MLED_BI: A Novel Business Intelligence Design Approach to Support Multilingualism

NEDIM DEDIĆ

A thesis submitted in partial fulfilment of the requirements of Staffordshire University for the degree of Doctor of Philosophy

August 2017

To my son, my wife, my mother, my father ….

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation, respect and sense of gratitude to my principal supervisor Dr Clare Stanier for her guidance, valuable advice and comments, constant support, and encouragement through the whole research process. Without your support and supervision, this work would not be possible. I would also like to thank my second supervisor Dr Russell Campion and Prof Dr Bernadette Sharp for their support, advice, comments and suggestions.

Despite the fact that words fail me to express my appreciation and denote my sincere thanks to my family, I express my deepest thanks to my beloved wife Zenina for her patience, understanding, constant support and encouragement through every stage of the research. I would like to thank to my son Daris for being my biggest motivation and inspiration for this journey.

Many thanks to my colleague Dr. Thomas Thalhammer for initial support and encouragement to start my research study at Staffordshire University. Many thanks to all colleagues as well (Dr. Günter Preuner, Dr Johann Strasser, Markus Rotter, Karl Reitshammer, Tomaž Kaštrun, Stefan Glas, Martin Wurz, and many others) who supported me during my research through various activities.

Finally, I would like to thank all the people who made the realization of this thesis possible through their involvement and support.

ABSTRACT

With emerging markets and expanding international cooperation, there is a requirement to support Business Intelligence (BI) applications in multiple languages, a process which we refer to as Multilingualism (ML). ML in BI is understood in this research as the ability to store descriptive content (such as descriptions of attributes in BI reports) in more than one language at Data Warehousing (DWH) level and to use this information at presentation level to provide reports, queries or dashboards in more than one language.

Design strategies for data warehouses are typically based on the assumption of a single language environment. The motivations for this research are the design and performance challenges encountered when implementing ML in a BI data warehouse environment. These include design issues, slow response times, delays in updating reports and changing languages between reports, the complexity of amending existing reports and the performance overhead. The literature review identified that the underlying cause of these problems is that existing approaches used to enable ML in BI are primarily ad-hoc workarounds which introduce dependency between elements and lead to excessive redundancy. From the literature review, it was concluded that a satisfactory solution to the challenge of ML in BI requires a design approach based on data independence the concept of immunity from changes and that such a solution does not currently exist.

This thesis presents MLED_BI (Multilingual Enabled Design for Business Intelligence). MLED_BI is a novel design approach which supports data independence and immunity from changes in the design of ML data warehouses and BI systems. MLED_BI extends existing data warehouse design approaches by revising the role of the star schema and introducing a ML design layer to support the separation of language elements. This also facilitates ML at presentation level by enabling the use of a ML content management system. Compared to existing workarounds for ML, the MLED_BI design approach has a theoretical underpinning which allows languages to be added, amended and deleted without requiring a redesign of the star schema; provides support for the manipulation of ML content; improves performance and streamlines data warehouse operations such as ETL (Extract, Transform, Load). Minor contributions include the development of a novel BI framework to address the limitations of existing BI frameworks and the development of a tool to evaluate changes to BI reporting solutions.

The MLED_BI design approach was developed based on the literature review and a mixed methods approach was used for validation. Technical elements were validated experimentally using performance metrics while end user acceptance was validated qualitatively with end users and technical users from a number of countries, reflecting the ML basis of the research. MLED_BI requires more resources at design and initial implementation stage than existing ML workarounds but this is outweighed by improved performance and by the much greater flexibility in ML made possible by the data independence approach of MLED_BI. The MLED_BI design approach enhances existing BI design approaches for use in ML environments.

PUBLICATIONS

Dedić N., & Stanier C. (2017). MLED_BI: A New BI Design Approach to Support Multilingualism in Business Intelligence. *TEM Journal*. Vol 6, Issue 4. Pp 771-782. DOI: 10.18421/TEM64-17

(Invited paper) Dedić N., & Stanier C. (2017). Measuring the success of changes to Business Intelligence solutions to improve Business Intelligence reporting. *Journal of Management Analytics*. Vol 4, Issue 2. Pp 130-144. DOI: 10.1080/23270012.2017.1299048

Dedić N., & Stanier C. (2017). Towards Differentiating Business Intelligence, Big Data, Data Analytics and Knowledge Discovery. ERP Future 2016 (Hagenberg, Austria). *Lecture Notes in Business Information Processing*. Springer. Vol 285. Pp 114-122. DOI: 10.1007/978-3-319-58801-8_10

Dedić N., & Stanier C. (2016). Measuring the Success of Changes to Existing Business Intelligence Solutions to Improve Business Intelligence Reporting. CONFENIS 2016 (Vienna, Austria). *Lecture Notes in Business Information Processing*. Springer. Vol 268. Pp 225-236. DOI: 10.1007/978-3-319-49944-4_17

Dedić N., & Stanier C. (2016). An Evaluation of the Challenges of Multilingualism in Data Warehouse Development. In *Proceedings of the 18th International Conference on Enterprise Information Systems* - ICEIS 2016 (Rome, Italy). Vol 1. Pp. 196–206. DOI:10.5220/0005858401960206

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

AA - Additional Attributes

ATS - Additional Table / Schema

BI - Business Intelligence

BICC - Business Intelligence Competence or Excellence Centre

CIF - Corporate Information Factory

CMS - Content Management System

CRM - Customer Relationship Management

CSV - Comma-Separated Values

DM - Data Mart

DQ - Data Quality

DSS - Decision Support Systems

DW - Data Warehouse

DWBSM - Data Warehouse Balanced Scorecard Model

DWD - Data Warehouse Design

DWH - Data Warehousing

ERP - Enterprise Resource Planning ERP

ETL - Extract-Transform-Load

EUCS - End User Computer Satisfaction

HBIF - A Holistic Business Intelligence Framework

HTML - Hypertext Markup Language

IS - Information Systems

IT - Information Technology

LIF - Language Identifier Field

MCMS - Multilingual Content Management System

ML - Multilingualism

MLED_BI - Multilingual Enabled Design for Business Intelligence

OLAP - Online Analytical Processing

OLTP - Online Transactional Processing

PoC - Proof-of-Concept

POS - Point Of Sale

RDBMS - Relational Database Management System

RLD - Reporting Layer Design

SSSD - Sample Source System Database

XML - Extensible Markup Language

Chapter 1: Introduction

1.1. Introduction

This chapter introduces the investigation into enabling support for Multilingualism in Business Intelligence and gives the motivation for the research. The aims and objectives of the research are explained together with the contribution to knowledge. The research philosophy, research design, methods of investigation and ethical issues are discussed and the chapter also outlines the structure of the thesis.

1.2. Research Motivation

With emerging markets and expanding international cooperation, there is a requirement to support Business Intelligence (BI) applications in languages other than English, a process referred to as Multilingualism (ML) (Dedić & Stanier, 2016a). Business users today expect to use software and applications, and to access information in the semantic web, which includes Business Intelligence Reports, in their own language (Gracia *et al.*, 2012; Hau & Aparício, 2008). The traditional dominance of English in computing (Hensch, 2005), sometimes referred to as the “linguistic hegemony” of English on the Web (Fairweather, 2003, p. 517) is giving way what has been described as networked multilingualism and linguistic diversity (Androutsopoulos, 2015). There is increasing recognition of the issues involved in support for user generated multilingual content (Dang *et al.*, 2014). Language barriers have been identified as a particular issue for multinational companies (Harzing *et al.*, 2011) although it has been argued that multilingual approaches in business are still in their infancy (Pierini, 2016). In some European countries where there are several official languages, such as Switzerland (Grin, 1998) and Belgium (Warren & Benbow, 2008), support for ML may be a legal requirement. Thus, organisations in those countries must support multilingualism in order to be able to operate. Business Intelligence is a fast-evolving field, and in addition to traditional activities such as data warehousing and reporting, the new generation of Business Intelligence focuses on data exploration and visualisation (Obeidat *et al.*, 2015; Anadiotis, 2013), which in the context of international Business Intelligence systems increases the demand for Multilingualism. ML is also seen as a data quality (DQ) requirement as the DQ dimensions of interpretability and ease of use require information to be available to users in formats and languages which they can interpret (Wang & Strong, 1996). Using automated translation tools to deliver BI content and BI reports in the local language offers insufficient and unreliable quality of translated content, as it can lead for example to the situation where are two or more different words in the original language have the same translation in the target language. There are also issues with the overhead of translation, particularly for large volumes of data. Access to information in the user's own language is particularly relevant in Business Intelligence where information is used to support decision making. This thesis focuses on the Data Warehousing (DWH) and Reporting components of BI and in the context of this research, Multilingualism in Business Intelligence is understood as the ability to store and manipulate descriptive content, such as descriptions of attributes and hierarchies at DWH level and to use this information at presentation level in more than one language.

The motivation for this research developed from the design and performance problems encountered when implementing ML in a real world commercial BI environment. It was identified empirically and described in detail in sections 2.5. and 2.6. that existing approaches to supporting multilingualism in a BI context created problems for business users, for example, slower information retrieval, delays in updating reports and difficulties in complying with legal requirements to provide data in more than one language. Additional problems resulting from existing solutions for ML in BI include the inability to enable, at reporting level, additional languages, which are not available in source systems, and the complexity of the processes required to change erroneous content in existing BI reports.

At a technical level, current strategies for enabling ML in BI present a number of challenges including the additional complexity of the Extract-Transform-Load (ETL) processes required to support ML, excessive resource consumption, content dependency between systems, and data and process redundancy. Examples of these problems include redundancy of descriptive information stored in dimensional tables, the requirement to iterate the complete ETL process to support small changes in descriptive content in business reports and a requirement to implement a language in full in the source systems to be able to use the language at reporting level, reducing flexibility.

As outlined in the literature in chapter two, section 2.6 existing approaches to enable Multilingualism in Business Intelligence, proposed by Kimball (2001), Imhoff *et al.* (2003), Kimball & Ross (2011), and Corr & Stagnittno (2014), are primarily ad-hoc workarounds that lack a theoretical basis in the data warehouse literature or are vendor specific. However, this literature, and in particular the work of Kimball (2001) and Imhoff e*t al* (2003), demonstrates that while support for multilingualism presents a significant challenge for the data industry, the literature does not sufficiently address the issues or provide a sufficient solution. It was identified that existing ML approaches did not sufficiently support the separation of logical and implementation level elements and that a design approach based around data independence would provide a more optimal solution to the challenge of supporting multilingualism in BI systems. This thesis introduces MLED_BI (Multilingual Enabled Design for Business Intelligence), a novel BI design approach which supports multilingualism in BI.

1.3. Aim and Objectives

The aim of this research is to investigate the issues involved in supporting ML in a BI environment, to develop a new design approach to support the optimal application of ML in a BI environment, to develop an implementation to support validation of the new design approach and to critically evaluate the outcomes and the research process.

The following objectives were identified to achieve the aim:

- To critically review the literature covering
    - Issues involved in ML in BI
    - Current BI and DW theories, tools and techniques and relevant data design concepts such as data independence
    - BI approaches used to support BI in a multilingual context
    - Validation and evaluation of BI systems
- To develop a novel Multilingual Enabled Design solution (MLED_BI) to the problem of supporting multilingualism in BI
- To initially validate that MLED_BI translates into functional implementation by establishing technical feasibility through a proof of concept implementation before considering other issues
- To further validate that MLED_BI translates into full-functional implementation by establishing technical feasibility through a large-scale system that simulates the full real world environment to support comprehensive validation of approach
- To conduct comprehensive validation of MLED_BI design approach by

- Comparison of performance metrics from a full MLED_BI implementation and implementations of existing solutions for ML in BI
- Validation of usability and acceptance with business and technical users

- To critically evaluate the research and the research process.

The literature review identified a number of gaps in the existing literature and in response to this, two further objectives were developed:

1. To develop and validate a novel BI Framework to support the analysis stage of MLED_BI
2. To develop an evaluation tool to provide evaluation criteria to measure the success of changes to existing BI solutions to support overall validation and evaluation of MLED_BI

1.4. Contribution to Knowledge

This research makes several contributions to knowledge. The major contribution to knowledge is MLED_BI, a novel BI design approach to support the optimal application of Multilingualism in the context of support for multiple languages in data warehouses for Business Intelligence.

Minor contributions include:

- A novel holistic Business Intelligence framework (HBIF)
- An evaluation tool which facilitates the measurement of the success of changes to existing Business Intelligence solutions
- A comprehensive review of the design issues relating to multilingualism in data warehouse design. Multilingualism in Business Intelligence is an understudied element and as far as is known, this thesis presents the first comprehensive review of existing approaches to support multilingualism in BI.

1.5. Research Approach

1.5.1. Research Philosophy

The choice of research philosophy is driven by the research questions (Borrego, Douglas & Amelnik, 2009; Onwuegbuzie & Johnson, 2004) and the identification of the research goals (Henze, Shirazi, Schmidt, Pielot & Michahelles, 2013). To critically evaluate research perspectives and philosophies relevant to this research, the concept of the research onion, shown in Figure 1-1, as defined by Saunders, Lewis & Thornill (2012) and refined by Saunders and Tosey (2013) was used. The research onion identifies the different research philosophies and the methods, strategies and techniques associated with them.

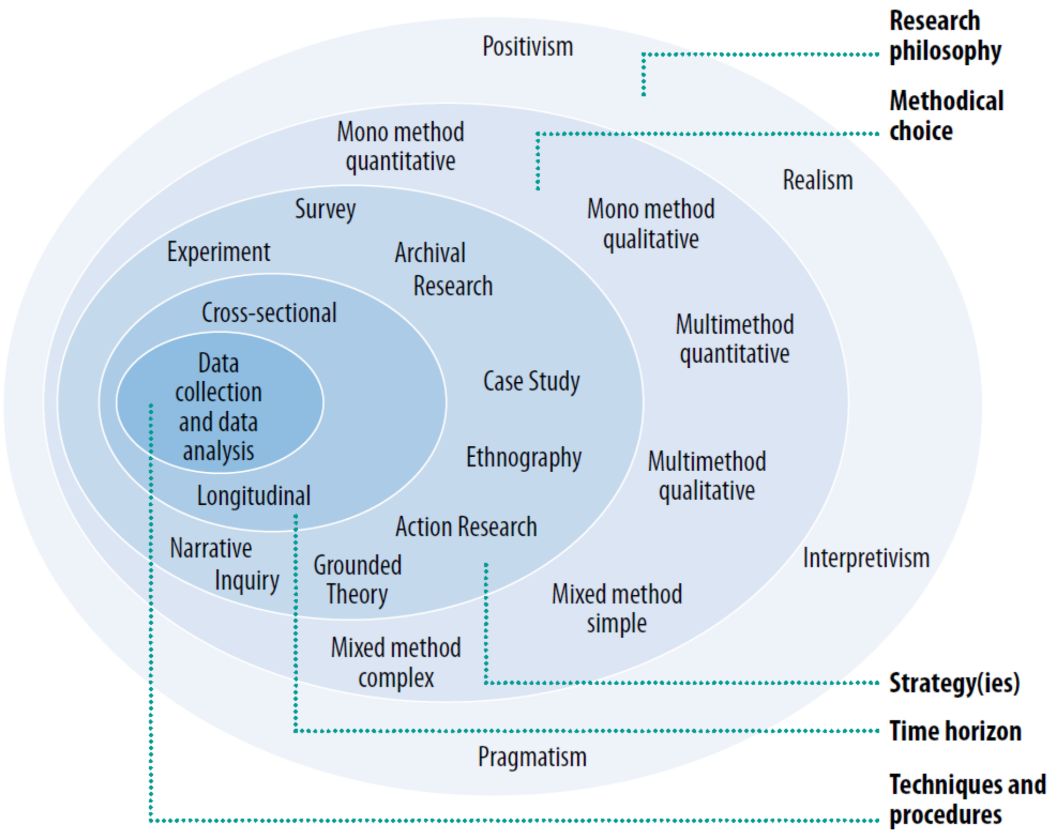


Figure 1-1: The Research Onion (Saunders & Tosey, 2013)

The focus of this research is to address the problem of ML in BI by developing, validating and evaluating a novel design approach. For this type of problem, which is founded on an examination of performance, sequential measurements of the quality attributes of the product or process are recommended (Florak, Park & Carleton, 1997). Experimentation is associated with the positivist approach; data which can be easily compared and evaluated are seen as one of the advantages of positivism (Didau, 2015; Mühl, 2014). Positivism is a research philosophy which regards reality as something which can be understood and ascertained objectively (Paltridge & Phakiti, 2015; Hair, Celsi, Money, Samouel, & Page, 2011; Orlikowski & Baroudi, 1991), supporting the use of metrics. In positivism, it is assumed that reality can be described through research

(Hair *et al.*, 2011) and that there are independent measurable criteria (Orlikowski & Baroudi, 1991). In this research, to support the development of MLED_BI, it was necessary to collect metrics about observable phenomenon such as speed of execution, memory consumption, the number of required processes, and similar measures. This experimental approach reflects the philosophy of positivism (Saunders, Lewis & Thornill, 2012).

Initially it was intended to adopt only a positivist approach. However, acceptance and usability are also key elements in evaluating the effectiveness of the MLED_BI approach and consequently there is an interest in exploring the feelings and attitudes of stakeholders. Thus, this research is also concerned with understanding the views of stakeholders through discussions using semi structured interviews, which according to Saunders and Tosey (2013) reflects the philosophy of interpretivism. Interpretivism is a research philosophy that claims our understanding of reality is socially constructed (Hair *et al.*, 2011), and "emphasizes an understanding of the meaning people attach to their experiences" (Schutt, 2012; Engel & Schutt, 2010, p. 40).

This research adopts the approach used by Niglas (2010) where research philosophies and approaches are seen as a multidimensional set of different continua, including those from positivism and interpretivism. The approach taken in this research is primarily positivist but also uses elements which, as shown in Figure 1-1, are linked to the interpretivist philosophy, particularly in the use of mixed methods.

1.5.2. Research Approach and Methodological Choice

Quantitative approaches to research employ strategies of inquiry, such as experiments, and collection of statistical data on predetermined instruments (Creswell, 2003) and are usually associated with positivism (Saunders, Lewis & Thornill, 2012). A quantitative approach supports the experimental nature of this research but applying a quantitative approach only would have some limitations. The development and evaluation of MLED_BI requires a richer insight into the views and experiences of relevant stakeholders than can be obtained from quantitative data alone. In this context, Creswell (2003) proposes the use of qualitative approaches which are associated with the interpretivist philosophy (Saunders, Lewis & Thornill, 2012). The strengths of qualitative approaches include data obtained from users' experience, in-depth analysis of attitudes and feelings of users, the possibility of revising direction as new findings emerge, and negotiability of findings to another setting. However, this approach can be time consuming, the quality of the data may be dependent on the skills of the researcher and visualisation of findings can be difficult (Anderson, 2010). Qualitative data is less easy to replicate than quantitative data but can add richness to the data obtained through quantitative methods.

This research uses mixed methods, which combines quantitative and qualitative research approaches (Saunders, Lewis & Thornill, 2012; Creswell, 2003; Bryman, 1998). Mixed method research is a subtype of multiple methods research design (Saunders, Lewis & Thornill, 2012). Data collection in a mixed method research project involves acquiring both quantitative data (e.g. statistical data from instruments by measuring) and qualitative data (e.g. interpretive data from interviews) (Creswell, 2003); this is the approach defined as mixed methods simple in the research onion, shown in Figure 1-1 (Saunders & Tosey, 2013). The benefits of a mixed method approach include a more in depth understanding of the problem, complementing the deficiencies and weaknesses of quantitative and qualitative approaches when used individually, and may provide possible explanations of causalities in processes. The motivation for using a mixed methods approach in this research was the need to evaluate MLED_BI both in terms of performance, which could be measured using quantitative data, and user acceptance which requires qualitative data.

1.5.3. Research Strategy

The experimental research strategy was initially seen as sufficient for this research. However, as the research developed, other strategies were also identified as relevant and useful. A proof of concept implementation, used for the initial validation of the technical feasibility of the proposed approach, identified the limitations of using only an experimental research strategy. The goal of this research was not simply to develop a technical solution but also to bring about a positive change in BI and DWH design concepts, thus conforming to software engineering research high level objectives (Runeson, Host, Rainer & Regnell, 2012). Adopting the MLED_BI design solution has implications for business end users as well as for technical users. One of the weaknesses of strategies based on the quantitative approach is lack of understanding of the context or environment in which people operate. It was therefore decided to use qualitative approaches as well as quantitative approaches, to provide a more complete understanding of the proposed solution and to obtain insights from stakeholders regarding the usability and acceptability of MLED_BI when implemented in a BI environment. There was a need to evaluate MLED_BI in a real-life context by obtaining views and individual experiences from relevant stakeholders (key users). The use of experimental data was enriched data gathered through semi-structured interviews and surveys. The research used a cross-sectional time horizon as explained by Saunders & Tosey (2013).

1.5.4. Research Design

The research design was developed based on five steps, adapted from the empirical cycle, proposed by De Groot (1969). The first step was an examination of the capability of existing BI solutions to support ML. The second step was the equivalent of the hypothesis formulation step, the development of a proposed new BI design approach that would support the optimal application of ML in BI. The third step included the definition of appropriate strategies and techniques to confirm or refute the previously defined hypothesis (design approach), which identified experimentation, semi-structured interviews and surveys as appropriate. Consequently, an artefact simulating a real-world environment was developed to enable testing and collection of relevant data as a fourth step. The fifth step covered evaluation and validation; this included the application of strategies and techniques identified in the third step and interpreting the data. Steps were iterated as necessary.

The same approach was used to support the minor contributions of the research, the development of the new holistic framework for BI (HBIF) and the development of an evaluation tool to measure success of changes in BI.

The first step in the development of the HBIF was the investigation of existing BI frameworks and DW design approaches. The second step was the development of the initial version of HBIF based on secondary research and discussions with domain experts. The next step was the initial validation by the means of pilot survey, followed by the iteration of the HBIF based on the feedback, and then a more comprehensive validation by the means of a larger scale survey, which provided the basis for the final version of HBIF.

When developing the evaluation tool, the first step was an investigation of the literature to identify and evaluate currently existing solutions. The next step was the development of the evaluation tool. The third step was the identification of a validation strategy, followed by validation of the tool through a survey. The survey process was iterated and a number of revisions were made to the tool.

The research was carried out in two stages, each stage consisting of four phases. Figure 1-2 shows the stages and phases of the research. The first stage of the research included all the activities related to the initial development of MLED_BI and the initial validation of technical feasibility through a Proof-of-Concept (PoC) artefact. Stage Two was dependent on successful completion of Stage One and included the activities related to the full implementation and comprehensive validation of MLED_BI using a much wider spectrum of measurements than those employed in Stage One. Most of the phases in Stage Two were based on the work already done in phases of Stage One. Stage Two also included the overall evaluation of MLED_BI.

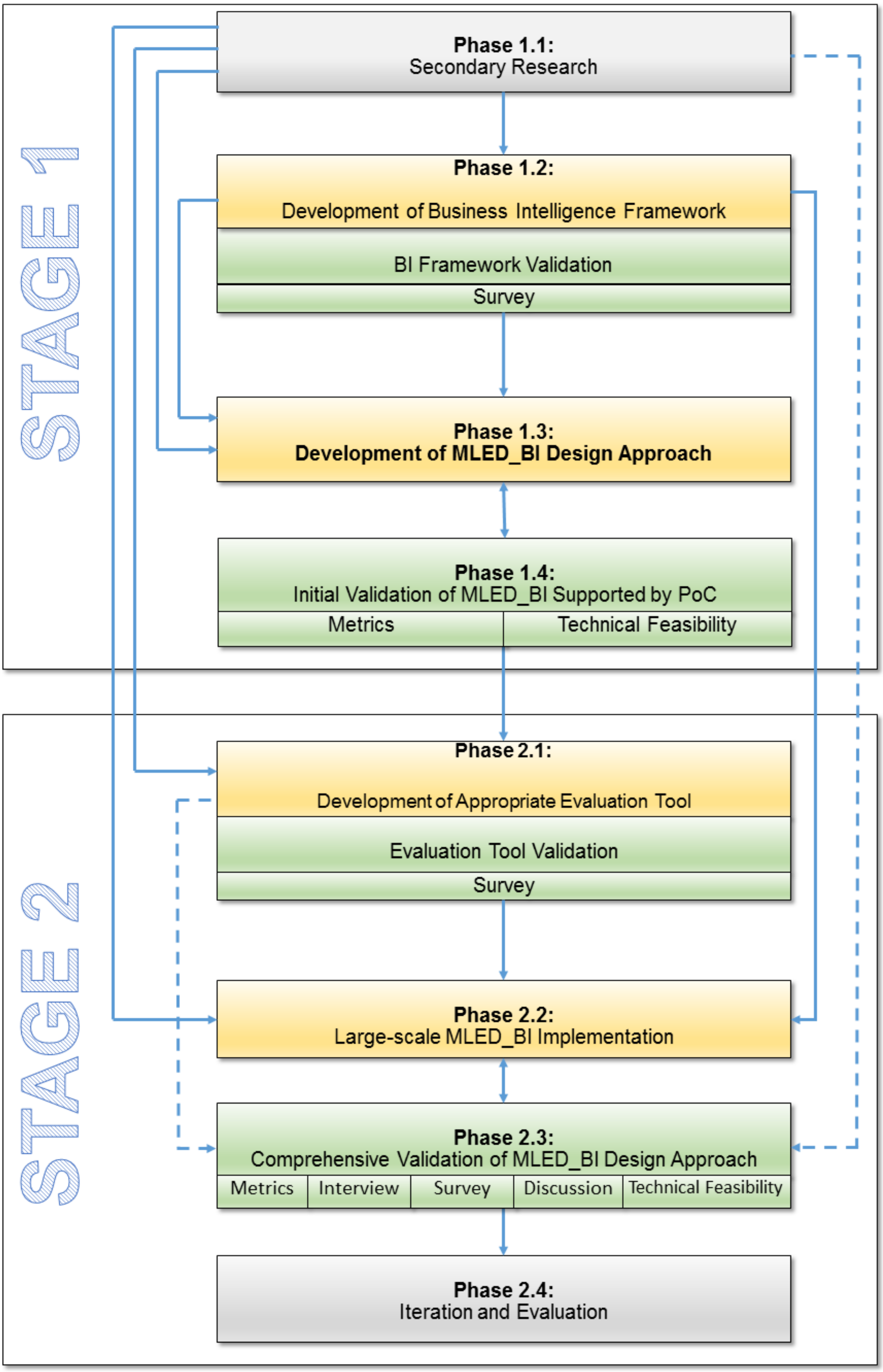


Figure 1-2: Research Phases

1.5.5. Stage one

- Phase 1.1 consisted of the initial literature review. Existing literature was critically reviewed to investigate current BI and DWH theories, tools and techniques and data design concepts and the design approaches currently used to support BI in a multilingual context. This enabled an in depth investigation of the problems and issues associated with the application of ML in BI in a real world environment and established the theoretical basis for the development of MLED_BI.
- Phase 1.2 was an evaluation of existing BI frameworks and DWH approaches in the context of their capability to identify relevant components and aspects when extending or modifying existing BI environments. The examination of BI frameworks was required because a prerequisite for addressing current issues associated with support for ML in BI was to identify the components and aspects of BI systems that are affected by ML, and which might need to be modified. The evaluation identified a gap in the literature as it showed that no current BI framework had the required capabilities. To support the development of MLED_BI, a new holistic BI framework (HBIF) was developed to address those limitations and to provide a clearer understanding of the BI environment. This phase also included the validation of HBIF with domain experts from academia and industry.
- Phase 1.3 was the design and development of the MLED_BI approach, grounded in the theoretical basis developed from the literature review and supported by the novel BI framework (HBIF) developed in phase 1.2.
- Phase 1.4 was the development and evaluation of a proof of concept (PoC) artefact to demonstrate the technical feasibility of the MLED_BI solution.

1.5.6. Stage two

The findings from the PoC artefact were encouraging and provided a basis for further work to validate the proposed MLED_BI design approach in a simulated real world environment, encompassing a wider spectrum of measurement criteria. The environment is referred as 'simulated real world' because although the structure of the data warehouse was based on a realworld data warehouse, data protection requirements meant that the data used was generated and not live customer data and the implementation was limited to the sales perspective and did not include all the data that would be used in a real world DW. Stage Two focused on the work that needed to be completed to comprehensively validate and then evaluate the MLED_BI approach.

- Phase 2.1: The initial literature review had indicated that there was lack of suitable evaluation tools and measures for this type of BI system. In phase 2.1, a more in depth review of evaluation tools and techniques was conducted and it was concluded that there was no existing tool which could satisfactorily be used to provide an evaluation of MLED_BI. An evaluation tool was developed and validated in this phase to support the validation and evaluation of MLED_BI.
- Phase 2.2 was a large-scale implementation of MLED_BI, in an environment designed to simulate a real world environment. This phase included also implementation of the three existing design approaches for enabling implementation of ML in BI for the purposes of collecting metrics for comparison .
- Phase 2.3 covered a comprehensive validation of the MLED_BI design approach using the evaluation tool developed in phase 2.1. This tool covered the use of appropriate metrics to compare MLED_BI with existing solutions, semi-structured interviews and surveys with business users, and discussions with domain experts.
- Phase 2.4 reviewed and revised MLED_BI, following feedback, and included evaluation of the approach and of the research project.

1.5.7. Data Collection Tools and Methods

This research utilised a mixed methods approach and employed a number of tools and methods for primary data collection and analysis. Metrics were used to gather information about the performance of MLED_BI compared to existing approaches. The data harvested was analysed using descriptive statistics. The measures collected were based on the metrics identified through the evaluation tool. Semi-structured interviews were used at a number of stages during the investigation. Exploratory discussions were held with seven domain experts from Germany and Austria as part of the initial development work for the novel Holistic BI Framework (HBIF). As a part of the qualitative validation of MLED_BI, discussions were held with six technical domain experts from Germany and Austria. Semi structured interviews took place with six business users from Austria, Slovenia and Croatia. Technical domain experts, for the purposes of this research, are understood as practitioners with expertise in Business Intelligence, Data Warehousing and Enterprise Reporting; business users are defined as individuals who interact with BI or DWH in the everyday business activities. The outcomes of the interviews and discussions were used for thematic analysis, as discussed in sections 8.3. and 8.4.

Surveys were used at a number of points during the research, to collect views from technical domain experts and business users. A pilot survey was carried out to elicit views from different categories of users about the first version of the HBIF which was developed to support the analysis stage of MLED_BI. The final version of HBIF was validated using an online-questionnaire, which received feedback from 109 BI and DWH domain experts from 25 countries, reflecting the international nature of BI. The same approach was used with the evaluation tool which was developed through a pilot survey of 10 BI domain experts; the final version of the evaluation tool was validated through a survey completed by 30 key users working in the field of BI.

1.5.8. Validation approach

In addition to verifying that the design approach can be translated into a functional artefact that simulates a real world BI environment, the MLED_BI validation process consisted of two phases: quantitative and qualitative validation. The quantitative validation benchmarked MLED_BI with existing BI design approaches by comparing metrics identified as appropriate through the evaluation tool outlined in section 1.3 and described in detail in chapter 6. The qualitative validation was carried out with technical domain experts and business users by the means of semi structured interviews and discussions; users were given the opportunity to compare MLED_BI with existing BI design approaches and then asked to evaluate the strengths and limitations of all the approaches. Other artefacts identified as minor contributions in this research, namely the evaluation tool and the BI framework (HBIF) were validated through use of surveys.

1.6. Ethical Issues

The main ethical issues in the research were around commercial confidentiality and participant consent. Commercial confidentiality was ensured by the use of randomly generated data samples for experimental purposes. It is for this reason that the validation is described as having been carried in a simulated real world environment as discussed in section 1.5.6. The structure of the data warehouse and the data used for validation purposes are based on a real world data warehouse and conform to commercial usage but data protection laws in Europe meant that client data, even anonymised, could not be used for the purposes of the investigation.

All the data acquired for use in the research complied with the Staffordshire University research code of practice. For surveys, personal information that could be used to identify participants was not stored or published, ensuring that individuals are not identifiable. Where appropriate, as for the semi-structured interviews, written permission was obtained from participants but the responses used in the thesis were anonymised. In the context of maintaining privacy, participation in any kind of communication was on a voluntary basis and users were able to withdraw from the interview process or completion of the survey at any stage. As required by professional and research ethics, all personal information obtained during the course of the research is treated as confidential.

1.7. Thesis Structure

This thesis is divided into nine chapters.

- Chapter one: introduces the research, provides the background and motivation for the research, gives an overview of aims and objectives and explains the research approach, including the research philosophy, strategy, design, data collection tools and validation. The ethical issues and contribution to knowledge are discussed and explained.
- Chapter two: critically reviews the issues involved in ML in BI, current BI and DW theories, tools and techniques and BI approaches used to support BI in a multilingual context.
- Chapter three: presents an examination of existing BI frameworks and DWH approaches with a view to using a BI framework to determine the components, which constitute BI and the relationships and dependencies between components to support the development of a design approach for ML. The chapter identifies

the limitations of existing frameworks and describes and justifies the development, evaluation and validation of a new framework, the holistic BI framework (HBIF).

- Chapter four: presents the development of MLED_BI, a novel BI design approach for ML. The chapter discusses the architecture of MLED_BI and differentiation in regard to conventional BI design approaches. Revised concepts of the DWH layer, data mart, and star schema are discussed together with a revised concept of the BI Reporting layer which provides additional possibilities in the MLED_BI environment.
- Chapter five: presents the Proof of Concept (PoC) implementation developed to verify the technical feasibility of the MLED_BI proposed design approach. The implementation approach is explained and the findings are presented.
- Chapter six: discusses the requirement for an evaluation tool to measure the success of changes to a BI reporting environment and gives the justification for developing a new tool. The chapter describes the development and validation of the tool and evaluates the results of the validation of the tool.
- Chapter seven: describes the development of the environment used for the comparative validation and evaluation of MLED_BI design approach. The chapter discusses the implementation of four different BI approaches; three of the approaches are based on existing methods for supporting ML in BI and the fourth approach is based on MLED_BI.
- Chapter eight: presents the validation of the MLED_BI design approach by discussing technical functionalities and user satisfaction aspects. The chapter describes the metrics used and the conclusions drawn from the examination of the metrics and also discusses the qualitative evaluation carried out with technical experts and end users.
- Chapter nine: draws conclusion from the research, discusses and evaluates the outcomes and the research as a whole and includes recommendations for future work.

1.8. Conclusion

This chapter introduced the investigation into support for Multilingualism in Business Intelligence and gave the motivation for the research. The aim and objectives of the research were explained together with the contribution to knowledge. The chapter discussed the research approach, including the research philosophy, the research design and data collection techniques. Ethical issues in the research were discussed and the chapter gave an outline of the structure of the thesis. The following chapter, chapter two, reviews the literature relating to data warehouse design, support for multilingualism in BI and discusses the concept of data independence.

Chapter 2: Literature Review and Theoretical Underpinnings

2.1. Introduction

This research is concerned with the development of a new design approach for Business Intelligence systems to support the optimal application of multilingualism in Business Intelligence. In this context, the first step in the literature review was a critical analysis of Business Intelligence concepts, philosophy, role and trends to identify the problem context. Subsequently, as the research focuses on the multilingual aspect of Business Intelligence, the next step included a critical review of the existing literature with regard to Business Intelligence in an international and multilingual context. The following step was concerned with the evaluation of the underpinning concepts of Business Intelligence. This identified the Data Warehouse as the core element and the heart of the Business Intelligence environment as discussed in this research. This led to an analysis of Data Warehouse design and concepts as the next step in literature review process. This stage included the examination of concepts such as Data Independence and Data Redundancy and the significance of these concepts in the data environment. In keeping with the focus of the research, the Data Warehouse modelling philosophy and the challenges triggered by application of multilingualism in Business Intelligence were identified and analysed. The analysis of Data Warehouse modelling methods, led to the star schema being identified as the most widely used and most relevant modelling element in the data mart context. Following on from this, the next step in the literature review included an analysis of the existing star schema solutions used to support multilingualism in Business Intelligence.

The focus of the investigation is on the design element and the role of data warehouses in storing and retrieving data to support analysis operations, rather than on the nature of the analysis operations. For this reason, data mining or OLAP processes and procedures are not considered except in relation to data storage and retrieval. The chapter defines what is meant by Business Intelligence and by Multilingualism in the context of this research. The underpinning concepts of BI including related elements such as data warehousing, data presentation and visualisation issues, data independence and data redundancy and strategies for DW design and development are discussed. The design approaches currently used to support ML in BI are evaluated and the implications for the performance and management of multilingual BI systems are discussed.

2.2. Business Intelligence

To survive in today's business environment, a company has to continuously improve productivity and efficiency, while management has to make decisions almost immediately to ensure competitiveness (Huff, 2013). Information is used to enable improved decision making and efficiency (Yrjö-Koskinen, 2013; Hannula & Pirttimäki, 2003). This process is supported by activities, processes and applications which are collectively known as Business Intelligence.

2.2.1. Definitions of Business Intelligence

The term Business Intelligence was first used in 1864 to describe the process by which one banker profited by analysing information in regard to his competition (Devens, 1864). In 1958, the term was adopted for Information Technology (IT) purposes by IBM and was defined as "the ability to apprehend the interrelationships of presented facts in such way as to guide action towards a desired goal" (Luhn, 1958, p. 314). Business Intelligence was later used as an umbrella term to describe "concepts and methods to improve business decision making by using fact-based support systems" (Power, 2002, p.128). BI, in the sense in which the term is often understood today, emerged in the 1990s and was initially used to describe activities and tools associated with the reporting, and analysis of data stored in data warehouses (Kimball, Ross, Thornthwaite, Mundy & Bob Becker, 2008).

Business Intelligence is sometimes defined as a managerial philosophy and a tool used to make business decisions more effective by managing and refining business information (Lönnqvist & Pirttimäki, 2006). The term can also be used more narrowly to refer to the relevant information and knowledge which describes an organisation and its business environment, its relationship to customers, competitors and the market, and to other economic issues (Lönnqvist & Pirttimäki, 2006). Brannon (2010) describes Business Intelligence as the successor to decision support systems (DSS) and BI is defined as the group of applications, technologies and methodologies that are used to gather, store, and analyse business data to provide access to meaningful information about organisational performance for decision makers (Jamaludin & Mansor, 2011; Brannon, 2010). An earlier and more formal definition is that BI is "an architecture and a collection of integrated operational as well as decision-support applications and databases that provide the business community with easy access to business data" (Moss & Atre, 2003, p.4.).

