The Emerging trend of Self-Service Business Intelligence: A sustainable solution for a large organization?

by

Thomas Schützler

LIU-IDA/LITH-EX-A–14/042–SE

2014-06-18
The Emerging trend of Self-Service Business Intelligence: A sustainable solution for a large organization?

by

Thomas Schützler

LIU-IDA/LITH-EX-A-14/042–SE

2014-06-18

Supervisor: Anders Fröberg, IDA Linköping University
Andreas Anderljung, Sigma
Zebastian Zaar, Sigma

Examiner: Erik Berglund, IDA Linköping University
Abstract

Business intelligence (BI) is an umbrella term used to describe the applications, infrastructure and tools, and best practices which organizations can use to analyze information in order to improve and optimize decisions and organizational performance. In the later years a new trend has emerged in the area of BI namely, Self-Service Business Intelligence. The purpose of Self-Service BI is to empower the users by allowing users to create reports and analyze data without support of the IT department.

This thesis have tested and evaluated Microsoft’s new Self-Service BI tool suite, Power BI, through a case study in a large organization. The main purpose was not only to conclude if it is possible to implement a complete Self-Service BI solution in a large organization, but also examine which parts of the Business Intelligence architecture are most suitable for implementing Power BI.

The result have shown that Power BI and Self-Service BI tools can’t meet the back-end requirements of a large organization and therefore it is not a suitable or functional solution. However, the front-end applications and best practises of Power BI and Self-Service BI are suitable for a large business. They support the users needs and empowers the users create better and more powerful analysis.
I would like to thank all the people involved in this master thesis at the county council of Östergötland, LiU and Sigma, especially, my supervisors Anderas Anderljung and Zebastian Zaar at Sigma and Anders Fröberg and examiner Erik Berghund at LiU for all their support. Secondly, I would like to thank my family and friends, in particular EG, for their support throughout my education. Last but not least, I would like to send a special regard to my loving and always supporting girlfriend. Her support have meant great deal to me in finishing this thesis but also my education.

Thomas Schützler
Linköping, June 2014
# Contents

1 Introduction .............................................. 2  
1.1 Background ........................................... 2  
1.2 Aim of the study ....................................... 3  
1.3 Problem definition ..................................... 3  
1.4 Demarcation ........................................... 3  
1.5 Disposition of the Report .............................. 3  

2 Theoretical Background ................................. 5  
2.1 Definition of Business Intelligence ................. 5  
2.2 Architecture of Business Intelligence ............... 5  
2.2.1 Data Sources ....................................... 6  
2.2.2 Extract Transform Load ............................ 7  
2.2.3 Data warehouse ..................................... 8  
2.2.3.1 Dimension modeling ............................ 9  
2.2.4 Mid-tier ........................................... 10  
2.2.4.1 OLAP Cube .................................... 11  
2.2.4.2 In-memory Approach ........................... 13  
2.2.5 Front-end applications ............................. 14  
2.3 Self-Service Business Intelligence .................... 15  
2.3.1 Microsoft Power BI ................................ 16  
2.3.1.1 Office 365 ..................................... 16  
2.3.1.2 Excel 2013 ..................................... 17  
2.3.1.3 Power Query ................................... 17  
2.3.1.4 PowerPivot .................................... 18  
2.3.1.5 Power View .................................... 19  
2.3.1.6 Power Maps .................................... 20  

3 Method .................................................... 21  
3.1 Case Study Method ..................................... 21  
3.1.1 Data Collection .................................... 22  
3.1.1.1 Interview ...................................... 22  
3.1.1.2 Tests .......................................... 22  
3.1.1.3 Observations ................................... 23  

vii
List of Figures

2.1 Business Intelligence architecture, modified (Chaudhuri, Dayal, & Narasayya, 1997) ......................... 6
2.2 ETL Process ........................................ 7
2.3 Star schema (Chaudhuri & Dayal, 2011) .............. 10
2.4 Cube for analysing products by date and cities. The axes are the dimensions and the measures are the values inside the cube. (Chaudhuri & Dayal, 2011) ......................... 12
2.5 OLAP Operations in a multidimensional cube (Han, Kamber, & Pei, 2012) ................................. 13
2.6 Decision matrix (Knigth, 2013) ........................ 18
2.7 Row Oriented database (Ferrari & Russo, 2013) ........... 19
2.8 Column Oriented database (Ferrari & Russo, 2013) ........... 19

4.1 Star schema for the data warehouse ......................... 29
4.2 Capture of Cognos PowerPlay: Choosing medicine ........ 30
4.3 The capture shows a star model used for creating Power View Reports. The fact and dimensions tables are extracted from the data warehouse. ................................. 32
4.4 Decision matrix: Assessed by BI team member ............ 34
4.5 Power View interaction ................................. 36
Chapter 1

Introduction

Initially in this chapter the background for the study will be presented. Furthermore the aim of the study, the problem definition as well as the demarcation of the study will be described, Followed by the disposition of the report.

1.1 Background

Business Intelligence (BI) is an area that have grown rapidly in the last decade. BI is often used as a generic term for services and software whose task is to support decision making in organizations and improve the quality of those decisions. During the later half of the 2000s, a new trend have grown strong, namely Self-Service BI. The concept of Self-Service is essentially that the users in an organizations should be able to generate reports and analysis based on the organization’s data without any significant involvement from IT departments. With this background, it would be interesting to investigate how Self-Service tools can be used to create business intelligence systems and also to investigate in which parts of the business intelligence architecture, Self-Service business intelligence tools are most suitable. Power BI is Microsoft’s answer to the growing demand of Self-Service BI.

A case study was done at the county council of Östergötland with the support of supervisors from a IT consultant company called Sigma It & Management.
1.2 Aim of the study

How can the phenomenon of Self-Service Business Intelligence be implemented in a large organization, through Microsoft’s Self-Service tool, Power BI? Furthermore, the study aims to explore if Power BI can be implemented as a complete solution or if there are certain parts in the Business Intelligence architecture that are more suitable for a Self-Service solution.

1.3 Problem definition

The following research questions will be guiding the report in order to reach the aim of the study:

- How can Power BI be implemented as a complete Business Intelligence solution?
- Which parts of the Business Intelligence architecture are most suitable for implementing Power BI?
- What are the possibilities and challenges of implementing Power BI?

1.4 Demarcation

- One of the limitations of this study is that it only aims to review and analyze if the BI tools are suitable for a large organization.
- The study only aims to review if Power BI may meet the requirements and needs of BI in a large organization.
- There is a time limit of 20 weeks on the master thesis. Therefore, only one case study have been conducted. To understand the potentials and challenges of Power BI, more scenarios should have been conducted.
- The case study has mainly been working with data structured according to dimensional modeling.
- Only the medical part of the BI system at the county council were used in the case study.
- Power Maps have not been tested or evaluated in the case study

1.5 Disposition of the Report

The report is divided into five chapters:
1.5. DISPOSITION OF THE REPORT

Chapter 1 - Introduction presents the reader to the background and the purpose of the report. It will also provide the reader with the research questions and the demarcations of the study.

Chapter 2 - Theoretical Background presents a thorough background of the area of Business Intelligence. Including both a historical background to better understand the needs and requirements of today and also the underlying techniques used in Business Intelligence solutions.

Chapter 3 - Method presents the case study method, including data collection, data analysis and validity.

Chapter 4 - Results presents the data collection and results of the case study done at the county council of Östergötland.

Chapter 5 - Discussion presents a discussion regarding the results in relation to the aim of the study and thereby the research questions. At the end of the chapter, there will also be a shorter look outside the scope and some recommendations for future research.
Chapter 2

Theoretical Background

The following sections will present an introduction to the area of Business Intelligence (BI). Initially, a conceptual description of BI will be presented and the different corner-stones of BI will be explained. After reviewing the overall idea of BI, each part will be explained in greater detail. Also introduce and explain the term Self-Service business intelligence.

