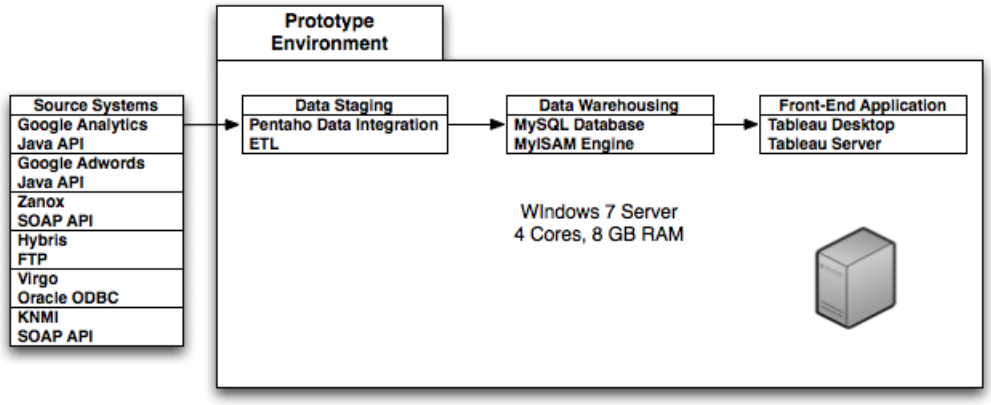

Figure A.11: Prototyping Environment at de Bijenkorf.nl

After the data audit is completed, the data model, that was developed using the MySQL Workbench, with its fact- and dimensional tables is forward engineered to the MySQL data warehouse. All tables have unique keys, that are preferably not the primary keys used in the source tables, in this manner, the history of the dimensions and fact can be kept. During the iteration steps the structure of the data model is very often changed, tables are altered, added and deleted. The first data model had a total of nine tables, the model used after the last iteration step contained 33 tables.

The next step is to develop the ETL processes used in the data staging phase.

A.3.3.3 Data Staging through ETL processes

The data model is in place and the mapping with the original data sources has been made. Moreover, the manner in which the data sources can be accessed is available. To fill the data warehouse, ETL processes now need to be be developed and executed (on a regular basis). It is common practice that all fact tables in the data model have one or more ETL processes. The data sources that need to be accessed during de Bijenkorf ETL processes are all accessed in a different matter. Google Analytics, Adwords and Zanox are accessed through an Application Programming Interface (API). Custom code has been written in the form of Java applications that can be used within the ETL tool as an input. Data concerning the online assortment is accessed through a set of flat files that are published to an FTP server daily, this can be directly accessed through

the ETL software. Order, customer and customer satisfaction data is accessed directly with the use of Open DataBase Connectivity (ODBC), these are transactional databases internal and external to de Bijenkorf. Connecting to the API's has proven to be the most challenging exercise, preferably API provided classes should be used because the API's change over time. When all data connections can be made the transformation steps for the data has to be developed.

The transformation steps depend on the structure of the source data and target data model. Moreover, transformation should include checking data and data quality. Using transformations operations such as grouping, filtering and aggregation can be performed. An important activity is looking up the dimensions in the dimensional tables and to delete the original fields from the data. During these processes memory management is an important issue, depending on the transformation, a large chunk of memory can be used, this makes the system prone to failure. Also because of the automated character of the ETL processes, in the transformation steps, error handling and logging should be used to minimize failure and to maintain the quality of the data.

The final step is to load the data into the data model. Depending on the data loading can be done by inserting or updating information. In the case of dimensions that are changing date stamps are essential to keep track of the history.

The ETL processes are then scheduled. Because a large part of the data sources only provide with daily update of the data the ETL processes of de Bijenkorf are ran every night (when no analytical processing is performed). Before the data warehouse is updated, a full backup is made and transferred to an external machine.