Business Intelligence is sometimes defined only as a process, excluding relevant applications from the definition (Dekkers, Versendaal & Batenburg, 2007; Lönnqvist & Pirttimäki, 2006; Golfarelli, Rizzi & Cella, 2004). Golfarelli, Rizzi & Cella (2004) argue that BI is a process, which turns data into information and then explicitly into knowledge, while Dekkers, Versendaal & Batenburg (2007) define BI as a continuous activity of gathering, processing and analysing data. The most detailed definition of BI as a process is given by Lönnqvist & Pirttimäki (2006, p. 32) who define BI as "an organized and systematic process by which organisations acquire, analyse, and disseminate information from both internal and external information sources significant for their business activities and for decision making".

Jourdan, Rainer & Marshall (2008) define Business Intelligence as being both a process and a product at the same time. Turban, Sharda, Delen & King (2010) regard BI as an umbrella term including computer architectures, tools, technologies and techniques which support decision making at the strategic level by exploiting historical data.

2.2.2. Definition of BI used in this thesis

Based on the discussion in section 2.2.1., which demonstrates that BI is a concept which covers many elements, but with a focus on producing information to support decision making, Business Intelligence in this research, is understood as a holistic umbrella term, which includes the concept, strategies, processes, applications, data, products, technologies and technical architectures used to support the collection, analysis, presentation and dissemination of business information (Dedić & Stanier, 2016b). As this understanding of BI includes a recognition of the role of data and the technical elements which contribute to BI systems, it is a helpful definition in the context of this research.

2.2.3. The Role of Business Intelligence

BI helps companies to out-think the competition through better understanding of the customer base (Brannon, 2010), which has been credited with helping to create a closer and stronger relationship with customers, leading to enhanced revenue (Alexander, 2014). BI has a critical role in terms of organisational development as BI can provide competitive advantage in the context of achieving positive information asymmetry, that is, unifying and making useful heterogeneous data (Thamir & Poulis, 2015; Marchand & Raymond, 2008). BI also contributes to the optimisation of business processes and resources, maximizing profits and improving proactive (Olszak & Ziemba, 2006) and strategic decision-making (Herschel & Clements, 2017; Popovič, Turk & Jaklič, 2010). Besides its strategic and tactical role, BI is also used at operational level. For example, Sandu (2008) argues that BI could enable operational staff to spot emerging trends, make faster decisions, take actions and cope with organisational problems as soon as they arise. Some of the areas of application of BI are for example fraud detection, customer retention, risk and customer satisfaction analysis, and actuarial analysis (Srinivasan & Kamalakannan, 2017). Key Performance Indicators (KPI) can be observed allowing immediate action to be taken. As Operational BI evolves into Real-Time BI, decision latency is reduced (Sandu, 2008). According to the American Institute of CPAs (2015), BI helps managers and decision makers to understand their organisations better, to make informed decisions, and to improve operational processes.

BI is used to extract meaningful information and hidden knowledge from data to help business stakeholders in variety of predictions, calculations and analysis (Kurniawan, Gunawan & Kurnia, 2014). Richards, Yeoh, Chong & Popovič (2014) claim that effective BI positively influences planning and analytics effectiveness, and through analytics indirectly positively influences the effectiveness of operational processes. In addition to being seen as the one of the most promising technologies in recent years in terms of value creation from perspective of IT executives (Fink, Yogev & Even, 2016), BI is already a well established approach which is very widely used in commerce and industry (Aufaure, Chiky, Curé, Khrouf & Kepeklian, 2015). For example, in retail, BI is used to support forecasting and marketing and to optimize the supply chain and logistics; BI is used in the insurance industry for claims management and risk analysis; in the banking industry for credit management and customer analysis; in telecommunications for customer profiling, segmentation and demand forecasting; and in manufacturing for logistics, transportation and inventory planning (Olszak & Ziemba, 2006).

2.2.4. Trends in BI

From both the academic and industry perspective, there is evidence of an increasing level of activity in the BI field in the last two decades (Wixom & Goul, 2014; Jourdan, Rainer & Marshall, 2008). As long ago as 2006, an industry based study concluded that it was not satisfactory only to apply conventional development models and system concepts to Business Intelligence (Gluchowski & Kemper, 2006), while a study of US CEOs from the same period found that BI projects were rated as the most important technology projects (Watson & Wixom, 2007). Despite the fact that IT management prioritized BI as one of the top topics (Luftman, Zadeh, Derksen, Santana, Rigoni & Huang, 2012; Pettey & Goasduff, 2011; Luftman & Ben-Zvi, 2010), Wixom, Ariyachandra, Goul, Gray, Kulkarni and Phillips-Wren (2011) identified that academic teaching was not properly aligning with industry practice (Wixom *et al.*, 2011). In 2014, BI technology was identified as the most significant current or near-future IT investment (Kappelman, McLean, Vess & Gerhart, 2014). Three years after the initial 2011 research paper, Wixom *et al.* (2014) found growing interest by academia, students and industry practitioners in the field of Business Intelligence. Conventional BI has focused on activities such as ETL, data warehousing and reporting (Dedić & Stanier, 2016a), but the new generation of BI has an additional focus on data exploration and visualisation (Obeidat, North, M., Richardson, Rattanak & North, S., 2015; Anadiotis, 2013). There is also evidence that the reporting function is moving from static reporting to interactive visualisations and from metrics overview to discovering the causes and effects of the phenomena the metrics express (Anadiotis, 2013). Increasing competitive pressure on existing businesses, new technology, new types of data streams, and new knowledge could be the factors underlying the emergence of new trends in this field, such as faster information delivery known as near real-time BI (Larson & Chang, 2016; Aufaure *et al*, 2015), text analytics (Chaudhuri, Dayal & Narasayya, 2011), self-service BI (Obeidat *et al.*, 2015), and mobile BI (Peters, Işık, Olgerta & Popović, 2016).

2.3. Multilingualism Business Intelligence

2.3.1. Definition of ML

Multilingualism is an individual and social phenomena that requires the acquisition, knowledge and use of several languages by communities or individuals, and usually implies more than two languages (Cenoz, 2009). However, individual and social bilingualism, or the use of two languages, is also considered as multilingualism (Cenoz, 2009). The European Commission defines ML as "the ability of societies, institutions, groups and individuals to engage, on a regular basis, with more than one language in their day-to-day lives" (2008, p.6.).

In the context of this thesis, Multilingualism in Business Intelligence is seen as the ability to store descriptive information at data warehousing level and to use this information at presentation level in the form of reports, queries or dashboards in more than one language (Dedić & Stanier, 2016a). It is the term used to describe the process of providing descriptive content in BI reports in more than one language. In Figure 2-1 the red border provides a visual example of descriptive content in relation to the *Country, Assortment Group* hierarchies and the *Article* attributes as stored in a DW and used for BI reports.

Country	Assortment Group	Article	Gross Profit (EUR)	Gross Profit Plan (EUR)	Gross Profit Difference (%)	Net profit (EUR)	Net Profit Plan (EUR)	Net Profit Difference (%)	Raw Profit (EURI)	Raw Profit Plan (EUR)	Raw Profit Difference (%)
Austria	Fruits and Vegetables	Apples	6033.05	3894.9	154.90%	2413.22	1752.71	137.69%	784.30	506.34	154.90%
Austria	Fruits and Vegetables	Oranges	6465.69	7371.61	87.71%	2586.28	3317.22	77.97%	840.54	958.31	87.71%
Austria	Fruits and Vegetables	Cherries	197.25	5659.29	3.49%	78.90	2546.68	3.10%	25.64	735.71	3.49%
Austria	Fruits and Vegetables	Cranberries	9943.91	304.07	3270.27%	3977.56	136.83	2906.91%	1292.71	39.53	3270.27%
Austria	Fruits and Vegetables	Grapes	4848.84	694.86	697.82%	1939.54	312.69	620.28%	630.35	90.33	697.82%
Austria	Fruits and Vegetables	Grapefruit	4509.29	6820.53	66.11%	1803.72	3069.24	58.77%	586.21	886.67	66.11%
Austria	Fruits and Vegetables	Pears	7445.22	6735.74	110.53%	2978.09	3031.08	98.25%	967.88	875.65	110.53%
Austria	Fruits and Vegetables	Pomegranates	2004.24	835.94	239.76%	801.70	376.17	213.12%	260.55	108.67	239.76%
Austria	Fruits and Vegetables	Raspberries	9693.8	3085	314.22%	3877.52	1388.25	279.31%	1260.19	401.05	314.22%
Austria	Fruits and Vegetables	Strawberries	9346.11	3867.01	241.69%	3738.44	1740.15	214.83%	1214.99	502.71	241.69%
Austria	Fruits and Vegetables	Watermelon	9988.17	4035.47	247.51%	3995.27	1815.96	220.01%	1298.46	524.61	247.51%
Austria	Fruits and Vegetables	Jaboticaba	2090.61	1030.88	202.80%	836.24	463.90	180.27%	271.78	134.01	202.80%
Austria	Fruits and Vegetables	Jackfruit	6053.59	5103.3	118.62%	2421.44	2296.49	105.44%	786.97	663.43	118.62%
Austria	Fruits and Vegetables	Jicama	6324.39	1395.78	453.11%	2529.76	628.10	402.76%	822.17	181.45	453.11%
Austria	Fruits and Vegetables	Jojoba	5456.73	4392.23	124.24%	2182.69	1976.50	110.43%	709.37	570.99	124.24%
Austria	Fruits and Vegetables	Asparagus	258.8	6092.92	4.25%	103.52	2741.81	3.78%	33.64	792.08	4.25%
Austria	Fruits and Vegetables	Atemoya	2969.24	4002.05	74.19%	1187.70	1800.92	65.95%	386.00	520.27	74.19%

Figure 2-1: Example of descriptive content in BI report

The full complexity of supporting ML in BI is visible in Figure 2-2 which shows that every layer of a BI system is involved in providing multilingual capability in Business Intelligence systems.

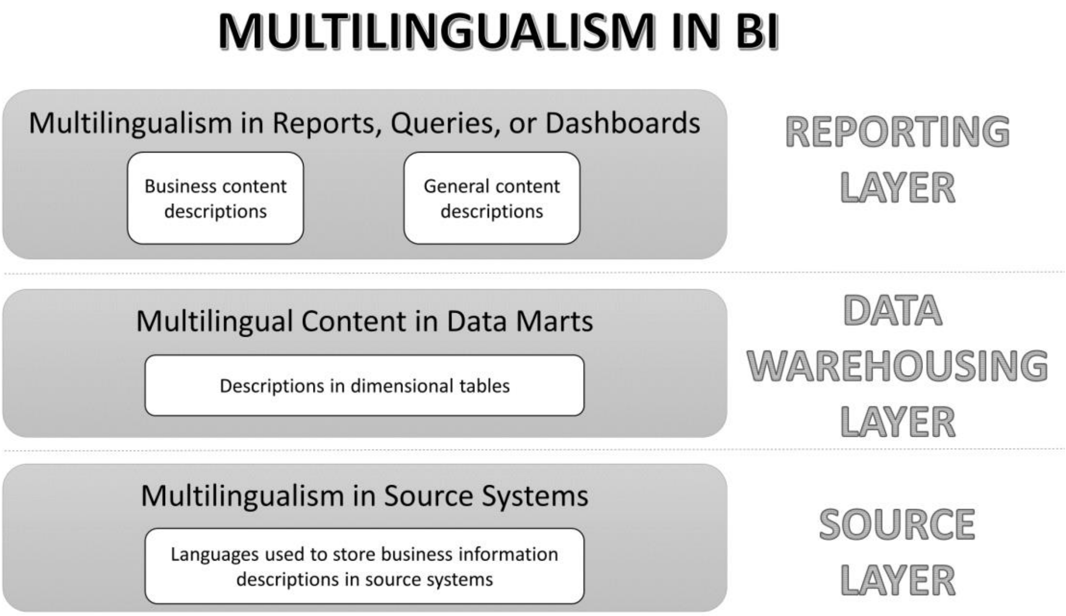


Figure 2-2: The complexity of Multilingualism in Business Intelligence

At the BI source layer, ML encompasses the concept of languages used to store business information descriptions in operational systems; this is conventionally known as master data (Talburt & Zhou, 2015; Kurbel, 2013; Ranier & Cegelski, 2010). In this research the terms business information descriptions and master data are used interchangeably as they represent the same concept. Master data are used to describe the entities, which are independent of and fundamental to the enterprise operations and because they describe things that are critical to organisation operations, such as products, persons, customers, locations, suppliers, or services, they are sometimes seen as "nouns" (Talburt & Zhou, 2015). According to the Ranier & Cegelski (2010), the purpose of master data is to categorize, aggregate, or evaluate transactional data. On the other side, transactional data describes activities and transactions of the business, and are generated by or from operational systems (Ranier & Cegelski, 2010). Transactional data are represented through numbers and are created during business processes, such as the placing an order by customer, or a purchase by supplier, while master data are independent of specific orders (Kurbel, 2013). As they represent descriptive content, the multilingual context of BI relates only to the application of master data, making transactional data, which are represented as numbers, out of the scope of this research.

At the data warehousing layer, ML is concerned with the dimensional modelling of business information descriptions (master data) and the storage of master data in dimensional tables at data warehouse (DW) or data mart (DM) level.

At the reporting layer there are two types of multilingual content possible: a) business information descriptions (master data), and b) general content/report descriptions. The focus here is on business information descriptions (master data). The business information descriptions used at the reporting layer are the same as the business information descriptions used at the source layer and business information descriptions saved in dimensional tables at DWH layer. Presentation data such as content and report descriptions, which provide data about reports but are not related to master data, are not considered as BI content, and are outside of the scope of this research.

2.3.2. Language issues in data interpretation

There is comparatively little discussion of the presentation issues of ML in a BI context but issues associated with multilingualism have long been a concern in the delivery of web content. It was early noted that the use of localized content on websites is regarded positively by native speakers of the languages in which content is presented (Ruffle, 2001). Language, including multilingualism, is a difficult issue in software localization (Collins, 2002). One of the methods used to deliver BI reports to end users is the presentation of information through dynamic and interactive webpages and dashboards intended for mobile use are becoming increasingly important (Firican, 2017). An examination of the BI systems of eight European companies confirmed that all eight companies used a web environment to deliver BI reports to end users. This section discusses web based delivery mechanisms as ML is a recognised issue in web development and web based delivery is widely used in BI systems. However, other methods for the delivery of BI reports are also possible and options include paper based, cloud based files and localized desktop applications.

Web-based business reporting technology was developing very quickly at the end of the 20th century (Lymer, Debreceny, Gray & Rahman, 1999) and many organisations at the international level were considering the impact of the Internet on the delivery and use of business information (Beattie & Pratt, 2003). The role of online reporting became more visible as the Internet developed (Rylander & Provost, 2006). The increased use of a web-based environment as the delivery method for reporting systems reflects the convenience that web systems provide; reports can be delivered via a web browser, there is no requirement to install additional software for every user or administrative access to specific machines; reports are immediately available to prospective users regardless of location and very little training is required to enable users to use reports in a web environment (Maxwell, 2008).

Managing multilingual websites, including those providing BI reports, and interpreting data in various languages presented through World Wide Web (WWW) is a challenging task. Localization of the website and resolving data interpretation language issues in a multilingual web environment requires a strategy that must consider relevant localization and the cultural markers of the intended audience. According to Sun (2001), those markers encompass not only pure content translation issues, but could include elements such as the meaning of colours, metaphors and language grouping conventions. In the context of ML in BI, however, the focus is on language issues in master data. Huang & Tilley (2001) identified two major perspectives to be considered when developing multilingual websites: content and structure. Managing content in multilingual websites faces consistency issues which are time consuming and error prone, while content localization has challenges in terms of the correctness and adequacy of translation (Huang & Tilley, 2001). These issues also apply to multilingual content in BI reports. From the technical point of view, providing web content and interpreting data in many languages has historically been challenging (Starr, 2005) and this is still the case. System support for the rendering and interpretation of data in different languages must be taken into account. Coding standards, such as Unicode, direction and the type of the text to be interpreted, and other language particularities that could raise issues in computing environment must be considered (Starr, 2005). In addition to coding issues, writing systems and text directions, Morgan, Luttrell & Liu (2001) add a number of issues, of which the most relevant in a BI context are average word length and content reproduction.

The expansion of BI systems to enable reporting in different languages is not trivial. In the context of multilingual websites, that deliver BI reports, several factors have been identified when presenting to different range of audience in different countries (Hiller, 2003). Creating and maintaining a web environment in a multilingual perspective creates special challenges, both cultural and technical (Huang & Tilley, 2001). Additional technical issues are identified when translating texts in computer-based environment (Hillier, 2003) and this is relevant for BI. Issues may range from different application environments to different implementation standards. To optimally apply multilingualism to existing BI environment it is necessary to identify the issues of multilingualism in a BI environment.

2.3.3. Regulatory issues around ML

As discussed in section 2.2.4., BI has developed in the last two decades and the expectations of business users have also evolved. In section 1.2., it was noted that multilingualism is a legal requirement in some countries (Europa.eu, 2015; Ulrich, 2006; Tilling, 2003; Grin, 1998;) and many European countries have laws on the official use of their respective languages in public communications (Italian Law No. 482, 1999; Federation Constitution, 1994; Constitution of Croatia, 1990; Spanish Constitution, 1978; Constitution of France, 1958). Where there is a need to support multiple languages, there is an imperative to enable the transfer and processing of textual accessibilities for localization purposes (Vazquez, 2013).

2.3.4. The move towards ML

From the early days of computing, computer technology and software has been associated with development in the English language (Hensch, 2005) and with what was described as the "linguistic hegemony" of English on the Web (Fairweather, 2003, p. 517). In 1990, English was found to be the predominant language for research communication (Rajan & Makani, 2016) (3). However, access to content in the user's own language was early recognised as a data quality issue linked to interpretability and ease of use (Wang & Strong, 1996). As web systems in particular have become more sophisticated, what has been described as networked multilingualism and linguistic diversity (Androutsopoulos, 2015) has developed and there is increasing recognition of the issues involved in support for user generated multilingual content (Dang, Zhang, Hu, Brown, Ku, Wang & Chen, 2014). Business users expect to be able to use software and applications, including BI, in their own language for the purpose of better productivity (Hau & Aparício, 2008) and users generally expect to access information on the semantic web in their own language (Garcia, Montiel-Ponsoda, Cimiano, Gómez-Pérez, Buitelaar & McCrae, 2012; Chung, Zhang, Huang, Wang, Ong & Chen, 2004).

Language barriers have been identified as a particular issue for multinational companies (Harzing *et al.*, 2011) although it has been argued that multilingual approaches to foreign business are still in their infancy (Pierini, 2016).

2.3.5. Requirement to support ML in BI

BI is a fast evolving field (Brichni, Dupuy-Chessa, Gzara, Mandran & Jeannet, 2017; Obeidat *et al.*, 2015) and although traditional BI focused on activities such as DWH and reporting, the new generation of BI has an additional focus on data exploration and visualisation (Obeidat *et al.*, 2015; Anadiotis, 2013), increasing the need for support for multilingulism. Globalization of the market and internationalisation of business through expansion to the other countries increases the demand for ML in BI as the number of languages supported by the businesses increases. This is particularly an issue for companies operating in Europe where there may also be legal requirements. Based on the online profiles of the biggest European companies (Forbes, 2015), most of these companies are international in their nature. Thus, to support operations in the global economy, enterprise database systems need to manage data in multiple languages (Kumran, Chowdary & Haritsa, 2006), and this also applies to DW and BI. As discussed in section 2.4., the seminal work in technical design for Business Intelligence systems took place at the end of the 20th century/beginning of the 21st century and BI design concepts are based on the assumption of a monolinguistic system. BI implementation was typically, although not necessarily, in English, reflecting the early work on BI and the importance of the US economy. The changing attitudes of business users, the importance of emerging and international markets and ever-growing local data warehousing communities are factors that support the application of multilingualism in BI. Multilingualism, however, presents challenges for design and reporting in BI; the following sections discuss concepts and approaches used in BI which are relevant to the use of ML in BI.

2.4. Underpinning concepts for Business Intelligence Design

2.4.1. The Context of BI design

Thamir & Poulis (2015) identified two strategies that underpin the development of BI: Business driven and Technical driven. The Business driven strategy approach is based on the view that the BI environment should be scoped to the business needs, meaning that there is a need for only so much BI as is required to support the actual business. In this approach, the technical aspects of the BI environment are important than business usability and the BI strategy must be aligned with business to better contribute to business effectiveness. This strategy is supported by Kimball *et al.* (2008). An alternative approach is the technical driven strategy, usually described in IT terms, where priorities are owned by the IT side (Thamir & Poulis, 2015). In this approach, greater importance is given to technical standards, conventions and requirements than to business needs. This contributes to IT efficiency by lowering the total costs of BI ownership and by achieving greater efficiencies in IT (Boyer, Frank, Green & Harris, 2010). This strategy, where the IT discipline plays a larger role than the business needs, aligns to the data warehouse approach proposed by Inmon (1992). Kimball *et al.* (2008) and Inmon (1992) are seminal authors in the field of data warehouse development and their work is discussed in detail in section 2.4.2. and 2.4.3.

In addition to the Business and Technical driven strategies identified by Thamir & Poulis (2015), Boyer *et al.* (2010) identify the Organisational and Behavioural strategy, which contributes to business efficiency through higher productivity and faster completion times. This strategy is concerned with understanding business culture, communicating the goals of BI solutions and projects effectively, the challenges of user adoption of technology and obtaining executive support (Boyer *et al.*, 2010) and is linked to the concept of a Business Intelligence Competence or Excellence Centre (BICC). Gartner Research and Oracle define BICC as a group of people, in the form of cross-functional team with specific tasks, roles and responsibilities working together established to promote collaboration and the application of BI conventions and standards across the organisation (Saporito, 2014; Oracle, 2012; Sabherwal & Becerra-Fernandez, 2011; Miller *et al.*, 2006). The BICC approach is seen here as an approach supporting the management and maintenance of BI systems since over time, the long term value of BI investment may begin to decrease due to issues related to data redundancy, quality and availability (O’Neill, 2011). The focus in this research is on the design element of BI systems rather than on the management aspect and the discussion is linked to the Business Driven and Technical Driven strategies associated respectively with Kimball and Inmon.

2.4.2. The concept of the Data Warehouse

Before the introduction of data warehousing (DWH), organisations used decision support systems (DSS) to support fact-based decision-making. In those environments and in the absence of DW architecture, large amounts of data redundancy were required to support functionality and decision-making (Hooda & Gill, 2012). In addition to redundancy, various other problems were connected to early DSS, such as high maintenance costs and lengthy response times. Data warehouses were developed in an attempt to solve these problems and make information more readily available for decision making. Data warehouses began to develop in the late 1980s as a single logical storehouse of all the information used to report on the business (Devlin & Murphy, 1988). The definition has not changed greatly over time although the size and scope of data warehouses has grown dramatically. Porter & Rome (1995) defined the DW as a separate store of data extracted from one or more production systems. Garani & Helmer (2012) define the DW as a repository used to archive and analyse huge amounts of data. A related definition is that a DW is described as a type of database, massive in its nature because it holds very large amounts of detailed and historical information (Breslin, 2004). According to Porter and Rome (1995), the main purpose of the DW is to support decision making in the organisation. Power (2002) extended this to include examples such as support for rapid online queries (reports) and summary data. A DW supports online analytical processing (OLAP), which is differentiated from online transactional processing (OLTP), because the DW works with historical instead with transactional data (Jensen, 2010). The data held in the DW can also be used to support data mining operations which in turn supports reporting.

A DW is seen as a core component of BI systems that use a database concept to store historical business information, later used for reporting and data analysis. However, there are a number of different views as to what constitutes a data warehouse. Inmon (2005) sees the DW as a subject-oriented, integrated, non-volatile, and time-variant collection of data replicated from the source system that could be stored in the DW to support current and future, currently unknown requirements. In the Inmon approach, data marts (DM) could be, but do not have to be, used as additional parts of the DW to serve the analytical needs of one group of the people in the enterprise, for example in finance department. This links back to the Technical Driven strategy discussed in 2.4.1.

The Data Vault model proposed by Linstedt, Graziano & Hultgren (2010) has a similar understanding of the role of the DW. However, the Kimball approach, which is more Business Driven, takes an alternative view which sees DMs as the core concept of the DW (Kimball *et al.*, 2008). In the Kimball approach, a single DM or a cluster of DMs represent the concept of a DW database. The different interpretations of the DW have led to different DW design approaches. However, the majority of data warehouses are ultimately based on either the Inmon or the Kimball approach, meaning that any strategy to support multilingualism in Business Intelligence based on a data warehouse must be capable of being integrated into both the Inmon and the Kimball design approach.

2.4.3. Data Warehouse Design and Development Approaches

There are a number of different possible architectures and design approaches for the development of the DW. Widely used approaches include the top down Corporate Information Factory (CIF) architecture (Inmon, 1992), the bottom up dimensional Data Mart approach (Kimball *et al.*, 2008), and the Data Vault approach (Linstedt *et al.*, 2010).

Inmon (2005) defines a DW as a collection of integrated databases designed to support the DSS function, with an architecture which is, or should be, almost the same as the source system. As shown in Figure 2-3, the Inmon DW also has data marts, or aggregated tables that are used for reporting and querying purposes.

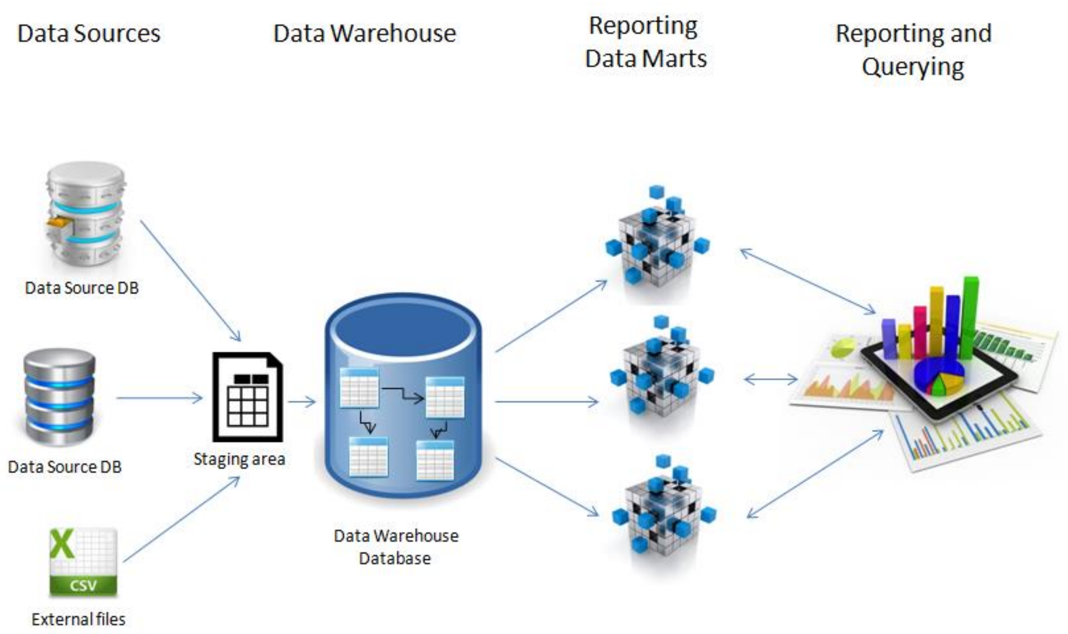


Figure 2-3: Simplified view of Business Intelligence based on Inmon's DW approach

Linstedt *et al.*, (2010) proposes very similar concept for the Data Vault approach. Differentiation is only in the context of modelling and storing information inside the data marts. In the Data Vault approach data is loaded from the source system as is, without any checks or manipulation (Linstedt *et al.*, 2010). The Data Vault approach is characterised by Hubs, Links and Satellites (Jovanović, Subotić & Mrdalj, 2014). Hubs represent source system business keys in the master table, links are associations between hubs with validity periods (from/ to date), and satellites point to the links containing attributes of transaction with the validity period (Orlov, 2014). As the structure of the data is highly normalized (4NF+), this approach to implementing the data warehouse is not adequate for direct reporting and requires additional dimensional data marts to enable reporting or querying (Orlov, 2014). Because of the complexity of the design, which includes very large amounts of historical data and complex joins, a direct query to a DW database based on the Data Vault approach would be highly demanding in time and CPU resources. Thus, DMs, as a form of focused and highly optimized database, are used in the Data Vault approach as an additional stage to support reporting.

A different approach was proposed by Kimball *et al.* (2008) who argue that a DW should be seen as a collection of the data marts which are used for querying and reporting and are connected used conformed dimensions. Conformed dimensions are standardized master data tables that describe the dimension, but which are intended to be used by more than one fact table, and/or by other dimensions for further detailing of existing attributes. Kimball argues that there is no need to replicate all the data from the source system, but only the data needed by the business. The Kimball approach is shown in Figure 2-4.

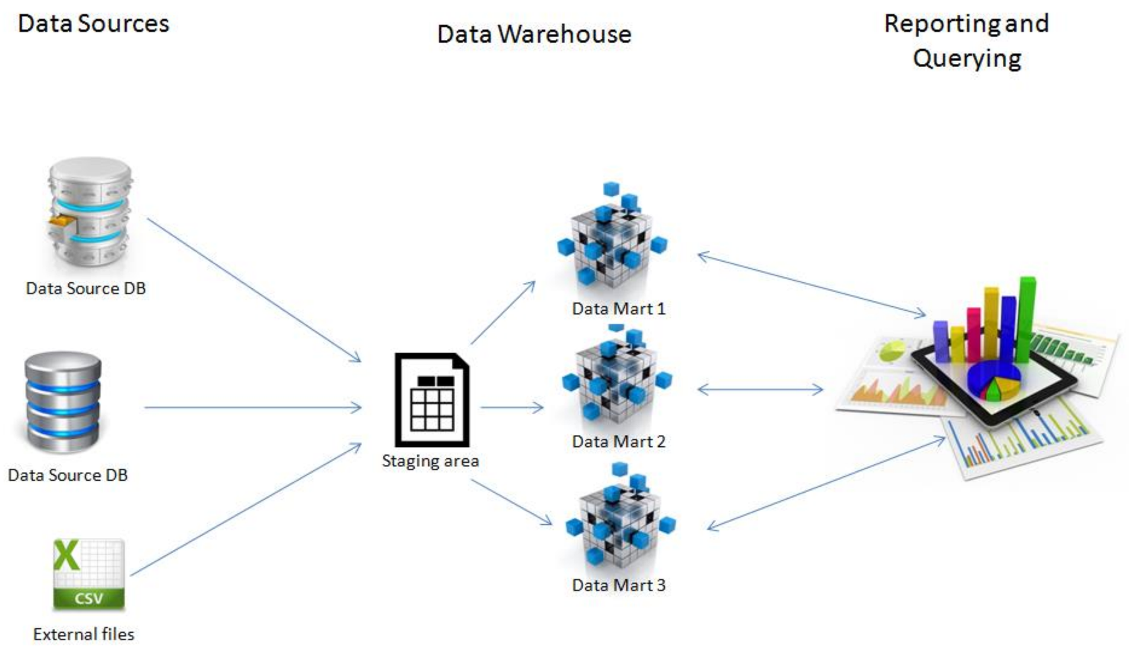


Figure 2-4: Simplified view of Business Intelligence based on Kimball's DM approach

By removing the "Data Warehouse Database" component from Inmon's approach shown in Figure 2-3, Kimball's concept (Figure 2-4) based on conformed dimensions would be produced.

The Inmon and Kimball strategies agree that no change to the data, master (dimensional) or transactional, should be made in the conceptual database/data marts that represents the DW. Any such changes could lead to consistency problems, as discussed further in section 2.6. However, transformation and extractions based on existing data are allowed in the conceptual database/data marts. It is accepted that change to master data and any correction of transactional data must done in source system and then sent to DW, rather than changes being made at DW level. This requirement has implications for ML in BI systems where reporting uses DW data as discussed in 2.6.

2.4.4. Design Concepts in the Data Warehouse

2.4.4.1. The Role of Schema

Relational Database Development relies on the use of schema and schema generation is still an active research area (DiScala & Abadi, 2016; Köhler & Link, 2016). In a relational database, schema are typically based on the ANSI-SPARC schema architecture, leading to the development of three schema, the conceptual, logical, physical schema approach. In conventional relational database design a logical schema can be seen as the technical translation of the database concept that describes the organizational structure of the collection of the related tables (Bouzeghoub & Kedad, 2001; Hainaut, Hick, Henrard, Roland & Englebert, 1997). The physical schema is the structure of the database developed on the basis of previously defined logical schema (Bouzeghoub & Kedad, 2001), and it represents the actual physical data modelling and physical database design (Yeung & Hall, 2007). Data warehouse development typically follows the 3 schema approach (Khouri, Bellatreche, Boukhari & Bouarar, 2012).

2.4.4.2. Data Independence

The ANSI-SPARC architecture has been described as having the goal of "setting a standard for data independence for RDBMS vendors" (Atzeni, Jensen, Orsi, Ram, Tanca & Torlone, 2013, p. 64). In the early days of database development, data independence was seen as one of the key advantages of the relational model as users were able to interact with the information content of the data, without needing to be concerned with how the data was represented (Chamberlin, 1976). Data independence has long been recognized as an important advantage of commercial relational database systems (Odysseas, Tsatalos, Solomon & Ioannidis, 1996; Fegaras & Maier, 1995) and as one of the major benefits of the relational model (Darwen, 2012). An early definition of data independence is that provided by C. J. Date, who described data independence as "the immunity of applications to change in storage structure and access strategy" (Date, 1975). This view of data independence refers to the separation of logical level and physical level implementation elements. A higher level definition of data independence is that data independence describes the immunity of applications at higher levels, such as the external view, to changes at lower levels (Singh, 2011). In this thesis, we adopt the higher level definition, seeing data independence as immunity from changes at lower levels, as this is not restricted to the consideration of storage and access strategies. Data independence is most often discussed in the context of (usually) relational database development; physical data independence describes the immunity of operations from changes at the physical level; for example, adding or deleting a physical level element such as an index, in the context of physical data independence, does not invalidate a query. Logical data independence is the immunity of applications at external view level to changes at logical level (Darwen, 2012). This is seen as a more challenging element; Curino, Difallah, Pavlo & Cudre-Mauroux (2012) linked failure to support logical independence in schema evolution with adverse impact on data and queries, problems of data integrity, expensive application maintenance and application downtime. Blurring the distinction between logical level and physical level design causes issues with maintenance (Atzeni *et al.*, 2013), particularly, we argue when it is necessary to expand a system as when adding additional languages to support multilingualism.

2.4.4.3. Data redundancy

In the seminal paper which introduced the relational model, Codd also introduced the design approach known as normalisation (Codd, 1970) (15) . There is an extensive and still developing literature on normalisation (Köhler & Link, 2016; Date, 2004; Codd, 1970) and in this section we consider only the issue of data redundancy. In database, data redundancy can be defined as the state of data repetition, meaning, where the same datum exists at two or more different places. The prevention of data redundancy is a key aim of normalisation. From a design point of view, data redundancy increases the risk of data anomalies (Codd, 1970) and can lead to reduced performance. Since the data warehouse is conceived as a historical repository of data, update and deletion anomalies related to data redundancy are not a significant consideration although performance considerations still apply. As discussed in in the following section, section 2.4.5, the star schema does not use full normalisation and allows redundancy; data redundancy exists in BI systems, independently of any multilingual issues.

2.4.5. Modelling the Data Warehouse

Much of the literature on the development of data warehouses, and particularly the seminal works by Inmon and Kimball, dates from the end of the 20th century/the first decade of the current century. There is a significant more recent literature on data warehouse development and optimization (Di Tria, Lefons & Tangorra, 2017; Khouri *et al.*, 2017, Cravero & Sepulveda, 2015; Dokeroglu, Sert & Cinar, 2014; Di Sano, 2014;

Graefe, Nica & Stolz, 2013) but there has been comparatively little recent work on DW design and schema development indicating that design concepts are seen as stable. Inmon and Kimball both propose dimensional modelling and the use of data marts for reporting (Orlov, 2014). The Data Vault approach introduced by Linstedt *et al.* (2010) also proposes data marts (using the star or snowflake schema) for reporting. Linstedt *et al.* (2010), Kimball (2008) and Inmon (1995) all recommend the use of the star schema as the most appropriate design strategy for the development of data marts. A survey paper by Sen & Sinha (2005) examined the approaches used by 15 data warehouse vendors and found that 12 of the 15 vendors supported the use of star schema (alone or in combination with others star schema based approaches). The Star schema is considered as the standard modelling paradigm in the DW (Nebot & Berlanga, 2016; Hossain, Islam, Karim & Siddique, 2014; Olaru, 2014; Chen, Zhang, Zou, Ding, Liu & Li, 2006) and as the most suitable basis for dimensional modeling in DW (Hossain *et al.*, 2014).

The star schema is a logical level schema (Nebot & Berlanga, 2016) based on the dimensional modelling concept that supports the storage of historical business information using relational concepts such as the primary key and foreign key without full normalisation (Garani & Helmer, 2012) which is not required given that the data is not expected to change. The star schema is based on a simple dimensional modelling approach (Hossain *et al.*, 2014; Chu, Tseng, Tsai & Luo, 2009; Menzel, Scherer, Schapke & Eisenblätter, 2002), and because of its simplicity, it is optimal for reporting and analytics purposes. This is partly because joining data from the fact table and a dimension table requires only one join while in a fully normalised system, more joins would be required (Garani & Helmer, 2012). As seen in Figure 2-5, the star schema is a collection of dimension tables and one or more fact tables (Cios, Pedrycz, Swiniarski & Kurgan, 2007). The fact table is a central table that contains transactional information and foreign keys to dimensional tables, while dimensional tables contain only master data (Jensen, Pedersen & Thomsen, 2010; Kimball *et al.*, 2008; Cios *et al.*, 2007). Dimensions have a key field and one additional field for every attribute (Kimball *et al.*, 2008; Jensen *et al.*, 2010). In visual model representation, the dimension model resembles a star (Figure 2-5), thus the name (Jensen *et al.*, 2010). The main benefits of the star schema design are ease of understanding and a reduction in the number of joins needed to retrieve the data (Cios *et al.*, 2007).