2.1 Definition of Business Intelligence

In the year 1989 Howard Dresner, who would later become a researcher at Garter Inc., coined the term Business Intelligence (Watson & Wixom, 2007; Smalltree, 2006). Noteworthy is that the term Business Intelligence was already stated in 1958 in a paper written by Luhn (1958). According to Chee et al. (2009) Dresner described BI as “a broad category of software and solutions for gathering, consolidating, analysing and providing access to data in a way that lets enterprise users make better business decisions”. There exists numerous definitions of BI. Gartner, Inc. which is a major information technology research and advisory company, defines BI as the following, “an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance” (Inc, 2013a).

2.2 Architecture of Business Intelligence

There exists numerous architectures for BI-solutions, one common architecture for supporting an enterprise is presented by Chaudhuri et al. (1997).
2.2. ARCHITECTURE OF BUSINESS INTELLIGENCE

The model consists of five components, \textit{data sources, extract-transform-load (ETL), data warehouses (DW), mid-tiers and front-end applications}. The first part of the model shows how data is collected from different sources. The second part, ETL, uses different back-end solutions to clean and conform data to support BI functions. The data is then loaded into a data warehouse which is the third element in the BI model. Information from the data warehouse is then used to create specialized functionality for different BI situations. An example of a mid-tier element is OLAP cubes. The last and final part in the BI architecture is front-end applications. In this step, information is turned into knowledge by analysing the data and visualizing it in multiple ways. In the figure below, an overview of the architecture is provided.

![Business Intelligence architecture](image)

Each layer in the model will be examined closer in the following paragraphs.

### 2.2.1 Data Sources

In a BI-solution, data can be derived from many different sources (Chaudhuri et al., 1997). In traditional BI applications, structured data is often the main source for analysis (Baars & Kemper, 2008). However, this is not optimal since a considerable amount of data are unstructured, or as Negash (2004) suggest; semi-structured, to acknowledge that most data have some structure. For example, reports have headers, sections and paragraphs. Enterprise resource planning systems are a good example of a data source with structured data. It is important to consider using both structured and semi-structured data as both sources provide information needed to perform accurate analysis of the business (Baars & Kemper, 2008; Negash, 2004).
Data can also be categorised into external and internal data. The definition of external versus internal data varies depending on a couple of aspects (Inmon, 2002). The author argues that internal data are derived from sources within the organization. For example, data that is generated within the organization and stored in the operational databases. In contrast, external data is generated outside the organization. For example, competitors reports.

As data are derived from different sources, the quality of data will vary. Data can be represented in various manners which may cause problems. If the same set of data is represented differently in the various sources, there will be challenges in the integration of data which are handled by the Extract Transform and Load system.

### 2.2.2 Extract Transform Load

The purpose of Extract Transform Load (ETL) is to integrate, clean and standardize the data, often for storage in Data Warehouses (DW) (Chaudhuri et al., 1997). According to Kimball and Caserta (2004) the ETL system is the foundation for a successful data warehouse. In case of a successful implementation, separate data sources can be used concurrently, delivering data in a format ready for presentation (Kimball & Caserta, 2004). The central objective of the extraction phase is to ensure that the required data is retrieved using as little resources as possible. In the following figure, an overview of the ETL process is presented.

![ETL Process](image)

Figure 2.2: ETL Process

The first sub-process focuses on retrieving data from numerous data sources. As mentioned in the previous section, each data source may have its own format or organization. Due to the different formats and context of data, architects often need to extract more data than needed because it is not possible to distinguish the specific subset of interest at the time (Lane, 2005). When designing an ETL system, the architects also need to take the performance of the source system into consideration. Extracting data from various sources often results in vast sizes of data (Lane, 2005).
avoid affecting the performance of the source systems, periodic extractions often occurs during idle or low-load periods, for example at night, when the systems are not used (Vassiliadis & Simitsis, 2009).

The second step in the process is to transform the retrieved data. According to Vassiliadis and Simitsis (2009) the ETL process often calls for multiple transformation of different characters. The authors presents a classification that divides the transformation and cleaning tasks into three sub categories; schema-level, record-level and value-level problems. Problems relating to the schema-level are naming conflicts, for example that different names are used to describe the same object or structural conflicts. Making sure that identical objects are described using the same representation. The second level of problems relate to the records. Common complications are contradicting or duplicate records as well as consistency problems. For example, the granularity of the data may differ between sources. One source may have collected data at costs per month in contrast to another source which granularity is at costs per day. Dealing with these problems of different aggregations levels are essential in order to conform the data for further analysis. The last category of problems that the authors presents is value-level problems. The example provided by the authors are value representations. For example the state of California may be expressed as its abbreviation CA or by full name, although the value is the same. The second example is interpreting values. The example given is the difference in how Americans expresses dates (mm/dd/yy) in relation to the Europeans (dd/mm/yy). Noteworthy is that popular publications in the area of ETL and DW categorizes the transformation step into cleaning and conforming data. For example Kimball and Caserta (2004).

The third and last step in the ETL process is loading the data into the target source. Although most people understand the purpose and mission of the ETL system. Namely to retrieve data from a source, conform and load it into the data warehouse. Building ETL systems are very challenging due to the constraint realities of data, and also very time consuming considering the complexity of the systems. According to Kimball and Caserta (2004) designing the ETL solution is by far the most resource demanding activity in the implementation and maintenance of a data warehouse.

2.2.3 Data warehouse

A data warehouse (DW) is often used in a traditional BI solution. It is a collection of data with business information derived from operational systems and it may also collect data from external sources as mentioned in previous section (Kimball & Ross, 2013). The goal of a data warehouse is to support business decisions by structuring the data interchangeable for analysis (Kimball & Ross, 2013).
There exist many definitions of DW, one commonly used and accepted is the one provided by Inmon (2002). Inmon is one of the pioneers in the area of DW and he is often referred to as "The father of DW". Inmon (2002) defines DW as the following:

A data warehouse is a subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management’s decisions” (Inmon, 2002, pp. 10)

There are some keywords to Inmon’s definition of a DW that needs clarification. Subject-oriented refers to that a DW supports analyzes of subject areas in a business, for example, marketing. The second keyword found in the definition is integrated. This describes the attribute that a DW may integrate or retrieve data from multiple sources, as described in 2.2.2. This aspect is maintained by the ETL system. Time-variant is the third keyword and it reveals that historical data are often to be stored in a DW. This is one of the most distinct differences between a DW and an operational transaction system. The last classification is non-volatile, meaning that once the data have reached the DW, it should never be altered or changed (Inmon, 2002).

Ralph Kimball, another pioneer in the area of DW provides a more concise definition which puts more emphasis on the functionality of a DW.

"A data warehouse is a copy of transaction data specifically structured for query and analysis” (Kimball, 1996, pp.310)

The data can be stored in a DW in different ways, one of the commonly used ways to model the data in a DW is by dimensional modeling.

2.2.3.1 Dimension modeling

Dimensional modeling is a approach for data modeling in a data warehouse. In contrast to ER modeling where the data is stored in a fashion which there is less redundancy, dimensional modeling requires more storage space. Dimensional modeling have three evident benefits, understandability, query performance and extensibility:

- Understandability - dimensional modeling delivers data that is understandable to the business users (Kimball & Ross, 2013). To understand why, the authors provide the following example of an executive describing their business: "we sell products in various markets and measure our performance over time”. The executive describes the business in terms of dimension, putting emphasis on products markets and time. This way of thinking is in line with dimensional modeling, where the measures or facts are analyzed in those types of dimensions.
• Query performance - due to the structure of dimensional models, the query performance is enhanced. As every dimension is an entry point into the fact table, this structure supports effective handling of the queries performed (Kimball & Ross, 2013).

• Extensibility - if new dimensions or measures needs to be added, it can be easily done by adding a new dimensions table or inserting new rows into existing tables. This is supported because there are not as many complex dependencies as in traditional normalized tables (Kimball & Ross, 2013).

An implementation of a dimensional model is often referred to a star schema. The reason for this is the star-like structure. The star schema divides the business process into fact tables and dimension tables (Kimball & Ross, 2013). The fact tables contain the measures of the business process. Typically, measurements can be transactions related to a specific event. The dimension tables hold data which describes the attributes to explain the measures in the fact tables. Examples of dimension tables may be time dimensions describing dates or periods. Another example of a dimension table is a article dimension, defining characteristics of all the products in the facts table (Kimball & Ross, 2013). The figure below provides an example of a star schema.