After a number of iterations, the ETL processes of de Bijenkorf have been improved and extended and now include 10 main processes. Because the data model and ETL processes are directly related, every update in either one has to be checked for impact on the other. Especially changes in dimensions can impact a large number of ETL processes. The development of the ETL processes is a cumbersome and timely activity but in the long run it is more durable then manually extracting data.

In de Bijenkorf data warehouse, also static data is used, such as meta data for locations and the product hierarchy. These were inserted manually using the MySQL querying browser.

A.3.3.4 Front-end prototyping

With the data model in place and the data warehouse filled with information through the ETL processes, the front-end applications need to be prototyped. This is the main and preferably the only touchpoint of the end-users with the entire prototype.

The front-end development is done by using Tableau software. The input for this phase is both the list of information requirements and the dimensional model. The information requirements can be visualized by using the dimensions and metrics in the data warehouse and corresponding data model. In Tableau a connection can be made to the data by querying directly to the database, the software automatically recognizes dimensions and metrics, similar to other BI software. The dimensions and metrics can then form the basis of views or drill down tables. Multiple views, tables and filters can be combined into interactive dashboards that can be manipulated by changing variables. Tableau has a very clear and easy to interpret visual layer.

During de Bijenkorf case, the incremental approach lead to the decision to first prototype the weekly report that is sent to all e-commerce employees and the board of directors of de Bijenkorf. The information requirements mostly stem from Key Performance Indicators and are on a strategic level. These requirements included:

- Profits, Turnover, Return, Conversion,Visitors, Bounces (drilling down into all categories, subcategories and brands of the product catalog)
- Year-over-year growth index for Turnover, Conversion, Basket Size and Basket Price
- Stakes of: Type of Shipping, Member Type, Referring Channel and Customer Frequency
- Average Basket Size, Average Basket Price
- Return-on-advertisement-spending

The grain and aggregation of the report is on a weekly basis. Because of the differences in fact tables, before the necessary data could be extracted from the data model, index or aggregate tables needed to be made. For example, visitor data is on a daily basis and needs to be joined with order data on a daily level. Also the costs of

advertisement spending is on a campaign level, but turnover is on a order level, therfor the turnover had to be aggregated by day and campaign and could then be joined with the costs. Finally, three different sets of dimensions and metrics have been joined into Tableau: a daily index, a set on the basis of orders and finally based on articles within an order.

Although most of these aggregates and indexes have been foreseen in the development of the dimensional model, the first iterations posed major adjustments to the data model, ETL processes and the manner in which the data is fetched by Tableau.

The first iteration round of the weekly report was in March 2011. During the weekly company briefing and e-commerce team meeting the report was used to convey the current status of the e-commerce business. The stakeholders were mainly the e-commerce manager of de Bijenkorf and all employees in the e-commerce team. The report was spread as a static PDF file through e-mail. In every e-mail message, it was clearly stated that this was a prototype and that any questions, remarks and adjustments should be notified through any possible channel. Moreover, the structure of the report was the subject of meetings with the managment of the e-commerce department.

In the months after the first"release" of the Weekly Report prototype until now, July 2011, the report has had 5 major iterations where the entire report has been redeveloped. The number of sheets went from an initial 8 to 11 and finally back to 6 key dashboards. In between these iterations small adjustments were made to the views. The stakeholders of de Bijenkorf were encouraged to give their opinion and to convey their needs during these iterations. Because the developer was internal to de Bijenkorf the communication lines were short and fast and short iteration cycles were possible, this helped significantly both in the quality of the dashboard and in the commitment of the stakeholders. On of the views in the weekly report can be seen in figure A.12

Two other information needs that arose from the unstructured requirements analysis process that have been prototyped in the same period are the overviews of sales down the category and product hierarchy and a top sales dashboard as seen from that same hierarchy and with respect to keywords. Both dashboards are very similar and provide large tables with data concerning categories, brands, products and even sku's. Because in the course of the operation of the webshop more than 1.2 million articles have been ordered data management and organization was essential. The end-user on