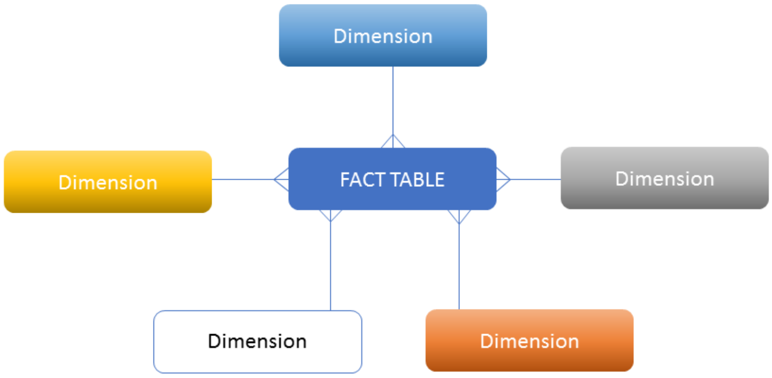


Figure 2-5: Star Schema

In dimension tables, the primary key is used to identify the dimensional value, while hierarchy is defined through attributes. Dimension tables do not conform to the relational model strategy of normalisation and may contain redundancy (Jensen *et al.*, 2010). The fact table, on the other hand, holds the foreign key to dimensional table values and as there is no redundancy it could be considered to be in 3NF (Jensen *et al.*, 2010). In the fact table, all the foreign keys to the dimensional tables build together to make the primary key for the fact table although a surrogate primary key approach with foreign keys is sometimes used.

There are other schemas used for the purpose of dimensional modelling, such as the snowflake (Figure 2-6) or the galaxy schema. The snowflake schema is a refinement of the star schema, where dimensional tables are normalized into a set of smaller tables (Garani &Helmer, 2012), Cios *et al.*, 2007). A collection of several snowflake schemas is known as a galaxy schema where multiple fact tables share same dimensions (Cios, Pedrycz, Winiarski *et al.*, 2007). Cios *et al.* (2007) consider the snowflake and the galaxy schemas as the variations of the star schema, while Inmon (1995), Kimball *et al.* (2008), Linstedt *et al.* (2010), Corr & Stagnittno (2014) and Jensen *et al.* (2010) consider it as a separate dimensional modelling philosophy and not as a variation of the star schema.

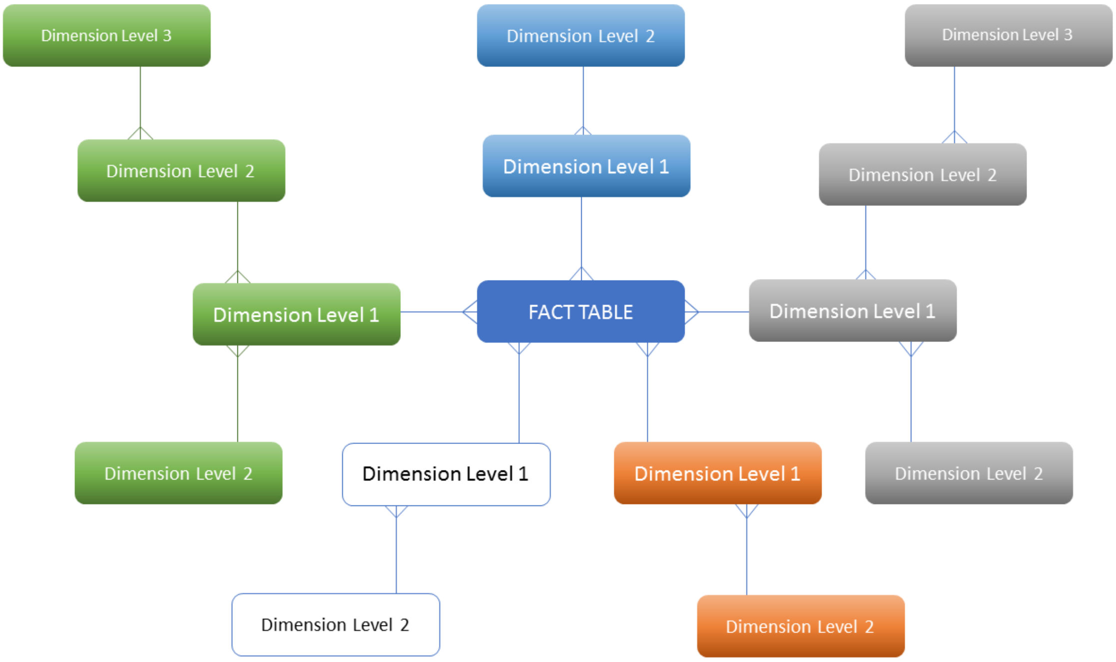


Figure 2-6: Representation of a Snowflake Schema

In a logical data warehouse based on the Inmon approach, Inmon emphasised the need to ensure that non-key data in the physical data warehouse was non redundant (Inmon, 2004).; however, Inmon himself proposes the use of Star Schema based data marts as the most appropriate form of design to support BI reporting (Inmon, 2005). The star schema is recommended as the most appropriate design strategy for the development of data marts (Linstedt *et al.*, 2010; Kimball *et al.*, 2008; Inmon, 1995) and is considered as a general dimensional modelling approach in the data warehouse (Nebot & Berlanga, 2016; Toumi, Moussaoui & Ugur, 2014; Hossain *et al.*, 2014; Olaru, 2014; Lord-Castillo, Mate, Wright, Follett, 2009; Chen *et al.*, 2006;). Thus, this research focuses on the issues of Multilingualism within the star schema.

2.4.6. ETL (Extraction-Transformation-Loading)

ETL is critical in the development of any DW (Jain, Garg & Sharma, 2015; Bansal & Kagemann, 2015; Song, Yan & Yang, 2009) and is discussed here because an understanding of the ETL process is required as part of the discussion of DW concepts (El-Sappagh, Hendawi & El Bastawissy, 2011). Appropriate ETL design is recognized as a key factor in the success of DW (Muñoz, Mazón & Trujillo, 2011) and can be highly complex depending on the needs of the solution (Awad, Abdullah & Ali, 2011).

El-Sappagh *et al.* (2011) describes ETL as a process that enables the extraction of data from data sources, the cleansing, customisation, reformatting, integration, and storage of data into a data warehouse. Extraction is the process of extracting data from source systems; Transformation is a process of cleaning data and transforming it into correct, consistent and compatible formats; Loading is the process that involves propagating the data into a target data mart or data warehouse (Jain *et al.*, 2015; Bansal & Kagemann, 2015). To develop an ETL process, it is necessary to focus on three main areas: the source area, the destination area, and the mapping area (El-Sappagh *et al.*, 2011). ETL supports data extraction, transformation and loading (Bansal & Kagemann, 2015; Song *et al.*, 2009); and is responsible for the integration of heterogeneous data sources within the DW solution (Jain *et al.*, 2015; Munoz *et al.*, 2011).

2.4.7. Data Presentation and Visualisation

In the BI environment, data presentation and visualisation happens at the reporting layer (Figure 2-3, Figure 2-4). The reporting function is one of the most important concepts in BI (Obeidat *et al.*, 2015; Anadiotis, 2013; Chu, 2013; Ranjan, 2009; Baars & Kemper, 2008; Kimball *et al.*, 2008; Watson & Wixom, 2007; Gluchowski & Kempner, 2006; Inmon, 2005; Imhoff, Galemmo & Gaiger, 2003). The reporting layer supports easier decision-making as it provides business users with aggregated and analysed historical data presented at the appropriate level (Mykitychyn, 2007). The reporting layer enables business users to see predefined queries in the form of standard reports, or to define their own reports, colloquially known as ad hoc reports, by using self-service BI capabilities (Rajesh, 2010).

Various applications at the presentation layer, such as reports, dashboards or queries communicate with the DWH layer using query language to retrieve the required information from the DW and to deliver and disseminate the information in a meaningful way. BI reports may take the form of a table or a grid holding mostly aggregated business information retrieved from the DW. However, BI dashboards are now widely used. The BI dashboard is intended to consolidate and present the most important information about the health of the business in an understandable format (Kianoff, 2010). For example, a dashboard may summarise the most important KPIs (key performance indicators) from numerous BI reports on a single page in graphical format. In this research, a BI query is understood as any query used to provide the data for a BI report, whether the report is presented as a standalone or through a BI dashboard. A BI query may be a simple code based query or a query developed by implementing complex objects.

2.5. The challenges presented by ML in BI

The next section examines three solutions which have been developed to enable ML in BI. We argue that these approaches are implementation fixes rather than comprehensive solutions based on a theoretical underpinning and are better understood as workarounds than as formal design approaches. The approaches also have a number of limitations and weaknesses. The first approach discussed requires including additional attributes in the dimension tables (Kimball & Ross 2011; Imhoff *et al.,* 2003); the second extends the primary key to include a language identifier (Imhoff *et al.* 2003); the third requires additional dimension tables/schema (Corr & Stagnittno, 2014; Imhoff *et al.*, 2003; Kimball, 2001). All these solutions, as discussed below lead to changes in the star schema and introduce problems such as extreme data redundancies leading to performance issues, and implementation and maintenance difficulties. The discussion uses an example scenario based on a Product dimension.

2.6. Existing Design Solutions to Support ML in BI

2.6.1. Additional Attributes

One approach to supporting ML, derived from Kimball's proposal for delivering country-specific calendars (Kimball & Ross, 2011), recommends that where there are new values for the dimension tables in star schema, new attributes should be added to dimensional tables. This method is also proposed by Imhoff *et al.* (2003) as a solution for simultaneous bilingual reporting. Imhoff *et al.* (2003) state that if we need to provide the ability to report in two or more languages within the same query, we need to store the data in multiple languages within the same row. When implementing dimensions using this method, attributes should be descriptive, added in the form of textual labels that consist of full words, without missing values, have discrete values and be quality assured (Kimball *et al.* 2008). This is illustrated by the simple Product dimension shown Table 2-1; the example includes data values to better illustrate the problem. The Product dimension attributes (Description, Code, Category and Subcategory) used in this example are textual fields and in a monolingual environment, a conventional star schema approach would support the development of the system.

Table 2-1: Simple Product dimension.

Key	Description	Code	Category	Subcategory	From_Date	To_Date
123	Apples	FA	Fruits	Fruits	01.01.2014	31012014
124	Beer	DB	Drinks	Alcoholic	01.01.2014	31012014

If the additional attributes approach is used to extend Table 2-1 to support multilingualism, the limitation would be extremely large dimension tables. For example, if there are ten descriptive attributes for the Product dimension, with five languages, there would be an additional forty columns. To demonstrate the problem, the product dimension table (Table 2-1) is converted to a logical view (Table 2-2). The sample Product dimension, based on Table 2-2, which includes the German, Italian and Bosnian languages in addition to English would look like Table 2-3.

Table 2-2: Logical view of the Product dimension.

Key (Primary Key)
Description
Code
Category
Subcategory
From_Date
To_Date

Table 2-3: Product dimension in English, German, Italian and Bosnian language.

Key (Primary Key)
Description
Code
Category
Subcategory
From_Date
To_Date
Description_DE
Code_DE
Category_DE
Subcategory_DE
Description_IT
Code_IT
Category_IT
Subcategory_IT
Description_BA
Code_BA
Category_BA
Subcategory_BA

In this attribute based approach, new attribute columns for, in this example, the German, Italian and Bosnian languages are added for every possible textual description. In Table 2-3, this is shown with the suffix **_DE**, **_IT** and **_BA.** This simplified example does not fully convey the scale of the problem. In implementation practice, the Product dimension might contain more than 20 textual attributes and the redundancy problem would be replicated in all dimension tables. A real-world example of a Product dimension would include descriptive attributes (*master data*) to be used as reporting aggregates; as an example, 15 typical descriptive attributes derived from an examination of an actual Product dimension are given here: *description, category, subcategory, assortment, assortment area, buying department, brand, brand origin, country, international categorization, product level, season information, product state, class* and *type*.

As they require large amounts of maintenance time and CPU (Poolet, 2008), large and wide dimension tables can be problematic, especially for rapidly changing dimensions such as a Customer dimension (Ponniah, 2004). Rapidly changing dimensions are those dimensions where *master data* (attribute or hierarchical values) change frequently (Boakye, 2012). To illustrate this, consider a Customer dimension with several million rows of data intended to be used in five languages. In this example, the Customer dimension has three descriptive attributes in all five languages. These categories are intended to be updated on a daily basis. This and similar scenarios creates system overhead on a daily basis. In addition, wide dimension tables require duplicate storage for *master data* and make ETL transformation complex as the language-based columns must be taken into account. More complex query statements are required with different language-based columns to change the language of data previews at the semantic level (reports, queries or dashboards). Moreover, queries that return data sets must be re-executed in the required language. There are other external, but related problems caused by using this approach. For example, consider the challenge of updating or changing *master data* that serves as the hierarchical attributes used as the basis for tables containing aggregated data. As an illustration, suppose a specific group of products change their category from non-alcoholic drinks to energy drinks, affecting also subcategories. It is necessary to update the dimension table to change the descriptive records for every language and also to re-aggregate the data in tables holding aggregated data. In this scenario, it would be necessary to delete all data in tables that hold aggregate data by category and re-aggregate. The process of re-aggregation could take several days if there are billions of records in the fact tables, which is not unusual; Wallmart.com sells more than 4,000,000 different products and Amazon.com more than 350,000,000 (Scrapehero.com, 2015). The situation is more critical with wide dimension tables that represent rapidly changing dimensions. The overhead would increase, as the company needs to store more languages meaning that this solution will present increasing problems.

2.6.2. Extending the Primary Key with Langue Identifiers

This approach to support ML in BI, discussed by Imhoff *et al.* (2003), proposes extending the primary key to include a language identifier. As shown in Table 2-4, the limitation in this case is duplication of the records with every new language. With five languages for the product dimension, which for example holds one million data elements, there would be five million records.

Table 2-4: Product dimension with extended primary key.

Key	Lang	Description	Code	Category	Subcategory	From Date	To Date
123	EN	Apples	FA	Fruits vegetables	Fruits	01.01.2014	31.01.2014
124	EN	Beer	DB	Drinks	Alcoholic	01.01.2014	31.01.2014
123	DE	Äpfel	FA	Obst und Gemüse	Obst	01.01.2014	31.01.2014
124	DE	Bier	DB	Getränke	Alcoholisch	01.01.2014	31.01.2014
123	IT	Mele	FA	Frutta e Verdura	Frutta	01.01.2014	31.01.2014
124	IT	Birra	DB	Beve	Alcolico	01.01.2014	31.01.2014
123	SI	Jabloka	FA	Sadje in Zelenjava	Sadje	01.01.2014	31.01.2014
124	SI	Pivo	DB	Pijače	Alkoholna	01.01.2014	31.01.2014
123	BA	Jabuka	FA	Voće i povrće	Voće	01.01.2014	31.01.2014
124	BA	Pivo	DB	Pića	Alkoholna	01.01.2014	31.01.2014

Larger dimension tables slow the process of query execution and make it harder to manage updates according to the rules of slowly changing dimensions. Slowly changing dimensions are dimensions whose attribute or hierarchical values change over time, but unlike rapidly changing dimensions, values are changed unpredictably and less frequently (Kimball *et al.*, 2008). The language identifier method is also problematic for rapidly changing dimensions (Ponniah, 2004), and as with the additional attributes approach, makes heavy increased demands in terms of maintenance time and CPU (Poolet, 2008). From a memory management perspective this method is less efficient than the additional attributes approach discussed in 2.5.1. as it doubles the storage requirements with every additional language. This method also suffers from the semantic layer problems previously discussed: to change the language of data preview at the semantic layer (reports, dashboards), query statements that return data sets must be re-executed. This method, unlike the additional attributes approach, does not lead to more complex ETL transformations and query statements. However, it produces similar problems in regard to rapidly changing dimensions and changing the structure of externally aggregated tables. For companies using several languages and holding millions of records in their dimensions, re-executing queries and re-aggregating data according to a specific language can be time, memory and CPU demanding. This impacts on the delivery of services to the end user.

2.6.3. Additional Tables / Schemas

A third method discussed by Kimball (2001), Imhoff *et al.* (2003) and Corr & Stagnittno (2014), proposes implementing one fact table and multiple dimensional tables. Different languages are saved in different database schema and/or in different tables. The approach is illustrated in Figure 2-5. For example, for five different languages, five product dimension tables would be implemented, one for every language. For the same example, if there are one hundred initial dimensions in the data warehouse, five hundred dimension tables would be required to satisfy the ML requirements for five languages.

This approach to supporting ML has numerous limitations. Since additional tables and possibly additional schemas are needed in the data warehouse, this approach makes ETL processes more complex as the language-based tables must be planned for. It requires additional transformations to every table for every additional language. The data to be used for aggregation and reporting is doubled and so is the metadata for tables and schemas. This approach requires more complex query statements than the two previous approaches, and changing the language of data preview at the semantic level requires the query to be re-executed.

Schema: BA

Key	Description	Code	Category	Subcategory	From_Date	To_Date
123	Jabuke	FA	Voće i porvće	Voće	01.01.2014	31012014
124	Pivo	DB	Pića	Alkoholna	01.01.2014	31012014

Schema: DE

Key	Description	Code	Category	Subcategory	From_Date	To_Date
123	Äpfel	FA	Obst und Gemüse	Obst	01.01.2014	31012014
124	Bier	DB	Getränke	Getränke	01.01.2014	31012014

Schema: IT

Key	Description	Code	Category	Subcategory	From_Date	To_Date
123	Mele	FA	Frutta e Verdura	Frutta	01.01.2014	31012014
124	Birra	DB	Beve	Alcolico	01.01.2014	31012014

Schema: EN

Key	Description	Code	Category	Subcategory	From_Date	To_Date
123	Apples	FA	Fruits vegetables	Fruits	01.01.2014	31012014
124	Beer	DB	Drinks	Alcoholic	01.01.2014	31012014

Schema: SI

Key	Description	Code	Category	Subcategory	From_Date	To_Date
123	Jabloka	FA	Sadje in Zelenjava	Sadje	01.01.2014	31012014
124	Pivo	DB	Pića	Alkoholna	01.01.2014	31012014

Fact Table

Fact_ID	Product_ID	Time_ID	Store_ID	Cashier_ID	Amount	Tax
0101	123	01	5	5001	1	7,00
0102	124	02	5	5002	1	23,00

Figure 2-7: Schema and multiple dimensional tables approach

Changing any descriptive data in dimensions requires re-aggregation of relevant tables holding aggregated data, which can be critical considering the ETL and query complexity of this method. If one part of the business (country), for example, changes the ID for a specific dimension value, this could lead to consistency problems. Having different IDs for the same data category in different languages causes significant issues with consolidated reporting for that aspect at the enterprise level. Other subtle problems that might arise when using this method as discussed by Kimball (2001) is the possibility of translating two distinct attributes as the same word in a new language causing ETL and reporting problems. To overcome issues in a multilingual context, this method requires additional programming, or the application of additional or surrogate keys as actual keys in fact table.

2.6.4. Vendor Specific Method: SAP Extended Star Schema

A review of the current data warehouse and BI software market found that the biggest vendors, such as Oracle, IBM and Microsoft, support one or more of the three methods discussed above. However, one of the biggest BI vendors, SAP proposes a SAP specific solution for ML, using the concept of an extended star schema, which also includes language as part of the key. In this method the dimensions and the fact table are linked to one another using abstract identification numbers (dimension IDs), which are contained in the key part of the respective database table (SAP, 2015). The representation of dimensions has similarities to the star scheme but is not represented in the same way. Dimensions are not represented as one table with redundant data as in classical star schema. In this case, one dimension can be seen more as an abstract idea. Values from the tables that hold information about a specific dimension attribute text or value are mapped to an abstract dimension key. Figure 2-8 shows an example of the "Product" dimension where we have a value which maps product number to the product "explanatory" tables.

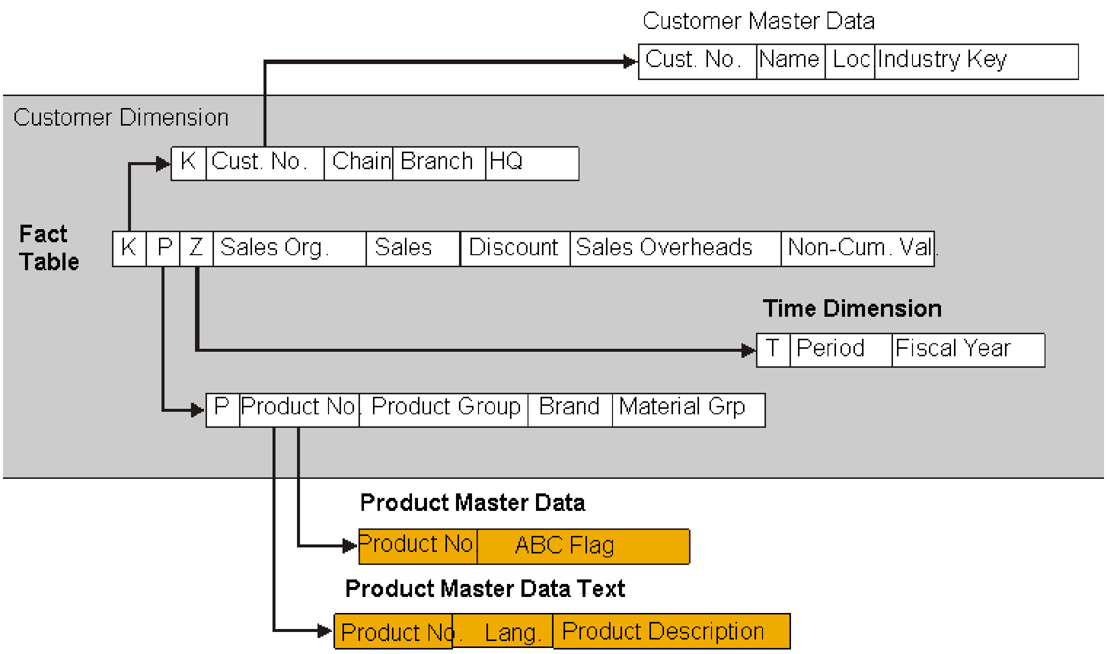


Figure 2-8: SAP BW extended star schema (Source: SAP, 2015)

The information about the product and its language dependant text are stored in "Product Master Data Text" table and follow the approach of including language as a part of the key. This is an implementation driven method which is only supported by SAP BW. This means it cannot be seen as a general design solution as it is a vendor specific proprietary solution, which relies on complex joins to retrieve content for reporting purposes.

2.6.5. Evaluation of existing ML solutions

As the discussion illustrates, although BI and DW concepts are well understood and extensively discussed in the literature, limited attention has been given to the problem of support for ML in BI. There is a lack of experimental data to demonstrate the effectiveness of the solutions currently proposed to address the challenges of ML. The literature did not provide examples of experimental testing or evaluation or comparison of the different approaches. Corr & Stagnitio, (2012), Kimball & Ross (2011), Imhoff *et al.* (2003), and Kimball (2001) present their proposals to overcome the issues of multilingualism or multinational data in BI but do not provide supporting evidence. For example, there were no comparisons between solutions, no performance metrics and no discussion of the possible effects of the proposed solutions in the future; the relationship to IT architectures and the fit to an existing BI environment and architectures were not analysed or evaluated. There was limited technical information regarding physical implementation aspects. Kimball (2001) provides a fuller technical description of his approach to handling multilingual content, and uses this to support his additional tables/schema approach however this is supported only with hypothesizing about possible effects and consequences. No real-life experimentation or testing is done. This is significant because as the examination of the solutions demonstrates, implementing ML using existing approaches creates performance and management issues. Extreme data redundancies, sluggishness, slow execution of reports and queries, implementation challenges and difficulties in maintenance are only some of the issues arising from the existing solutions. One limitation common to all three approaches is that physical implementation elements are introduced into the logical level schema. Although the star scheme remains a logical level element, existing solutions for multilingualism in data warehousing mean that the size of dimension tables is increased, leading to performance issues. Language elements are built into dimension tables, meaning that changes have to be propagated throughout the system. This in turn means that elements and processes in the BI system are not immune from changes at lower levels. There is a further disadvantage that as the business environment changes, for example as more languages are introduced, it is necessary to amend the logical level design.

Although the accepted design approach for DW development, and regarded as a good fit for business requirements (Purba,1999), the traditional understanding of the star schema presents issues when handling multilingual BI systems. The Star Schema has historically been designed to support a monolinguistic environment in which, for example, it is acceptable to store descriptive content at logical level since in a monolinguistic environment, attributes in dimensions will have only one occurrence; ‘category’ for example, as a column that represents attribute of ‘product’ dimension, will only be implemented once. The significance of this for ML is that although redundancy is accepted in a design based on the star schema, the amount of redundancy in a monolinguistic environment will be limited. In a multilingual environment, adopting any of the existing solutions for ML in BI, as discussed above, requires additional elements or tables or additional columns in dimensional tables in the Star Schema, creating greater redundancy. Implementation issues, such as the need to support descriptive content in more than one language, become part of the logical level design. Change immunity is lost since in a DWH BI context, for example, enabling new languages in the star schema in existing BI environments requires modifications at physical, conceptual and application level. As a further example relating to the provision of an optimal service to the end users, changing even the smallest error in descriptions requires iteration of the whole data load (ETL) process from source systems to reporting data marts. The primary weakness of the star schema in the context of existing solutions to support ML in BI is the introduction of implementation considerations into the logical design of dimensional tables. This increases the coupling between elements and this in turn raises real world challenges in terms of performance and maintenance. The issues are not dissimilar to those originally identified by Chamberlin (1976) with respect to data independence.

The solutions to support multilingualism in BI discussed in this chapter are ad-hoc workarounds without an underpinning theoretical basis in the context of BI/DW design or are vendor specific. The additional attribute approach, the extension of the PK approach and the additional tables/schema approach all present performance, management and extensibility issues and do not provide optimal support for ML in a BI context. One reason for the use of ad hoc solutions may be that BI is resource heavy and large multinational companies will typically already have some form of BI infrastructure in place. Any solution for ML in BI will therefore need to be compatible with existing BI frameworks and structures and should be evidence based, supported by experimental data. The proposed solution should be grounded in the theory of DW development to avoid the limitations of the ad hoc solutions discussed in this chapter and should be generic in nature, not limited to a vendor specific solution and capable of supporting both the Inmon and the Kimball approach to DW development.

2.7. Conclusion

This chapter defined BI and multilingualism as it is approached in this research and critically reviewed the underpinning concepts for BI, which included consideration of BI strategies, DW concepts, strategies for DW design and development, ETL processes, modelling in DW, and data and visualisation issues. The challenges presented by ML in BI were discussed and the strengths and limitations of existing approaches to the implementation of ML were reviewed. The discussion showed that existing solutions have serious limitations and a more efficient solution is required to provide optimal support for the application of ML in BI and DW. To address the issues associated with support for ML in BI, the following chapter, chapter three, examines existing BI frameworks with the aim of using a BI framework to determine the components which constitute a BI system and the relationships and dependencies between components, in order to support the identification of the elements of BI systems that are affected by the implementation of ML.

Chapter 3: A Holistic Framework for Business Intelligence

3.1. Introduction

This chapter discusses the development, validation and evaluation of a new Business Intelligence framework, the Holistic Business Intelligence Framework (HBIF). The HBIF is one of the minor contributions to knowledge of this thesis. The analysis stage of the development of MLED_BI required the identification of the components of a Business Intelligence system that would be affected by an extension of the system to support Multilingualism. As discussed in this chapter, existing Business Intelligence Frameworks and Data Warehousing approaches were analysed to determine their capability to identify and communicate the aspects and components which would be affected when extending or modifying an existing Business Intelligence environment to support Multilingualism. The evaluation of existing Business Intelligence Frameworks revealed that no existing framework has the required capabilities. For this reason, the Holistic Business Intelligence Framework was developed to address the limitations of the existing frameworks and to provide a clearer understanding of the Business Intelligence environment. The framework presented in this chapter is described as holistic; in the context of Business Intelligence Frameworks and this research, the term "holistic" is understood as describing a framework which represents all the core components of the Business Intelligence environment that might be affected by changes to components and shows the interactions between components. In addition to addressing the limitations of existing Frameworks and providing support for the development of MLED_BI, the Holistic Business Intelligence Framework developed in this chapter is generalisable and, as already noted, represents one of the minor contributions of the thesis.

3.2. Existing Business Intelligence Frameworks

3.2.1. The Role of Business Intelligence Frameworks

As discussed Chapter 2, section 2.2., Business Intelligence is understood as an umbrella term, which includes the strategies, processes, applications, data, products, technologies and technical architectures used to support the collection, analysis, presentation and dissemination of business information (Dedić & Stanier, 2016b). Because of the complexity and range of Business Intelligence, adapting or extending specific components in a Business Intelligence environment is a challenging task. Changing content requirements at the presentational level requires modification and alteration of the relevant Business Intelligence components through all the data journey processes, from extraction to presentation. For example, extending an existing Business Intelligence report to add a new key figure, descriptive characteristic, or to enable a new language may require modifications to data sources, data warehouse and data mart design, adaptation of Extraction-Transformation-Loading processes and modification of existing queries and reports. Business Intelligence Frameworks can be used to support the identification of components, and elements that need to be modified or extended to support changes to the Business Intelligence system. There is also a need to identify relationships and dependencies between the different elements of the Business Intelligence system to ensure that changes to one element do not have unintended consequences for other elements. In the context of this research, a prerequisite for addressing the issues associated with support for multilingualism was to identify those elements and relationships which would be affected by a design solution for multilingualism.

3.2.2. Review of Existing Business Intelligence Frameworks

The literature review identified 12 existing Business Intelligence Frameworks and Data Warehousing approaches and a detailed evaluation of each of the Frameworks is presented in APPENDIX A. The existing Frameworks were analysed with regard to their capability to identify and communicate aspects and components of Business Intelligence systems which would be relevant when extending or modifying an existing Business Intelligence environment. Special attention was given to identifying from the frameworks which components of the Business Intelligence environment would be relevant to multilingualism in Business Intelligence. From the evalution of the existing fameworks in the literature, five perspectives *concepts*, *users*, *software (applications)*, *data types* and *hardware* were identified as the core components of the Business Intelligence environment that might be affected by changes to Business Intelligence processes, such as, for example, the inclusion of multililingualism. The development of the perspectives is discussed further in section 3.3.1. In the proposed Holistic Business Intelligence Framework, the term "*concept*" refers to the grouping of Business Intelligence components or ideas with similar purpose into appropriate clusters; for example, elements related to data sources are grouped together. The *Concept* perspective is a term also used by Inmon (2005), Kimball *et al.* (2008) and in the Data Vault approach (Linstedt *et al.*, 2010). The Holistic Business Intelligence Framework provides a visual representation of the elements that constitute the Business Intelligence environment. *Users* refers to the different types of users of Business Intelligence systems, *applications* refers to the software applications which operate on the data, *types of data* refers to the different kinds of data present in the Business Intelligence system. *Hardware* provides the basis for the Business Intelligence system, enabling acquisition of local content at operational level and visualisation at presentational level. In addition, a clear indication of relationships, dependencies and connectivity between elements in the different data layers was required. This is due to the fact that, for example, certain types of data might require specific software which might in turn require specific hardware. Evaluating the implications of changes to the Business Intelligence environment against a comprehensive and holistic Business Intelligence Framework supports a better understanding of the implications of the changes and the interactions between elements.

The existing frameworks and approaches evaluated as a part of this research were grouped into three categories, High Level and Conceptual approaches, Data Oriented Approaches and Business Oriented Approaches. A comparison of the Frameworks is presented in Table 3-1 with more detail about each of the frameworks given in APPENDIX A. Frameworks in the High Level and Conceptual and Data Oriented Frameworks support the description and explanation of Business Intelligence and aspects of Business Intelligence functions, and provide a useful overview of the Business Intelligence environment in general. However, frameworks in these categories do not fully support the identification of relevant aspects and components and do not capture multiple perspectives. Some of the data oriented approaches provide visual insight into the data journey from source to presentation but it is difficult to clearly identify or to separate components of the Business Intelligence environment, such as hardware, concepts, user groups and applications or to define the relationships between components. Frameworks belonging to the business oriented category have a specific business focus; they therefore tend to include some elements which are outside the scope of Business Intelligence implementation and exclude some required elements. For this reason, business oriented Frameworks are not considered as holistic in the sense defined in section 3.1. Frameworks in this category can, however, support the partial identification of components and aspects in a scenario where the existing Business Intelligence environment is to be extended or modified.

The Inmon and Kimball philosophies, while not officially defined as a Business Intelligence framework, seem to offer the most generic and also the most comprehensive overview of the Business Intelligence environment and have been included in the High Level and Conceptual category. Frameworks extracted from the Inmon/Kimball approaches provide good insight into most of the relevant aspects and components of Business Intelligence but it is difficult to identify functional relationships between users, hardware and applications in the context of a holistic overview of the Business Intelligence environment. Functional relationships, for example, the relationship between data and software and software and hardware, are important in the Business Intelligence context. Frameworks extracted from the Inmon/Kimball approaches do not support identification of which user categories (technical, business, management or other) are relevant for which components (applications, types of data, hardware or concepts).

Table 3-1: Comparison of the Business Intelligence Frameworks

Framework	**Focus**	**General Category**	**Holistic**	**Which user groups can benefit from this framework?**	**Supports the identification of all relevant components in the BI environment?**
Business Inteligence Framework - (Watson & Wixom, 2007)	Process	* High Level	No	Technical, Business, Management	Partially
RAP: A Conceptual Business Intelligence Framework – (Laha, 2008)	Activity, Data	* Conceptual	No	Business, Organizational	Partially
SBI: A Semantic Framework to support Business Intelligence - (Sell et al., 2008)	Data, Semantics, Ontologies	* Conceptual	No	Technical	No
A Conceptual Framework for Delivering Cost	Cloud	* Conceptual	No	Technical, Management	No

Effective Business Intelligence Solutions as a Service - (Muriithi & Kotzé, 2013)					
Inmon's approach: A Business Intelligence framework for holistic view of enterprise data - (Inmon, 2005)	Data, Activity, Applications, Business, Processes	* High level * Conceptual	Yes	Technical, Business, Management, Other	Partially
Kimballs's approach: A Business Intelligence framework with the focus on business needs – (Kimball et al., 2008)	Data, Activity, Applications, Business, Processes	* High level * Conceptual	Yes	Technical, Business, Management, Other	Partially
Three-layer framework - (Baars & Kemper, 2008)	Data, Layers	* High Level	No	Technical	Partially
Business Intelligence architecture – (Ranjan, 2009)	Data, Processes, Applications	* Conceptual	Partially	Business, Technical, Organizational, Management	Partially
Business Intelligence Layers Architecture - (Gluchowski & Kemper, 2006)	Applications, Layers, Processes	* Conceptual	No	Business, Technical, Organization, Management	Partially
Process Mining: A framework proposal for Pervasive Business Intelligence - (Guarda et al., 2013)	Process	* Business	No	Business, Technical	No
Business Intelligence Systems Implementation in Manufacturing - (Chu, 2013)	Process, Data, Applications,	* Business	No	Business, Technical	No
A Dynamic Capability-Based Framework for Business Intelligence - (Olszak, 2014)	Capability	* Other	No	Other	No

The evaluation of Business Intelligence Frameworks showed that existing frameworks were not sufficient to identify all the elements of a Business Intelligence system or modification of existing systems, particularly with reference to multilingualism as they did not sufficiently identify the components of Business Intelligence systems and the relationship between Business Intelligence components. As a preliminary to developing a design solution for Multilingualism in Business Intelligence, the Holistic Business Intelligence Framework was developed to support understanding of the Business Intelligence environment and the implications of changes to reporting. The following sections in this chapter discuss the development of the Holistic Business Intelligence Framework. The motivation for developing the framework was to support the development of a design solution for multilingualism - MLED_BI. However, the Holistic Business Intelligence Framework also provides a generic representation of the Business Intelligence environment and can be used to support exploration and understanding of the Business Intelligence environment in a range of contexts.

3.3. Development of the novel Holistic Business Intelligence Framework (HBIF)

This section describes the development of the Holistic Business Intelligence Framework, based on the previous work of Inmon and Kimball and uses the concept of a three-layered framework which is widely supported in the literature (Inmon, 2005; Gluchowski & Kempner, 2006; Kimball *et al.*, 2008; Baars & Kemper, 2008; Laha, 2008; Ranjan, 2009; Chu, 2013).

3.3.1. Development of the Framework

The structure of the Holistic Business Intelligence Framework is provided by the principle of separation of data layers. For easier understanding, the framework is presented as a 2D matrix consisting of horizontal elements. There are three Layers, as shown in Figure 3.1, separated according to data functionality: (i) *Source Layer* that covers all components for data collection,(ii) *Warehousing Layer* that includes all components relevant for data storage and analytics and (iii) *Presentation Layer* that encompasses all components associated with the retrieval and presentation of information from the *Warehousing Layer*. Analytics is sited at the *Warehousing Layer* because the aggregation, transformation and partial calculation of the data happens at this layer. However, from the business users' perspective, analytics could be seen as a component of *Presentation Layer*, as some types of calculations, such as summation of information in reports, happens at this layer. This is represented by the dashboards and queries element in the *Presentation Layer*.

Using an iterative approach, all components from each of the evaluated frameworks (APPENDIX A) were analysed and grouped into categories to identify *Perspectives* which would support another view of the Business Intelligence environment. This process identified five *Perspectives*: *concepts*, *applications*, *types of data*, *users*, and *hardware* (Figure 3-1). The Holistic Business Intelligence Framework went through several iterations as part of the validation process and although the perspectives identified did not change, the ordering of the perspectives, as discussed in section 3.4., was revised in later versions of the framework, based on the feedback received from users.

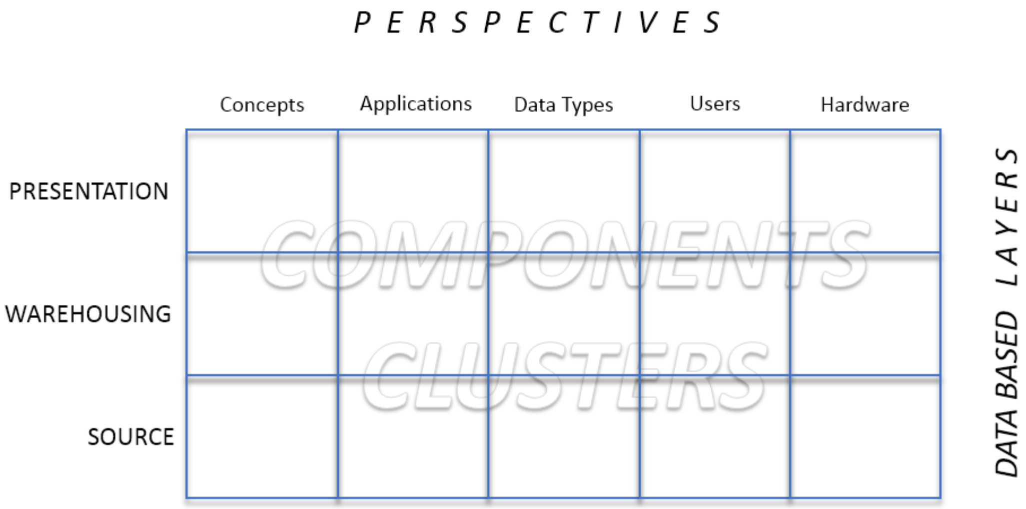


Figure 3-1: Sketch of the first version of the Framework to be proposed

The next steps consisted of allocating components from different *Perspectives* to the relevant *Layer*, creating component clusters, removing redundancies and clarifying terms. Every component cluster represents an intersection of a *Perspective* and *a Layer* encompassing a group of similar components. Figure 3-2 depicts the initial version of the Holistic Business Intelligence Framework; the *X* axis captures the *Perspectives*, while the *Y* axis represented *Layers*. Components are embedded into appropriate fields (components clusters).