![Star schema](image)

Figure 2.3: Star schema (Chaudhuri & Dayal, 2011)

### 2.2.4 Mid-tier

A mid-tier can be seen as a complement to the data warehouse. Mid-tier techniques, such as Online Analytical Processing (OLAP) cubes or in-memory approaches, provides the user with specialized functions and customized solutions for different situations and needs (Chaudhuri et al., 1997).
For example, an OLAP cube provides the user with an understandable view to the multidimensional data from the data warehouse. It also provides the user with frequently used BI operations such as filtering or drilling down (Chaudhuri et al., 1997). The following section will provide a deeper understanding of OLAP cubes and their functionality.

### 2.2.4.1 OLAP Cube

OLAP cubes is a method for storing multidimensional data. The main purpose with cubes is to support reporting. Therefore query performance is essential (Chaudhuri & Dayal, 2011). OLAP cubes improves the query performance by pre-calculating over the dimensions. So when the user performs a question, the cube just retrieves the information and presents it to the user (Chaudhuri et al., 1997). In contrast, consider retrieving data from the data warehouse, where joins to be made by the database manager.

As OLAP cubes are built upon multidimensional data, the OLAP database contains two types of data, measures and dimensions. Measures are a set of values, often numerical, from the fact table in data warehouse. These values are the objects of analysis. As mentioned in previous sections, common examples of measures are sales and expenses.

Dimension describes the measures and categorizes it. Time and markets are examples of dimensions. Moreover, every dimension can also be described by a set of attributes. For example the dimension time may be described by attributes such as year, quarter and months. These attributes may then be associated with each other by hierarchy relation. (Chaudhuri et al., 1997). The figure below illustrates a multidimensional cube as Chaudhuri and Dayal (2011).
2.2. ARCHITECTURE OF BUSINESS INTELLIGENCE

Figure 2.4: Cube for analysing products by date and cities. The axes are the dimensions and the measures are the values inside the cube. (Chaudhuri & Dayal, 2011)

**OLAP Operations**  As a result of the structure and organization of the cube, BI operations are provided to the user to view the data in different ways and perspectives (Han et al., 2012). The operations that the cube support are, *filtering, pivoting, slice-and-dice, roll-up and drill-down* (Chaudhuri et al., 1997; Han et al., 2012).

- *Filtering* is a operation which hides some of the data, thereby enabling the user to focus on specific data.
- *Pivoting* uses a visualization operation which rotates the axes of the data in order to support an alternative perspective.
- *Slice-and-Dice* operations are used to create sub-cubes. Performing a section on one of the dimensions is one way of creating a sub-cube, this operation is called slicing. Creating a sub-cube by selecting more than one dimensions are known as dicing.
- *Drill-down* is the operation of going down in the hierarchies of a dimension. The result is a view which describes the data in a more detailed manner. For example, drilling down from year to month.
- *Roll-up* is the opposite of drill-down, navigating from a more detailed level to a more conceptual.

In the figure 2.5 provided by Han et al. (2012), all operations except filtering are explained in further detail.
2.2.4.2 In-memory Approach

According to Microsoft (2012) in-memory computing is one of the largest growing trends in business intelligence. In-memory analytics is an approach for analysing data held by the main memory, the Random Access Memory (RAM). In contrast, conventional techniques hold the data in disk storage. In-memory database systems are not a new technology. However because 64-bit computing has become more frequently used and because the price of RAM memory has decreased, it is now realistic to hold and analyze large data sets. The data is compressed in a non-relational way and there after loaded into the memory (Yellowfin, 2010). One of the benefits of an in-memory approach is that the performance of querying and interacting with the data will be notably faster compared to retrieving the data from
2.2.5 Front-end applications

There are several front-end applications available when implementing a BI solution. The front-end applications has a decisive role in the success of business intelligence. Without a working front-end solution it does not matter how perfectly the underlying data warehouse is implemented. The users will not be able to work with the system. Both appearance i.e, how the tools look-and-feels, and ease of use together with the technical capabilities are essential to create a successful front-end implementation (Howson, 2008). Common front-end tools are spreadsheets, dashboards and Ad hoc query in which users can perform BI operations (Chaudhuri et al., 1997).

Spreadsheets are one of the front-end tools that has been around the longest. According to Howson (2008), spreadsheets was at the beginning the only interface in which one could work against an OLAP cube and they continue to be an important tool for working against OLAP cubes (Chaudhuri & Dayal, 2011).

Dashboards are another front-end tool often used to monitor an organizations performance. A Dashboard provides numerous indications or charts in a highly visual fashion, much like car dashboards displaying essential information needed when monitoring objectives. The purpose of a dashboard is to arrange information in a manner so that it can be quickly monitored and analyzed (Howson, 2008). Depending on the department and situations, dashboards should be easy to build, displaying the information needed for the specific user, without having to involve the IT department (Howson, 2008).

According to Howson (2008) the term ad hoc query are often related to business queries and reporting tools. The term is somewhat misleading as ad hoc refers to being spontaneously created. However they are often fixed reports. The main difference is that the reports have been created by business users rather than an employee from the IT department. Moreover, Howson (2008) also stresses that ad hoc queries are a key factor in delivering Self-Service BI and information access. Finally, as the users explore the data by using ad hoc queries the outcome of the queries and the question may become a standardized report.

The front-end tool of a BI implementation will be essential for the success of the BI project (Howson, 2008). Even though front-end tools are only one aspect in the BI solution, it is the front-end tools that support the business users to access the data and make analysis. Therefore it is essential to have a front-end solution that enables the business users to leverage the data collected (Howson, 2008). The authors also stresses that a business focused
meta data layer is provided to support easier interaction from the front-end tools. Lastly, Howson (2008) argues that it is important to follow and upgrade to emerging technologies that can offer users a more user friendly interface.

2.3 Self-Service Business Intelligence

A growing trend in the area of BI is Self-Service BI. Inc (2013b) defines Self-Service BI as: "... end users designing and deploying their own reports and analyses within an approved and supported architecture and tools portfolio." Before Self-Service BI, BI applications were build by BI developers and these applications were often built for custom requirements. Once developed and deployed, these applications often became static. In case of new business requirements or user needs new applications or reports had to be rebuilt. According to Harinath, Pihlgren, and Lee (2010), analyst making business decisions who did not have the competence to built new applications or reports, could easily be tempted to use workarounds. For example collecting the data needed and saving it into a local spreadsheet. The result of of such action could be that analysis were made upon outdated data or incomplete data. In addition, as new data models emerged, the importance of the data warehouse delivering one single truth, as Kimball and Ross (2013) identify, could vanish.

To better understand Self-Service, Harinath et al. (2010) provides the readers with an analogy. Comparing Self-Service BI and more traditional BI solutions to the movement in operating systems going from command-line interfaces to graphical user interfaces. In the earlier versions of operating systems, the users were provided with a very simple interface containing just a command line. There existed more user-friendly solutions however these applications had to be developed by IT-professionals (Harinath et al., 2010). The result was that non-advanced users had to rely on the IT-department staff to develop applications in order to perform the task or wait for the IT-personell to provide them with the information needed. Later on when the operating systems started to provide a richer interface, allowing the user to graphically interact with the computer, tasks like connecting a computer to a printer device became possible for the users. Before, they had to rely on the IT-personell to write low-level printer drivers, now they could perform the same task in a graphical interface (Harinath et al., 2010).

Analyzing the definition provided by Inc (2013b), one can see similarities in how Self-Service BI aims to support the users with the ability to perform certain tasks without the support of the IT-department, in a similar way that modern operating systems supported end-users. According to Shailesh (2008) numerous business users and decisions-makers are dependent on the IT department to provide them with the information needed. Furthermore,
if the organization does not provide the users with the ability to create standardized reports and ad hoc reporting, it will result in a higher dependency on the IT department and also increasing costs (Shailesh, 2008). The aim of Self-Service BI is to provide the users with these capabilities and thereby reducing the dependencies on the IT department and also accelerate the business efficiency by making accurate and fast decisions (Harinath et al., 2010; Shailesh, 2008).