A. APPENDIX: INFORMATION REQUIREMENTS AT DE BIJENKORF.NL

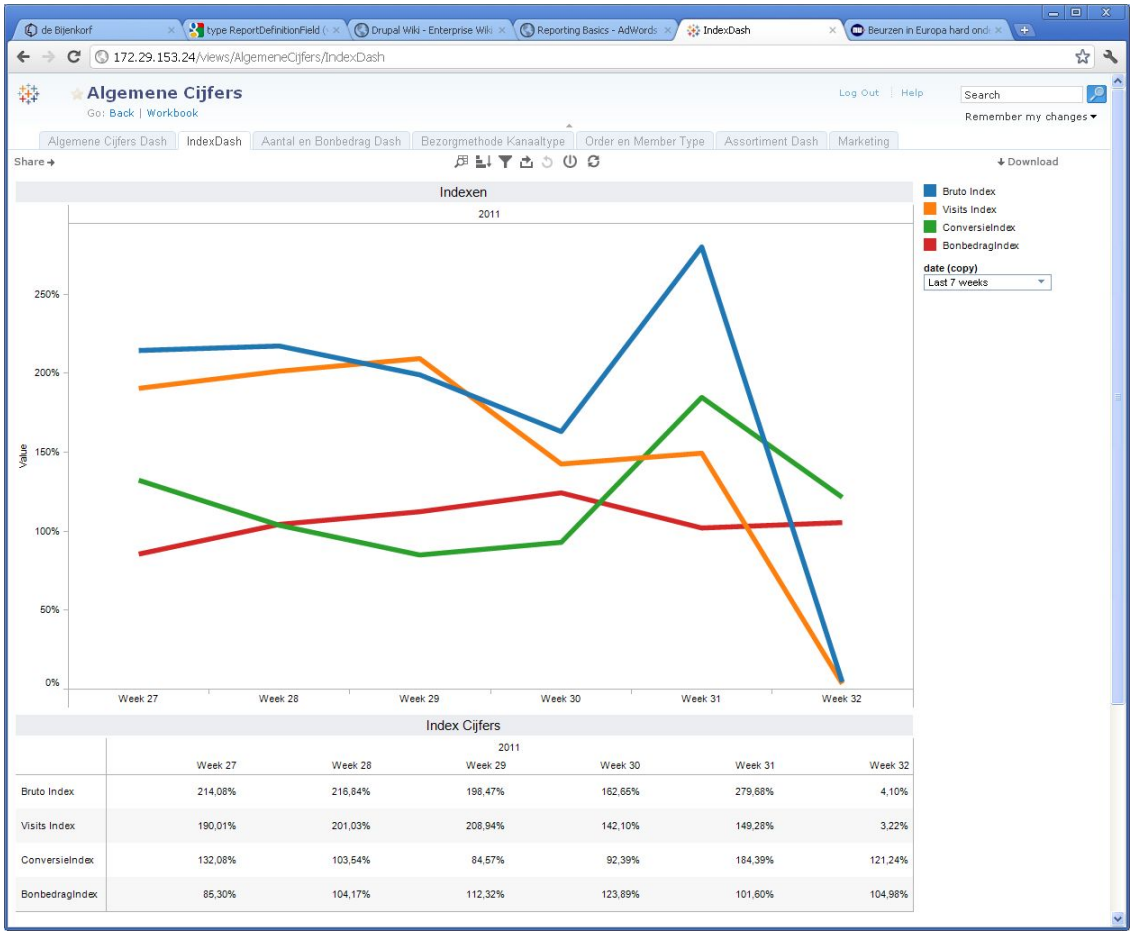

Figure A.12: Week Report Prototype

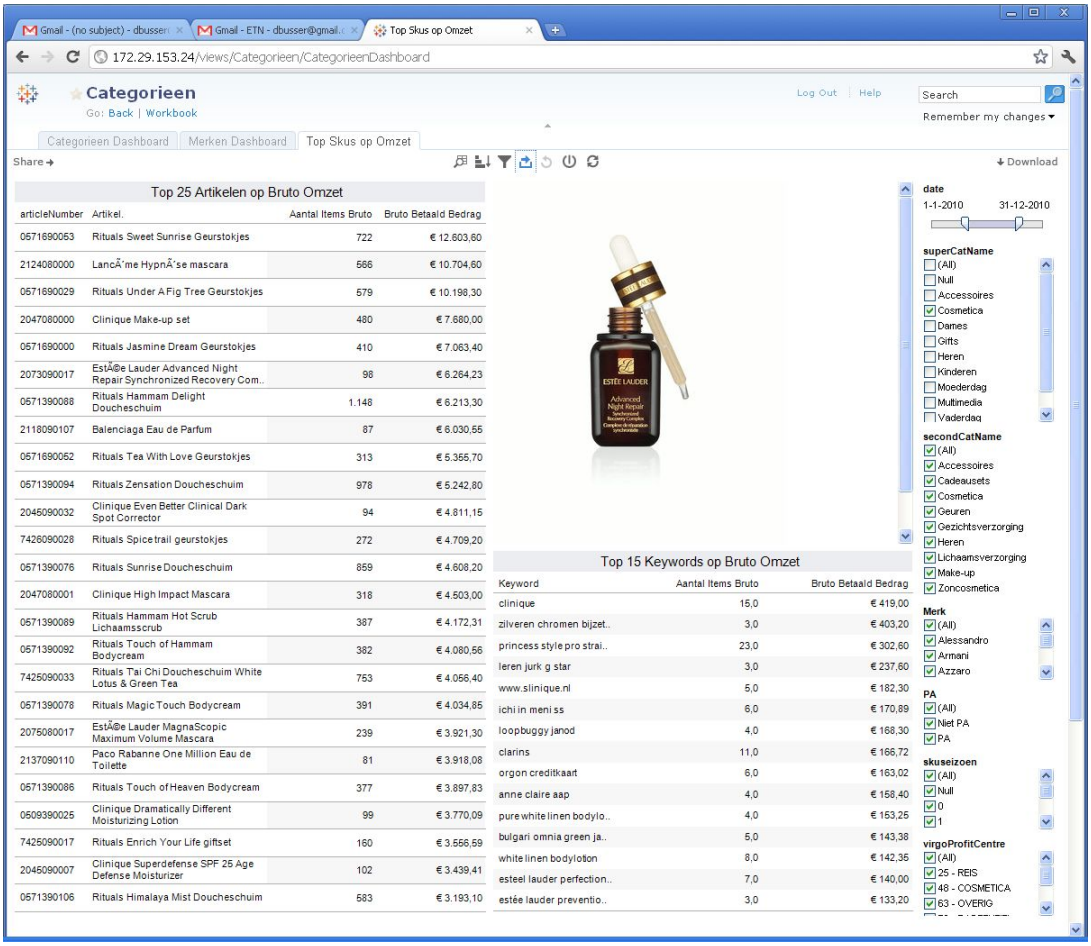

Figure A.13: Top Articles Prototype

the other hand requires a large form of autonomy in drilling down into the data and making selections alongside the dimensions that are available. Also, because of the large amounts of data, the dashboard would get a more spreadsheet like feeling with meaningless numbers, words and codes. To give the end-users more feeling into the data, the images of products were added, this can be seen in figure A.13. The stakeholders started using these two dashboards heavily because they got a true impression of the meaning and context of the data. The interactivity and use of filters in these dashboards required a number of iterations to fully meet the requirements as set by the end-users.