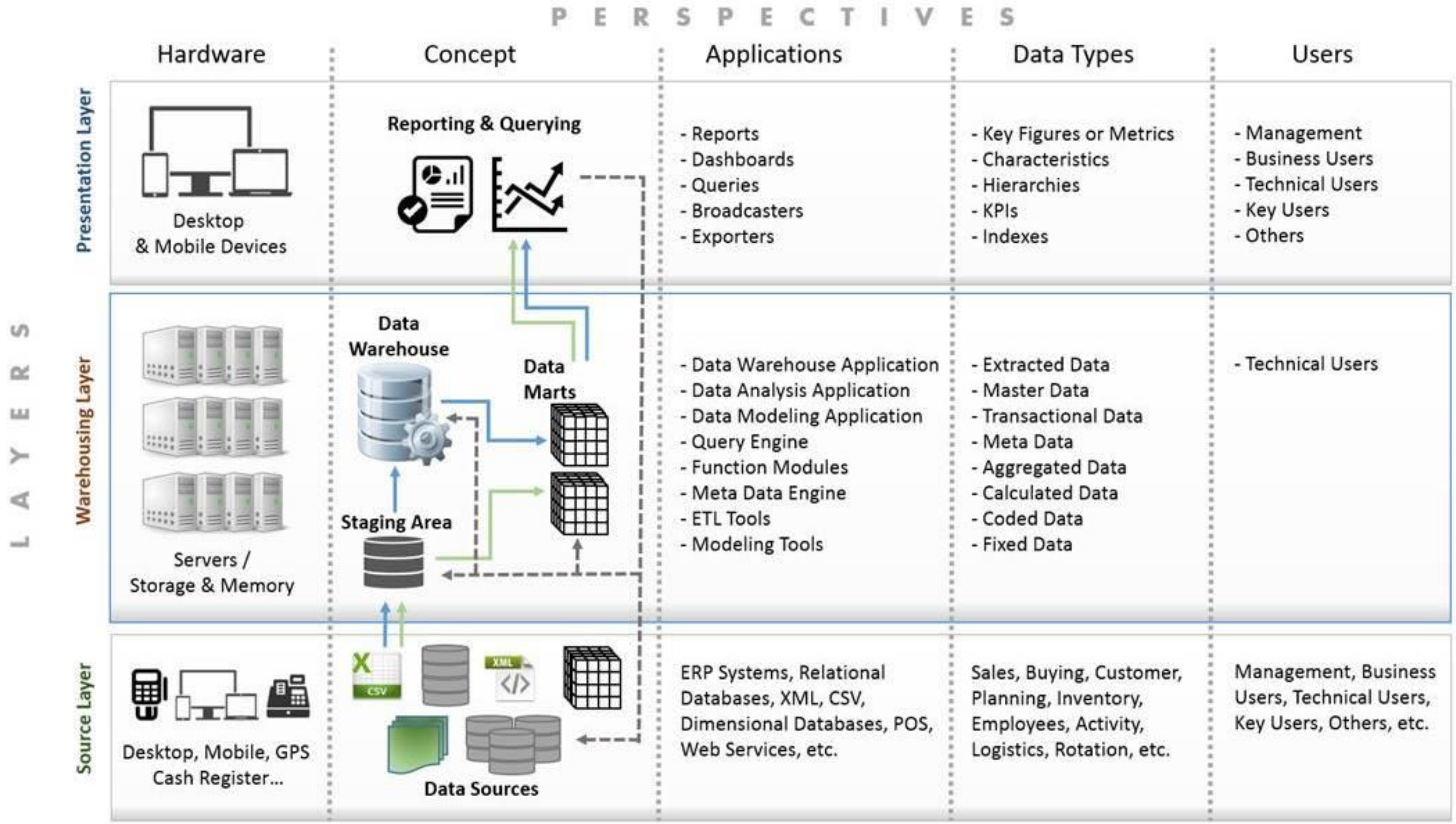


Figure 3-2: First version of Holistic Business Intelligence Framework

The initial design of the Holistic Business Intelligence Framework was influenced by discussion sessions held with seven domain experts from the fields of Business Intelligence and Data Warehousing. The Framework was modified based on the feedback obtained from the experts and was then validated by means of a survey of Business Intelligence users. The Holistic Business Intelligence Framework was revised following feedback from the survey.

3.3.2. Pilot Validation of Framework

The validation of the Framework was designed to cover two elements. First, there was a wish to review whether the Framework was holistic, in the sense of covering all core components of a data warehouse based Business Intelligence system. The other element of the validation focused on usability. The Holistic Business Intelligence Framework was created to support development work in the Business Intelligence field and designed to be used by different categories of end users, including some users who might not possess technical skills. An initial survey of users was carried out, which acted as a pilot for the second, larger, survey. Using a web-based questionnaire, given in APPENDIX B, the initial framework as shown in Figure 3-2, was presented and users were asked to review the framework in terms of elements covered, comprehension and usefulness. To ensure that the respondents had relevant domain expertise and to capture differences in the requirements of different users, respondents were asked to provide information about their role and expertise with Business Intelligence systems and to indicate what type of Business Intelligence users they were (*technical*, *data-centric*, *business*, *management* or *other*). The user information was analysed to identify which type of users had issues with which parts of Holistic Business Intelligence Framework.

The aim of the pilot survey was to identify component clusters in the Framework which might be difficult to understand, to improve them according to the feedback provided, and to reconceptualise if necessary to improve comprehension. Respondents were asked to assess how easy it would be for them to identify from the Framework diagram, the *Perspectives* (Hardware, Concept, Applications, Data Types and Users) and *Layer* components (Source, Warehousing and Presentation Layer) that might be involved in a data warehouse based Business Intelligence project. A Likert scale was used and respondents could select one of the seven options to check understanding (*impossible*, *very hard*, *hard*, *undecided*, *easy with help*, *easy without help* and *very easy)* for each

Perspective and each *Layer* presented in Figure 3-2. A threshold was set in advance that for any component cluster identified by any user as *impossible* or by at least 5% of users as *very hard* or *hard* to understand, the aim would be to improve the usability of the cluster.

Respondents were also asked to comment on the framework concept and usefulness. The pilot survey included an open question to allow for additional comments and suggestions, allowing respondents to comment on the components included in the Holistic Business Intelligence Framework and to identify any issues or omissions. The pilot survey received 29 responses from business, management, technical and other users who work with Business Intelligence on a daily basis. The questionnaire captured feedback from users from seven different countries (Austria, Australia, Bosnia & Herzegovina, Croatia, Hungary, Slovenia and United Kingdom), reflecting the international nature of Business Intelligence.

Feedback indicated that the rationale of the proposed Framework was easy to understand as 28 out of 29 users found the Framework useful in supporting understanding of the Business Intelligence environment and the components that might be involved in Business Intelligence related project.

Reviewed against the scenario of implementing a Business Intelligence related project, no user found the *Concept* perspective *impossible* or *very hard* to understand, while one respondent out of 29 found it *hard* to understand. No user found the *Applications* perspective *impossible*, *very hard* or *hard* to understand. Only one out of 29 users in the pilot study found the *Users* perspective *hard* to understand, while no user found it *very hard* or *impossible* to understand. Three out of 29 users identified the *Hardware* perspective as *very hard* to understand and two as *hard*, highlighting the need to improve this element to enable easier understanding. However, none of the users identified this perspective as *impossible* to understand. One out of 29 users identified the *Types of Data* perspective as *impossible* to understand, one as *very hard* and one as *hard* to understand, again suggesting the need for further review. No other cluster was identified as *impossible* to work with by any user.

None of the data based Layers were identified as *very hard* or *impossible* to understand. Only two out of the 29 respondents found the *Presentation* Layer *hard* to understand, while three users found *Warehousing* and *Source* Layer *hard* to understand.

Based on the feedback and additional comments from users, difficulty understanding some of the definitions was identified as the biggest issue with the framework. It was possible to test this, since following explanations of the definitions, users who had found elements *impossible, very hard* and *hard* to understand, reclassified these elements as *easy (to understand) without additional help* or *very easy*. Based on this feedback, additional components were refined and embedded into the clusters. Nomenclature and phrasing used in the initial Framework diagram were also refined. For example, according to the comments received in the pilot survey, the name "Data Types" that is used for one perspective in Holistic Business Intelligence Framework was identified as confusing. Thus, the name of that perspective was changed to "Types of Data" to avoid any confusion between data types, in the sense of string or numeric, and types of data in the sense of transactional/master data.

3.3.3. Final Evaluation and Modification of Proposed Framework

Following the iteration of the Framework based on the feedback from the pilot study, a second, larger scale survey was conducted to validate the revised Holistic Business Intelligence Framework. Details of the survey are given in APPENDIX C. This larger survey used the same approach as the pilot survey but was based on a revised version of the Holistic Business Intelligence Framework, given here as Figure 3-3. The survey was extended to provide opportunity for respondents to discuss the components included in the framework and to propose components which should be added or removed from the Holistic Business Intelligence Framework.

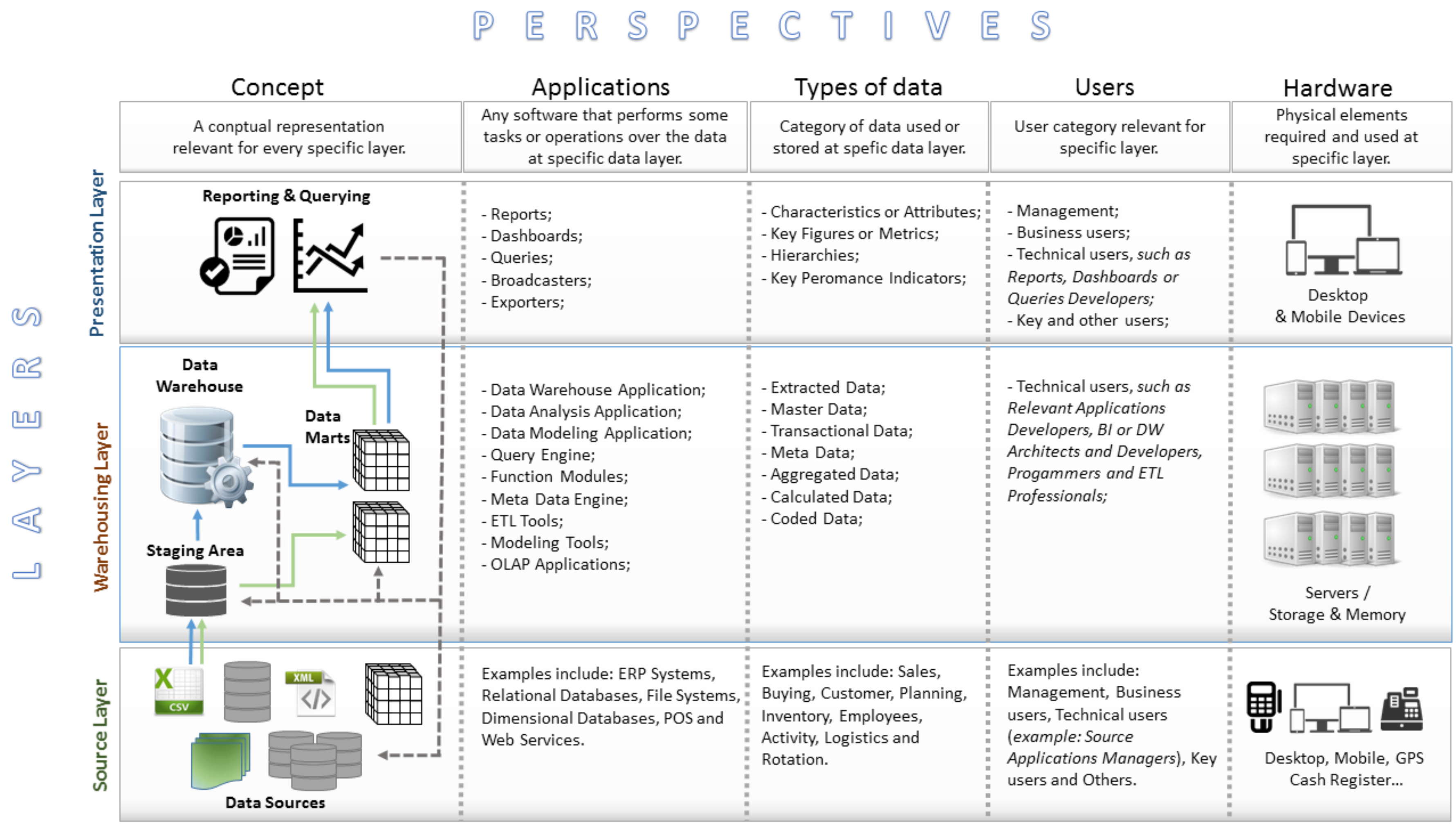


Figure 3-3: Holistic Business Intelligence Framework presented in the second survey

In the second survey, 109 Business Intelligence and Data Warehousing domain experts from 25 different countries took part, again reflecting the international nature of Business Intelligence. More than 95% of the domain experts who provided feedback agreed that the Framework would be useful in identifying the components that might be involved in a Business Intelligence project. More than 93% of respondents agreed that they found the Holistic Business Intelligence Framework as shown in the diagram (Figure 3-3) easy to understand.

For the revised version of the Framework, no component was described as *impossible* to understand. On average, across all perspectives, 89.7 % of respondents found the components of every *Perspective very easy*, *easy without help* or *easy with help* to identify. 4.7 % of respondents were *undecided* with some perspectives, and 5.6 % found it *hard* or *very hard* to identify components of any *Perspective.* On average 92.3 % of respondents found it *very easy*, *easy without help* or *easy with help* to identify the components of any *Layer.* No single *Layer* was marked *very hard* to identify; this was expected given that the 3-layer separation is a widely accepted concept in Business Intelligence and the majority of the Business Intelligence frameworks evaluated as part of this research, as detailed in APPENDIX A, use the same approach. Only 3.6 % respondents across all three layers found it *hard* to identify some component of the *Layers*, while 3.9 %, were *undecided*.
Based on the feedback and suggestions received in the second survey, Figure 3-3 was extended by adding Self-Service Business Intelligence to the *Concepts* perspective at the *Presentation Layer* (Figure 3-4). This required the inclusion of Data Feeds used by Self-Service Business Intelligence to the *Application Perspective* at the *Presentation Laye*r. Responding to the feedback received, Project Sponsors and Decision Makers were included in the *User Perspective* in the *Presentation Layer*, as the suggestion that these titles represent different types of users was accepted. Business Analytics Applications were added under the *Applications Perspective* at the *Warehousing Layer*. A recommendation to add WWW and Cloud Services as possible applications at the *Source Layer* was accepted. A suggestion that Modelling tools should be removed from the *Application Perspective* in the *Warehousing Layer* was accepted as the “Data Modelling Application” component is already included in the same cluster. Finally, in response to comments received, Master and Transactional Data in the *Types of Data*

Perspective under the *Warehousing Layer* were defined as a subcategory of Extracted Data (Figure 3-4).

Several respondents proposed additional perspectives, such as Planning, Communication, Security, Data Quality, Governance, and even perspectives representing specified tools such as Tableau, PowerBI, SAP BW, MicroStrategy, Headoop, and SSRS. The Holistic Business Intelligence Framework is extensible and it would be possible to include these elements if required in a specific business context. However, those elements are not included in the Holistic Business Intelligence Framework discussed here since the aim is to present a generic Framework which presents core elements but is capable of being tailored to users' needs.

It was also suggested that Big Data should be included in the Framework. Big Data is concerned with large-volume (Dhote *et al.*, 2015), complex and ever growing data, often from autonomous sources (Wu *et al.*, 2014), which are often unstructured (Lokhande & Khare, 2015). Business Intelligence, however, according to the definitions from Power (2002), Moss & Atre (2003), Golfarelli *et al.* (2004), Lönnqvist & Pirttimäki (2006), Dekkers *et al.* (2007), Kimball (2008), Jourdan *et al.* (2008), Brannon (2010), Jamaludin & Mansor (2011), focuses on the collection, analysis, presentation and dissemination of business information coming from sources that mostly hold structured data. In that context, we regard Big Data and Business Intelligence as two separate, although related concepts (Dedić & Stanier, 2017a), and including Big Data in the Business Intelligence Framework would reduce the clarity and comprehensibility of the Holistic Business Intelligence Framework in a data warehousing Business Intelligence context. However, the approach used here could be adapted to provide the basis for a framework that encompasses both of these concepts.

3.4. The Holistic Business Intelligence Framework

Figure 3-4 presents the final version of the framework. The Holistic Business Intelligence Framework comprises two views. In one view, the Framework is separated into three Layers: *Source Layer*, *Warehousing* and *Presentation Layer* (Figure 3-4). As discussed in 3.3.3, the separation of Layers is a well-established approach with a basis in the theoretical foundations of Business Intelligence. The three-layered approach enables identification of components and aspects at a specific data layer when working in a data warehouse based Business Intelligence environment. For example, it enables identification of relevant concept, applications, hardware, types of data and users at data source level.

In the proposed Framework, the traditional three-layered separation was extended with a horizontal presentation of the Business Intelligence environment/ecosystem (Figure 3-4). This allows visualisation of the layers in the wider context of the Business Intelligence environment. As an example, the Resource Manager for a Business Intelligence project needs to understand the hardware, applications and user requirements of the project in order to be able to plan those resources. Each perspective must be clearly defined in order to support optimal acquisition and supply. The Holistic Business Intelligence Framework enables an overview of the resources required at different stages, such as implementation at the *Warehousing* (storage) *Layer* or *Presentation Layer*. The framework structure can support users with different requirements. Information Technology Management, for example, might be interested only in a high level view while implementation teams, and in particular, teams dealing with hardware infrastructure and those providing applications can use the Framework to focus on their field of interest and expertise.

The sequence of the layers is fixed, based on the well-established three layer approach. However, other components are not fixed and can be changed to suit the requirements of users. The sequence of the perspectives shown in Figure 3-4 is based on the feedback from IT domain experts; it broadly indicates the complexity of the element on a right to left scale. Applications, for example, are seen as more complex than hardware. The Holistic Business Intelligence Framework is extensible and could be amended to include additional perspectives as required, for example, for a specific business context and the sequence of perspectives could also be amended if appropriate. The framework presented here uses only those perspectives that have been identified from the literature and the survey results as generic and are therefore applicable to any Business Intelligence project.

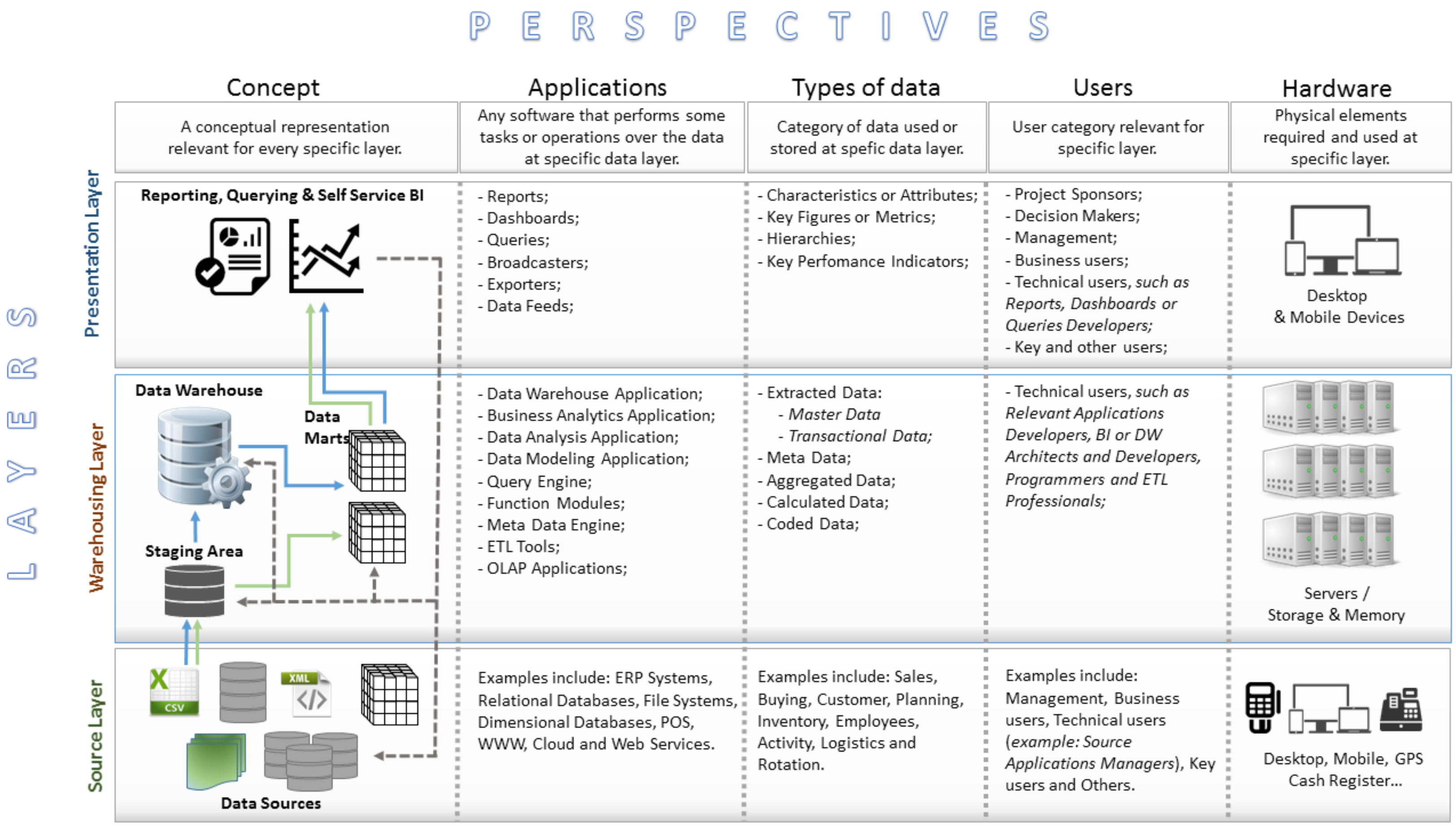


Figure 3-4: Holistic Business Intelligence Framework

3.4.1. Layers in the Framework

The *Source*, *Warehousing* and *Presentation Layer* are separate layers. The *Source Layer* deals exclusively with data sources (Figure 3-4). From the *Hardware* perspective, the *Source Laye*r encompasses all possible devices that enable the physical collection of source data, such as desktop PC, mobile devices, bar code readers and any similar devices. Applications in the *Source Layer* are also linked to the collection of source data and are represented for example by Enterprise Resource Planning (ERP) systems, databases, web applications such as Point of Sale (POS), Web services and many others. Types of data in this layer include sales, buying, customer, planning, inventory, logistics, employees and similar data. All user groups, including management, business and technical users, are present in this Layer. Examining the *Source Laye*r by reading across the perspectives, reveals the interconnectivity of components. For example, business users collect sales and purchase data using Enterprise Resource Planning and Point of Sale systems via desktop and other similar devices, and that data is used as source data for further processing.

There was an issue as to whether to name the middle layer in the Framework *Storage Layer* or *Warehousing Layer.* Following discussions with Business Intelligence experts, it was concluded that the term *Storage* might be ambiguous as it may be also understood as referring to the source system databases. As the term *Warehousing* is widely adopted in a Business Intelligence context, and because of the extensive use of the phrases "data warehouse" or "data warehousing", it was decided to describe the middle layer as the *Warehousing Layer.* This layer includes the objects that constitute the Data Warehousing element and includes all objects that hold data extracted from source systems. This includes staging areas, Data Warehouse and data marts. Servers and other memory storage devices are part of the *Warehousing Layer* when viewed in the *Hardware Perspective*. The *Applications Perspective* includes relevant components such as data warehouse applications, Extract-Transform-Load tools, meta data engine, data analysis applications and modelling tools. This layer is concerned with operations on data extracted from source systems, which are mostly defined as master and transactional data in Data Warehouse approaches. Metadata, aggregated, calculated and coded data also belong to this layer. The *Warehousing Layer* is managed by different groups of technical *Users*. The layer demonstrates the interconnectivity of perspectives and components. A component in this case represents any constituent part of a component cluster. Technical users define and manage all types of data in this layer, using elements such as data modelling or Extract-Transform-Load tools, and store data using Data Warehousing objects (data marts, Data Warehouse object or staging areas), which are physically stored on servers or other memory and storage devices.

*The Presentation Laye*r includes elements such as reporting and querying. Desktop and mobile devices are used as hardware components to present application outputs such as reports, queries and dashboards. Reports, dashboards and queries use Types of Data such as key figures, metrics, hierarchies and KPIs to describe characteristics. The *Presentation Layer* may include users from all categories, including management, business users, technical users, key users and others. This layer also illustrates the interconnectivity of components.

3.4.2. Perspectives in the Framework

The Holistic Business Intelligence Framework supports both a bottom up and top down view of the Business Intelligence environment. When analysing the *Concept* perspective in the *Warehouse Layer* using a bottom up approach as presented in Figure 3-5, the diagram shows that staging areas are initially supplied with data from data sources and the data may be in different formats (relation, dimensional, CSV, and others). Using a bottom up approach, following the blue arrows and moving upwards from the staging area to Data Warehouse and then to data marts, the framework supports, that is, can be used with, Business Intelligence environments developed on the Inmon (2005) and Linstedt *et al.*, (2010) data warehouse and data mart design approaches. In the Inmon/Linstedt approaches, as discussed Chapter 2, section 2.4.2., and in APPENDIX A, the Data Warehouse is almost the same as the source system(s) and is implemented as a separate and not-changeable database. The Data Warehouse has data marts, or aggregated tables that are used for reporting and querying purposes. Following the green arrow from the staging area directly to the data marts, the framework supports the Kimball *et al.* (2008) design approach in which the Data Warehouse is a concept built up from different data marts and connected with conformed dimensions. This meant that the Holistic Business Intelligence Framework is compatible with the two most widely used Data Warehouse design approaches.

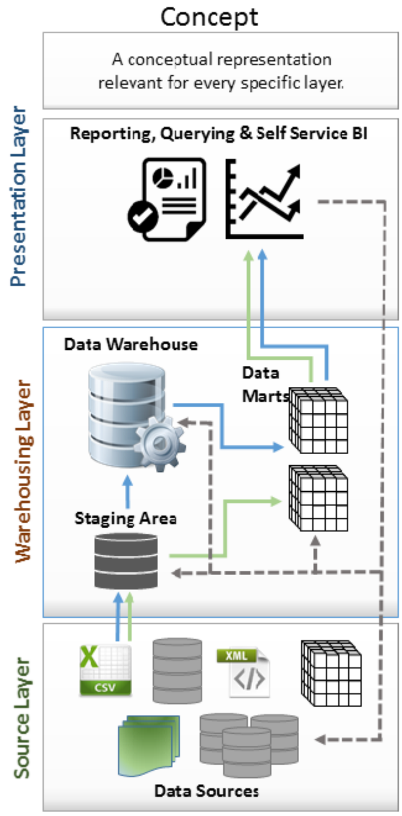


Figure 3-5: Holistic Business Intelligence Framework Concept Perspective

It is important to note that the proposed Framework supports data flow in other directions. If following the flow of the dashed grey arrow, it can be seen that the *Presentation Layer* reporting and querying can be used as input to provide the *Warehouse* and *Source Layers* with additional data. For example, the metrics that show average inventory status per week can be used by the source systems as a guide for planning future inventory stocks, and then again in Data Warehouse to calculate results for planned vs actual. Metrics can be used as a basis for the calculation of new constant values to be stored in the Data Warehouse, and then again used by reports to support advanced metrics.

The other perspectives can also be understood in a vertical view. In the *Applications Perspective*, systems such as Enterprise Resource Planning or Point of Sale are used to collect information. This information is stored and manipulated using Data Warehousing or Extract-Transform-Load applications, and then presented using, for example, reports, dashboards and queries. In the *Hardware Perspective* desktop, mobile and other devices are used to collect information. The information is stored on servers or other memory devices, and then presented using desktop, mobile and other devices. From the *Types of data Perspective*, sales, purchases and other data are collected, then stored as master, transactional or other data and presented as metrics, KPIs, characteristics or hierarchies. Examining the *Users Perspective*, all user groups are involved in activities at the *Source Layer* and at the *Presentational Layer,* always depending on the context. Only technical users are involved at the *Warehousing Layer.*

3.5. Discussion and evaluation

The motivation for the development of the Holistic Business Intelligence Framework was that work in the field of Multilingualism in Business Intelligence, required identification of the components and the relationships between components that must be taken into account when developing a new Business Intelligence system or modifying or extending an existing Business Intelligence system and that there was no existing framework which had this capability.

The Holistic Business Intelligence Framework presented in this chapter supports the investigation into Multilingualism in Business Intelligence by providing a high level view of the core components in the Business Intelligence environment, and the relationships between these components. This understanding of core components and relationships is also relevant in the general, data warehouse based, Business Intelligence environment and can be customised to suit the requirements of users. Based on the feedback from Business Intelligence practitioners and developers, the Holistic Business Intelligence Framework presented in Figure 3-4 is regarded as holistic, in the sense that it covers all core components and supports the identification of the relationships and dependencies which must be taken into account when developing or extending a Business Intelligence system, not only to support Multilingualism, but in a generic context as well.

The framework provides for a vertical separation of the Business Intelligence environment and also a horizontal (component and user based) view of the Business Intelligence environment. This enables an overview of the resources required at different stages of the data journey, for example implementation at the *Warehousing Layer* or *Presentation Layer*. This enables the Framework, as discussed in section 3.4, to support users with different requirements and different levels of technical knowledge. In a multilingual context, the Holistic Business Intelligence Framework supports users in identifying which elements of Business Intelligence, such as applications, or data types are affected when enabling multilingualism in Business Intelligence. The Holistic Business Intelligence Framework provides both an overview and a cross sectional view of the Business Intelligence environment. This facilitates the understanding of different aspects of the Business Intelligence environment and means the Framework can be used as a communication tool, particularly to support discussions between technical and non technical users. The use of perspectives means that the framework is customisable as components can be added or removed, as required for domain specific reasons. The feedback received on the final iteration of the Holistic Business Intelligence Framework included a number of suggestions for additional components, indicating that there is a demand within the Business Intelligence community for an extensible Business Intelligence Framework.

A limitation of the Holistic Business Intelligence Framework is that it focuses on the data journey in a data warehouse context and does not cover wider aspects of the Business Intelligence environment. It is proposed that the Holistic Business Intelligence Framework can be regarded both as a stand alone representation of the core Business Intelligence environment and also as a building block of the larger picture. The Holistic Business Intelligence Framework could be used, for example, as the starting point for a Framework which encompasses both Big Data and Business Intelligence.

3.5.1. The HBIF in Relation to Multilingualism

The focus of this research is on multilingual presentation of business information descriptions in a data warehouse based Business Intelligence context where Business Intelligence reports are presented in the form of descriptive characteristics, attributes or hierarchies and are colloquially known as master data. Characteristics, attributes, hierarchies, and master data were identified as the starting point for the identification of interfaces relevant for this research. Developing the Holistic Business Intelligence Framework enabled these elements to be identified as the part of the *Types of Data Perspective*. Visualising components and relationships through the Holistic Business Intelligence Framework made it possible to identify that the relevant concepts related to multilingualism in Business Intelligence encompass *Reporting*, *Querying*, *Data Warehouse*, *Staging Area*, and *Data Marts,* where *Data Marts* are a critical interface for any type of content, including descriptive. Applying the Framework, it was possible to identify Business Intelligence users and applications such as *Reports*, *Dashboards*, *Queries*, and *Data Warehouse Application* which would be relevant in the multilingual context. Analysing Business Intelligence processes and operations in terms of the *Layers* and *Perspectives* identified in the Holistic Business Intelligence Framework revealed that despite the important role that Hardware plays when collecting localised content at operational level, hardware was not a significant component in relation to support for Multilingualism.

The identification of relevant elements through the Framework supported the development of the new design approach for multilingualism in Business Intelligence (MLED_BI). As *Data Marts* are a critical interface and are used as a basis for Business Intelligence reporting, the design approach needed to preserve the independence of the *Warehousing* and *Presentation Layers* in regard to the *Source Layer*. Existing workarounds for Multilingualism, as discussed in Chapter 2, affect star schema design and require design changes at the warehouse layer. Changes at the *Warehousing Layer* have implications for the *Presentation Layer.*

3.6. Conclusion

Business Intelligence Frameworks can be used to support the identification of elements that need to be modified or extended to support changes in the Business Intelligence environment. In this chapter, existing Business Intelligence Frameworks and Data Warehousing approaches are analysed with regard to their capability to present the core components of the Business Intelligence environment and to support the identification of relationships, dependencies and connectivity between components at different data layers, whether developing a new Business Intelligence environment or extending or modifying the existing Business Intelligence environment to support requirements such as Multilingualism. Evaluation of the Business Intelligence frameworks contained in the literature, showed that although existing Frameworks can be used to identify some aspects of the Business Intelligence environment, none of the existing Frameworks provide a holistic representation of the Business Intelligence environment and enable identification of elements that might be relevant in multilingual context. Thus, a new Business Intelligence Framework, the Holistic Business Intelligence Framework, was proposed which includes all the core components of a generic Business Intelligence system. The Holistic Business Intelligence Framework supports the analysis required to develop a new approach to enable multilingualism in Business Intelligence but the framework is generic and can be used both as a stand alone representation of the generic Business Intelligence environment and as the basis for an exploration of the wider Business Intelligence environment. Setting the newly proposed Framework in the context of the comparison of the Business Intelligence Frameworks given in Table 3-1, the focus of the Holistic Business Intelligence Framework is on Applications, Concepts, Data, Hardware, and Users, while also supporting additional aspects such as Activity, Business, Layers, and Processes. The proposed Framework is categorized as High Level and Conceptual, and considered as Holistic because of the elements that are covered. The Holistic Business Intelligence Framework would be benefitial for Business, Management, Organizational, and Technical user groups, and it supports the identification of all relevant components in the Business Intelligence environment. In the context of this research, as discussed in the following chapter, the Holistic Business Intelligence Framework was used to support an investigation of the components which would be affected by support for multilingualism. The next chapter, chapter four, discusses the development of the MLED_BI design approach and the use of a Proof-of-Concept (PoC) implementation to assess the technical feasibility of MLED_BI.

Chapter 4: MLED_BI: New BI Design Approach

4.1. Introduction

This chapter introduces MLED_BI, a new BI design approach to support ML in BI. The chapter gives an overview of MLED_BI and evaluates the MLED_BI approach in relation to BI design approaches and existing solutions to support ML. MLED_BI is grounded in the design theory which underpins the technical design of data warehouses, as discussed in chapter 2 and is based on a critical evaluation of the components in a BI system which would be affected by support for multilingualism, as identified from the HBIF developed in chapter 3. In the present chapter, the findings from existing work and the information obtained from the HBIF are used to identify critical elements in the development of MLED_BI. After defining the context of the investigation, the requirements of MLED_BI are discussed. A new concept for the star schema that reintroduces data independence and supports immunity from changes for multilingual BI systems is discussed, together with the implications of the new approach and the MLED_BI design stages, the delivery of reports in the new environment, and data manipulation aspects. The expected benefits and limitations of the MLED_BI design approach are discussed and the conclusion summarizes the content of the chapter.

4.2. Context of the investigation

The literature review (section 2.6.) identified that existing approaches to support ML in BI did not provide a sufficient solution as they cause usability issues, extreme data redundancies, sluggishness, implementation challenges, and difficulties in maintenance. It was argued that the underlying cause for these limitations was the ad-hoc nature of the solutions which are more solutions to an implementation problem than design approaches with a theoretical underpinning. The primary weakness, as discussed in Chapter 2, was that existing approaches to support ML in BI re-introduce data dependence since, as discussed in 2.6., additional languages are modelled by making changes to the Star schema. This loses the benefits of data independence, such as immunity of applications to alterations at different levels and this in turns means that the system does not provide the most optimal support for end users. As an example related to immunity to alterations, enabling a new language in an existing BI environment by applying existing approaches, would require modifications at all levels of the BI system. As a second example, changing even the smallest erroneous description in a BI report requires iteration of the whole data load process from source systems to reporting data marts.

The literature review also identified that there is a lack of experimental data to demonstrate the effectiveness of currently existing ML solutions and relationships to IT architectures. There was no evidence in the literature that the fit to an existing BI environment and architectures had been analysed or evaluated. Taking into account the issues identified in relation to Star schema design and data dependency, it was concluded that there is a need for a new BI design approach to support ML. The new approach should support immunity from changes, consequently enabling more technical flexibility and provision of a better service to end users. The following section describes the requirements that should be met in the newly proposed BI design approach.

4.3. Requirements for the MLED_BI design approach

This section outlines the issues which the MLED_BI design approach will address.

- The discussion in Chapter 2 identified that the problems with existing solutions to support ML stemmed from the lack of data independence that results from including additional language elements within the star schema. Thus, the first requirement of the newly proposed BI design approach was to re-introduce data independence in BI design approaches that support ML. In practical terms, this means that a new BI design approach for ML must be based on the concept that the implementation of language requirements must provide alteration immunity at all levels. For example, it must be possible to add new languages without the requirement to change ETL process, data mart concepts, dimension table design, or BI reports.
- As an additional requirement, it should also be possible to change erroneous master data descriptions in existing BI reports without the requirement to iterate the complete ETL cycle on relevant data in BI system.
- As section 2.4.3. identified that the data warehouse design and development approaches of Inmon and Kimball are the most widely used approaches in industry and in academia, the proposed new BI design, MLED_BI, must support both approaches - meaning that ML implementations based on the MLED_BI

design approach must be capable of being integrated with BI environments based on the Kimball and the Inmon approach. This will allow the MLED_BI approach to be implemented in existing BI environments.

- Section 2.4.4. identified that the star schema is the schema approach most widely used to design data marts in data warehouses and is used in both the Inmon and Kimball approaches. Thus, to be considered as a generic solution, the concept of the data mart in MLED_BI must be based on the star schema
- The complexity of the processes required to change erroneous content for master data in existing BI reports, the inability of end users to perform such activities immediately and by themselves, the complexity of ETL processes required to support ML in BI, and difficulties in complying with legal requirements to provide data in more than one language were identified in section 1.2. as some of the motivating factors for this research. One of the aims of MLED_BI was seen preventing these problems by providing greater possibilities for data manipulation. Section 2.3.2. identified web based reporting as the most appropriate mechanism for delivering reports for companies working internationally. In that context, a new BI design approach should be able to support the delivery of BI reports via the web and a content management system where this meets the requirements of the organisation.