Imhoff and White (2011) have found that there are four key factors in creating successful a Self-Service BI solution.

1. Firstly the users have to be able to easily access the different data sources needed. This can be a challenging task because the users might require information from multiple sources. The authors stress the fact that if the users can’t access the information needed, it is impossible to create a working Self-Service environment.

2. The second key factor is that the BI tool set have to be easy to use. Easy-to-use applications are essential in making Self-Service BI work. If the applications are difficult to handle, the user will find workarounds or even make decisions without any confirmation that the data supports that decision.

3. Thirdly, the authors found that accomplishing a DW solution that is fast to deploy and easy to manage also is important in creating a desirable Self-Service solution.

4. Lastly, the authors found that making the results from the BI system easy to consume and enhance is one of the most important objectives for the business users. If the users do not understand the information presented or finds the information difficult to interpret they will stop using the system.

2.3.1 Microsoft Power BI

Power BI is the latest BI solution from Microsoft and it is an umbrella name for multiple features and services. It includes the features/service, Power Query, PowerPivot, Power View and Power Maps (Microsoft, 2014b). This product series from Microsoft is a Self-Service BI solution which lets the users create data models and visualizations. Also, through Office 365, reports can be shared for collaboration in teams.

2.3.1.1 Office 365

Microsoft Office 365 is a platform for web applications. It include tools for everything from tools for collaboration to and document management.
Power BI sites is an application for Power BI in Office 365. It enables the users to visualize and dynamically view and share reports created with PowerPivot or Power View. Furthermore, it enables the users to share data and insights. 2 gigabyte is the maximum upload file size for a worksheet, including the data model and visualizations (Microsoft, 2014b).

Another feature in Power BI for office 365 is Power BI Q&A. The feature allows users to use natural language queries in order to analyze the data. The results of a question is presented in a visualization, by using a in-memory storage, the results are shown fast and dynamically as the question is formulated. When a question has returned a result, the user can modify the visualization by choosing a different chart or filtering the data further. Lastly, once the user is satisfied with the results, the question can be saved and shared with the team member (Microsoft, 2014b).

2.3.1.2 Excel 2013

Excel is a commonly used tool for ad hoc reporting and analysis (Warren, Neto, Misner, Sanders, & Helmers, 2013). The spreadsheet application helps organizations organize and understand their business. According to the authors, Excel is already used by most analysts in organizations as well as by casual users. Therefore, most users in organizations are already familiar with this tool. Warren et al. (2013) describes that it can be challenging to stop the users from using excel and difficult to get it out from the organizations business. Because it is so widely used in organizations, the authors claims that depending on the requirements and needs of the organization, creating a BI solution based on Excel could be the path of least resistance. One reason is that the users won’t have to learn a new tool. Excel is the foundation of the power BI suite.

2.3.1.3 Power Query

Power Query is a Self-Service tool to extract data from different sources, transform them and load them into a PowerPivot data model, similar to an ETL tool in a traditional solution (Microsoft, 2014b). Power Query can extract data from various sources, stretching from data warehouses to webpages, for example, Wikipedia articles. However, as mentioned in section 2.3, allowing the user to create their own data models can result in more than one truth and thereby causing the user to create misleading analysis and reports. Knigth (2013) presents a table in which different aspects have been identified and weighted in order to give a better understanding in which situations Power Query is appropriate to use. The table is presented in the figure below:
2.3. SELF-SERVICE BUSINESS INTELLIGENCE

Figure 2.6: Decision matrix (Knigth, 2013)

The decision matrix can be used to understand if Power Query can be a suitable solution for the organization. The Importance score may vary depending on the situation and the organizations needs. For example if Data Quality is the most important factor in a certain BI solution, it will be given the value 5.

### 2.3.1.4 PowerPivot

Microsoft PowerPivot is a native add-in for Excel 2013 which aims to provide Self-Service BI. The technology objective is to provide the users with the ability to connect to different data sources, create data models, create their own reports and perform complex analysis without the involvement of BI personal (Ferrari & Russo, 2013). There are two versions of PowerPivot: one 32-bit version and one 64-bit version. The two versions have different limitations on how large a data model may be. The 32-bit environment can only use 2 gigabytes of virtual address space in contrast to the 64-bit version which is only limited by the available virtual address space and the system resources (Harinath et al., 2010).

In PowerPivot, the users are able to create data models. The data model is a set of tables and it describes the relations between the organizations business functions and processes, for example how products relates to costs (Warren et al., 2013). The add-in utilizes an in-memory database technique called xVelocity analytics. Most databases stores the data in a row-oriented way, however xVelocity uses a in-memory technique in which the data is stored in a space-saving columnar database. Instead of saving the data row by row for each table, xVelocity structures every column as a separate entity and the for each column the data is stored in an abstracted way (Ferrari & Russo, 2013). This structure supports very good query performance but higher computational efforts.
CHAPTER 2. THEORETICAL BACKGROUND

PowerPivot have the possibilities to work as a complete BI solution and thereby be an alternative to traditional BI solutions (Harinath et al., 2010). However according Ferrari and Russo (2013) PowerPivot should not be seen as a replacement for more traditional BI solutions, but rather a complement or as a part of an existing BI solution. According to Warren et al. (2013), PowerPivot supports the users in a way that makes it easier to create a Self-Service BI solution. However, PowerPivot is mainly for power users. End-users with less experience however will benefit the most from PowerPivot when they can take part of workbooks and reports shared in Office 365.

2.3.1.5 Power View

Power View is a reporting engine in which users can build interactive reports and visualizations. Once created, the reports will support interactive data exploration and inspire ad hoc reporting (Ferrari & Russo, 2013). The tool mainly targets nontechnical business users as it is designed to be easily used (Warren et al., 2013). There are some limitations in Power View for Excel. Power View reports can only be built upon PowerPivot data models. Excel 2013 does not provide the user to work against data sources as OLAP cubes or directly against the data warehouse (Ferrari & Russo, 2013). However, Power View can connect to a multidimensional model if the user is operating on an Office 365 instance (Microsoft, 2014a). Power View reports support all the operations described by Han et al. (2012) in section 2.2.4.1. The interactive interface provides the users with an easy way to drill-down in the charts. Just by double-clicking on the data, the user can drill-down into a predefined path which allows the users to analyse data from a different perspective (Ferrari & Russo, 2013). In addition, there are cross-interaction
functions in Power View. The charts in the same presentation area are connected and interaction with one chart will affect the data in the other visualizations. This function makes it easy for the user to explore the data from multiple perspectives, while remaining in the same view (Warren et al., 2013).

2.3.1.6 Power Maps

The tool Power Maps is also an add-in to Excel in which 3D data can be visualized. The tool lets the user to explore traditional 2D charts and visualizations in a new powerful way, discovering new perspectives to analyse and present the data. One can also create interactive presentations in the tool and sharing them with others. Geographical data in spreadsheets or data models are automatically recognized by Power Maps, stretching from coordinates to city names and then plotted on maps provided by Bing (Price, 2013). Note that Power maps haven’t been evaluated in the case study.
Chapter 3

Method

The following chapter will describe the method used in this study. First of all, theory describing case study method and data collection will be presented. After that the case study approach will be explained, which covers a section of pre-study, implementation and evaluation. Additionally, the data analysis and validity will be emphasized in this chapter.