Another information requirement that was formulated during interviews and ses-

A. APPENDIX: INFORMATION REQUIREMENTS AT DE BIJENKORF.NL

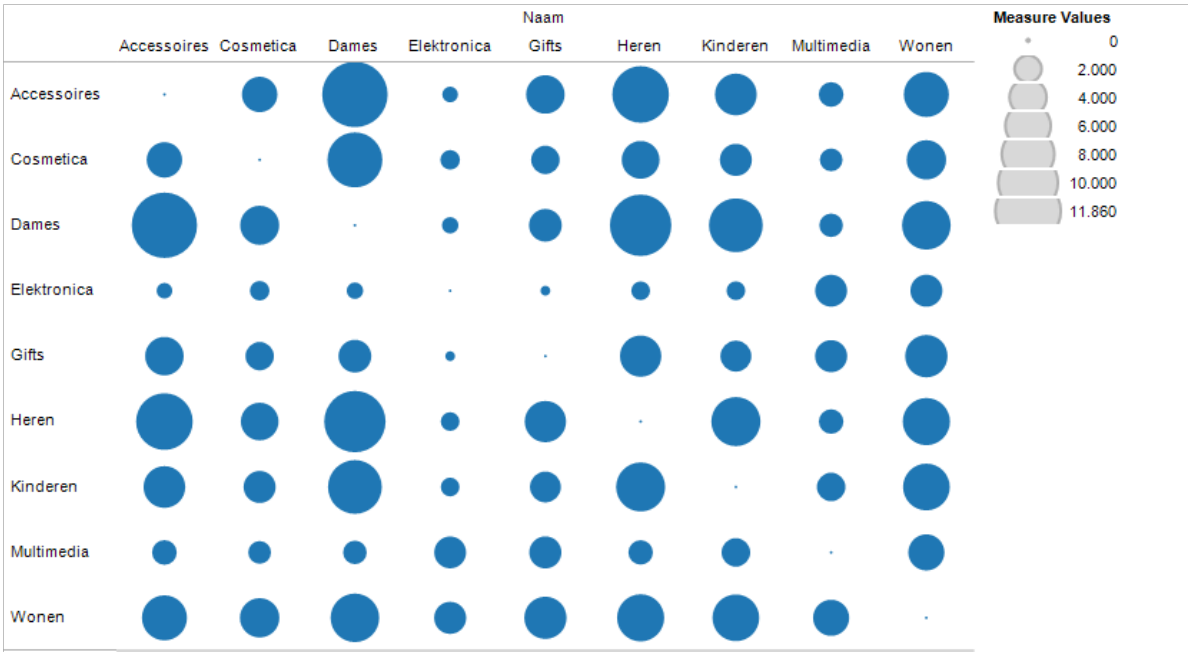

Figure A.14: Cross Sales Prototype

sions was information and intelligence on cross-sales. Cross-sales are defined as orders or transactions in which multiple types of artikels, brands or categories are sold. In the online shop cross-sales are used as business rules to decide what products are shown as recommendations on product pages. Currently the category specialists base these business rules mostly on the characteristics of the products, so more based on similarity in looks, size, fit and price. The information requirement was set because the category specialists wanted to have a better sense of what products sell together to base their business rules upon. In the dashboards and reports found during the previous phases, no such view was found, thus prototyping and experimenting was needed. The result was a matrix graph where the size of the data points indicates the volume of cross-sales. The dashboard on the intranet makes it possible to drill down into sub-categories and brands, possibly into products. One of the first versions of the prototype can be seen in figure A.14.

So far the information requirements that were elicited using the unstructured, demand-driven phase of the model have been covered. For the structured Goal oriented phase two the front-end applications still need to be developed. Because of the

efforts during the prototyping of the requirements that were found during the demand-driven phase the majority of the dimensional model required was already in place. The only addition to the model was the Commercial Calendar Dimension, this dimension contains the dates and attributes describing when and what the company wide strategy is concerning customer contacts. The information requirements that have to be prototyped are mostly concerned with customer behaviour, product sales history and trends to support the decision making process when sending out promotional emails to customers.

During an ad-hoc assignment prototypes for this set of requirements have already been made. The assignment was to find the customer behavior of different customer types with the goal of developing a customer loyalty program to encourage customers to sign up for the loyalty program. The views and dashboards that were created for this assignment can be used as a basis for the prototyping phase. Moreover, the first assignment was an analysis of the customer base, this analysis has also been used as an input for prototyping.