Based on the requirements identified in this section, the rest of this chapter describes the development of the novel BI design approach to support multilingualism in BI environments, MLED_BI.

4.4. Development of MLED_BI

4.4.1. Identification of Components

To start addressing the issues associated with support for ML in BI, existing BI frameworks were evaluated in chapter 3, with the aim of determining the components of BI systems that would be affected by the implementation of ML, including relationships and dependencies between components. As the evaluation showed that none of the existing frameworks provided sufficient support for the identification of components affected by ML, a new Holistic Business Intelligence Framework (HBIF) was developed.

The newly proposed Holistic Business Intelligence Framework was used to support the analysis required to develop a new approach to enable optimal application of multilingualism in Business Intelligence environment. Chapter 2.3. identified *master data* as the critical element when considering content relevant in the multilingual context of Business Intelligence. Thus, the first step of the analysis of the elements relevant to the development of the new design approach was the visual identification of the component cluster from Holistic Business Intelligence Framework that includes *mater data*. As seen in Figure 3-4, the relevant component cluster that includes the master data element intersects the *Types of data* perspective and the *Warehousing layer*. The Business Intelligence environment, as discussed in the HBIF in chapter 3, can be understood in terms of the Source layer, the Warehouse layer and the Presentation layer (Figure 4-1). Visualising the relevant components and relationships of master data through the HBIF framework from its horizontal aspect it was identified that the Warehousing layer, and specifically the Data Warehouse, plays a crucial role in the BI environment in the context of multilingualism for any type of business content, including descriptive data. From the vertical aspect, the Holistic Business Intelligence Framework clarified the role of data marts as the crucial asset to deliver descriptive content from Data Warehouse to Business Intelligence reports. Additionally, in section 3.4.1., the Holistic Business Intelligence Framework defined the scope of the Business Intelligence Source layer as a source of data delivery to the Warehousing layer. In this context, it was concluded that to enable effective delivery of multilingual content at the Business Intelligence Presentation Layer, the focus of the research should be directed towards evaluating and re-engineering existing concepts of data mart. As the chapter 2.4. identified the star schema as the most widely used approach for the design of the data mart concept in Data Warehouse, the MLED_BI design process primarily impacts at the Concept perspective of the Warehouse layer, shown by a solid outline on Figure 4-1, as it relates to the design of the fact and dimensional tables which provide the design basis for the development of the data warehouse and star schema based data marts. The Holistic Business Information Framework supported the identification of the scope and components of MLED_BI.

The MLED_BI design approach makes possible the development of a multilingual content management system (MCMS), which provides flexibility in data reporting and manipulation and can therefore also impact at Presentation layer. This is shown as a dashed outline on Figure 4-1.

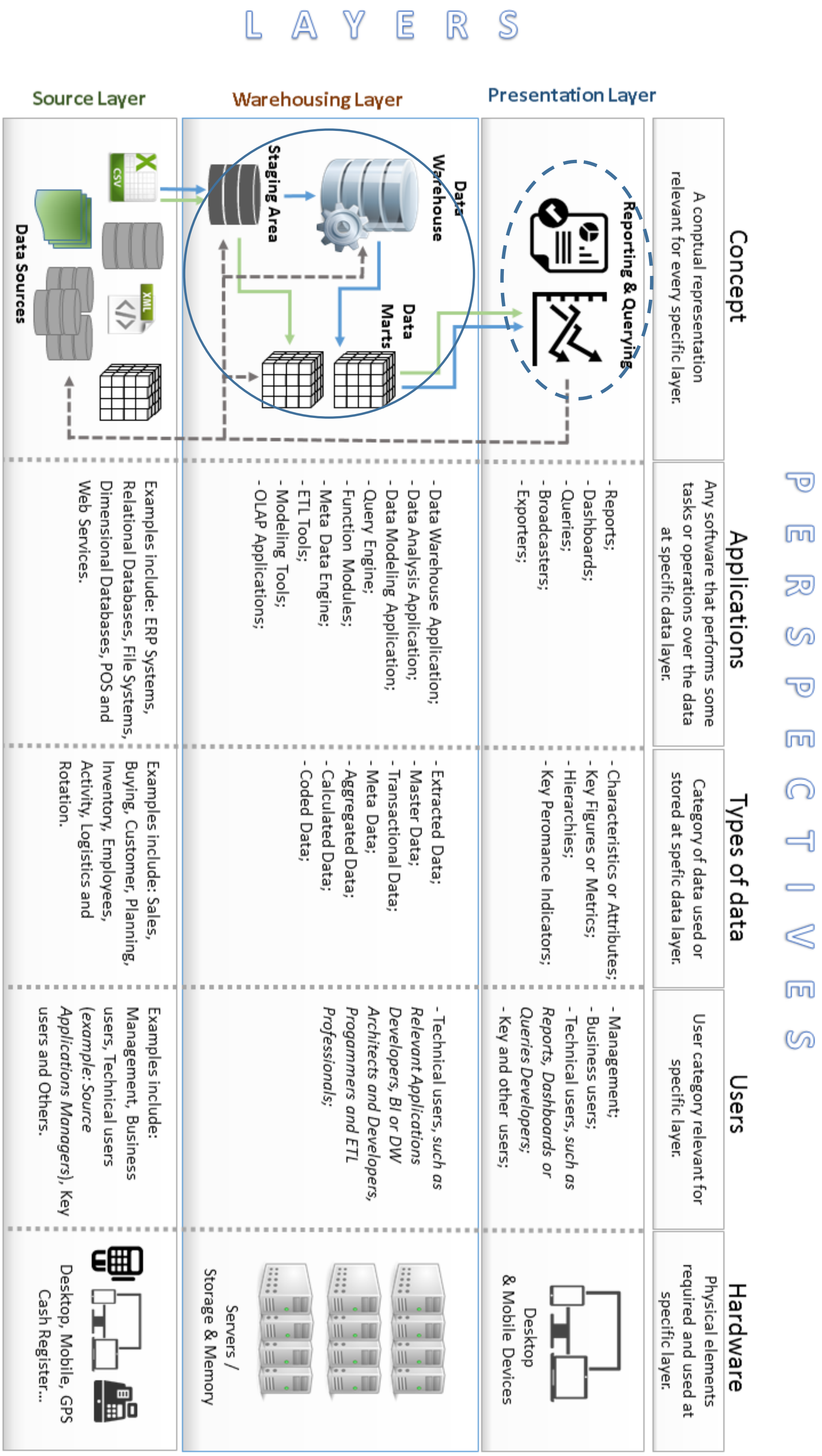


Figure 4-1: HBIF showing the elements impacted by MLED_BI

4.4.2. Content Change Issues

The analysis of DW design and development approaches in sections 2.4.3., and 2.4.4., and the discussion of existing ML solutions in section 2.6., showed that current solutions for ML introduce data dependency and that this affects the whole BI system. In existing DWH design approaches, content changes are permitted only in the source system and changes are not made at DW level. This restriction on content change excludes additional data generated directly in DW itself through transformations and operations on existing data. The star schema uses attribute descriptions and hierarchies as a basis for data aggregation and representation in reports, thus, in existing approaches, this "no data change philosophy" is reasonable. The following scenarios illustrate the problem and the need for the "no data change" philosophy in current design approaches and the implications of the "no data change" approach.

Suppose there is a dimension *Product* in the DW that has an attribute, *Category*. The dimension holds a *Category* 'Sweets'. The business users wish the description in the future to be 'Candies'. If the description of an attribute is changed only in the dimensional table in the DW and not in the source system, any subsequent full load of master data from the source system would overwrite changes made in dimensional table, thus reintroducing the former name "Sweets". Suppose that in the *Product* dimension in the DW, the attribute *Category* was changed from "Sweets" to "Candies" for all relevant products and that this change is made only in the DW and not in the source system. If a new product is added to the source system, the new product would belong to the *Category* "Sweets" as categories have not changed in the source system. Loading master and transactional data for this new product into the DW would create two different categories -"Sweets" and "Candies"- for products which should in fact belong to the same category. When running a BI report that aggregates data on product categories, two categories would be shown rather than one. There are additional factors that justify the "no data change" at DW level philosophy. One example is that of using a parallel system to control the accuracy of the data arriving in the DW. In this approach, the parallel system receives data aggregated by product category from the same source system for DW control purposes. Changing the description of a specific category only in the DW dimensional table would lead to data differences in regard to the control system. As already discussed in 2.6.1., existing approaches to enable multilingualism in BI require modifications and extensions to DM design to support new languages, modifications and extensions to existing ETL processes and changes to applications that deliver BI reports. The implication of the “no data change” rule is that where changes on master data are permitted only in the source system, to enable a new language in the BI environment, the language must first be enabled in the source system, meaning that source systems also require modification to support ML.

4.5. The Star Schema

Reengineering the concept of the star schema to better support multilingualism was identified as the first step in the development of the MLED_BI BI design approach. For the reasons discussed in 4.3., the star schema concept provided the underpinning basis for the development of MLED_BI. In existing approaches to support ML in BI, data dependency for multilingual content is introduced at star schema level. Thus, a new approach to the design of the star schema was needed that would re-introduce the concept of immunity to changes. Removing implementation detail for the support of new languages from the star schema would also support a new approach to changes to descriptive content. Previous work identified that the main challenge to providing support for ML in BI in the context of the star schema is that attribute and hierarchy descriptions are saved inside the dimensional tables of data marts (Dedić & Stanier, 2016a). This leads, as previously discussed in section 2.6., to performance problems and problems of dependency and coupling. A snowflake design approach reduces redundancy but is highly normalised which introduces other performance issues. To avoid the problems that have been identified, MLED_BI treats the star schema as a higher level entity and extends the design process to the design of separate language files. Textual descriptions from attributes and hierarchies (master data) are held in language files and are stored outside the dimension tables. Despite the fact that language descriptions are saved outside dimension tables, they are still part of the concept of dimensions in the star schema as this is understood in MLED_BI. Because master data descriptions are saved outside dimension tables, the manipulation, addition or removal of master data descriptions does not affect the structure, content, or architecture of the tables. This provides immunity from changes since, for example, enabling an additional language does not require modification to the dimensional tables, the source systems or reporting applications; an additional language file appropriately designed would be sufficient.

Using the MLED_BI approach, before data is stored in reporting data marts, the analysis and design process allows descriptive information (master data descriptions, such as attributes and hierarchical descriptions) to be extracted and held in language files. This use of language files simplifies the design of dimension tables which no longer contain descriptive information but hold identifiers (Master Data IDs) to support relationships with the language files. In implementation, as attributes and hierarchical descriptions (master data) and their IDs are extracted to separate language files, only numerical values (master data IDs) are stored in dimensional tables. This allows aggregation to be based on identifiers which has parallels to the concepts that underpin the relational design strategy of normalisation. This separation avoids redundancy and description-based aggregations and also removes source system language dependency: a new language can be added by providing a new language file and the language in the language file does not have to be available in the source system. It is important to note that the MLED_BI design concept is based on the use of language files and does not use additional dimension tables to store descriptions of master data. Using additional dimension tables would have implications such as introducing a normalised snowflake schema and would lose the main benefit of the star schema, discussed in section 2.4.4., namely, reduction in the number of joins. Figure 4-2 provides an overview of data marts based on the star schema and supported by existing approaches while Figure 4-3 provides an overview based on the MLED_BI approach.

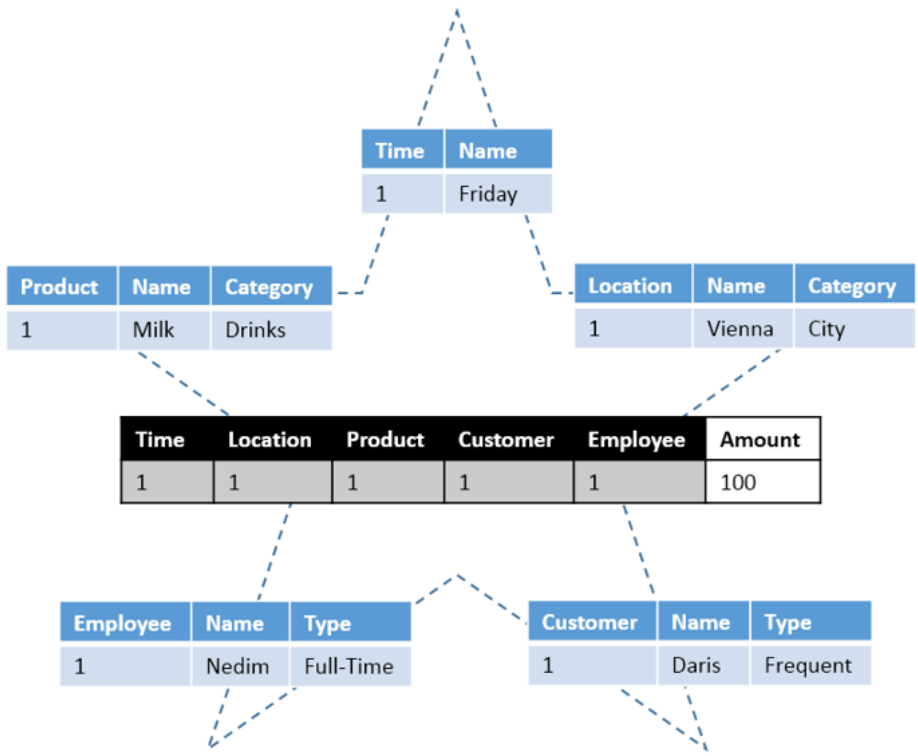


Figure 4-2: Data mart design based on the established Star schema approach

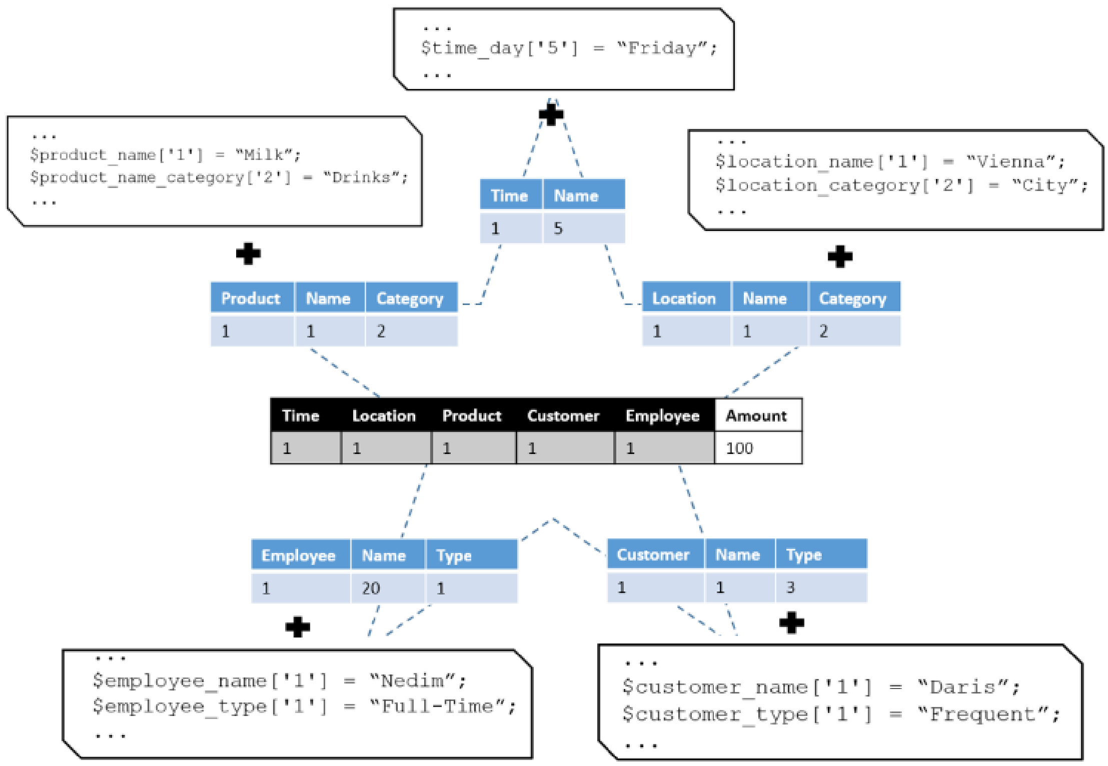


Figure 4-3: Data mart design based on MLED_BI approach

As shown in Figure 4-2, a data mart based on the established star schema approach consists of a fact table holding transactional data and foreign keys to dimensional tables holding descriptive master data. A star schema based on the MLED_BI approach, as shown in Figure 4-3, also consists of a fact table holding transactional data and foreign keys to dimensional tables. However, the dimensional tables hold only master data IDs and this links to language files with arrays holding descriptive information.

4.5.1. MLED_BI Design Process

The design approach proposed in MLED_BI can be integrated into existing BI environments. This is an important feature of MLED_BI since BI systems are expensive to develop and have organisation wide implications, meaning that complete redesign of existing systems would not be a practicable proposition. Figure 4-4 shows that MLED_BI can be integrated into systems developed using the Kimball philosophy where DW design includes only the dimensional modelling concept (red arrow in figure 4-4), and into systems developed on the Inmon philosophy where DW design includes both the design of DW database and dimensional modelling. (blue arrow in Figure 4-4).

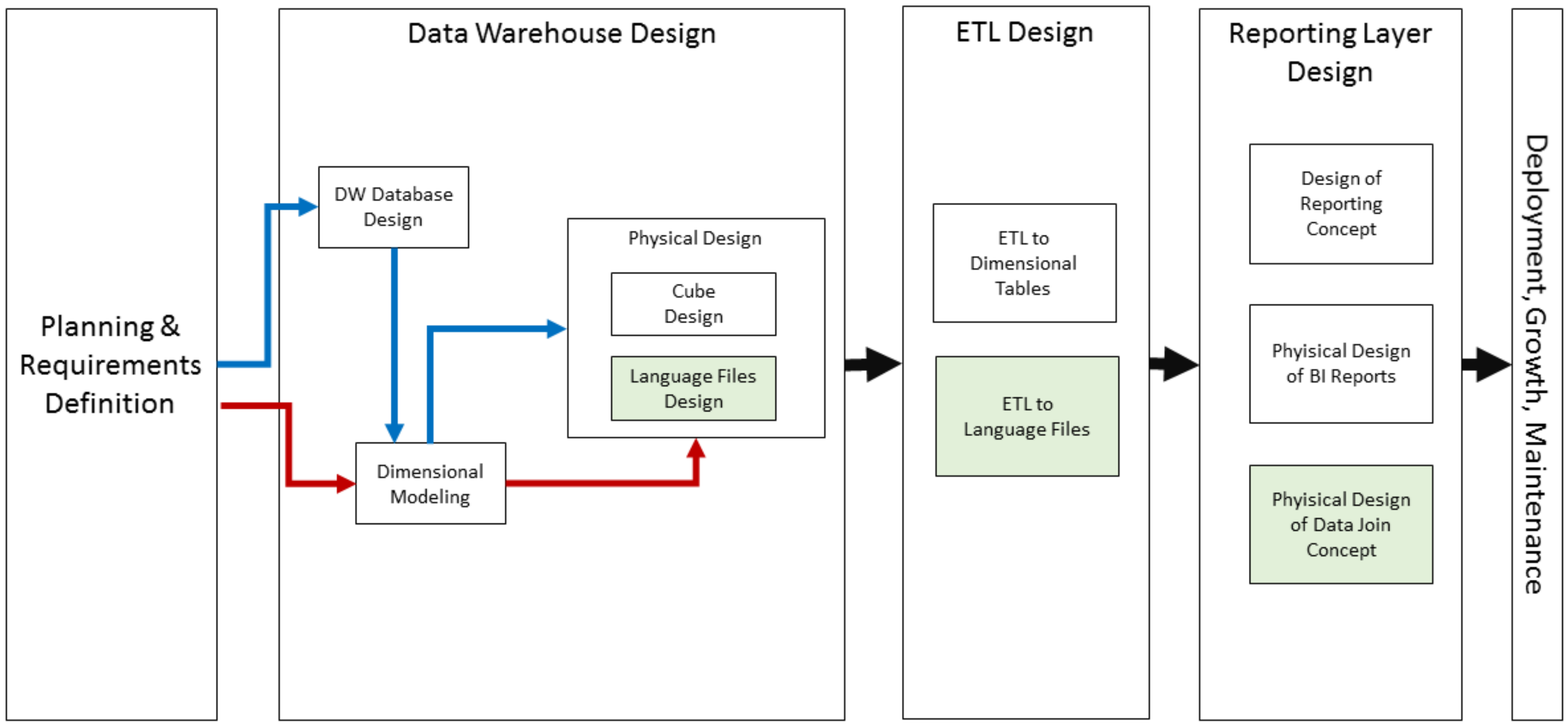


Figure 4-4: Integration of MLED_BI design process

4.5.1.1. Planning and Requirements Design Stage

The first phase in the MLED_BI design process includes Planning & Requirements Definition. Activities in this phase are the same as those proposed by Kimball *et al.* (2008) and are concerned with defining the business requirements to be met through the introduction of BI systems. This phase also covers planning activities in regard to the designing, developing, deployment, and maintenance of BI environment.

4.5.1.2. Data Warehouse Design Stage

The second phase of the MLED_BI design process is Data Warehouse Design (DWD). This phase includes the Dimensional Modelling and Physical Design stages (where Kimball's BI/DWH approach is used), and also includes a DW Database Design (Inmon's approach). In both the Kimball and Inmon approaches, Dimensional Modelling and Physical Design stages are required. The MLED_BI star schema introduces a revision of established approaches to the design of data marts and this has implications for the Physical Design stage. As seen in Figure 4-4, MLED_BI Physical Design belongs to the DWD phase and includes two design processes: Cube Design and Language File Design (shown in the light green box). Cube Design is a required part of DM physical design in established BI design approaches, and includes the physical design of dimension and fact tables. Language File Design is the process introduced by MLED_BI, and covers design of the language files used to support dimension table based DMs with master data descriptions, and design of appropriate web environment to serve as a storage facility for the language files.

4.5.1.3. ETL Design Stage

The MLED_BI design process also affects the ETL design stage. As MLED_BI requires two processes at the DM Physical Design stage, this has to be reflected in the ETL design stage. Established ETL design covers definition of activities, processes, and functionalities to support delivery of data to dimension and fact tables. However, as shown in Figure 4-4, in the MLED_BI approach, the ETL design stage is extended with an ETL-Language Files process. The ETL design phase of MLED_BI includes definition of activities, processes, and functionalities to support data delivery to language files.

4.5.1.4. Reporting Layer Design Stage

Data Marts are used to support effective and fast BI reporting (Inmon, 2005; Kimball *et al.*, 2008); thus, the changes introduced by the MLED_BI star schema have direct implications for the Reporting Layer Design (RLD) phase as well. The first activity in the RLD phase, namely Design of Reporting Concepts, defines the types of reporting artefacts to be used, such as queries, reports or dashboards, and their operation in the BI environment, such as type of delivery processes, or usability aspects. In addition to the activities of Designing Reporting Concepts and physical design of BI reports, MLED_BI introduces a further design requirement (shown in light green box) to support joining data from language files and from cube tables.

4.5.1.5. Other Design Processes

The MLED_BI approach has minimal implications for processes such as providing for deployment, growth or maintenance which form part of the standard design and development activities for BI systems. One issue would be that where MLED_BI is used with a web based reporting system, it will be necessary to define additional folders in the web environment to serve as a storage facility for language files at Deployment Design phase. Not all implementations of MLED_BI will necessarily use a web based environment and this is not seen as a sufficiently large task to be a separate phase in the MLED_BI design processes. Where additional folders that would serve as a storage facility for language files are provided, it is recommended that these folders are used to provide a better overview of the system structure, however, any other existing folder in web environment could serve the same purpose.

4.6. Extending the Reporting Layer Design

4.6.1. Optional Web Component

The application of the MLED_BI design approach in the BI environment has a number of implications in the context of the delivery of BI reports via a web environment. It should be noted that the MLED_BI design approach does not require web based delivery of reports, but the MLED_BI design process includes an optional web reporting module to support web based report delivery as shown in Figure 4-5.

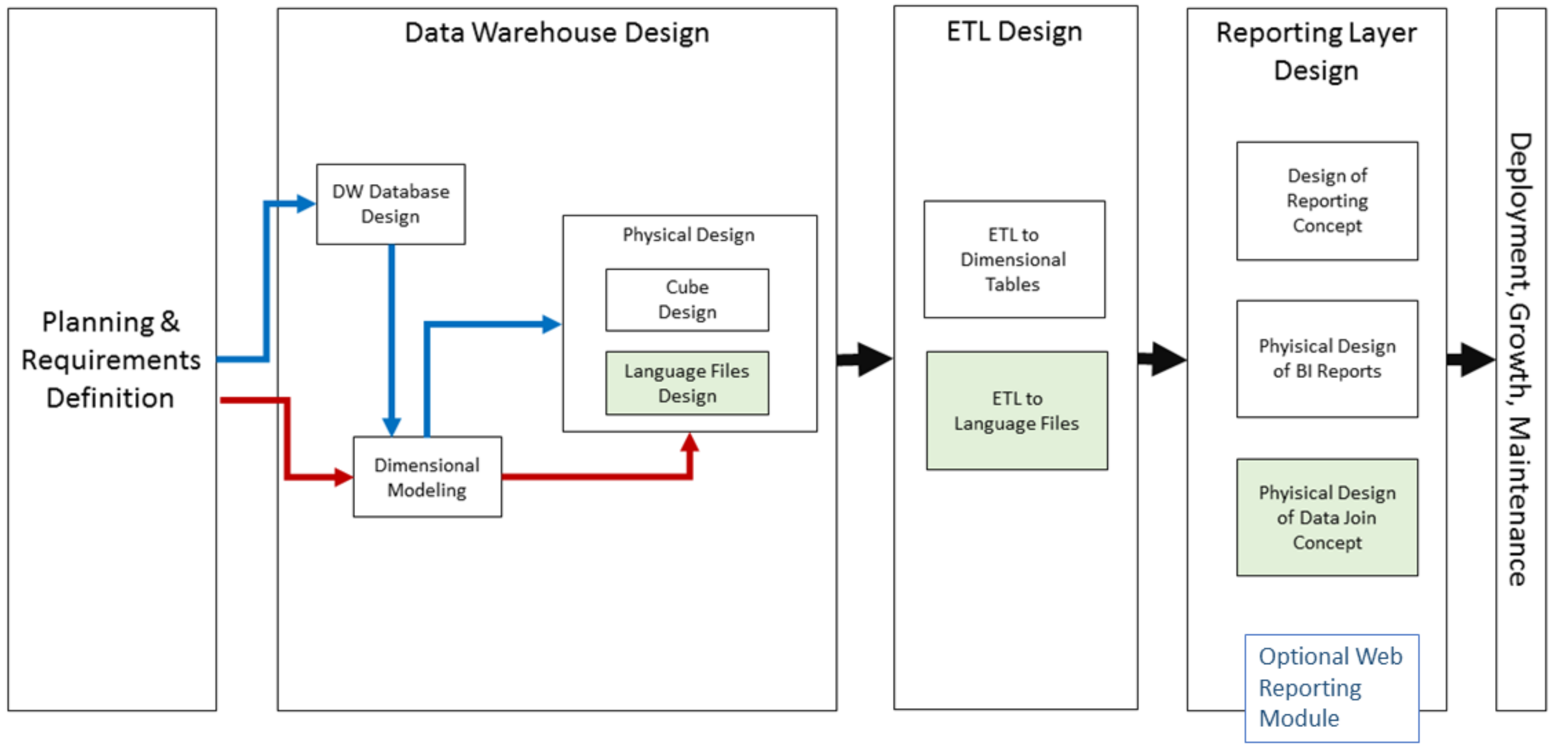


Figure 4-5: MLED_BI design process with optional Web Reporting Module

In section 2.3.2., use of a web based interface was identified as the most appropriate way to deliver BI reports in the context of this research to end users. In established relationally based BI design and development approaches, database query languages, such as SQL or MDX, are used as the tool to deliver the result set to the web application and to create BI reports, including business information descriptions. Also, in established design, all business information (master and transactional data) is stored in database tables representing data marts, thus this is a very straightforward process. However, as MLED_BI proposes saving textual descriptions from attributes and hierarchies outside dimension tables as language files, using database query languages to return a result set to be used in web-based BI reports was not sufficient. A functionality that assigns master data descriptions from languages files to a result set acquired by means of querying the data mart is needed.

The design of the reporting process used a modified version of the dynamic content concept for multilingual websites patented by Kumhyr (2001). Kumhyr proposed the use of content strings identified by content keys with values retrieved from a data store based on language preference. Referencing the MLED_BI star schema design, Kumhyr’s concept, and the idea of using separate HTML language files to overcome issues of ML in web as proposed by Lepouras & Vassiliakis (2000), a design process for web based BI reports that supports the MLED_BI approach, was developed. In the optional web design module, descriptions of attributes and hierarchies are associated with relevant IDs from the dimensional tables during report or query execution (on the fly), depending on the default language or language selected (Figure 4-6).

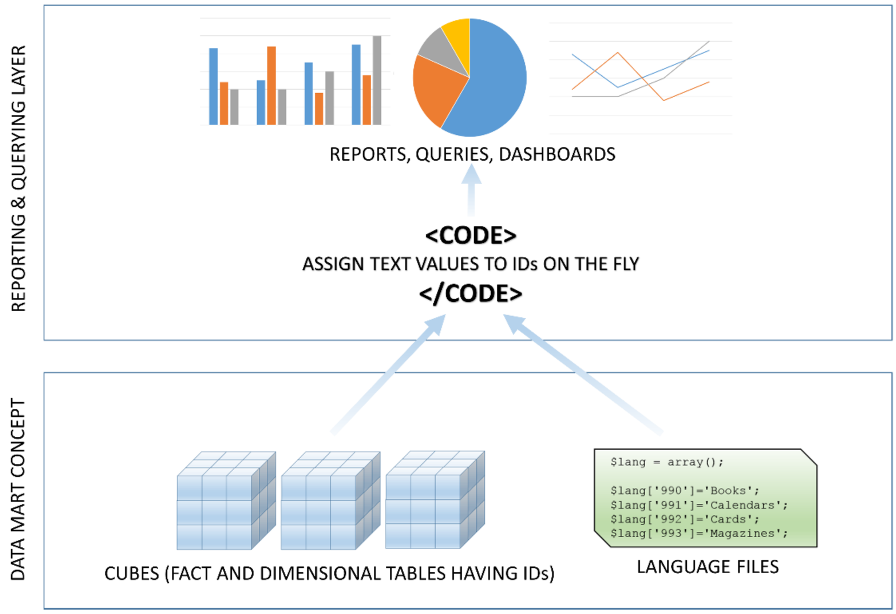


Figure 4-6: The process of assigning descriptions of attributes and hierarchies on the fly

4.6.1.1. End User Data Manipulation in MLED_BI

Section 4.3. suggests a greater facility in data manipulation and easier management of erroneous content directly by end users as one of the requirements for the MLED_BI. As explained in section 4.4.2., in established BI design approaches, any change of master data descriptive content at DM level is not recommended. This has the consequence that as changes are made in source system applications and then transferred to DMs, the use of web-based Content Management System (CMS) to manipulate master data descriptions at DM level in BI systems is not supported. However, MLED_BI is based on the use of IDs, meaning that attribute descriptions and hierarchies are not used as the basis for data aggregation. This makes it possible to change descriptions of master data in language files without creating unrelated or inconsistent data. It also makes it possible to enable new languages in BI reports by simply adding additional language files holding appropriate descriptions for the required language. There is no need to extend existing tables in data mart cubes, ETL processes, or to modify existing BI reports. There is no prerequisite to enable the new language in source systems because the language element resides in the language files. This therefore makes possible the use of a web-based CMS. Multilingual CMS (MCMS) have been identified as an important mechanism for overcoming the limitations of managing multilingual content from the technical perspective (Arefin, Marimoto & Yasmin, 2011).

Existing BI/DWH design approaches support only the following activities in the reporting layer: viewing of reports and associated activities such as drill up/down, selection, filtering, other analytical operations and browsing; re-execution, sharing and changing languages. However, applying the MLED_BI design approach would make it possible to incorporate the MCMS concept into the BI System. In addition to reporting layer activities supported by existing BI/DWH approaches, an MCMS would support the provision of further functionality such as editing descriptions for existing languages directly via a web interface and adding new languages and their variations directly by business users, independently of the existing languages in source systems. Figure 4-7 provides an overview of differences between the established understanding of the BI reporting layer and the functionalities that could be made available as a consequence of introducing the MLED_BI design approach.

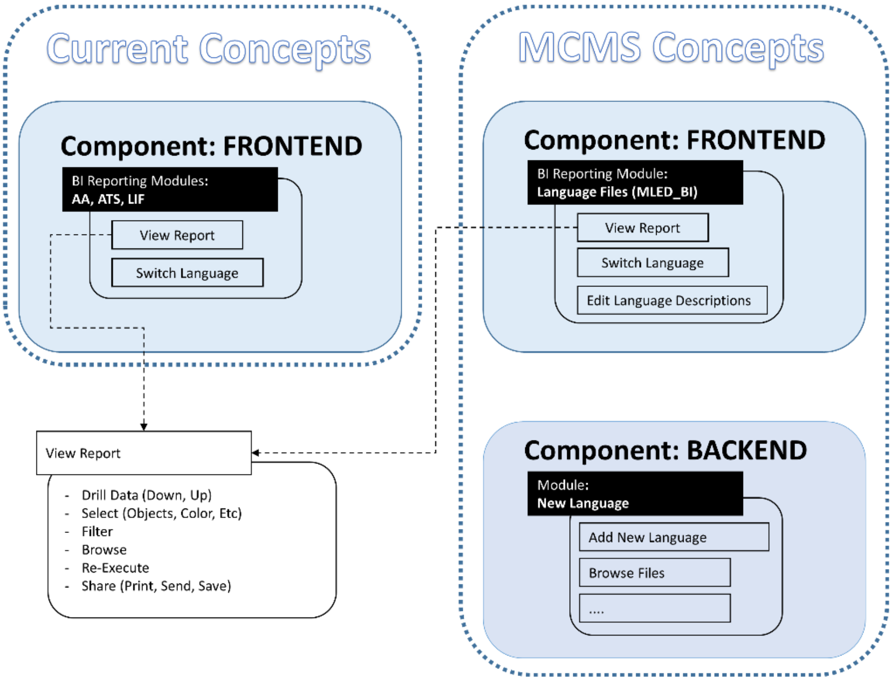


Figure 4-7: Comparison of conventional reporting layer functionality and functionality provided by a MCMS supported by MLED_BI

The MCMS made possible by MLED_BI would include a frontend element that enables execution of the standard activities found in established BI reporting approaches and a backend element that provides additional functionality, allowing the editing and management of content and the inclusion or removal of additional languages. It would be possible to extend the backend functionality with additional modules, for example to enable the execution of ETL processes by business users or to edit various aspects of the web interface. This would lead to the fulfilment of one of the requirements identified in 4.3., namely greater flexibility of data manipulation and easier management of descriptive content.

4.6.2. Extended MLED_BI Design Process

Including the web reporting element and the MCSM in MLED_BI has implications for the Reporting Layer Design phase shown in Figure 4-4. Figure 4-8 shows an extended version of MLED_BI, incorporating the optional web element to show the design of MCSM at the Reporting Layer Design phase.

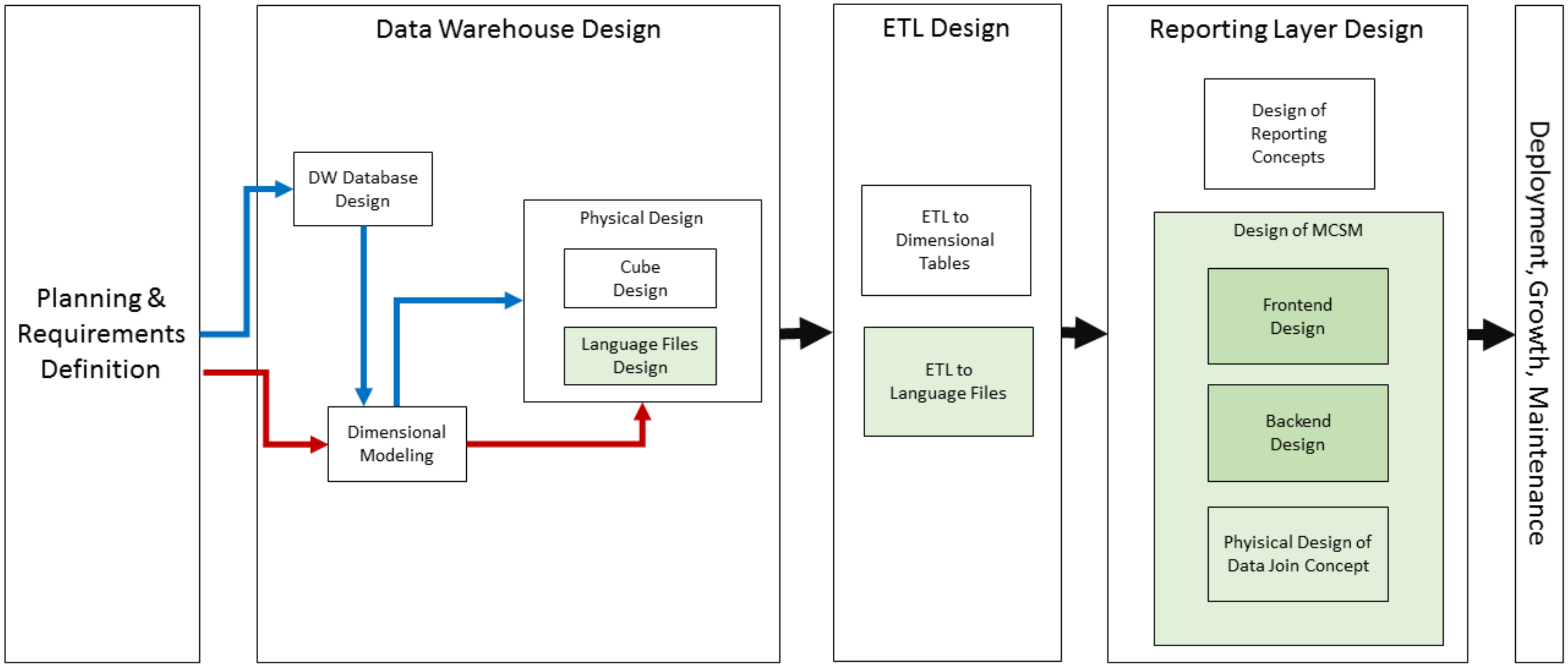


Figure 4-8: Revised MLED_BI design process that includes MCMS concept

Incorporating MCMS design activities in the Reporting Layer Design phase, as shown in 4-8, means that the following elements are now included

- Design of reporting concepts to define the kind of BI reporting artefacts to be used, such as reports, queries, dashboard
- Design of MCMS which includes physical design of data join concept
- Frontend design for MCMS, that is design for delivery of BI reports
- Backend design for MCMS, that is the design of the data manipulation environment

A simplified overview of the MLED_BI environment is provided in Figure 4-9.