3.1 Case Study Method

Performing case studies has become a frequently used method in software engineering research papers. Case studies in software engineering all have a common denominator, namely that they study specific cases (Runeson, Höst, & Rainer, 2012). The focus in a case study is on studying the contemporary phenomenon in everyday situation and context (Yin, 2003). Derived from the leading researchers in the area and rephrased to be specifically for software engineering case studies Runeson et al. (2012, pp.12) provides the following definition of a case study:

"...is an empirical enquiry that draws on multiple sources of evidence to investigate one instance (or a small number of instances) of a contemporary software engineering phenomenon within its real-life context, especially when the boundary between phenomenon and context can’t be clearly specified"

That is, a case study method can be suitable when reviewing correlations between contextual conditions, assuming they are related to the phenomenon of the study. In contrast to an experiment, where the phenomenon is separated from the context so that the variables can be controlled and study how they each affect the phenomenon (Yin, 2003). Furthermore, there may be
different purposes behind a case study. Originally, case studies were mainly conducted in exploratory purposes. However, today they can be used for other purposes as well. For example in a Descriptive purpose, describing a situation or in Explanatory purposes, searching for an explanation of a phenomenon (Runeson & Höst, 2009). As seeking new insights is the main purpose of this thesis, the more traditional exploratory approach have been chosen in this case study.

3.1.1 Data Collection

Data collection is the process of collecting and mapping relevant information in a systematic and controlled practise. The goal of the process is to be able to answer research questions or evaluate other presumptions based on quality evidence. According to Runeson and Höst (2009) it is important to understand what data to collect as well as how to collect it, in order to achieve a valid result. The data collected can be either quantitative or qualitative. Quantitative data deals with numbers which can be measured. For example, volume, sales or time. In contrast, qualitative data relates to descriptions and words. This type of data can not be measured however it can be observed. Qualitative data is more frequently gathered in a case study due to its structure. It provides a deeper understanding of actions, thoughts and experiences of the phenomenon studied (Runeson et al., 2012).

3.1.1.1 Interview

In case studies, interviews are an important source for data collections. Interviews can be categorized into, unstructured, semi-structured and fully structured. In a unstructured setting, the interview questions are formulated so that the interviewee may speak openly around the area, thereby allowing the interviewer to understand the general picture and interests (Runeson & Höst, 2009).

In a semi-structured setting, questions are prepared in advance. However the questions do not have follow a specific structure and the interviewee in order can be more flexible for example following up on questions that came up to better understand the interviewee’s thoughts and perceptions. The last setting, fully structured, uses prepared questions and they should all follow a particular order. Fully structured interview are similar to questionnaire-based surveys in some aspects.

3.1.1.2 Tests

In this case study tests have been made in an exploratory purposes. The tests was performed in Excel 2013 32-bit with the Power BI add-ins. The
computer had an Intel Core i7 processor with a dual core, 8 gigabyte of RAM and Windows 7 Enterprise edition as the operating system. The goal of the test was to try to build the applications and reports based on the requirements from the conducted interviews.

### 3.1.1.3 Observations

The aim of observations are to study how particular tasks are performed and how the system behaves. In this case, the observations have been made when interacting with the tool observing the results of the interaction Runeson and Höst (2009).

### 3.1.2 Case study approach

This section will highlight the approach of the case study. First of all the pre-study will be clarified followed by the implementation as well as the evaluation of the study.

#### 3.1.2.1 Pre-study

Firstly a semi-structured interview with a controller of medicines was conducted in order to understand the requirements and goals of the case study. The interview also included questions to understand the current situation at the county council to better understand how they worked with the existing tool and if there were any challenges from a user perspective regarding the existing BI solution. In addition, semi-structured interview were also conducted with a member of the Business Intelligence team. The aim of this interview was partly to understand how the existing system worked today but its intention were also to understand common problems with the existing BI solution. During the pre-study, literature was read to understand the concepts of both BI and Self-Service BI.

#### 3.1.2.2 Implementation

Backups of the data warehouse and multidimensional models were created from the production environment in order to have an up to date instance of the database while testing as well as to avoid any problems with the database in the working environment. These backups were then restored on the test machine.

The first step in the case study was to try implement Power BI as a complete solution for the county council. The process of using Power Query as implementation for the ETL system was tested. After exploring if Power BI
could be implemented in all the steps of a complete Self-Service BI solution, the project continued with importing the data with PowerPivot into Excel in order to try the different front-end tools of the suite. The importation of data revealed a number of limitations in both the hardware and the software of the test machine. Due to the 32-bit installations of excel, the in memory model could only contain 2 GB data. In addition the test machine only had 8 GB ram and during the importation of the larger fact tables, containing many rows, the processes of SQL server instances together with other processes on the computer, reached its maximum capacity resulting in crashes of Excel. To handle these problems selections of data was imported by writing SQL commands that only selected every eight row from the fact table.

After the data was imported into PowerPivot the work continued with building a star model in PowerPivot. This was done by looking at the requirements and mapping from the data warehouse to understand how the relations between the fact table and the different dimensions were to be connected. Specifications on how the data were structured where provided during the interview with the member of the BI team. This data model was then used in building the different Power View charts and reports. Power View was tested by trying to create the reports from the requirements received in the interviews with the controller.

The two main goals were to build reports, firstly compliance in procurement of different drugs and secondly be able to follow the volume development of selected drugs. There were two different approaches in creating the reports. The first approach was to use the data model that were built in PowerPivot and then create the reports and dashboards in Power View. Due to limitations in the existing version of Power View for excel, connections couldn’t be made directly to multidimensional models, therefore reports were also made in PowerPivot. The main technical difference was that PowerPivot could be used to build visualizations with connection directly to the multidimensional cubes. Creating reports in Power View and PowerPivot were very similar and there is no major difference how the measures and dimensions are presented to the user. Power View was the main tool used for the front-end implementation.

Furthermore as described in the background, the organization has projects that aims to use office 365 as a platform for the front-end in the BI solution. Therefore Office 365 has briefly been tested by uploading and interacting with the reports in an evaluation edition for Power BI Office 365, which is a light-weight version of SharePoint. Observations were made while interacting with the reports in Office 365.

During the implementation and testing of Power View and Office 365, continuous dialogues were held with the controller. The deliveries of reports were made in three iterations, presenting the current reports to the con-
troller in order to receive feedback on how to visualize the data and gather information to understand what was interesting to visualize. These dialogues provided useful information in how to build the reports.

### 3.1.2.3 Evaluation

To evaluate the different solutions, the results were compared to the requirements of the county councils need on a BI solution and also how the Power BI implementation could compare itself with the existing solution. The tools were evaluated by running a number of test cases described in the previous section and evaluating these finding by comparing the observations and experiences from the test with both the information from the interviews and also with the literature.

### 3.1.3 Data Analysis

Data analysis is the process of examining the collected data with the goal of reaching useful conclusions and informations. According to Yin (2003), data analysis is particularly difficult in a case study due to the absent of proven techniques for data analysis. Furthermore, Yin (2003) recommends that each case study should follow a generic analytic approach, specifying priorities for what to analyze and why. In line with Yin (2003), Runeson et al. (2012) provides some bullet points that are generic for analysis techniques.

- Identify abstractions regarding patterns, sequences and relations in the data.
- The analysis should be performed in an iterative manner.
- Being systematic in order provide a clear chain of evidence to the readers.

Specific to a data analysis in a qualitative case study is that the analysis made in parallel with the data collection. The reason behind this is that analyze of data may provide insight that additional data collections are needed (Runeson et al., 2012). As the overall objective is to derive conclusions from the data, being systematic is an important factor in maintaining an understandable line of argument.

### 3.1.4 Validity

The validity of a case study is a measure of how reliable the result and findings are. It also reflects the unbiasedness of the researcher (Runeson et al., 2012). According to the authors, the researcher should regard the validity
throughout the case study although it should be evaluated in the analysis phase. According to Runeson and Höst (2009); Yin (2003) there are several aspects of validity; construct, external and reliability. Construct validity reflects if the study really reviews the phenomenon that the researcher claims to be investigating. External validity are related to what degree the findings are generalizable and relevant to other researchers. The last one mentioned, reliability is concerned with the dependency to the author. If the study would to be reconstructed by another author, the outcome should be equivalent with the previous study (Runeson & Höst, 2009; Yin, 2003).
Chapter 4

Results

This chapter will present the results of the case study performed at the County council of Östergötland. The result is divided into three different sections: pre-study, implementation and evaluation.

4.1 Pre-study

In the following section there will be give a description of how the studied organization work with Business Intelligence.