After the dimensional model was modified, the front-end prototypes for the original information requirements for the off-site marketeer could be developed. The first step was constructing the KPI for the overall goal of the off-site marketeer, the traffic and conversion index, these were already in place, the e-mail conversion could be constructed by using the overall conversion and adding the source and medium dimensions to the dashboard. These dimensions can split the measures into the source and medium of the customers, one of the attributes is email. The information requirement of customer behavior encompasses a number of different views for the stakeholders. For example, they would like to know the conversion for email as seen from the perspective of different customer types and their characteristics. But also the type of products that these customers have bought. Additionally, for the conversion of previous emails need to be used as a basis for the content of new emails.

The segmentation rules are used to divide the customers into (five) different customer types. These customers can then receive an email based on their customer type, sex and type of membership. At de Bijenkorf, there is a large customer loyalty program, the BijCard, which is both a loyalty card and a credit card, currently more than 400K cards have been distributed. The segmentation rules are applied to the sales history of clients, for example customers that spend above average and return very frequently are

called"Top Shoppers" and are seen as most valuable. By applying the segmentation rules, the attribute customer type is filled in the customer dimension. To be able to identify the churn rate of customers, this attribute can also be calculated based on the moment of a transaction, to calculate for example the churn rate of customers. The sex and type of membership are already attributes in the customer dimension.

The views and dashboards that contain information about customer behavior, such as sales history in product groups, types, brands and sizes, can be analyzed according to these dimensions. This approach makes it possible to re-use basic reports and views for different analysis.

One of the reports that is used is the product sales history. Using the commercial calendar, the filter of the product sales view can be adjusted. For example, if the focus of the commercial calendar is on sales, the filter selects products types and articles that have been sold in previous sales periods. Then the data is segmented according to the type of customer, sex and if he is a member or not.

The prototype was thus an extension of the existing product sales dashboard that was made for the category specialists. Because of the addition of a number of new dimensions, the query, and thus the dashboard, was rebuilt. The set of dashboards that cover the customer behavior information requirements have currently been through three iteration cycles. One of the main challenges and needs that were identified in these cycles is that the dashboards should be highly flexible to be able to analyze a variety of combinations of dimensions and filters. The static dashboards and reports would not be able to provide this. Interactivity can be seen as a non-information requirement to the system.

The iteration cycles did not include a complete run through of the AKGDI model. Similar to the demand driven requirements phase, the iteration was done through (in)formal sessions with the stakeholders. After the development of a prototype, the stakeholder would get access by logging into the front-end bi server. The request to the stakeholders was to test and use the prototype as they would if it were in production. Any comments, recommendations, errors, etc could be communicated directly. A number of sessions were planned to evaluate the set of prototypes in full.

Alongside the functional information requirements that were adjusted and added during the prototyping process, non-functional requirements have also been a subject that was discussed between the developer and stakeholders. These requirements are

most notably data quality and availability. Data quality is an essential requirement to maintain the commitment of the stakeholders and their participation in the development process. If stakeholders can not rely on the data and information that they receive from the Business Intelligence solution the success of the system is at risk. As to availability, the initial report was only sent out once a week, placing the document in a shared folder helped the stakeholders to locate the report. The last iteration introduced even better availability of the report by publishing it to the intranet of the e-commerce department. This also made the report more interactive through the use of custom filters, the ability to change dimensions and drill-down capability. This also encompasses another important requirement to the system that the end-users have a lot of appreciation for, interactivity, they want to be able to perform analysis from certain perspectives. Interactivity should however not conflict with the ease of use and simplicity of reports and dashboards.

The prototyping phase at de Bijenkorf was done by going through a number of iteration cycles. Although the model presents the possibility to go through the requirements eliciting phases again, this was not done during the case study. A probable reason for this was because the case study was done internally at de Bijenkorf. The communication lines were very short and any modification and altering was done almost on the spot. When development is done externally, project planning should be used more intensively and evaluation meetings need to be planned regularly to support and monitor the process.