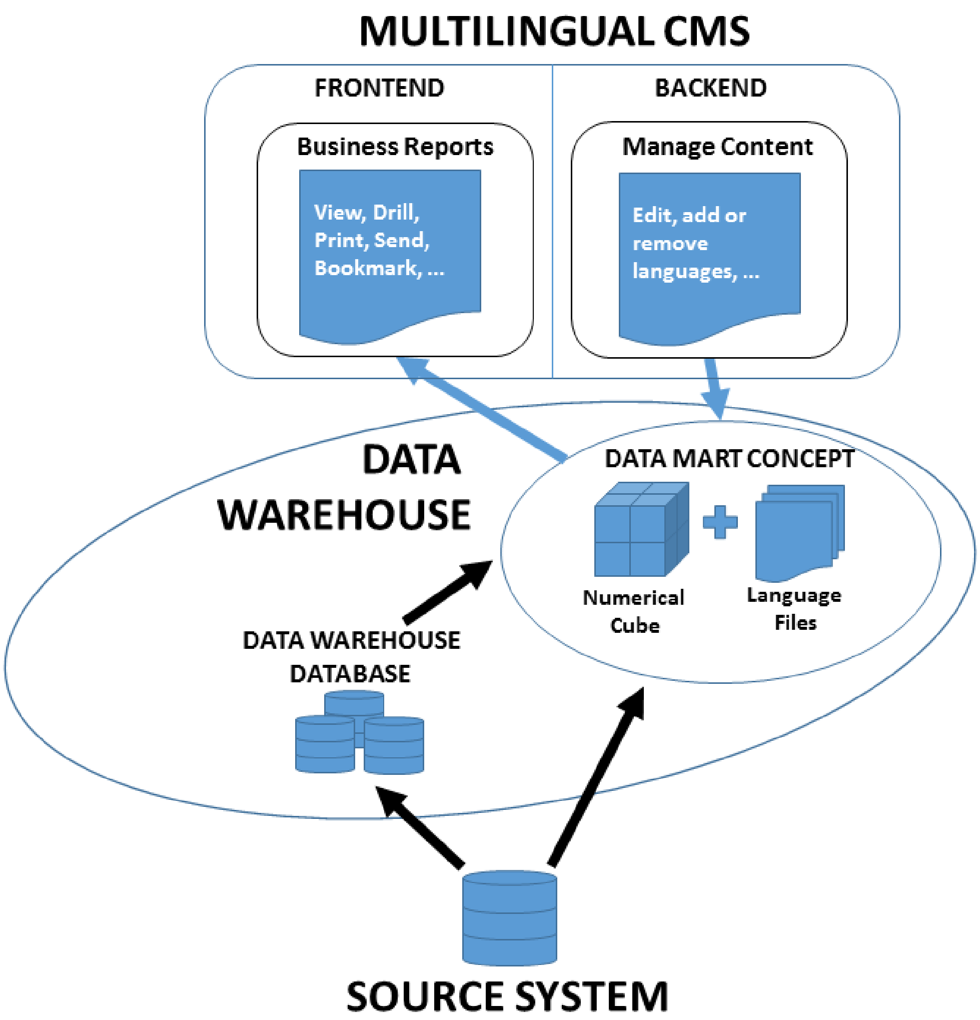


Figure 4-9: BI environment based on the MLED_BI BI design approach

MCMS concepts, such as backend and frontend belong to the presentation layer in the BI environment. The language files used by the MCMS, or any alternative reporting system, belong to the warehousing layer, as do the data marts that hold fact tables and dimensions with numerical values. The concept of source system is the same in MLED_BI as in established BI design approaches. Figure 4-10 shows MLED_BI components mapped to the layers identified in HBIF.

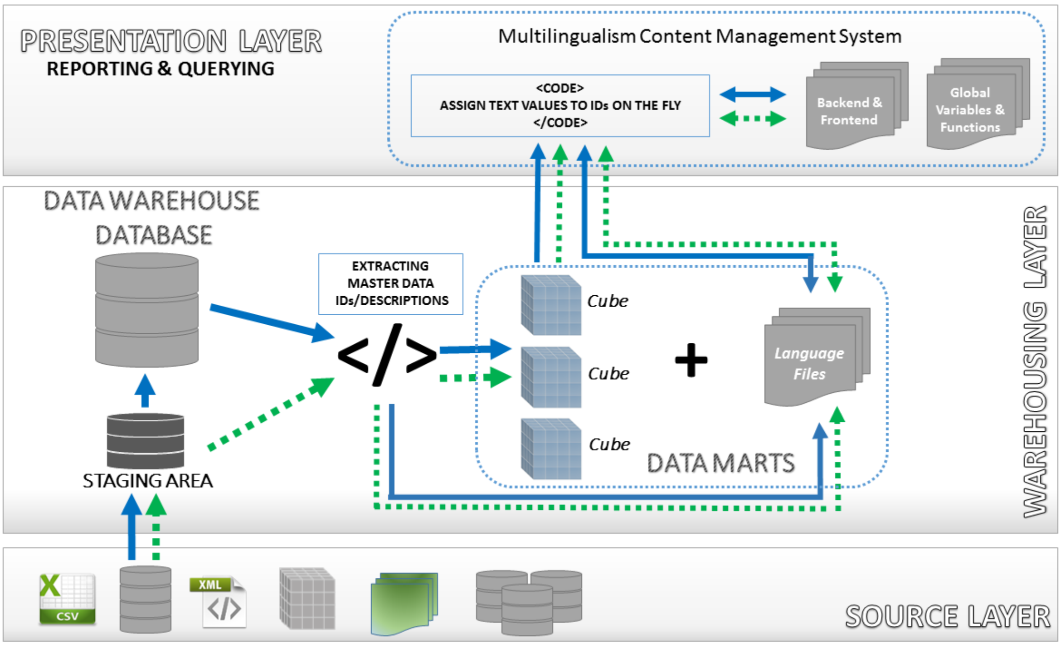


Figure 4-10: MLED_BI viewed against the HBIF

4.7. Comparison of MLED_BI with existing BI design approaches

This section summarises the differences between MLED_BI and established design approaches and outlines the benefits expected to result from adopting MLED_BI as a solution for ML in BI

4.7.1. Overview

As outlined in 4.4.4 and shown in figure 4.4, the MLED_BI approach can be integrated into both the Inmon and Kimball design approaches, which was one of the requirements identified in section 4.3. Figure 4-11 provides a visual overview of the design differences between the Inmon/Kimball approaches and MLED_BI. A tabular comparison of the differences is given in APPENDIX D.

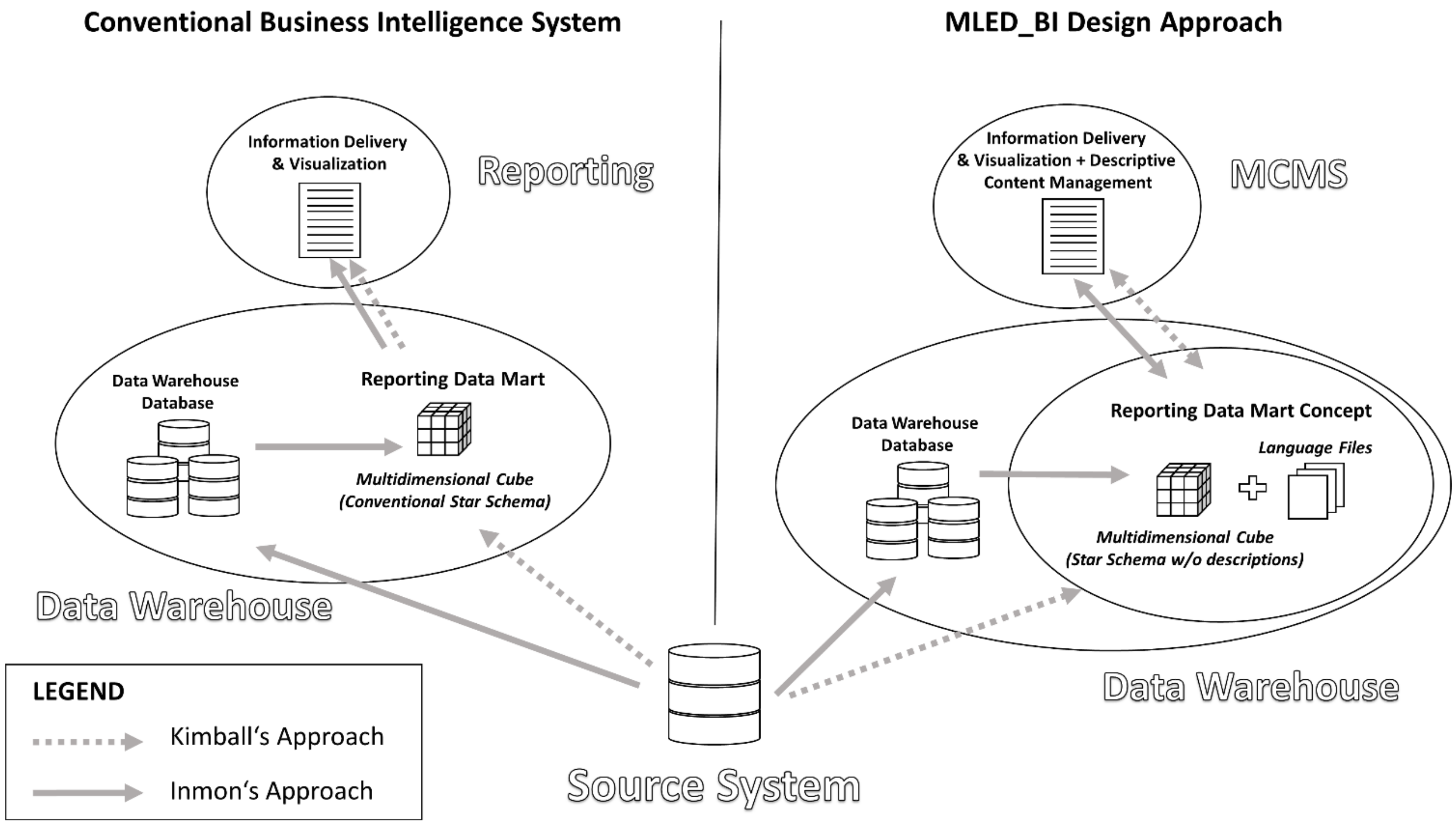


Figure 4-11: Main design-based differences between BI system based on

The MLED_BI approach differs in several respects from established approaches based on a star schema design, as shown in Figure 4-12. Only the initial phase, namely the Requirements Analysis phase, has no differences between MLED_BI and existing BI design approaches. This is due to fact that requirements analysis can be seen as a part of the requirements engineering process (Somerville & Sawyer, 1997) which would be same in all BI environments based on any BI design approach. However, as shown in Figure 4-12, there are differences in the Design and Development phases**.**

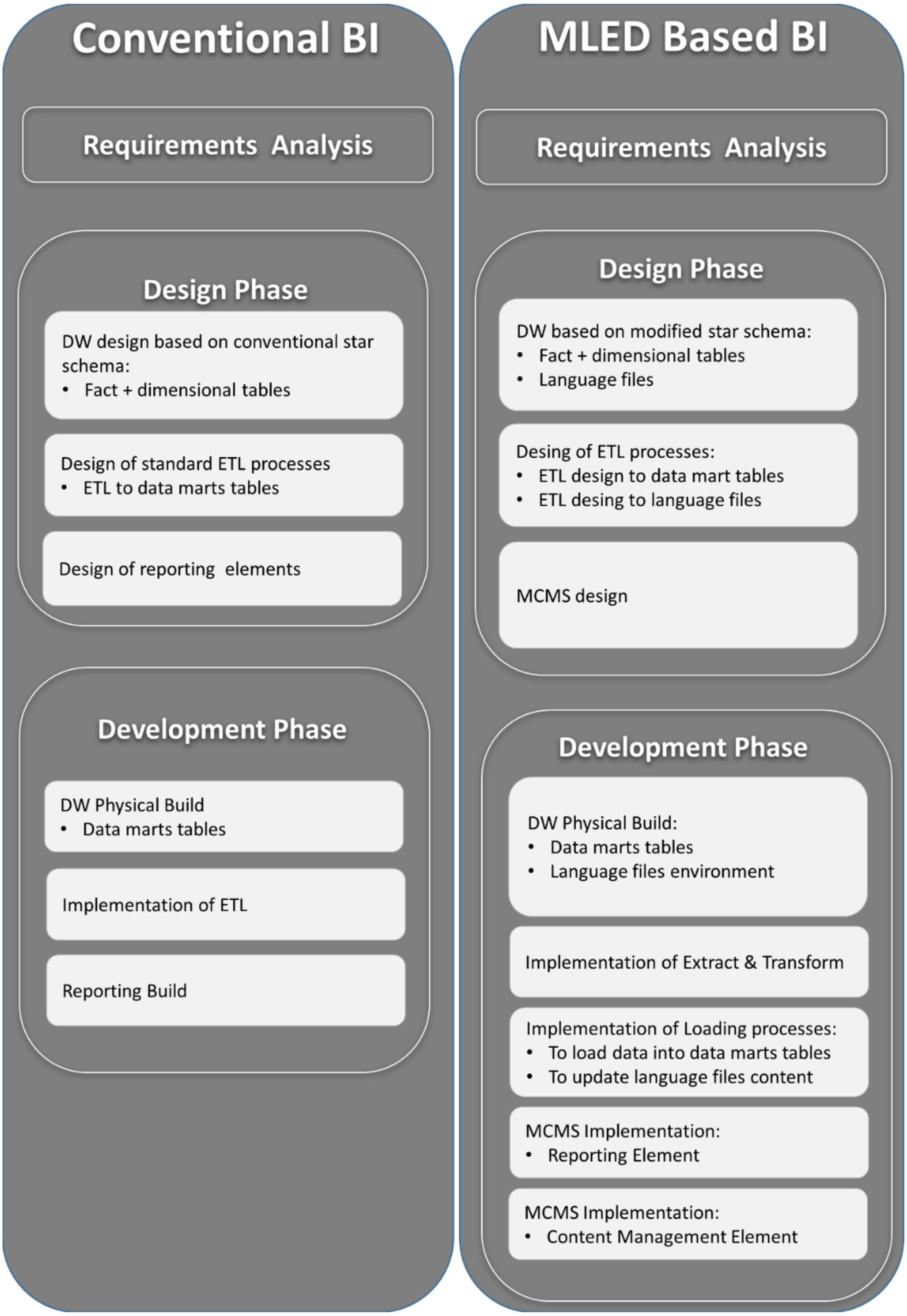


Figure 4-12: Differences in development phases between established BI design approaches and MLED_BI

4.7.2. Design Phase Changes and Implications

The Design phase in established BI implementation includes DW design including design of the star schema and data marts based on the star schema, ETL design and the design of reporting applications. As the established star schema approach proposes storing all business information, master and transactional data in data mart tables, design covers only dimensional and fact tables. ETL design includes the design of ETL processes. The design of reporting applications covers reporting strategies and delivery formats such as queries, reports and dashboards. In comparison, the MLED_BI design phase is more complex. It extends DW design to define data marts based on MLED_BI, which covers design of the fact and dimension tables, plus definition of language files. ETL design is also more complex, reflecting the use of language files. If an MCMS is included in the BI system, the reporting design will include design of content manipulation as well as reporting applications.

4.7.3. Development Phase Changes and Implications

There are also several differences in the MLED_BI development phase. Established BI development has three main phases: DW physical build including the implementation of data mart tables; the second phase is ETL development, the third phase is the build of the reporting component. The development phase in MLED_BI is more complex since the first phase includes not only implementation of data mart tables, but the development of an environment to support language files. The ETL phase is also more complex as it covers not only the implementation of ETL processes which support the loading of business information from source system to data mart tables, but implementation of separate processes to load descriptive information into language files. Implementing an MCSM, if this additional component of MLED_BI is developed, covers not only the implementation of reporting elements in the form of queries, reports or dashboards, but implementation of additional modules to enable the manipulation and management of master data content directly from the web interface.

4.7.4. Reporting Phase Changes and Implications

Reintroducing data independence at the star schema level, incorporating the MCSM concept, and delivering master data descriptions on the fly to BI reports supports the optimal application of ML in BI systems for business users. As there is no dependency between attributes, given that the attributes in the star schema are separated from language representation, there is no requirement in MLED_BI to limit changes of language content exclusively to the source system; thus, it is possible to support content manipulation by business users immediately through the web interface by a MCSM. Supporting new languages or dialects at the reporting layer in MLED_BI, does not require the source system to be modified and extended to enable those languages. This seen as a benefit for companies operating in countries where several dialects or regional language variations are used, but where the source system is in only one language. In this case, an additional language file that holds appropriate descriptions is sufficient to fulfil the requirements of providing an additional language at the reporting layer. This also produces technical benefits: as there is no need to enable new language in the source system, there is no need to modify and extend existing ETL processes, or reporting applications to handle the addition of a new language.

In MLED_BI, a query that delivers the result set to a reporting application will aggregate the result set on the basis of numerical IDs from dimensional tables, while master data descriptions from language files are assigned on the fly; thus, faster BI reports delivery and faster language switch in already executed reports are expected.

MLED_BI will make possible reduced technical and operational maintenance costs. For example, the process for changing erroneous content in a BI report in a system developed on established design approaches, is that when a business user notices an error in a BI report, he/she needs to inform the relevant business department responsible for maintenance of the master data in source system. After the error has been corrected in the source system, it is necessary for the BI or DWH team to be informed of the change and for a request to be made to start the ETL process to transfer the amended data from source system to DW and to the relevant DM. Immediate execution of the ETL process is rare in a business environment especially if re-aggregation of existing data is required; to avoid problems with overload, it is usual to wait until scheduled ETL processes are executed. In a standard BI environment, which provides BI reports on data for the following day, ETL processes are usually executed every 24 or 36 hours. Although most of the ETL processes to load master data might already be scheduled, it is still necessary to inform the BI or DWH team if business information descriptions in source systems have been changed as in some cases, there will be no scheduled ETL processes for specific master data, where these do not change often. After successful execution of the ETL process, a member of the BI or DWH team will inform the original business user of the successful change. Changing descriptive content of master data using existing BI design approaches, is a complex and time-consuming process. The MLED_BI design approach supports a MCMS, making it possible for business users to change content themselves, improving speed and reducing the resources involved.

4.8. Limitations of the MLED_BI design approach

MLED_BI supports a more flexible and efficient design solution to the challenge of supporting the use of multiple languages in Business Intelligence. The limitations of the MLED_BI approach, compared to established design solutions, relate to the greater initial design and implementation effort. As discussed in section 4.7.2. and 4.7.3. and shown in Figure 4-12, MLED_BI requires more resources at the initial design and development stage than established BI design approaches. The requirement to provide language files means the data mart design and development stages are more complex. The initial ETL design and implementation is also more complex as it is necessary to load languages from the language files. However, it is argued that this initial extra design and development cost is outweighed by the performance and extensibility benefits provided by MLED_BI. Similarly, if an MCMS is included, the intial design and development cost for reporting applications is greater but this will be outweighed by the greater flexibility available to end users. It is expected that for larger organisations, the future benefits of using MLED_BI for data management in multilingual environments will justify the initial increased resource demand. For smaller companies, however, and particularly those that do not operate in multilingual environment, MLED_BI might not be an appropriate solution.

4.9. Conclusion

This chapter presented the MLED_BI design approach. Findings from the literature review and the HBIF were used to critically evaluate the BI environment and informed the development of MLED_BI. The context of the investigation was discussed together with the requirements for a multilingual design approach. MLED_BI was described and justified and the MLED_BI design approach was reviewed against established BI design approaches. The expected benefits of MLED_BI were identified as support for immunity from changes through data independence, support for the star schema and enhanced data manipulation and reporting abilities. It was shown that MLED_BI fits well into BI environments developed on existing DW philosophies. The expected limitations of MLED_BI, in terms of increased design and development effort in the initial stages were discussed. The following chapter, chapter 5, describes a proof-of-concept (PoC) for MLED_BI, developed to validate the technical feasibility of the approach and the performance implications.

Chapter 5: MLED_BI Initial Validation and Technical Feasibility

5.1. Introduction

The previous chapter introduced the MLED_BI design approach. This chapter describes the initial validation of the MLED_BI design approach through the development of a Proof-of-Concept (PoC) artefact. A PoC is defined as a small-scale implementation of a proposed approach through the development of an appropriate artefact based on an incomplete design, to demonstrate the feasibility of the approach (Perry, 2011). The PoC discussed in this chapter was developed to demonstrate that the MLED_BI design approach translates into implementation before proceeding to a full validation which uses additional criteria such as user satisfaction. The PoC examines the technical feasibility of MLED_BI and also measures report execution speed compared to performance in other ML in BI design approaches. The chapter describes the design and development process, the use of the report execution metric and discusses what further work is required to validate the MLED_BI design approach.

5.2. Design of the PoC

5.2.1. PoC Requirements

For the initial validation of the MLED_BI design approach, a PoC artefact was designed and developed following the steps of the MLED_BI design processes discussed in section 4.5.1. The MLED_ BI design approach recognises Planning & Requirements Definition (4.5.1.1) as the first phase of the MLED_BI design process. Thus, it was necessary to define appropriate business requirements for the PoC. As identified in the literature review and described in chapter 2, a common use for BI in the retail sector is maximising revenue through the analysis of sales, based on historical information. For this reason, sales data as used in a retail environment was chosen to support the implementation of the PoC artefact. Based on information provided by one of the major retailers in central Europe (personal communication) two of the most important aspects of sales data are products descriptions and sales information (transactions). The Product dimension has been described as one of the most common dimension tables in the BI environment and is one of the two or three primary dimensions found in almost every data mart (Kimball & Ross, 2011; Kimball & Ross, 2013). For this reason, the Product dimension and the Sales fact table were identified as the modelling objects required for the PoC artefact. The development of a BI report that provides an overview of the product sales per product category was identified as an appropriate business requirement for the PoC.

5.2.2. Logical level design

The next step in the PoC design process was the design of an appropriate data warehouse environment. Following the definition of business requirements, the design process requires the modelling of dimensions. The Product dimension and Sales fact table, for the purposes of the PoC were identified as constituting a single data mart, completing the dimensional modelling stage in the DW design. Figure 5-1 shows the structure of the PoC artefact.

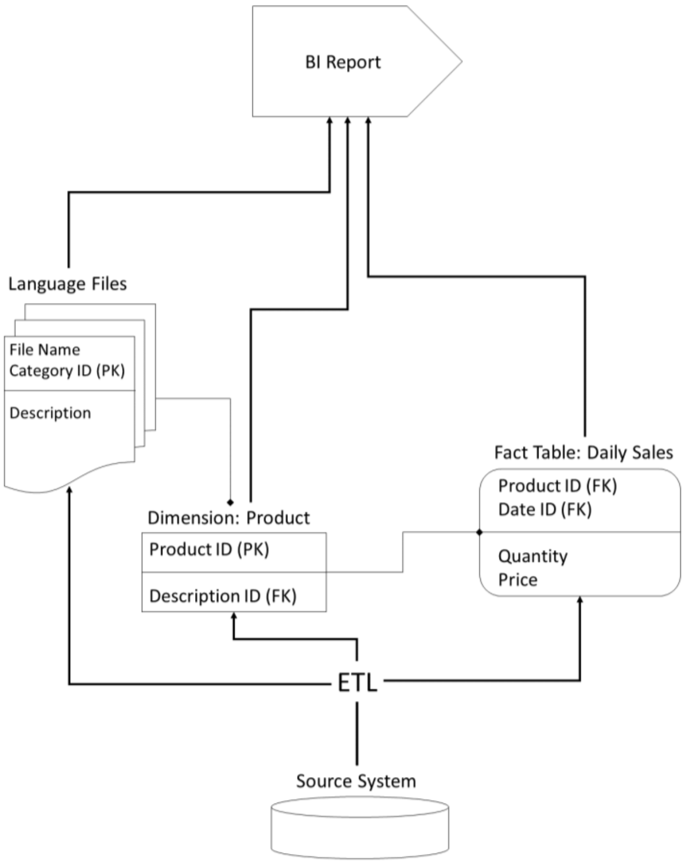


Figure 5-1: Structure of the PoC artefact

As the PoC had only one data mart and does not require replication of the whole source system, the Kimball data warehouse design philosophy was followed. As seen in Figure 5-1, the data mart based on the MLED_BI star schema included design of the Daily Sales fact table, Product dimension table, and appropriate Language file as an integral part of dimension design. Compared to other methods of DM implementation to support

ML in BI (Figure 5-2), the greater complexity of the MLED_BI star schema is evident. While other methods require the extension of existing tables by adding additional columns, or even adding new tables, the MLED_BI star schema requires the design of additional language files to hold master data descriptions.

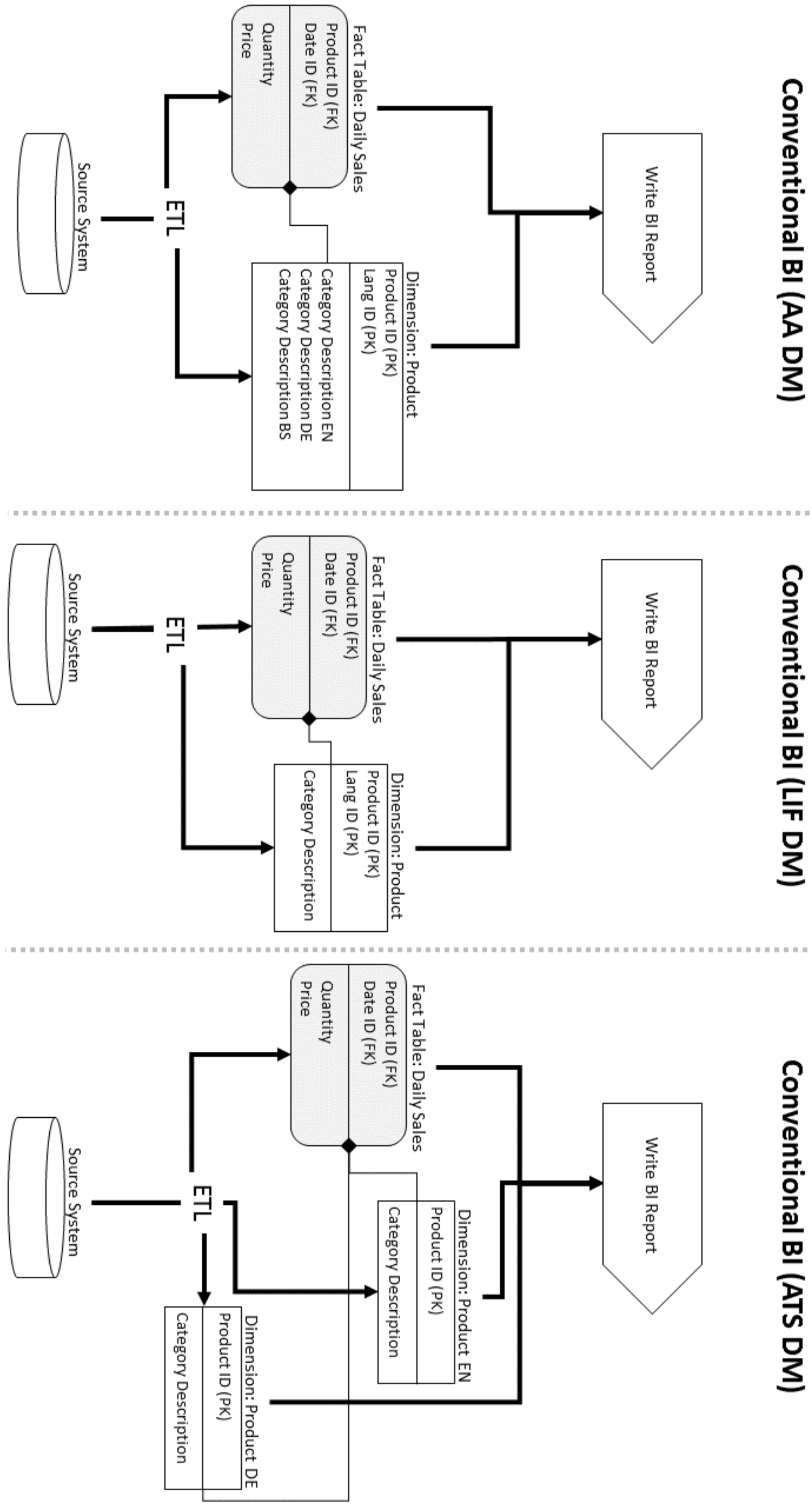


Figure 5-2: Existing ML design approaches

The next phase in the PoC design process included the definition and development of ETL design processes to fulfill the MLED_BI data extraction, transformation, and data loading requirements. As the PoC represented a closed world implementation, that is the incomplete implementation of a real world artefact, it had only one dimension and one fact table and required only three simple ETL processes to extract data from the source files into: a) Daily Sales fact table, b) Product dimension table, and c) Product dimension Language files.

The fourth phase in the PoC MLED_BI design process included reporting layer design. General reporting concepts, such as design of the components that constitute the BI report were designed, followed by the physical design of the BI report. This included the physical design of join concepts to support assigning master data descriptions from language files to master data IDs from dimensional tables, and the development of the visual elements that make up the BI report.

5.2.3. Physical level design

The first step in the PoC physical level design included the physical design of the MLED_BI star schema. As shown in Figure 5-3, the physical design of the star schema in MLED_BI has three main elements: a) Daily Sales fact table holding foreign keys to the Product and Time dimensions and transactional information about quantity and price of products sold, b) Product dimension having only Product primary key and Product Category Description ID, and c) Language files holding actual Product Category Descriptions as an array of variables. While established approaches for supporting ML in BI store all information in data mart tables (shown on the right side of Figure 5-3), in the MLED_BI approach the data mart tables, represented here by the Product dimension and Daily Sales fact tables hold only numerical values, while language files hold any kind of master data descriptions. The Product dimension holds identifier values represented through numeric values, while the fact table holds transactional data in addition to identifier values.

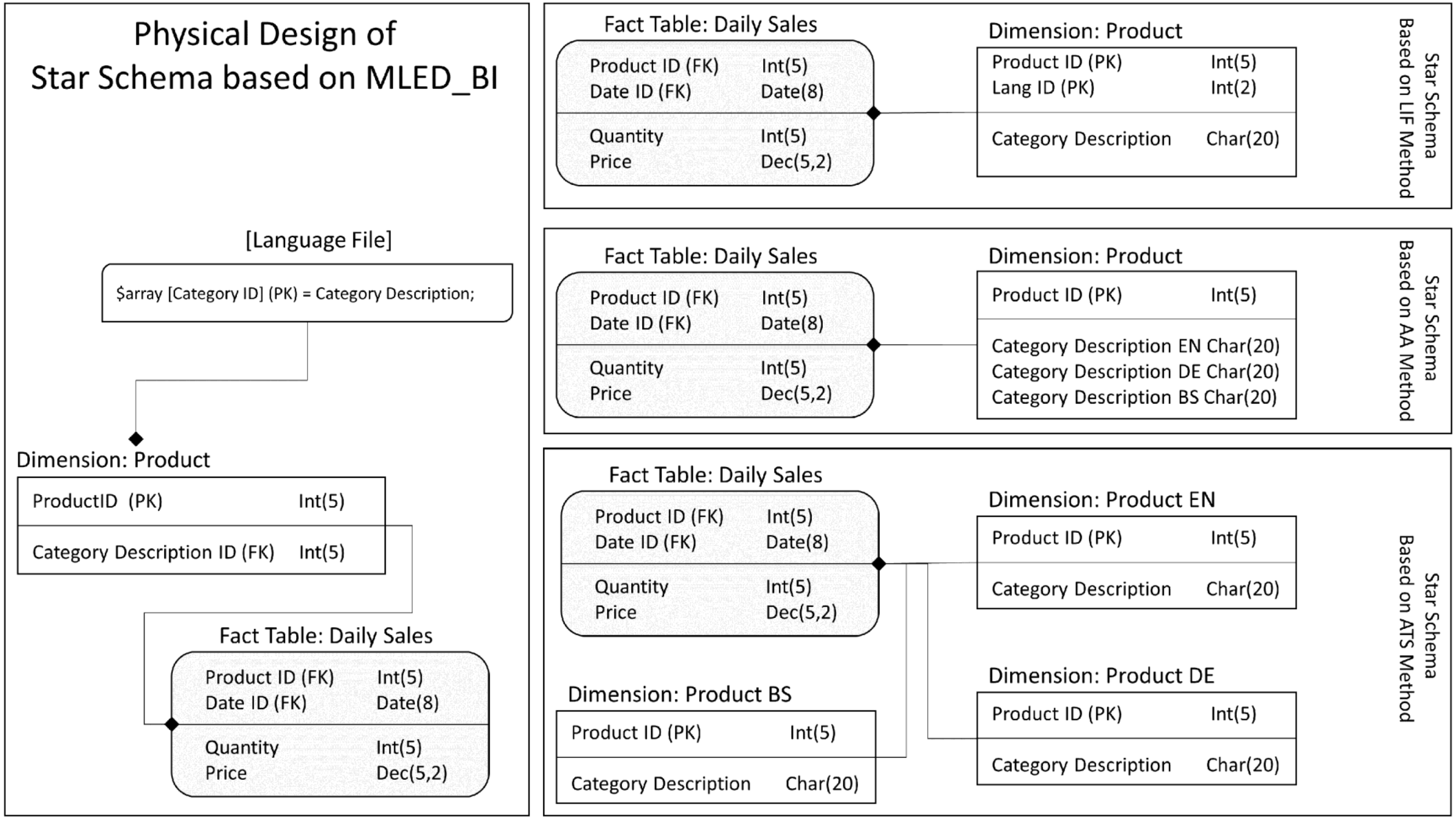


Figure 5-3: PoC star schema compared to other star schema designs to support ML in BI

The MLED_BI design approach means that it is possible to design dimensional Language files in any appropriate implementation form, such as an array of variables, CSV, or XML file. For the purposes of the PoC, a PHP-based array was selected as an appropriate form of physical implementation for the Language file. The PoC supported two languages, English and German. The implications of the concept of one file per language holding all master data descriptions was addressed in the PoC. When the new master data arrives from source system, this is appended to the existing language file array. If there is a need to provide a historical overview of the changes of master data descriptions, additional files or tables could be used to store such information or this could be stored in the existing language files. For example, when the master data description is changed, the old value could be commented out while appending new values to the array.

The next step in physical level design included the physical development of ETL processes. For the purpose of the PoC three separate SQL statements were used to select, extract, transform, and enable the load of data into appropriate objects. Two select statements selected the appropriate data from source system and inserted this into fact and dimensional tables, while the third select statement selected master data descriptions from the source system and generated the appropriate language files in the two languages, English and German.

Section 2.3.2. identified a web environment as the most appropriate environment for the delivery of BI reports and this approach was used as a basis for the PoC BI report design. As noted in 5.2.1., the PoC included a BI report based on product sales per product category. This report was designed as a part of the PoC Reporting layer (Figure 5-4).

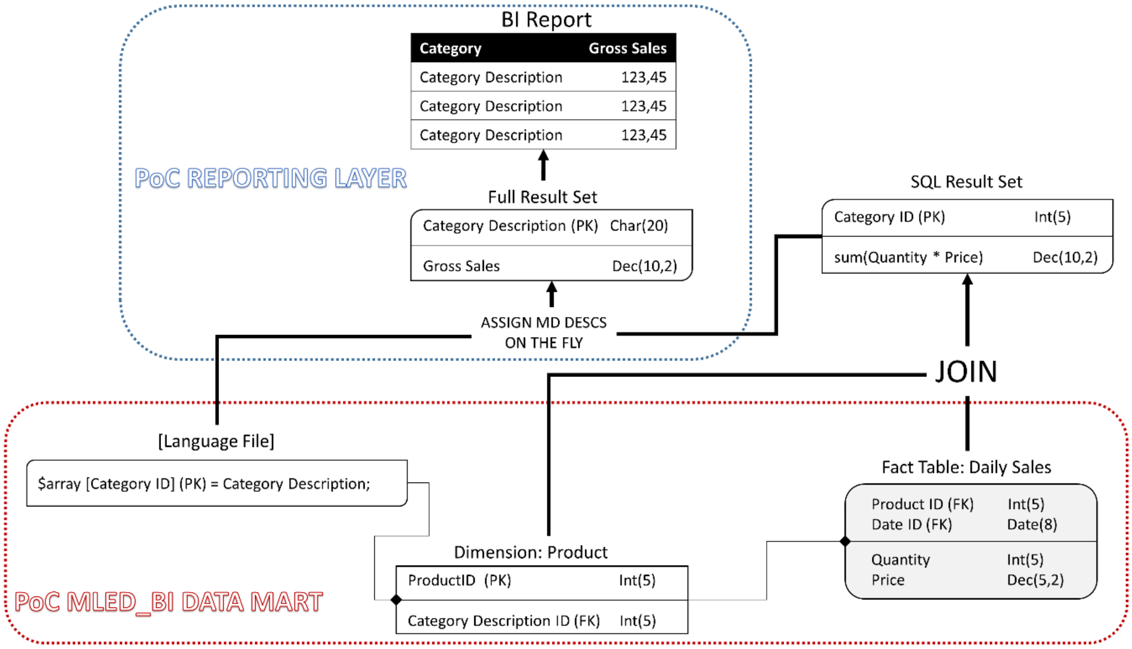


Figure 5-4: PoC physical level design for BI report delivery

The MLED_BI design process uses language files and for this reason, as shown in Figure 5-4, the PoC Reporting layer included the functionality required to assign master data descriptions from the relevant Language file to the SQL Result Set retrieved by querying the tables in the data mart.

Figure 5-5, gives a high level view of the physical design of the PoC and supporting functionality and illustrates the data journey process from the source system to the BI report.