4.1.1 County council of Östergötland

The county council Östergötland (LiÖ) is a large organization with 12,000 employees. Their main responsibility is to provide health and dental care for Östergötland, about 70 percent of the budget is used for health care. The rest is used for public transportation, research and development, education and regional development. The cost of the County Council’s operations are financed primarily by the tax incomes from the residents of Östergötland and contributions from the government. LiÖ is one of the most efficient county councils in Sweden. This means that the organization has been performing well in the areas of medical quality, patient experience and availability, this while costs are low compared to most other counties in the country. (Bjäresten, 2014) One of many reasons for this might be explained through the usage of business intelligence systems. Since 2005 LiÖ has been working with a business intelligence system. Today there is a business intelligence team with 17 employees including 2 external consultants whom works with management and development of the existing solution.
The existing business intelligence system has a mature IT architecture, however, the user interface is based upon an old tool which results in a low usability and challenges in creating reports and charts. In addition, there are some analysis that cannot be done with the existing tools which have resulted in employees taking workarounds. For instance, extracting data and build individual data models and reports in Excel. One of the dangers with workarounds is that mistakes can be made creating new data models, and thereby result in impaired analyzes. In 2012 LiÖ started a new BI initiative with the goal to replace some parts of the existing BI solution. One of them is the front-end tool which the end-users work with.

4.1.2 Background of the case study

Semi-structured interviews have been conducted with one of the controllers of medicines who is a member of the group of medicines and also with a member from the Business Intelligence team. The controller is responsible for the development and management of budget and management model for therapeutics, as well as, prognosis and follow-up. Furthermore, the controller works with support for politicians and the business regarding questions of medicines. The interviewee is a member of the management staff responsible for monitoring activities related to medicine. The group of medicines is part of the management staff at LiÖ. The group has overall responsibility for strategic and administrative issues relating to medication. Some of the drug group’s main tasks are to:

- Provide support for prescribers, operations managers, officials and politicians in the pharmaceutical field.
- Encourage local quality assurance in the field of medicines.
- Responsibility for medical and financial monitoring of the pharmaceutical field.

Members of the group of medicines are users of the BI system. The tool provides them with necessary information for analysis and is an important tool for monitoring different aspects in the area of the pharmaceutical field. In addition to the interviews held with the controller of medicine, interviews were held with a member of the BI team who works with the development and management of the medicinal part of existing BI system. In this case study, as mentioned in the demarcations, only the medicinal part of the BI system has been targeted.

4.1.3 Existing back-end solution

There exist a mature and well managed data warehouse in the county council of Östergötland BI solution today. The database related to medicine con-
tains large amounts of data. In just one year more than 10 million rows are inserted into the existing tables. This reveals that the system handle large amount of data every month. Therefore, the ETL system which handles the data from the various suppliers are one of the most crucial components in the back-end structure. The illustration below shows a star schema of how the major fact tables and dimension are related to each other in the data warehouse.

![Star schema for the data warehouse](image)

Figure 4.1: Star schema for the data warehouse

All the main three fact tables have been used in the case study. However only a number of the dimensions have been used in the process of testing Power BI. Those dimension are:

- Medicinal preparation
- Period
- Prescribing Unit
- Item description including ATC-Codes which is a classification system for medicines.
- Geography

In addition to the data warehouse, there exists multidimensional cubes that have been build for specific purposes and periods. Each cube contains data over a three year period.
4.1.4 Existing front-end solution

The top management at LiÖ is interested in following up on different aspects of their organizations performance. One large area in the organization is the monitoring of medicines. They are both interested in monitoring volume and cost related to prescriptions of certain medicines. Also, they would like to be able to follow for example the volume development of a procured medicine in comparison to a similar medicine previously prescribed. The existing BI solution supports some of these tasks. However, the front-end tool is quite old and not very user-friendly.

In the following capture, the user is trying to find the right medicine for analyzing the volume development. For additional captures of the existing front-end tool see appendix A

As one can see in the capture, traversing the list of medicines order to find the right one is very demanding. The tool can only hold a certain amount of elements in one list and there are thousands of medicines in the list, resulting in an abundance of sub-lists. If the user were to make a small mistake in the navigation, the process of finding the right one would start over. Furthermore, another feature which is missing is the possibility to choose multiple medicines to follow. Because of that, the user have to create multiple charts, one for each medicine, and saving them as non-interactive pictures or pdf:s in order to compare them. The outdated front-end tool together with employees being used to work in Excel have resulted in employees creating workaround in Excel. In order to create these charts employees extract data from various data sources and saves the data in spreadsheets on which the charts are built upon.
4.2 Implementation

This section will investigate the implementation of Power BI as a complete BI solution. Followed by explaining Power BI as a front-end solution.

4.2.1 Power BI as a complete BI solution

The implementation and testing of Power BI started by extracting data using Power Query. A selection of the raw text files from the suppliers were chosen for the test. To understand how to extract the information, specifications of the structure of the raw text files were provided by the BI team. The process of extracting the data from the text file was very demanding. Power Query extracted the data into a table, however as the columns in the text file were separated after a number of varying character, the separation of the column had to be done manually in a repeating fashion. In the next step, transformations procedures should have been done. However, the operations provided couldn’t perform the transformations needed. Some transformations, like merging columns. However there were no features for locating new data in order to compliment the dimension tables.

The extracted data was then loaded into a PowerPivot data model. Additionally, the dimensions were loaded into the model as well, extracted from the data warehouse backup. In the next step, a dimensional model was created with the help of the mapping specification, see figure 4.1. However, the model couldn’t be completed. The reason for this was that the fact extracted from the raw files needed transformation procedures which could not be done in Power Query. Only a few fact columns were able to relate to the dimensions. However, many of them in a incorrect way which led to an impaired data model. The test of creating a data model from the raw file was repeated with some help from one of the members of the BI team however the results were the same. Power Query could not meet the ETL requirements of the county council.

4.2.2 Power BI as front-end solution

As Power BI could not work as a complete BI solution, the front-end tool in Power BI was tested in order to see if they could work together with existing back-end solution. In order to test Power View a data model had to be build in PowerPivot. The reason for this were that Power View for Excel did not support building visualizations directly from multidimensional models. Although there are technological differences in how the users connects to the data sources, the result for the end-user is essentially the same. Building the data model in PowerPivot resulted in some complications. The test machine only had a 32-bit installation of Excel, restricted the data model to the size
of 2 gigabyte. The importation of these fact tables were made by SQL queries, removing rows containing null values and invalid values etc. Also, lastly choosing only every eight row with a module where clause.

After that, the collections of tables were given relations according to a dimensional model resulting in a data model supporting the same functions as the multidimensional cube. This was done by looking at the requirements and data mapping of the data warehouse which described how the fact tables were related to the dimensions, the data models can be viewed in the capture below.

In the first test of Power View, reports and dashboards for monitoring the volume development of selected medicines and for analyzing the development of medicine procurements were build. Different solutions and charts were tested and to understand how the visualizations worked in Power View, observations were made when interacting with the charts. The end results can be viewed in appendix B. In addition the reports built in Power View, test were also made by using PowerPivot as the visualization tool. The main difference in using PowerPivot was that it supported visualizations to
be built from a multidimensional source.

The last tests were to explore the function of Power BI in office 365. A Power BI site was set up and the reports created in Power View for Excel was uploaded. In Office 365 there is a function for asking natural language queries against the report. The feature was tested by trying different questions in order to understand how to formulate questions. The results were mixed and the tool behaved inconsistently. For example, when a question is asked, the feature interprets the question and reformulates it into a question understandable for the data model. When writing the same question as proposed by the feature, a different result was presented. This indicates that both how the question is formulated and in which way the question was entered affects the results. Once a question is formulated and the user is satisfied with the results, it might be saved. The interviewees found this feature very helpful for collaborating within a team. By saving important questions, all the member of the team could consume the report in the same way and interesting questions could be shared with the team members. Captures of Power BI in office 365, can be viewed in Appendix C.

4.3 Evaluation

This section will review the evaluation of Power BI as a complete BI solution. Followed by explaining Power BI as a front-end solution.