A.4 Conclusions

This chapter covered the validation of the new developed information requirements model for business intelligence development projects. The validation was done by using the model in a case study at the e-commerce department of de Bijenkorf. Although de Bijenkorf only operated for two years when the case study started, a large number of different information systems were used to gain insights and intelligence into the business and to base decisions upon. Because of those different systems, data and intelligence was distributed across the department and a large amount of time and money was lost because analysis needed to be done on a ad-hoc basis.

A. APPENDIX: INFORMATION REQUIREMENTS AT DE BIJENKORF.NL

As part of the Business Intelligence development project at de Bijenkorf, the information requirements analysis was done by going through the model. Because the structured AKGDI model was under development, the project started with the unstructured demand driven requirements analysis phase. Before starting with modeling, the stakeholders were introduced into business intelligence, its capabilities and into the project process. The developer got a feeling of the current organization and processes within the organization by starting with an ad-hoc analysis assignment.After this introduction the demand driven process was performed. With input coming from the existing applications and external sources, dashboards and reports extensive sessions were held with the stakeholders of de Bijenkorf.

A low-cost prototype environment was developed using cheap hardware and mostly open-source software. The prototyping environment included a data warehouse, ETL processes and a front-end application server that could be accessed by the stakeholders. The underlying data model for the data warehouse was developed according to the dimensions and metrics that could be found from the information requirements that were formulated. The prototypes were developed after filling the data warehouse with sample data from a large array of different sources.

All prototypes were developed according to the information requirements that followed from the interviews and sessions with the stakeholders. They were asked to communicate any comments, adjustments, errors or further needs with regard to the prototypes as soon as possible through all possible channels. The first large prototyping development cycle was centered around the weekly department report sent out on mondays. The report gives an overview of the KPI's per team on a weekly basis. In total, the weekly report had eight major revamps because of changing requirements and the corresponding changing underlying data model. In between these major updates, a large number of small adjustments and modifications were made.

The structured analysis phase using the AKGDI model was executed after the first prototypes were finished. The focus of the analysis was on a small part of the organization due to time constraints. This limitation did not however constitute that a full evaluation of the method could be made. During modeling sessions with stakeholders of the e-commerce department the complete AKGDI model was drawn up. The information requirements that were a result of this method have been prototyped similarly to the demand-driven requirements phase. The iteration cycles did not reach back to

the modeling phase as it was experienced by both developer and stakeholders that this was not nessacery to succesfully develop the prototypes. Moreover, the data model and corresponding ETL processes, that were developed during the prototyping for the requirements that were gathered with the unstructured method, did not require major modifications.Thus the time spend on actual prototyping was much lower.

Overall, a number of key lessons have been learned. First of all, data quality is essential for acceptance and engagement of the stakeholders. If data is incomplete or wrong the end-users will not believe in the system and will not be encouraged to help developing the system further. Also communication is important, the end-users need to be kept in the loop at all times, this stimulates the process and makes the stakeholders commit to the project and the system.

When looking at the model, a number of changes and modifications can be recommended. Although during de Bijenkorf case study the order of the process was the introduction, then the unstructured method, the structured method and finally the validation step, the two analysis phases should switch. The goal-oriented requirement analysis method AKGDI gives a structured view of the operations and business of an organization and includes the essential and common information requirements. To complete the information requirements the demand-driven method is suited. This method also covers the information requirements that are not immediately related to any decision, goal or KPI that is documented within an organization. Moreover, it gives creative and less structured employees a platform to think of great information requirements that they will use.

Another learning is that iteration during validation does not immediately imply going back to the requirements model and going through the entire model again. By keeping short communication lines and small bariers to make modifications and adjustments the validation phase was very successful.

Because research and the case study were done in parallel, good structure and process management was lacking. Especially when projects become large, complex and include a large number of stakeholders, good project- and requirements management is essential.

A. APPENDIX: INFORMATION REQUIREMENTS AT DE BIJENKORF.NL

Appendix B

Appendix: Scientific Paper