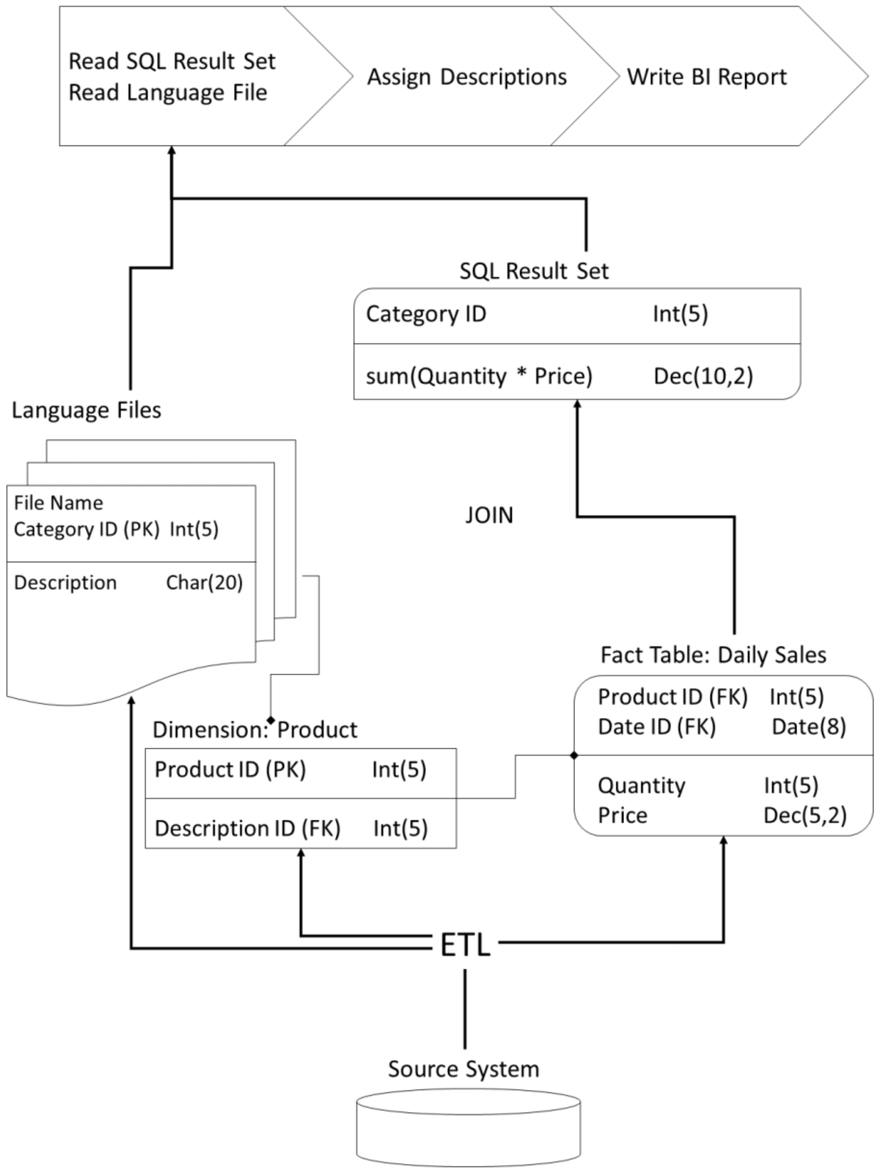


Figure 5-5: High-level overview of the physical design of the PoC artefact

5.2.4. PoC System Environment

Open Source solutions were used to support the implementation. MySQL was used as a supporting database, the web interface was developed in PHP and the web server was Apache HTTP. All applications were installed locally (localhost) on a machine running on the Linux Operating System.

5.2.5. Architecture

The architecture of the PoC is shown in Figure 5-6.

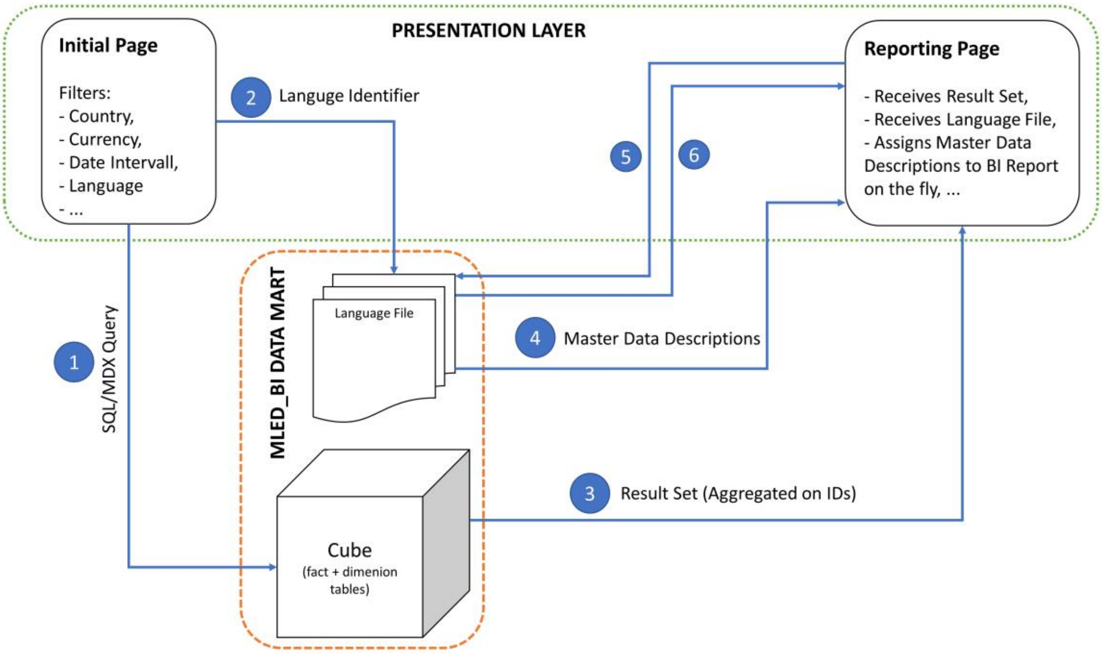


Figure 5-6: Architecture diagram showing the MLED_BI PoC

As seen in Figure 5-6, the Initial Page and Reporting Page elements represent the presentation layer and are used to deliver BI reports to the end user. Cube and Language Files together constitute the data mart developed in accordance with MLED_BI. In the PoC web environment, the user first sees the Initial Page where he/she is able to define the criteria for the execution of the BI report. Following selection of filter criteria, the user executes the report, where the Initial Page carries out two actions: sends the appropriate query to cube and selects the appropriate language file by sending the language identifier. A result set based on the query from step one is then sent to the Reporting Page, followed by delivery of master data descriptions for the language previously selected in step two. The Reporting Page takes the master data descriptions and assigns them to the result set as a part of the data generation process and provides the BI report to the end user. To change the language of master data in the report that has been executed, it is not necessary to re-execute the query, but only to call a different language file (steps 5 and 6 in Figure 5-6).

5.2.6. Designing the test environment

As described in section 1.2., the performance problems encountered when applying existing methods to support ML in BI were one of the motivating factors for this research, especially in the context of information retrieval speed during execution of BI reports in ML environment. Thus, in addition to demonstrating the technical feasibility of MLED_BI through the PoC artefact, it was also necessary to measure performance as compared to a more physically coupled ML design approach. For the PoC validation, the measure used was execution speed although this is only one of a number of possible metrics. A range of other criteria were used for the full validation and evaluation of MLED_BI, as discussed in chapters 6 and 7.

The purpose of the test was to compare report execution speeds between the MLED_BI and other ML design approaches. Figure 5-7 summarises the different strategies and methods.

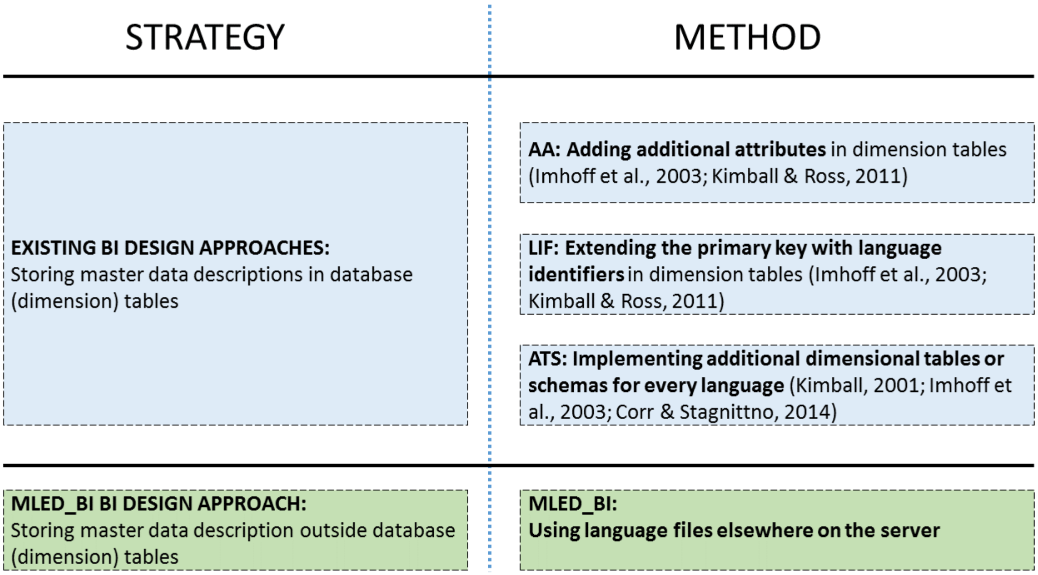


Figure 5-7: Approaches to enable multilingualism in DW data marts

As a first step, existing ML implementation methods from the “storing master data in dimension tables” strategy were evaluated to identify which approach provides the fastest data retrieval. All three approaches were evaluated using SQL queries that returned the same result set. This identified that extending the primary key to include a language identifier (LIF) gave the fastest data retrieval. It was therefore decided in the PoC, to compare the report execution speeds of MLED_BI and LIF since that if this produced useful information, a full comparison against all three methods would be carried out in the next stage of the validation. The design of the BI report and supporting environment used for the LIF data mart implementation method is shown in Figure 5-8.

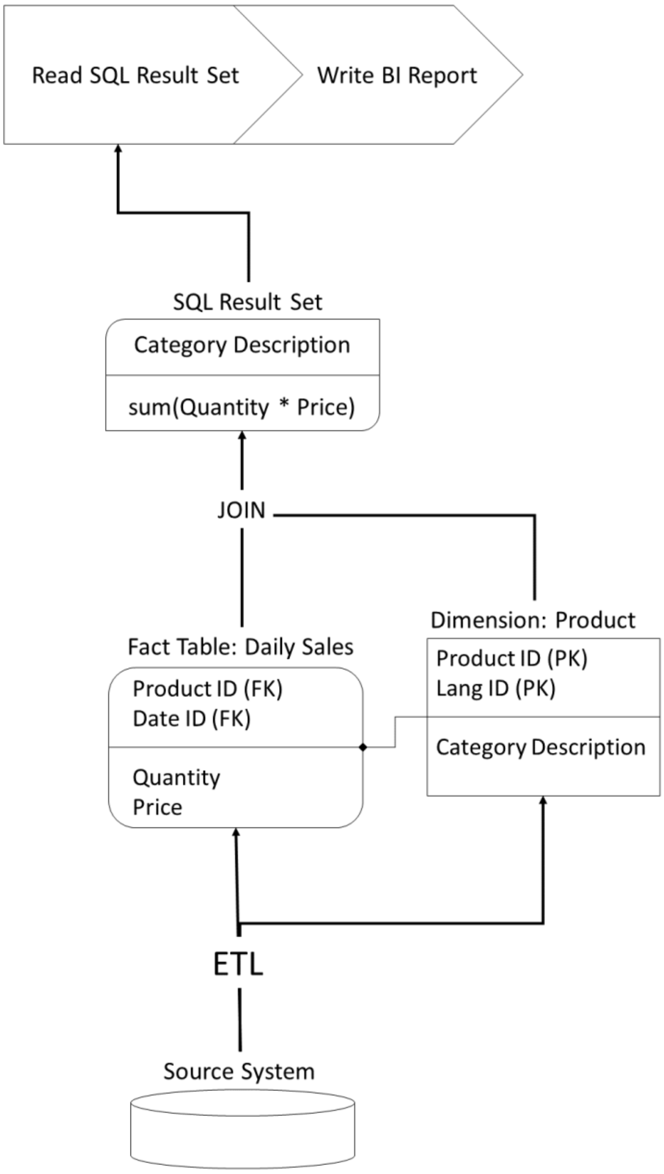


Figure 5-8: Design of the LIF DM

The LIF design process (Figure 5-8) was less complex than the MLED_BI design (Figure 5-5). The LIF approach included the same business requirements and the same dimensional modelling requirements as MLED_BI, simpler physical DM design compared to MLED_BI as it was only necessary to design the cube, simpler ETL processes as there was no requirement to load data to language files, and a simpler process for the reporting layer. From the end user perspective, there were no differences in usability between the PoC BI report based on MLED or the PoC BI report based on LIF. However, as can be seen when comparing Figure 5-6 and Figure 5-9, technical differences exist in the context of the data retrieval process. While MLED_BI requires two separate processes to deliver the BI report (Figure 5-6), the LIF approach requires a single process (Figure 5-9). However, to change language in a previously executed BI report, MLED_BI requires only a call to the appropriate language file without the need to return to the initial page or to re-execute the SQL query for new language. This is not case with the LIF approach or other existing BI ML workarounds which require re-execution of the query.

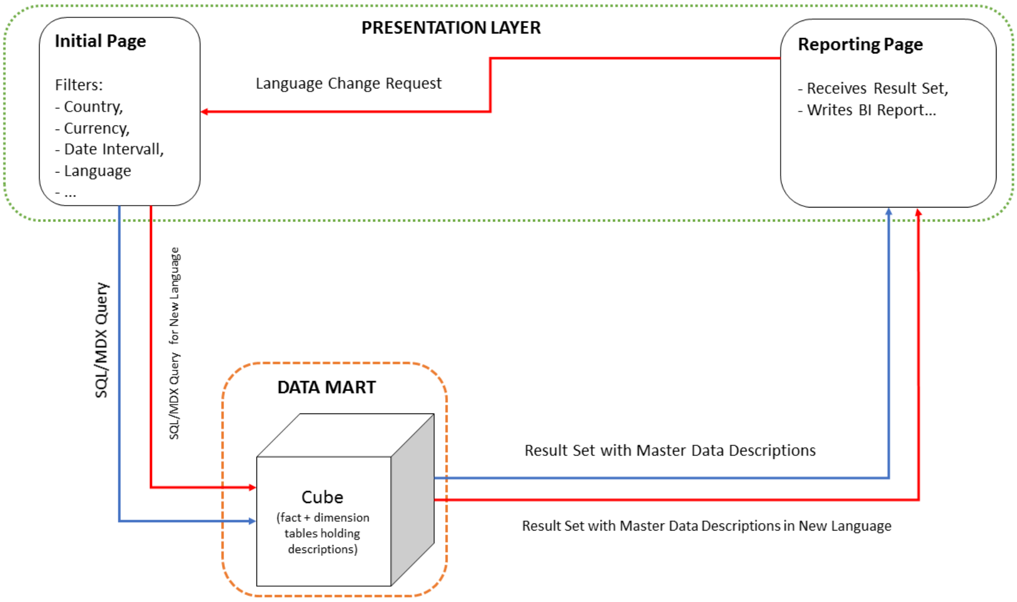


Figure 5-9: Architecture diagram showing the LIF PoC

To enable comparison of report execution speeds, a single web environment was implemented containing a report developed on the MLED_BI approach and a report developed on the LIF approach. Both reports returned the same result set to end users.

5.3. Implementation of the PoC

The PoC has one fact table (*daily_sales_fact_table*) that stores product sales information at daily level (Table 5-1). In this table, the fields *Day* and *Product IDs* served as the *Foreign Keys* and also as a compound *Primary Key* for fact table, while *Quantity* and *Price* held transactional data.

Table 5-1: Daily sales fact table (daily_sales_fact_table)

Field	Type	Part of PK
Day ID	INT (5)	Yes
Product ID	Date (8)	Yes
Quantity	INT (5)	No
Price	DECIMAL (5,2)	No

The table named *product_mled* was implemented to represent the MLED_BI dimension table (Table 5-2), and had the *Product ID* and *Category ID* (short for *Category Description ID*) field as descriptive attributes. *Product IDs* from the fact (Table 5-1) and dimension tables (Table 5-2) are used to establish relationships between the two tables, while *Category ID* was used as a basis for aggregated measurements from the fact table.

Table 5-2: MLED_BI dimension table (product_mled)

Field	Type	Part of PK
Product ID	INT (5)	Yes
Category ID	INT (5)	No

A language file (*language_en.php)* holding a variable array of descriptions of product categories in English was implemented to provide master data descriptions for MLED_BI Product dimension (Figure 5-10).

```
<?php

$lang = array();

$lang['990']='Books';
$lang['991']='Calendars';
$lang['992']='Cards';
$lang['993']='Magazines';
$lang['994']='Journals';
```

Figure 5-10: Part of the language file holding descriptions in English

To support the LIF approach, an additional table named *product* was implemented (Table 5-3). This table had *Product ID* and language identifier (*Lang*) fields as the compound *Primary Key,* and the *Category Description* field as a descriptive attribute. *Product IDs* from the fact (Table 5-1) and dimension tables (Table 5-3) are used to establish relations between those two tables, while the *Category Descriptions* field was used to enable aggregation and to provide meaningful descriptions for aggregated measurements from the fact table.

Table 5-3: Dimension describing the products (product)

Field	Type	Part of PK
Product ID	INT (5)	Yes
Lang	INT (2)	Yes
Category Description	VARCHAR (20)	No

5.3.1. Measuring Report Execution Times

To support the comparison of the design approaches, an index PHP- file was developed and used as the initial page to send the query request to the data mart. It included/called other PHP-based configuration files and stored information regarding processing time later used for comparison in an additional table. A drop down menu was used to enable users to select the required report execution method (Figure 5-11).

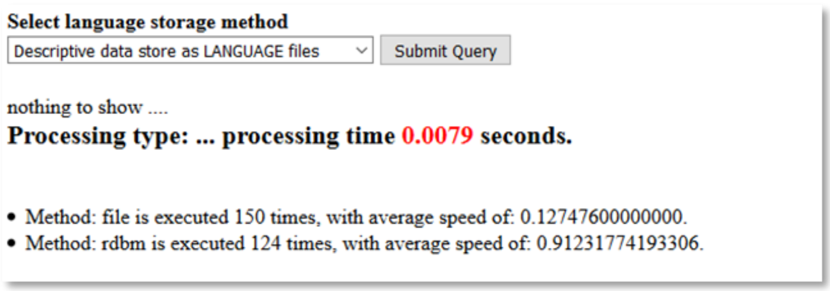


Figure 5-11: Index page including drop down select option

When the MLED_BI language files method is selected, the system checks the language required, retrieves the appropriate language file and then executes the query and retrieves the result set without master data descriptions. Master data descriptions in the required language are then assigned.

When the user selects the LIF approach, where master data descriptions are saved in the dimension table, the system first checks the language requested. If no language is specified, English is assigned. The SQL query is then executed to retrive the result set from the *daily_sales_fact_table* fact table and *product* dimension table. Master data descriptions are taken directly from dimension table according to the language defined in the previous step. After acquiring the result set, the aggregated data is returned for each row togther with master data descriptions.

As one of the aims of the PoC was to measure differences in the execution speed of BI reports based on MLED_BI and on the LIF approach (5.1), a reporting module was implemented in the web environment. The reporting module stores and returns information about individual and average processing times for the BI report together with the report. This was considered more appropriate than manually writing values acquired during processing time.

5.4. Test Results

The initial MLED_BI PoC demonstrated the technical feasibility of the MLED_BI approach in that the implementation showed that the use of language files was a possible solution for the challenge of ML in BI. The PoC implementation also examined the performance benefits of the MLED_BI data independence approach as compared to performance using the LIF approach. The metric used to evaluate performance in the PoC was report execution speed.

The PoC was tested with 107.768 records in the *daily_sales_fact_table* fact. The Table *product_mled* used to support MLED_BI design approach, and the table *product* used to support the LIF approach both held 97 records in their tables. In terms of real world BI systems, this is a trivial data set but this was considered sufficient for initial testing and validation. Descriptions of the product categories for the LIF method where the primary key is extended to include a language identifier were saved directly in the *product* table itself. In the case of MLED_BI PoC, the table *product_mled* held only IDs of categories, while descriptions were saved in an array of variables stored on the web server as language file. The language file held the same type and volume of category descriptions as those stored in *product* table in the LIF approach.

To execute the BI report in the MLED_BI environment, *quantity* and *price* from the fact table are multiplied and aggregated using *Category ID*. Master data descriptions of categories are then assigned to the BI report from the appropriate language file during report execution. During execution of the BI report based on the LIF DM implementation method, the same multiplication operation is carried out; however, the descriptions of categories in the *product* dimension table are used directly for aggregation purposes, and to provide descriptions for BI final BI report.

The BI reports from PoC artefact were tested on the localhost web server using the Firefox (45.0.2) web browser. Figure 5-12 shows the results when executing using the MLED_BI approach method, while Figure 5-13 shows the results using the LIF approach.

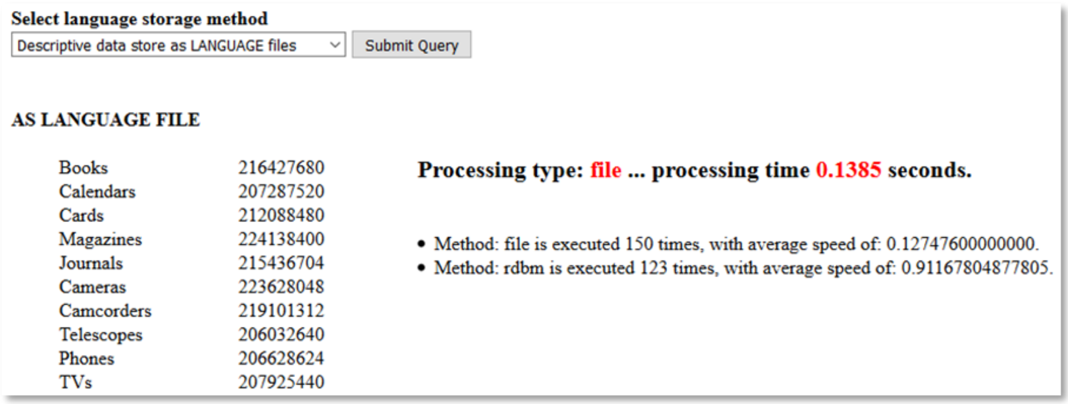


Figure 5-12: Screenshot of PoC report based on MLED_BI

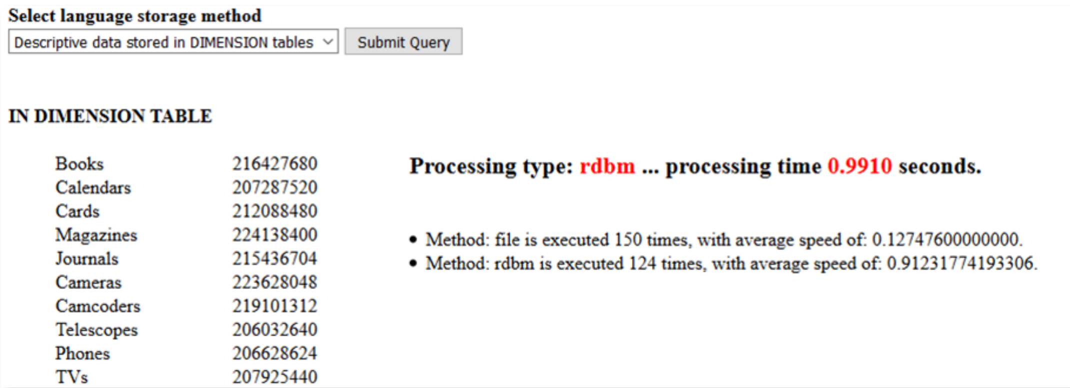


Figure 5-13: Screenshot of PoC report based on LIF

Results from the PoC showed an enviable improvement in BI report execution speed performance when using the MLED_BI design approach. The average execution time of the MLED_BI based BI report was on average 7.6 times faster than using LIF method to support ML in the star schema. The BI reports used in the PoC were executed 300 times each: 150 times for MLED_BI and 150 times for the LIF method: the average execution time for the MLED_BI approach was speed of 0.12 seconds; the average execution time for the LIF approach was 0.92 seconds.

5.5. The strengths and limitations of the PoC approach

The PoC artefact demonstrated the technical feasibility of MLED_BI and showed a performance enhancement compared to an implementation using one of the existing ML design approaches. However, the PoC presented in this chapter was based on a small-scale implementation with an incomplete design and used a single metric for test purposes. It was not possible to assess whether the results from the PoC would scale in a real world environment. To fully evaluate the usefulness of the MLED_BI design approach, a more complete implementation in a simulated a real-world environment was required, using a wider range of evaluation criteria. One of the motivations for this research was the problems experienced by business users when interacting with existing approaches to support ML in BI. Business user input is therefore required as part of the full validation of MLED_BI. As business users interact with the BI environment only at presentational layer, through applications such as reports, dashboards, or queries, the presentation layer is the focus of evaluation with business users. Validation is also required in the context of the strengths and limitations of the functional implementation and this requires input from BI domain experts who can evaluate issues such as performance, maintenance and usability at both the presentation and the data warehouse layers, together with any implications for the source layer.

The PoC was developed on the basis that speed of execution of BI reports would be a sufficient metric for the PoC stage, but would not be sufficient to assess performance in the full implementation and that other technical measurements, as well as user assessments, would be required. The results from the PoC made it possible to move to a full validation but a pre-requisite for the full validation of the MLED_BI approach was identification of the metrics, technical and end user, required for the full validation.

5.6. Conclusion

This chapter described the initial validation of the MLED_BI design approach through the development of the PoC artefact. It was shown that the MLED_BI approach is technically feasible and translates into functional implementation. It was also shown that the MLED_BI PoC implementation provided faster report execution compared to the LIF approach. These findings enabled the research to move to the second phase of validation, a large scale implementation in a simulated real world environment. It was identified that full validation would require input from business users and domain experts and would require an appropriate set of evaluation criteria. The following chapter, chapter 6, discusses the process of identifying relevant evaluation criteria and developing an evaluation tool which measures the success of changes to existing BI environments.

Chapter 6: Development of an Evaluation Tool and Validation Design

The work described in this chapter was presented at CONFENIS 2016, (International Conference on Research and Practical Issues of Enterprise Information Systems), Vienna, Austria, and published in *Lecture Notes in Business Information Processing (Springer)*. The conference presentation then became an invited paper in the *Journal of Management Analytics (Taylor & Francis).* The conference paper was the most downloaded paper from CONFENIS 2016 with more than 200 downloads via SpringerLink at the time of writing and in the same time it has achieved more than 500 reads on Researchgate. The evaluation tool developed in this chapter addresses a gap in the literature and is seen as one of the minor contributions of the thesis.

6.1. Introduction

The aim of the MLED_BI design approach is to enhance existing BI environments by improving support for multilingualism in data warehouse based business intelligence. Chapter 5 identified the need for a comprehensive set of evaluation criteria which would include technical and end user metrics. As discussed in this chapter, a suitable and comprehensive evaluation tool, able to measure the success of changes to reporting in the BI environment, could not be found in the literature. This chapter describes the development of an evaluation tool which was created to support the overall validation of the MLED_BI design approach. The chapter also identifies relevant users, from the perspective of evaluating changes to BI systems and defines what is meant by satisfaction in this context from both a user and a technical perspective.

Improved decision-making (Popovič, Turk & Jaklič, 2010), competitive advantage (Thamir & Poulis, 2015; Marchand & Raymond, 2008), increased profit and efficiency (Olszak & Ziemba, 2006) are some of the potential benefits of improving the performance of analytical applications, such as Business Intelligence (BI), within an organisation. However, to measure the success of changes to existing applications, it is necessary to evaluate the changes and compare satisfaction measures for the original and the amended versions of that application. The PoC artefact discussed in chapter 5 demonstrated the technical feasibility of MLED_BI and showed performance enhancement. However, the PoC was a small-scale implementation of an incomplete design which used only technical feasibility and speed of execution of BI reports as evaluation criteria. For a comprehensive validation of the MLED_BI design approach, the development of a complete BI system simulating a real-world environment was needed and it was also necessary to develop a tool to support validation of the implementation by BI domain experts. One of the motivations for the development of MLED_BI was the problems experienced by business users when interacting with BI systems, thus it was also necessary to validate the MLED_BI implementation with business users. This in turn produced a requirement to develop an appropriate evaluation tool encompassing a wide spectrum of relevant metrics to evaluate the MLED_BI approach in a real world environment.

6.2. Measuring success in BI

6.2.1. Defining success in the BI context

MLED_BI has a theoretical basis in the BI and DWH design literature and complies with the two most widely used philosophies for the development of DW and BI systems, namely the Inmon and Kimball approaches; MLED_BI can be seen as an extension to existing development approaches. In this context, developing an instrument to measure user satisfaction with the MLED_BI approach, is actually concerned with measuring the success of alterations to an existing BI environment. For the purposes of this thesis, the definition of success provided by Işik, Jones & Sidorova (2013) is adopted and success is understood as the positive benefits of BI reporting which the organisation could achieve if modifications were implemented to the BI environment. BI modifications are considered successful only if they provide or improve a positive reporting experience for users. The focus in the evaluation of MLED_BI is to determine whether the MLED_BI approach provides a positive and improved experience for users.

There is a need to define the criteria to be used as measurements of success in this context. DeLone and McLean (1992) proposed the well-known D&M IS Success Model to measure Information Systems success. According to Sabherwal, & Chowa (2006), the D&M model was based on a comprehensive literature survey but was not empirically tested. In their initial model (DeLone, & McLean, 2003), which was later slightly amended (Petter, DeLone & McLean, 2013), DeLone and McLean aimed to synthesize previous research on IS success into coherent clusters. The D&M model, which is widely accepted, considers the dimensions of *information quality*, *system quality*, *use*, *user satisfaction*, *organisational* and *individual aspect*. The most current D&M model provides a list of IS success categories identifying some examples of key measures to be used in each category (Petter, DeLone & McLean, 2013); for example, the category *system quality* could use measurements such as ease of use, system flexibility, system reliability, ease of learning, flexibility and response time; *information quality* could use measurements such as relevance, intelligibility, accuracy, usability and completeness; *service quality,* measurements such as responsiveness, accuracy, reliability and technical competence; *system use*, measurements such as amount, frequency, nature, extent and purpose of use; *user satisfaction* could be measured by a single item or via multi-attribute scales; *net benefits* could be measured through increased sales, cost reductions or improved productivity. To identify the IS success variables and critical success factors relevant in the context of changes in BI reporting, there must be a focus on BI activities, phases and processes.

Lönnqvist and Pirttimäki (2006) propose four phases to be considered when measuring the performance of BI: (1) *identification of information needs*, (2) *information acquisition*, (3) *information analysis*, and (4) *storage and information utilisation.* The first phase considers activities related to discovering the business information needed to resolve problems, the second relates to the acquisition of data from heterogeneous sources, and the third to the analysis of the data and conversion to information products (Lönnqvist & Pirttimäki, 2006). The first three phases are outside the scope of this chapter as the focus is on BI reporting. However, the fourth phase, namely *storage and information utilisation*, is relevant to the discussion on changes in BI reporting as this phase is concerned with the storage, retrieval, sharing and use of knowledge and information through BI technologies, such as queries, reports and dashboards. Those aspects cover two clusters of measurements, those relevant to *business/end-users satisfaction*, and those relevant to *technical functionality*.

6.2.2. Business/End User Satisfaction

User satisfaction is one of the most extensively used measures in the evaluation of IS systems (Sedera & Tan, 2005), is widely recognised as a critical measure of IS success (Dedić & Stanier, 2016b; Rahman, 2013; Petter, DeLone & McLean, 2013; Hou, 2012; Dastgir & Mortezaie, 2012; Davison & Deeks, 2007; DeLone & McLean, 2003, 1992), and has been used as a surrogate measure of IS effectiveness (Gatian, 1994). User satisfaction has been defined as “an affective attitude towards a specific computer application by someone who interacts with the application directly” (Doll & Torkzadeh, 1988, p.261). For example, positively influencing the end user experience, such as facilitating easier decision-making, can lead to a positive increment in user satisfaction. User satisfaction is also seen as the sum of feelings or attitudes of a user toward factors relevant for a specific situation (Bailey & Pearson, 1983). In a BI context, Data Warehouse (DW) performance needs to be acceptable to the end user community (Rahman, 2013). To be regarded as successful, BI solutions, such as reports and dashboards, need to meet criteria that lead to positive user satisfaction.

6.2.2.1. Identifying Users

It is important to define what is meant by user in this context. Davis & Olson (1985) distinguished between two groups of users: users making decisions based on outputs from the system, and users entering information and preparing system reports. According to Doll & Torkzadeh (1988) end-user satisfaction in computing can be evaluated in terms of both the primary and secondary user roles, thus, they merge these two groups defined by Davis and Olson (1985) into one. However, in modern BI and DW, user is a more complex concept than that defined in the previous century and in developing the evaluation tool, it was necessary to define the users, and user roles, which would be relevant when assessing whether reporting changes led to user satisfaction.

Following an analysis of staff roles in eight large European companies using BI, and based on feedback from BI and DW domain experts, 4 groups and 10 different user roles relevant to BI were identified. For consistency, as roles are named differently in different companies, the categorisation is based on activities. Table 6-1 presents user groups and roles, and descriptions of associated activities. Measuring user satisfaction with BI reporting processes requires insights from those using reports to make business decisions or complete operational activities and requires technical elements to be taken into account. Thus, the user roles Management and Business Users, together with the Technical User group, are relevant to the evaluation of the effectiveness of changes to the BI environment.

Table 6-1: User groups, roles and relevant activities in Business Intelligence

User group	User role	Activities
Business	Management	- Use reports & dashboards to make decisions at enterprise level;
Business	Business Users	- Use reports & dashboards to make decisions at lower levels (departments, cost centres, etc.); - Use reports & dashboards for operational and everyday activities (controlling, planning, etc.); - Control the content of the reports & dashboards and require changes or corrections if needed; - Optimal participation in Business Intelligence Competency Centre (BICC) activities;
Organisational	Key Users	- Communicate requirements of Business Intelligence (BI) reports and systems between business and technical groups of users; - Communicate BI project implementation phases between business and technical groups of users; - Actively participate in BICC activities;
Organisational	BI Team Manager	- Organisation, motivation and further development of BI team; - Anticipatory care of new projects and technologies in the field of BI; - Monitoring and optimization all BI Team quality-related processes and procedures; - Control cost of BI resources and work on profit maximisation;
Organisational	Project Manager	- Communication, organisation and supervision of the BI project implementation phases with technical users;
Conceptual	BI Architect	- Define BI strategy and processes at enterprise level; - Analyse and design architecture of BI environment; - Ensure compliance of BI architecture with other enterprise systems; - Initiate, develop and/or lead BICC;
Conceptual	BI Solution Designer	- Analyse and design BI system components and applications; - Communicate design of BI system components and applications to Project Managers and technical users for further implementations; - Define development standards and naming conventions in cooperation with other technical users, such as BI Product Manager; - Actively participate in BICC activities;

Technical	BI Application or Product Manager	- Manage BI applications from the technical perspective, such as dealing with processes, upgrades and other technical issues; - Work on continuous improvement to BI applications and systems, such as analysing current problems and identifying opportunities for optimization; - Implement objects, modules, functions and procedures required by BI system or other BI applications; - Actively participate in definition of development standards and naming conventions from software or tool perspective; - Optional participation in BICC activities;
Technical	Report Developer	- Develop reports according to Solution Designer specification; - Communicate implementation status with BI Solution Designer, Project Manager and BI Application or Product Manager; - Actively participate in definition of development standards and naming conventions from Reporting perspective;
Technical	Data Warehouse Developer	- Analysis, design and implementation of Data Warehouse (DW) environment, such as ETL processes, transformations, staging areas and data marts; - Communicate implementation status with Report Developer, Project Manager, BI Solution Designer and BI Application or Product Manager and other IT people responsible for source systems; - Actively participate in definition of development standards and naming conventions from DW perspective;

6.2.2.2. Measuring End User Satisfaction

Doll and Torkzadeh developed a widely used model to measure End User Computer Satisfaction (EUCS) that covers key factors relating to the user perspective (Hou, 2012; Doll & Torkzadeh, 1988). The approach includes twelve attributes in the form of questions covering five aspects of satisfaction: *content*, *accuracy*, *format*, *ease of use* and *timeliness*. This model is well validated and has been found to be generalizable across several IS applications; however, it has not been validated with users of BI (Hous, 2012). Petter, DeLone & McLean (2013) provide several examples of measuring user satisfaction as part of IS success based on the D&M IS Success Model. In this approach, single items can be used to measure user satisfaction, or semantic differential scales can be used to assess attitudes and satisfaction with the system, or multi-attribute scales can be used to measure user information satisfaction. However, in the context of evaluating user satisfaction with changes to BI reporting systems, three issues have been identified with this approach. First, the discussion is about methods of measuring, rather than relevant measurements; Petter, DeLone & McLean (2013) focus on how measuring is done rather than on what is measured. The second issue is that this approach is designed for IS rather than the narrower spectrum of BI. As IS is a higher-level concept that encompasses BI, the approach covers a wider spectrum of measurements and goes beyond the BI scope and requirements. The third issue is that, in the context of evaluating the success of changes to BI reporting, the approach does not identify explicit measurements and there is no clear definition of what to measure in the given scenario. Considering the D&M model in the context of MLED_BI, *ease of use* and *flexibility* are identified as the measures of *system quality* which are relevant.

In the Data Warehouse Balanced Scorecard Model approach (DWBSM), the user perspective, understood as user satisfaction with data quality and query performance is defined as one of four aspects to be considered when measuring the success of the DW (Rahman, 2013). The DWBSM considers data quality, average query response time, data freshness and timeliness of information per service level agreement as key factors in determining user satisfaction. As data warehouses are at the heart of BI systems (Dedić & Stanier, 2016a; Olszak & Ziemba, 2006), these factors are relevant to the evaluation of the success of changes to BI reporting, but are not comprehensive enough as they cover only the DW element of a BI system.

Elements from different approaches were combined to develop a tool for measuring user satisfaction with changes to BI reporting systems. As the EUCS model is well validated and widely used, EUCS was used as a basis for the user satisfaction element of the measurement tool. Aspects and attributes from the EUCS model were cross-tabulated with the phases proposed by Lönnqvist and Pirttimäki (2006). Table 6-2 shows the results of the cross tabulation with areas of intersection marked with '✓'. Categories and questions in the left-hand column of Table 6-2 present aspects and attributes from EUCS model. The numbers in the right-hand column relate to the four phases ((1) *identification of information needs* (2) *information acquisition*(3) *information analysis* (4) *storage and information utilisation)* proposed by Lönnqvist and Pirttimäki for use when measuring the performance of BI systems.