4.3.1 Power BI as a complete BI solution

The results show that ETL system and process is very important for the county council of Östergötland. The organization needs to be able to handle large data sets from several different suppliers. Furthermore, the data files from the suppliers varies both in how the data is structured, as well as in the data quality. Handling these data files is one of the tougher challenges for the organization. According to Kimball and Caserta (2004) the ETL system is the foundation for a successful data warehouse and it is one of the most resource demanding activities in an organizations BI solution. Therefore this is one of the most crucial factors for the BI system’s success. The importance of the ETL system was also confirmed from the interview with the BI team member.

In the Power BI, Power Query is the tool for extracting, transforming and loading data. This Self-Service ETL tool however can not meet the requirements and needs of LiÖ. The add-in have functions for extracting data from various sources and it would be possible to extract the needed data from the different suppliers.
Even though it is possible to extract the information, each extraction have to be made manually and this is far too time consuming with the amount of data files the county council receives. The major problem arises in the transformation step. There are many kinds of complications that must be handled and the data sets are too large for it to be realistic to manage manually. For example, when extracting the data from a comma separated file, the slightest deviations in the file, like a top row explaining the document, can impair the transformation. In the existing ETL system at LiÖ there are transformation packages filled with code which automatically handles this logic and thereby resolving all the problems that Vassiliadis and Simitsis (2009) described. For example schema-level problems like naming conflicts. These kinds of conflicts have to be handled manually in Power Query by example using the search and replace tool. The organization requires that the BI system can handle problems in all of the levels that Vassiliadis and Simitsis (2009) described.

Moreover, there are also complications in the load step using Power Query. All data that have been extracted and transformed in the previous steps can only be loaded into PowerPivot’s data model or as tables in a spreadsheet. The data model in PowerPivot uses an xVelocity which is an in-memory database solution. Storing the data in xVelocity can be suitable when the data sets are quite small. However in the case of LiÖ, the data sets are very large and continuously growing. Even though there is a 64-bits Excel version installed with enough memory to hold the data model, it exists a 2 GB limitation of the worksheet in order to share it on office 365.

In addition, to better understand why Power BI can not fulfill the requirements and needs of the organization, Knigth (2013) decision matrix can be used. The matrix is used to prioritize different factors that are essential for the organizations BI solution. The matrix then provides the organization with a recommendation which kind of solution fits the organization best. When a consultant from the County Council’s BI team assessed which factors were the most important, the result was the following:

<table>
<thead>
<tr>
<th>NEED</th>
<th>IMPORTANCE SCORE (1-5)</th>
<th>SELF-SERVICE BI</th>
<th>CORPORATE BI</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATA QUALITY</td>
<td>5</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>SHORT DEVELOPMENT CYCLE</td>
<td>1</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>SINGLE VERSION OF TRUTH</td>
<td>5</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>USER DEVELOPMENT</td>
<td>4</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>SCALABILITY</td>
<td>4</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>SCORE</td>
<td></td>
<td>5</td>
<td>14</td>
</tr>
</tbody>
</table>

Figure 4.4: Decision matrix: Assessed by BI team member

As the result shows, some key factors for the organization are data quality,
CHAPTER 4. RESULTS

single version of the truth and scalability. These results provides additional confirmation that the back-end structure supported by Power BI can not meet the demands and needs of the County Council. All the highly prioritized factors are strongly related to the back-end solution of BI systems. Therefore, a complete implementation of a Self-Service BI based on Power BI would not be a realistic solution for LiÖ. The organization requires a powerful ETL system and a robust data warehouse for their business intelligence solution to function well. However, the decision matrix also shows that self development is important for the organization. Therefore, the front-end solutions could work well together with the existing back-end structure.

4.3.2 Power BI as front-end solution

As presented in pre-study, the existing front-end tool is outdated and some of the employees finds it difficult to use. This has resulted in employees creating workarounds, extracting data and building their own data in Excel. By doing this, key factors such as data quality and one single version of the truth can be impaired, resulting in incorrect analysis and in the end users making the wrong decisions. Some of the factors found in why the existing front-end does not support the end users are; navigation difficulties, not user-friendly and does not support some analysis, for example following the volume development of two specific medicines. In addition, some of the operations described by Han et al. (2012) are not possible in the current tool, and operations such as roll-ups are non-intuitive and difficult to use. For example, when the user have performed a drill-down on a period, going from viewing the data by year to looking at a the month of a specific year, the operation for rolling up is then to navigate to the appropriate dimension in the list above the chart. After that the user have to remember which level they were on previously and choosing that item from the list. This procedure for roll-ups is not user-friendly according to the interviews. Another drawback with the existing front-end tool is that it can not display more than one chart in the same view.

In Power View many of these drawbacks described with the existing tool are solved. As Imhoff and White (2011) describes, the BI tools have to be easy for the users to work with. This is one of the key factor in achieving a Self-Service BI solution. The first observation made when working with Power View is that it is very easy to use. This is mainly to the drag-and-drop functions of the program. The users do not have to be certain of how to create the reports. Dragging the measures and dimensions into the workspace will result in Power View generating the appropriate table by interpretation the data.

Moreover, when building a report in Power View the end-user is able to create their own hierarchies. Therefore it is possible for the user perform
4.3. EVALUATION

drill-downs that can go across dimension. This function turned out to be very helpful when presented to the controller of medicine. Members of the medicinal group often wanted to first analyse the data. For example a period perspective, viewing the volume development by month. However, after viewing the data by month, drilling across the dimension, viewing data by prescribing units. These kinds of analysis are very powerful for the county council in order to understand how the organization is performing. According to Imhoff and White (2011), making the results easy to consume and enhance is also an important factor in creating a working Self-Service Solution. This possibility for users to create customized hierarchies and thereby exploring the data in new exiting ways is an example of how Power View enhances the users ability to make accurate analysis. Another feature which the controller of medicine found helpful was that the charts were highly interactive. As shown in the capture below, when the users click on a bar, it filters all the charts in the workbook providing the user to see correlations between data.

From the tests and observation from working with Power View, creating reports is highly possible for users without any interaction from the IT-department and it support the users to make accurate and fast decisions. Another benefit from working with Power View and in an Excel based environment is that the users are familiar with the settings. Both interviews confirmed that the tools are easier to work with because of the previous experiences with Excel. It also confirms the reasoning by Warren et al. (2013)
that Excel, which is a familiar environment for many users makes it easier for the users to adapt and use the tool.

Office 365 will be an important platform if the organization wants to implement Power BI as their front-end tool. It provides the organization with a tool to share and consume the reports that have been created. The interviewees also found that the feature of Power Q&A rather helpful and powerful. Being able to share the questions with the team members will enhance the exploration of the reports. It would also save time for the team because only one member needed to find the answer in order for everyone to view it.
Chapter 5

Discussion

In this chapter the result of the study will be discussed in relation to the aim of the study and the problem definition. Furthermore it will analyze the result in relation to the theoretical background. The method critic will also be presented in the end of this chapter.

5.1 Results

This section will present a discussion regarding the results of the thesis.

5.1.1 Implementation of Power BI as a complete BI solution

Implementing a complete Self-Service solution is not realistic in an organization as large as the county council of Östergötland. The results show that Power BI can’t meet the requirements that LiÖ has on the back-end components of their BI system. The main reason why Self-Service BI is not a sustainable solution for the county council is the demands for high data quality as well as that it can only exist one version of the truth are decisive. Furthermore, the management of the systems which handles large data sets are too demanding; it requires people with IT expertise to maintain the system. The results from the case study are supported by Knigth (2013) decision matrix. A more traditional BI system is more suitable in organizations where data quality and one version of the truth are the most important. These finding strengthens Ferrari and Russo (2013) reasoning that these tools should not be seen a complete solution but rather as an efficient compliment or substitute for selected parts.
5.1.2 Most suitable parts for the implementation of Power BI

The results show however that some parts of the Self-Service tools can be suitable for the county council’s needs. For instance, Power View along with Office 365 can be a good solution for front-end part. The results also showed that the existing system was difficult to use and did not support the end-users needs. Implementing Power BI as the front-end solution and combining it with the existing back-end structure could be very successful for LiÖ. The existing back-end solution would ensure a high data quality and one version of the truth and Power BI would provide the users with a more user-friendly front-end tool which would support their needs better. Power BI also provides three of the key factors in achieving a Self-Service environment. Because the existing tool is not user friendly and does not fulfill the users requirements, workarounds have occurred in the organization. Which is supported by Imhoff and White (2011).

5.1.3 Potential effects of Power BI

The finding from working with Power View is that it is easy to work with, implementing Power View could therefore eliminate the presence of harmful workarounds. In addition, the results have found that Power View, along with Office 365, enhances the users ability to understand and analyze data. The reports are highly interactive and the users can easily perform the operations illustrated by Han et al. (2012). Moreover, the report area supports multiple charts and tables. These charts are all connected to each other making it possible for the user to view the data from multiple perspectives simultaneously. Finally, the finding also shows that sharing the reports in Office 365 and enabling Power BI Q&A will further improve the users ability to consume the results and share it among the teams. Reviewing a new solution for the front-end is an important task for the county council. As Howson (2008) described, the front-end tools are essential for the BI solution to enhance the organizations performance.

From findings of the case study, changing the front-end solutions could result in a number of positive effects for LiÖ. Firstly, the quality of the users analysis would probably improve resulting in a higher performance of the organization. Moreover, a new front-end solution which is easy to use and familiar to the employees in the organization could also increase the usage of the BI system, culminating in a higher awareness of the organizations situation among the employees. It could also empower new users to perform their own analysis, gaining insights for their angle in the organization. There are also several benefits with employing a collaboration platform such as Office 365. Making it possible for the users to share their insights will
probably result in improved communication, additional knowledge and in the end, an increased ability to make the right decision for LiÖ.

Although a full implementation of Power BI is not realistic in a large organization such as LiÖ. The experience gained from testing Power BI and discussion with the member of the BI indicates that the tools could be a good solution for smaller organizations. In a smaller organization where the competitive situation is tougher, the short development cycles becomes more important in order to stay flexible. Smaller data set could also make it possible to built complete data models, which could store all data of the organization. It is also realistic to handle the ETL process manually, in Power Query, if the data sets are smaller. Another function that could be useful in Power Query is the extraction of external data. This function wasn’t very helpful for LiÖ because they mainly based their analyzes on internal data. However, Power Query allows the user to extract external data and then combine the results with their own. An example could be to extract financial statements from competing businesses. Visualizing data from competitors in correlation with their own, thus provide a better understanding of the situation.

5.2 Method Critic

In this section the method critic is presented.

5.2.1 Literary criticism

Even though the demands for Business Intelligence grows rapidly, the area remains unexplored by the researchers. Many references used in this thesis mentions the lack of research in the area. The following citation provides an example of this: ”...research in this field is, to put it charitably, sparse” (Negash, 2004, pp.1). Even though the publication is quite old, the problem still exists. Research regarding the new trend with Self-Service is even more thrifty. Most of the major Business Intelligence vendor offer some versions of Self-Service BI and the demand is growing rapidly. However, very few publications have been done in the area. Most literature available are provided by the vendors in form of white papers or from studies sponsored by these vendors. To really understand the effects of Business Intelligence and the Self-Service trend, more research have to be done.

5.2.2 Construct Validity

According to Yin (2003) subjective judgement is often a problem as data are collected in a case study. To increase the construct validity interviews
have been done with employees with different situations in the organization. In addition, findings of previous researchers have been used to measure the organizations needs and requirements, for example Knight (2013), in order to increase the construct validity.

### 5.2.3 External validity

External validity are related to what degree the findings are generalizable and relevant to other researchers (Runeson & Höst, 2009; Yin, 2003). The findings from this case study are difficult to generalize. The primary reason is the nature of the studied organization. Their situation, regarding several factors for example their competitive situation, affects to what degree the study can be generalized. However, the findings can be helpful for studies of large organizations. In summary, the findings should not be seen as general for every situation and organization but useful for understanding the BI needs of a larger organization. To increase the external validity, more than one organization should have been studied.

### 5.2.4 Reliability

The aim of reliability is ensure that if the study would be reconstructed by another author, the outcome should be equivalent with the previous study (Runeson & Höst, 2009; Yin, 2003). The main data collection method in this case study was semi-structured interviews and observations made during testing of the tools. Therefore, this case study could be difficult to replicate. To increase the reliability, quantitative collection methods could have been used. However, it was not possible due to the time restriction of the thesis and due to the situation with the organization.

### 5.3 Future Works

The area of Business Intelligence remains quite unexplored by the academics however it is widely used by most organizations. In addition, the concept of Self-Service BI should also be researched more as it has the potential to increase the users ability to take advantages of the benefits of BI. It would be really interesting to do a case study in a smaller organization, with no or limited back-end structured for BI, to better understand if a complete implementation of Self-Service BI really is possible in any organization.

In addition, an interesting area for future research would be to compare the long-term results and effects of traditional BI compared to Self-Service BI to conclude if the later truly empowers the users and improves the organization decision making and performance.
Chapter 6

Conclusion

The purpose of this study was to investigate and review how Self-Service BI could be implemented in a large organization, through Power BI. Reviewing the research questions will conclude this thesis:

- How can Power BI be implemented as a complete Business Intelligence solution?
- Which parts of the Business Intelligence architecture are most suitable for implementing Power BI?
- What are the possibilities and challenges of implementing Power BI?

The major finding of the case study showed that an implementation of Power BI was not realistic in an organization as large as the county council of Östergötland. Power BI could not meet the requirements and expectations that LiÖ had on the back-end solution for their BI system. Therefore, discarding Power BI as an independent solution for LiÖ BI system.

However, the results show that changing the front-end solution to Power BI would be beneficial for the organization. Users would then be equipped with an user friendly tool in which they could create reports and dashboards in order to analyse the data and gain new insights. In addition, introducing Power View would probably reduce the routines of workarounds as well, eliminating potentially harmful analyzes. There are also several advantages with employing a collaboration platform such as Office 365. Office 365 would create an environment in which users can share their reports and thereby their insights, resulting in an improved communication, additional knowledge and in the end, an increased ability to make the right decision.
References


43


ing case study research in software engineering. *Empirical software engineering, 14*, 131-164.


Appendix A

Existing front-end tool

Figure A.1: The capture shows number of prescriptions over time, legend by unit. There is no possible way to filter the legend so that it only shows 2 or more units. This is one of the reasons for workarounds.
Figure A.2: The capture shows the result of the number of prescriptions over time for a specific medicine.

Figure A.3: In the capture the user is trying to navigate the list of medicines. However, navigating the list is very demanding as the user have to go through countless sub-lists in order to choose the right medicine.
Appendix B

Reports and dashboards using Power View

Figure B.1: Dashboard: Monitoring volume and net amount development of ATC groups

48
APPENDIX B. REPORTS AND DASHBOARDS USING POWER VIEW

Figure B.2: Report: Analysing procurements

Figure B.3: Report: Volume growth for osteoporosis
Figure B.4: Report: Compliance in procurement of two medicines
Appendix C

Office 365 Power BI

Figure C.1: Power BI site for team. Shows uploaded reports and the questions saved.
Figure C.2: Shows a result from a natural language query and also how the user can modify the results.
På svenska

Detta dokument hålls tillgängligt på Internet – eller dess framtida ersättare – under en längre tid från publiceringsdatum under förutsättning att inga extra-ordinära omständigheter uppstår.

Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickekommersiell forskning och för undervisning. överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns det lösningar av teknisk och administrativ art.

Upphovsmannens ideella rätt innefattar rätt att bli nämnd som upphovsmann i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart.

För ytterligare information om Linköping University Electronic Press se förlagets hemsida http://www.ep.liu.se/

In English

The publishers will keep this document online on the Internet – or its possible replacement – for a considerable time from the date of publication barring exceptional circumstances.

The online availability of the document implies a permanent permission for anyone to read, to download, to print out single copies for your own use and to use it unchanged for any non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional on the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity,
please refer to its WWW home page: http://www.ep.liu.se/

©Thomas Schützler