Table 6-2: Cross-tabulation of EUCS attributes and phases of measuring BI performance

EUCS aspects and their attributes (Doll and Torkzadeh, 1988)		Phases of measuring BI performance (Lönnqvist and Pirttimäki, 2006)			
		1st	**2nd**	**3rd**	**4th**
Content	Does the system provide the precise information you need?	✓	✓	✓	
	Does the information content meet your needs?	✓	✓	✓	✓
	Does the system provide reports that seem to be just about exactly what you need?	✓	✓		
	Does the system provide sufficient information?	✓	✓		
Accuracy	Is the system accurate?				✓
	Are you satisfied with the accuracy of the system?				✓
Format	Do you think the output is presented in a useful format?				✓
	Is the information clear?	✓		✓	
Ease of use	Is the system user friendly?				✓
	Is the system easy to use?				✓
Timeliness	Do you get the information you need in time?				✓
	Does the system provide up-to-date information?				✓

As discussed in section 6.2.1., only the *storage and information* utilisation phase (phase 4 in Table 6-2) from the Lönnqvist and Pirttimäki approach is considered relevant when measuring the success of changes made to BI reporting systems, meaning that the focus in Table 6-2 is on the intersection of EUCS elements and phase 4. The eight key measures identified for phase 4 in Table 6-2 were adapted for use in a BI context and used as the basis for a user satisfaction questionnaire. This follows the EUCS model, which also uses a question-based approach. Table 6-3 presents the questions developed from Table 6-2; the questions themselves were later revised following feedback during the initial phase of validation of the measuring tool.

Table 6-3: User satisfaction questions

1	*Does the information content of the reports meet your needs?*
2	*Are the BI system and reports accurate?*
3	*Are you satisfied with the accuracy of the BI system and the associated reports?*
4	*Do you think the output is presented in a useful format?*
5	*Are the BI system and associated reports user friendly?*
6	*Are the BI system and associated reports easy to use?*
7	*Do you get the information you need in time?*
8	*Do the BI system and associated reports provide up-to-date information?*
9	*Are you satisfied with the changing descriptive content (CDS) functionality?*
10	*Is the BI system flexible enough regarding CDS functionality?*
11	*Is CDS functionality fast enough to fulfil business requirements in a timely fashion?*

The EUCS elements were extended to include three additional questions related to changing the descriptive content (CDS) of BI reports. CDS issues are common with large and rapidly changing dimensions (Dedić & Stanier, 2016a) and are a significant issue in managing BI reporting. Descriptive content is conventionally known as *master data* and is used to describe entities, which are independent of, and fundamental to, enterprise operations such as products, persons, customers, locations, suppliers, or services (Talburt & Zhou, 2015). An example of descriptive content (*master data*) is provided in Figure 6-1, in the Country, Assortment Group and Article columns. The most common cause of CDS change requests are errors in the descriptions. The issues were also discussed in section 2.3. and 4.5.

Country	Assortment Group	Article	Gross Profit (EUR)	Gross Profit Plan (EUR)	Gross Profit to Plan (%)
Austria	Fruits and Vegetables	Apples	7.728,00	9.782,00	79,0
Austria	Fruits and Vegetables	Oranges	9.348,00	7.949,00	17,6
Austria	Fruits and Vegetables	Cherries	6.140,00	2.562,00	239,6
Austria	Fruits and Vegetables	Cranberries	3.279,00	8.784,00	37,3
Austria	Fruits and Vegetables	Grapes	1.022,00	4.133,00	24,7
Austria	Fruits and Vegetables	Grapefruit	3.005,00	9.590,00	31,3
Austria	Fruits and Vegetables	Pears	8.297,00	9.324,00	88,99

Figure 6-1: Example of descriptive content in BI report

6.2.3. Technical Functionality

The nature of BI systems mean that user satisfaction alone is not a sufficient measure of success and it is also necessary to consider technical factors. In section 6.2.1., technical functionality is identified as the second cluster of measurements that need to be considered when measuring the success of changes to BI reporting systems. *Reporting & BI query runtime* was identified from the DWBSM approach (Rahman, 2013) as relevant in the context of BI reporting. From the D&M IS success model (Petter, DeLone & McLean, 2013), the *response time* measure was extracted from the *system quality* cluster of IS success variables. *Reporting & BI query runtime* and *response time* both belong to the cluster of measurements dealing with time and were evaluated from a BI reporting perspective to identify appropriate measurements. Table 6-4 shows the elements identified as a result of this process and includes additional elements identified empirically, related to memory use and technical scalability.

Table 6-4: Technical measurements

1	*Initial BI report or dashboard execution time*
2	*Query execution time*
3	*Re-execution time when changing report language, currency or unit*
4	*Time required to change erroneous descriptions of descriptive attributes / hierarchies*
5	*Database memory consumption*
6	*CPU memory usage during execution of: a) Initial BI report or dashboard; b) Query; c) Re-execution of report when changing language, currency or unit;*
7	*Technical scalability and support for integration of proposed solution in regard to existing environment*
8	*Flexibility and extensibility in regard to possible extension of the system in the future*
10	*Is the BI system flexible enough regarding CDS functionality?*
11	*Is CDS functionality fast enough to fulfil business requirements in a timely fashion?*

6.3. Development of the Evaluation Tool

From the literature, two clusters of measurements, one relating to end user satisfaction and one to technical factors, were identified. Determining the success of changes requires the same measurements to be taken, first in the existing BI environment, and secondly, in the new environment. The results can then be compared and used for evaluation. The *user satisfaction questions* and *technical measures* were combined into a single evaluation tool, in the form of a questionnaire. The evaluation tool was tested in a pilot survey with 10 BI domain experts/report users and following the pilot, a number of revisions were made: questions 2 and 3 were merged, the wording of questions 5 and 6 was modified and the original question 9 was removed. In response to comments, two additional questions, one user focused, one technical, were added. The user question related to the exporting and sharing of content functionality; the technical question related to the speed of execution time when drilling-down, conditioning, removing or adding columns in reports. The final list of factors is shown in Table 6-5.

6.4. Validation of the Evaluation Tool

Thirty users working in the BI field took part in the final survey. Respondents were selected through a professional network. Fourteen of the respondents were business users with a technical focus; sixteen were business users having an exclusively business focus. All users completed the user factors element of the survey. Technical functionality may be relevant or understood only by technical users; hence, this part of survey was optional and completion depended on the respondent's expertise. A Likert scale was used, scoring each factor on a scale of 1 – 5 (where 1 is less important and 5 is most important). In the original Likert scale approach, responses are combined to create an attitudinal measurement scale, supporting data analysis on the combined results (Boone & Boone, 2012). However, the intention was to score each individual question or statement separately and to examine the views of users regarding each separate factor. This meant that most of the bi-and multivariate inferential statistical tests, such as those seeking relationships or group membership, were not relevant to the analysis of responses to the evaluation tool.

Two groups of users were identified in the survey: business users with a business focus and business user with a technical focus. Consideration was given to using the chi square or t-test but as there were no expected frequencies for the answers, the use of chi square test was inappropriate in this context. To analyse each individual item from the Likert scale properly, the discrete nature of responses must be acknowledged, otherwise analysis can lead to inferential errors (Clason & Dormody,1994). The t-test was not used as the t-test ignores the discrete nature of responses. It was noted at the beginning of this section, that the part of the survey relating to the technical measurements cluster was optional and completion depended on the respondent's expertise, as it was expected that only business users with technical focus would provide answers to those questions. Differences between two groups of users in the survey could be identified by simple summation of the number of responses given by each of the groups. As the aim of the chapter is to examine the views of users regarding each separate factor/item in the evaluation tool, the use of central tendency statistical tests was identified as the most appropriate approach. Likert-type items fall into the ordinal measurement scale, thus *mode* or *median* are recommended to measure central tendency (Boone & Boone, 2012). The results of our survey are presented in Table 6-5, and are grouped into two clusters of measurements, namely *user satisfaction* and *technical functionality*. Table 6-5 shows that for the user satisfaction section, no question had mode or median value less than 4, indicating that the factors identified in each question were considered important. For the technical factor section, no question had a mode or median value less than 3, indicating that all the technical factors identified were seen as relevant, confirming the factors including in the evaluation tool.

As expected, a larger percentage of business users with a technical focus commented on technical aspects than business users with exclusively business orientation. Users with a greater business orientation rated user satisfaction questions as more important than users with a greater technical role, and the same effect was found in relation to users with a greater technical role commenting on technical functionality elements.

Table 6-5: Survey results based on Likert-type items

	Business Users			***Technical Users***			***All Users***		
User Satisfaction	*Nr.*	*Mode*	*Median*	*Nr.*	*Mode*	*Median*	*Nr.*	*Mode*	*Median*
- Information content meets your needs?	16	5	5	14	5	5	30	**5**	5
- The information provided in the reports is accurate?	16	5	5	14	5	5	30	**5**	5
- Output is presented in a format that you find useful?	16	5	5	14	4	4	30	**5**	4
- The system and associated reports are easy for you to use?	16	5	4.5	14	4	4	30	**5**	4
- Information in the reports is up to date?	16	5	5	14	5	5	30	**5**	5
- Reports have the functionality that you require?	16	5	4.5	14	4	4	30	**4**	4
- The BI system is flexible enough to support easy change of *descriptive content"*?	16	4	4	14	4	4	30	**4**	4
- Is the change of "descriptive content"* fast enough to fulfil business requirement?	16	4	4	14	4	4	30	**4**	4
- Exporting and sharing content functionalities meet your needs?	16	5	4.5	14	3	3	30	**5**	4
Technical Functionality									
- Speed of execution time for Initial BI report or dashboard	10	4	4.5	13	4	4	23	**4**	4
- Speed of execution time for SQL query	8	4	4	13	4	4	21	**4**	4
- Speed of re-execution time when changing report language, currency or unit	11	4	4	13	4	4	24	**4**	4
- Speed of execution time when drilling-down, conditioning, removing or adding columns in reports	10	4	4	13	5	4	23	**5**	4
- Amount of Time required to change erroneous descriptions of descriptive attributes and hierarchies	7	3	3	12	3	3.5	19	**3**	3
- Database memory consumption	4	4	3	13	3	3	17	**3**	3
- CPU memory usage during execution of initial BI report or dashboard	3	3	3	12	3	3	15	**3**	3
- CPU memory usage during execution of SQL query	4	4	4	12	3	3	16	**4**	3.5
- CPU memory usage during re-execution of report when changing language, currency or unit	4	4	3	12	4	3	16	**4**	3
- Technical scalability of proposed solution in the existing environment	6	4	4	13	5	4	19	**5**	4
- Support for possible extension of the system in the future	7	3	4	12	4	4	19	**4**	4

The questions given in Table 6-5 represent the core evaluation tool. Two additional user satisfaction questions were suggested by users in free text comments, relating to the availability and accessibility of key figures and to whether support for further consolidation of existing information is available. An additional technical question relating to the platform independence of BI reports was also suggested. The evaluation tool can be extended for use in other contexts by including additional questions and other factors as identified by stakeholders, but the survey indicated that the evaluation tool covered the relevant core measures for the validation of MLED_BI

6.5. Conclusion

The previous chapter, chapter 5, described a limited evaluation of MLED_BI based on the PoC. This chapter described the motivation for developing an evaluation tool to support a detailed evaluation of the MLED_BI implementation as part of the overall validation of the MLED_BI design approach. The chapter defined success in the context of changes to the BI environment and identified relevant user groups. The process by which the evaluation tool was developed from a literature review was described and the validation and evaluation of the tool was discussed. The development of the evaluation tool presented in this chapter was seen as a prerequisite for the development and evaluation of the Business Intelligence systems used to evaluate the MLED_BI design approach. In order to support a comparison of different design approaches, Business Intelligence systems, based on conventional and MLED_BI design approaches were developed. The evaluation tool not only identified the measurements and clusters which would be used to evaluate the success of the different approaches relevant in this context but also identified which elements of the test artefacts should be included in the implementation and the focus and direction of the implementation. The evaluation tool provided clear input as to which elements should be implemented to successfully support the comparison of measurements.This in turn made the next stage of the research possible as it provided a structured basis for the actual development of the artefact used to evaluate the MLED_Bi design approach. The following chapter, chapter 7, describes the development of a large scale BI environment, developed for use with the evaluation tool and designed to support a detailed evaluation of MLED_BI.

Chapter 7: Implementation of BI Design Approaches

7.1. Introduction

This chapter describes the development of a BI environment to be used with the evaluation tool described in chapter 6, to support the validation and evaluation of the MLED_BI design approach. The BI environment included four BI systems which support ML in BI: three of the systems were based on existing ML BI design approaches, each using a different data mart implementation approach, and one BI system implemented the MLED_BI design approach. This enabled MLED_BI performance and functionality to be compared with existing approaches. The chapter discusses the motivation for developing a comparative implementation of ML in BI and gives an overview of the system. The design and implementation of the source system, the data warehouse layer and the presentation layer, together with the ETL processes are discussed and the chapter explains the role of each element. Technical information about the implementation of the system is given in APPENDIX E.

7.2. Motivation for Developing a Comparative Implementation of ML in BI

The PoC artefact described in chapter 5 was developed to check technical feasibility and was a small-scale implementation of the MLED_BI design approach based on an incomplete BI system. In the PoC implementation, the MLED_BI design approach was compared to only one of the existing design approaches used to support ML. It was concluded in section 5.5. that a large-scale implementation of the BI environment was needed to support a comprehensive validation of the MLED_BI design approach and that it was necessary to review MLED_BI against all existing ML in BI approaches. A further reason for developing a large scale implementation was to support a more comprehensive evaluation of the implications of the greater up front design and development effort of using the MLED_BI approach. The evaluation tool described in chapter 6, showed that technical measures and measures of end user satisfaction were appropriate mechanisms for validating BI reporting systems. To validate MLED_BI, it was necessary to have a BI environment which supported comparison of MLED_BI metrics with metrics from BI systems developed using existing ML BI design approaches and which was substantial enough to allow end users to experience the differences between approaches. For this reason, an implementation which supported comparisons between systems and enabled end user evaluation was needed. The data used was identical for all systems except where, as discussed further in section 7.3, the design approach required modification to the data.

7.3. Overview of the System

The complete system included four different multilingual BI systems, simulating a real-world BI environment. Three of the systems covered three existing approaches to support ML in BI where all business information including descriptive information is stored in data marts tables, while the fourth BI system was based on MLED_BI. The three existing approaches used were:

- the AA approach, based on additional attributes in data mart dimensional entities
- the LIF approach, based on language identifier field in data mart dimensional entities
- the ATS approach based on additional schema/entities for dimensions

These approaches are discussed in detail in section 2.6. The fourth approach used was the MLED_BI design approach discussed in chapter 4. The visual difference between existing BI design approaches and a BI system based on MLED_BI was shown in Figure 4-12, given again here as Figure 7-1 for ease of reference. In all BI systems based on existing BI design approaches all business information (master and transactional data) are stored in dimensional tables while MLED_BI uses a higher level star schema/language file approach.

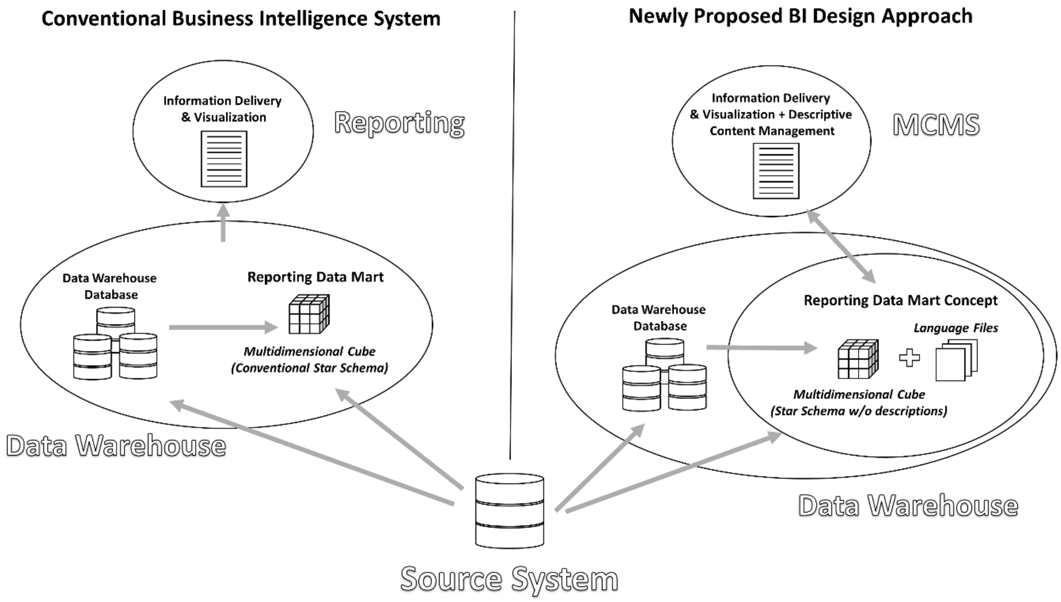


Figure 7-1: Existing vs Newly Proposed BI Design Approach

The BI environment developed for the full evaluation followed the HBIF framework approach and comprised three main layers:

(i) Source Layer in the form of the Sample Source System Database (SSSD): the same SSSD was used for each implementation.

(ii) Data Warehousing Layer: four different dimensional modelling approaches to enable ML in BI were implemented. All were based on the Star schema.

(iii) Reporting Layer: each of the four approaches used had an implementation at the Reporting Layer. The output was a total of four BI reports: one report per approach based on existing ML BI design approaches reflecting the three different approaches of data mart implementation (AA, LIF, and ATS) plus an MCMS reflecting the MLED_BI approach. To implement the MLED_BI reporting environment, the extended version of the MLED_BI design process presented in Figure 4-9 which includes the MCMS concept was followed.

7.4. System Development

7.4.1. Requirements Definition

The first phase in BI design is Planning and Requirements Definition. From the business content perspective, BI reports that provide overview of *sales per year, product area*, *category* and *subcategory* and include *gross sales*, *net sales* and *profit* as appropriate metrics were identified as the main business requirement for the experimental system. There are two justifications for this: section 5.2.1. identified sales information (product descriptions and data about transactions) as the most common type of reporting in BI systems used with commerce, and discussion with 28 BI domain experts conducted via the social business network Linkedin identified location and time as the next most important and most used attributes after product information. It was a requirement of the validation that all the reports provided the same data based on the same source system. An additional functionality requirement was that it must be possible to change the report language but still provide the same transactional data.

7.4.2. Development of the Source System

The source system was developed first, followed by design of the data marts. The SSSD was designed to simulate data in a Customer Relationship Management (CRM) system used by a major European retailer. The CRM system was simplified from the original and only the elements needed to support the validation were implemented. It is important to note that that source system design is not part of the MLED_BI approach. However, a SSSD was required for validation since for reasons of data protection and commercial confidentiality, it was not possible to use a live source system. The SSSD functioned as the source system for each design approach used in the validation. The design of the source system is shown in Figure 7-2 on the next page.

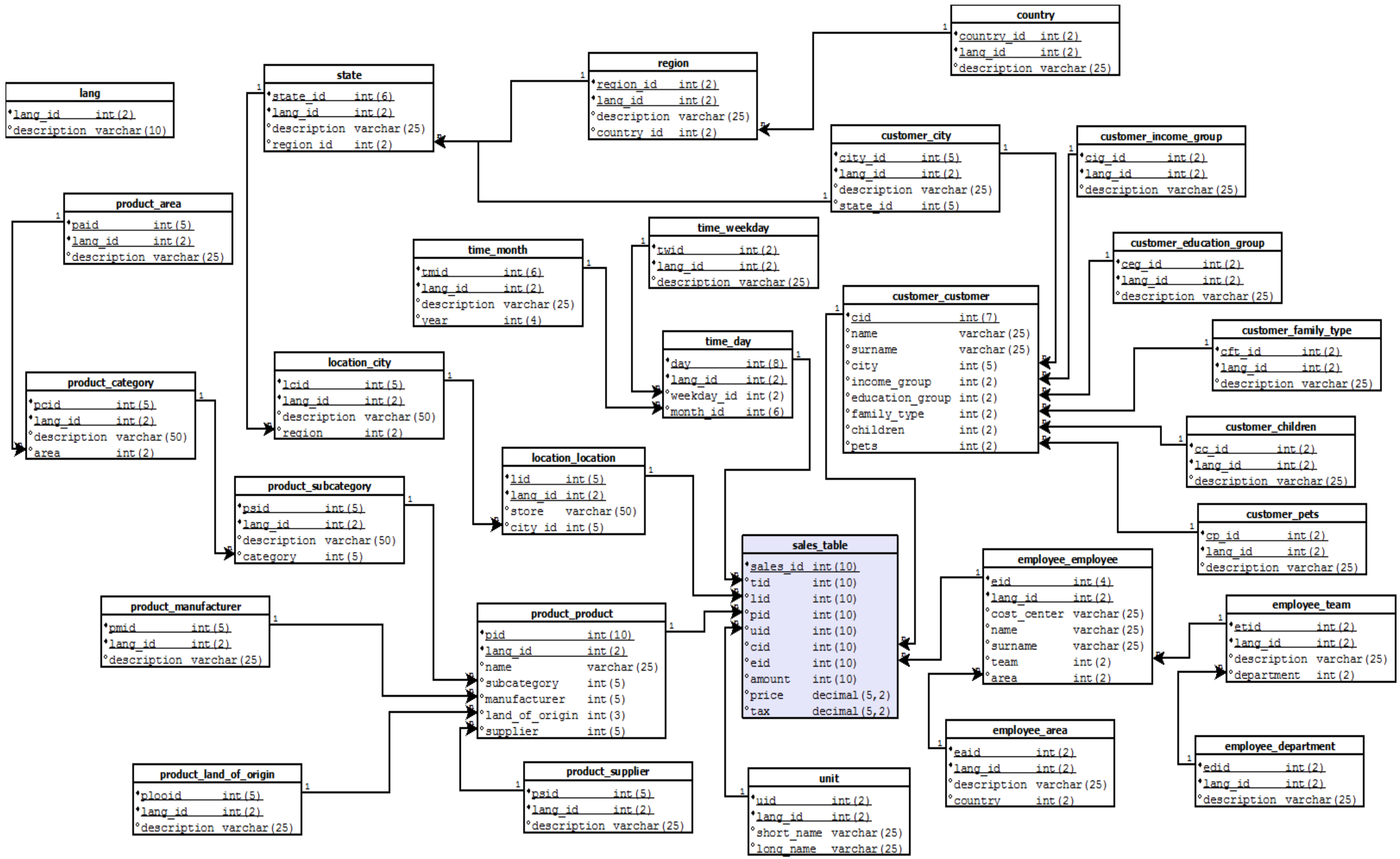


Figure 7-2: Design of the Source System

In addition to the sales_table entity, shown in Figure 7-2, which holds information about sales and holds 1,199,989 sales records, there are a further 28 entities which hold data about customers, employees, products, locations, time and unit. Only three dimensions (product, location and time) were required for the BI reports used for validation; however, the intention was to reflect as much as possible the real CRM source system, thus, other master data was included in the SSSD and for use in ETL processes. The master and transactional data used to populate the SSSD were generated using data generation software (Data Generator) but the structure was based on real world data. Figure 7-2 includes an entity named lang. To improve readability of the diagram, the relationships of lang are not shown since lang represents the language identification entity and is related to all the entities that have a lang_id field as a part of primary key. At physical level, a SSSD database called phd_project_source, simulating a multilingual CRM database was implemented in MySQL.

7.4.3. DWH Layer Design

Once the SSSD had been implemented and sample data loaded, the next phase process was DW design. To support the validation, the DWH layers based on the four different BI design approaches were designed and implemented as part of the same BI environment. This encompassed four different dimensional modelling approaches to support ML in BI. As there was no requirement to replicate all the data from the SSSD in the DW for the purposes of validation, the Kimball DW design approach was used for all four approaches. The star schema designs for all four approaches are given in this chapter to demonstrate in outline the difference between approaches. To support readability, larger versions of the diagrams are given in APPENDIX F.

7.4.3.1. MLED_BI Approach

As discussed in 4.5., the MLED_BI design approach treats the star schema as a high-level design entity in which textual descriptions from attributes and hierarchies are modelled not in dimensional entities but are designed to be held elsewhere as language files. In this view of the star schema, the fact and dimensional entities hold only identifiers, stored as numerical values. The fact entity contains foreign keys to support relationships with dimension entities and it holds transactional data. Thus, as attributes and hierarchical descriptions and their identifiers are extracted to separate language files, at implementation level, only numerical values (identifiers) are stored in dimensional entities. A DW *phd_project_files* was designed. The star schema design for the

MLED_BI data mart has seven entities: one fact table and six tables representing customer, product, location, employee, time and unit dimension, as shown in Figure 7-3. A web environment in the form of server folder named “files” was also designed to store the language files that hold master data descriptions but is not shown.

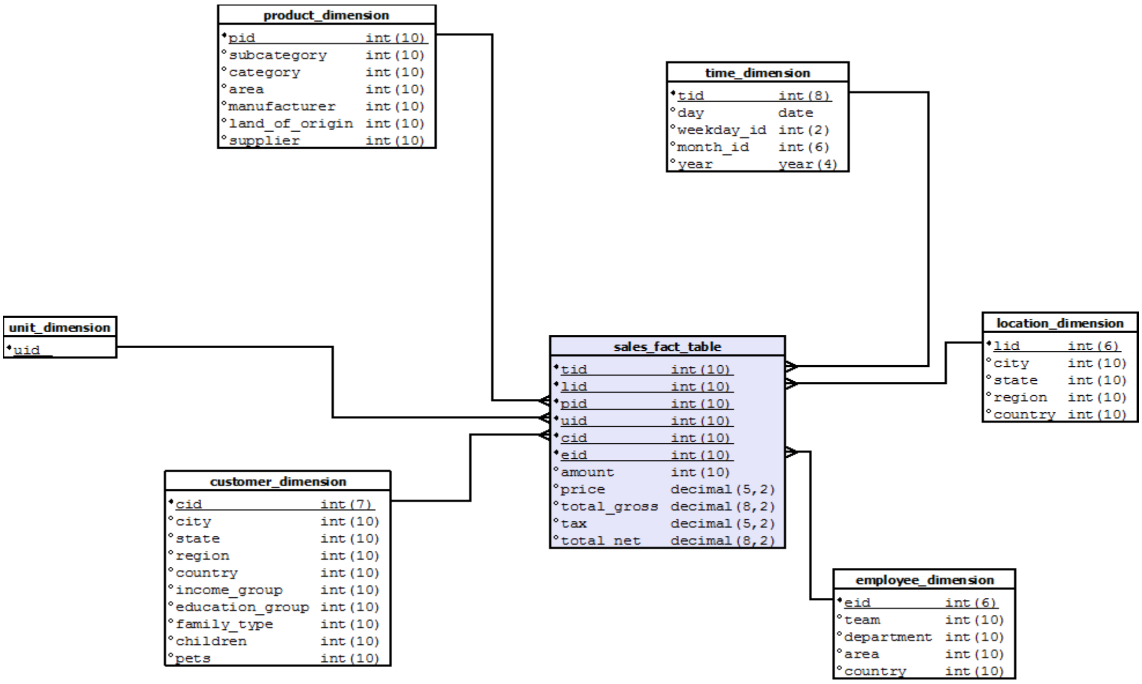


Figure 7-3: DM Star Schema based on MLED_BI approach

7.4.3.2. Design of the Additional Attributes (AA) Data Mart Approach

The additional attributes approach, as discussed 2.6.1. and 7.3., proposes that where there are new values for the dimension tables in the star schema, new attributes should be added to dimensional tables. To support the AA approach, a new data mart named *phd_project_aa* was designed. This data mart has seven entities composing the star schema; this is the same number of entities as the MLED_BI approach. In addition to the fact entity (sales_fact_entity), this star schema has customer, employee, location, product, time, and unit dimension entities as shown in Figure 7-4. As discussed in 2.6.1, the AA approach does not require additional entities or schemas, but uses additional fields in existing entities for additional languages.

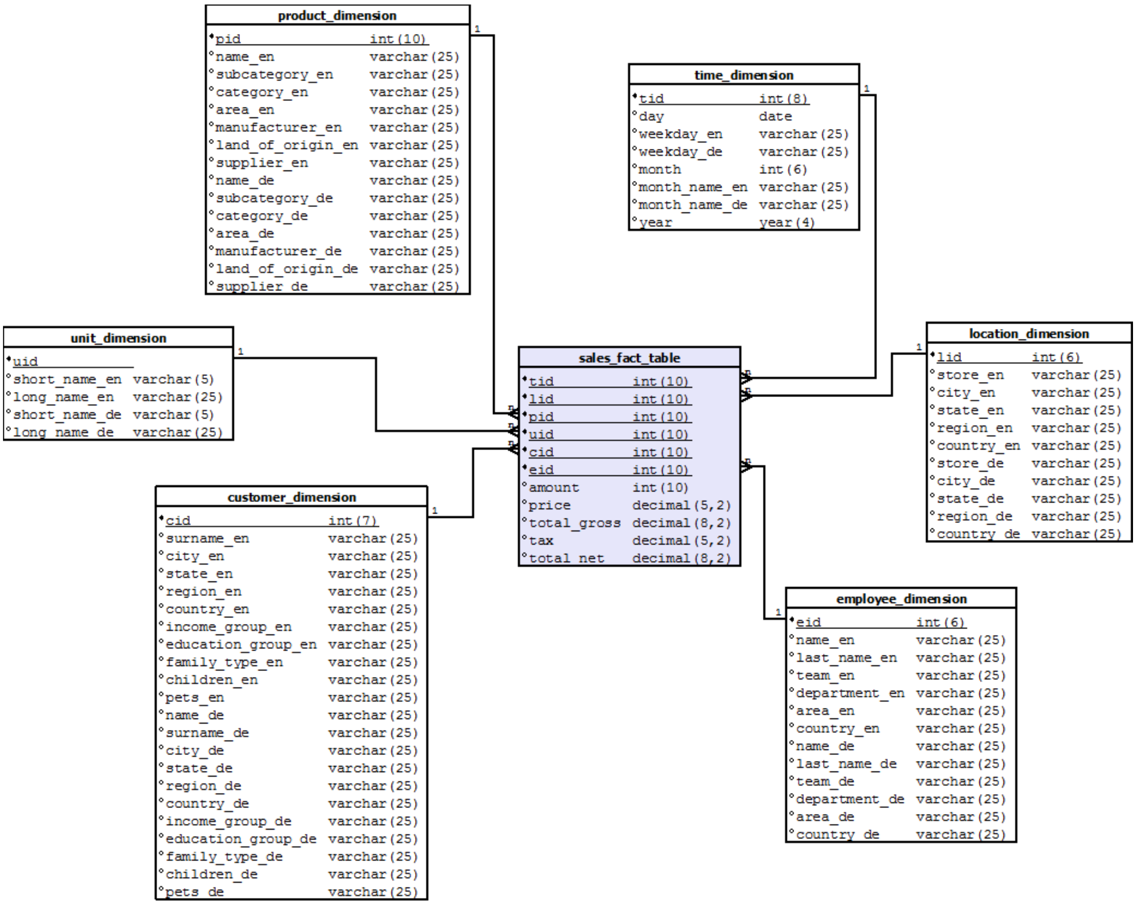


Figure 7-4: DM star schema based on the AA approach

7.4.3.3. Design of the Language Field Identifier (LIF) DM Approach

The LIF approach, discussed in section 2.6.2. and 7.3. also uses the strategy of saving transactional and master data in dimensional entities. The LIF approach uses the same number of entities to compose the star schema as the AA and the MLED_BI approaches. The LIF approach, however, results in fewer attributes than the AA approach. The LIF data mart had one fact table and six entities representing customer, employee, location, product, time and unit dimensions, shown in Figure 7-5 overleaf. This approach uses a lang field to support the identification of the language used for each row in every dimensional entity.

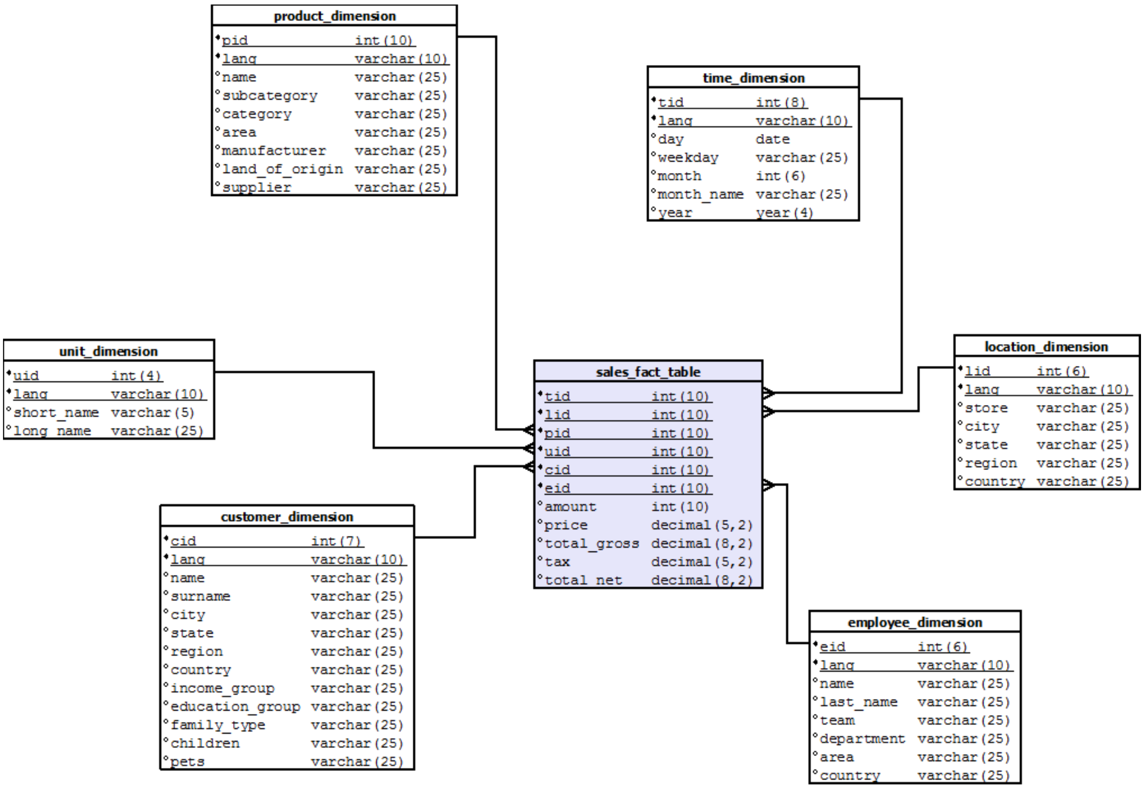


Figure 7-5: DM Star Schema based on the LIF approach

7.4.3.4. Design of the Additional Entities or Schema (ATS) DM Approach

The ATS approach discussed in section 2.6.3. and 7.3., like the AA and LIF approaches, saves both transactional and master data in dimensional tables. The ATS approach results in a larger number of entities than any other approach. As shown in Figure 7-6 on the following page, to support the application of ML in BI, the ATS approach needs twice the number of dimensional entities required by AA. In the design based on ATS, as shown in Figure 7-6, twelve dimension entities represent six actual dimensions in two different languages: English and German. A language prefix was used to identify each dimension in each specific language: de_ for German and en_ prefix for English.

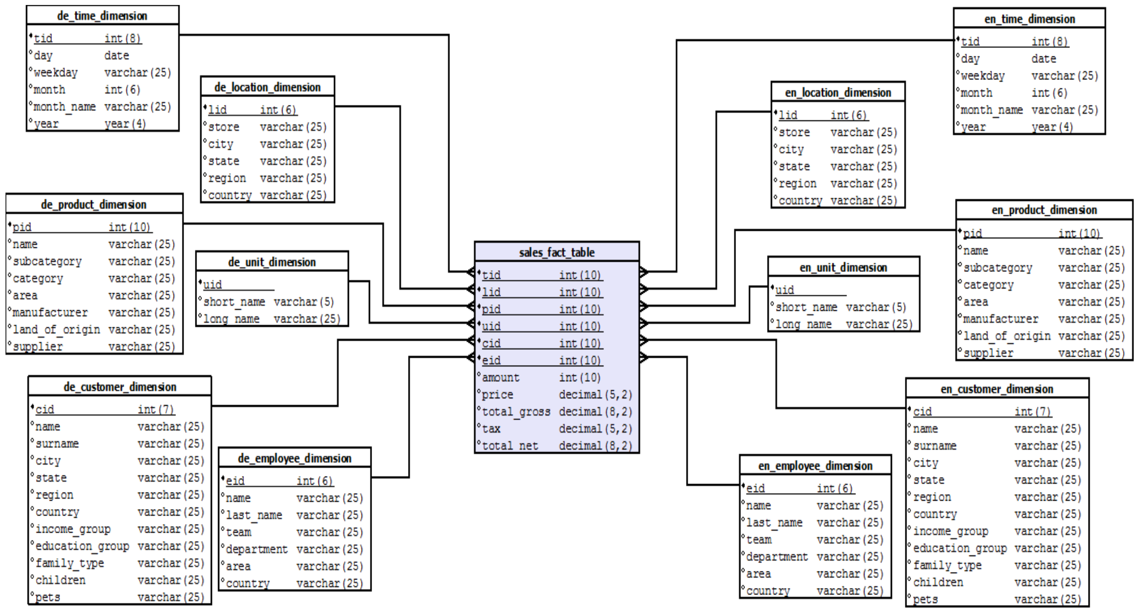


Figure 7-6: DM Star Schema based on ATS approach

7.4.3.5. DWH Layer Summary

All the data marts hold the same data and support two different languages, English and German. The ATS approach had the largest number of entities and hence the largest number of attributes. The MLED_BI design approach, the AA and the LIF approaches had the same number of entities. The MLED_BI implementation had fewer attributes than AA and LIF. However, the MLED_BI requires language files which are not shown in the star schema design presented in figure 7- 3. The source system and all the entities in the different data marts were implemented in MySQL. The implementation is discussed further in APPENDIX E, section E.2.

7.4.4. Development of ETL Processes

The existence of four separate data marts required the implementation of four different ETL processes; one to enable data delivery from the SSSD to the MLED_BI DM, and three to enable data delivery from the SSSD to each of the three DM developed based on existing approaches. As the MLED_BI data mart required processes not only to deliver data from SSSD to DM entities, but also to extract, modify and load data from SSSD to the appropriate language files, the MLED_BI ETL processes were the most complex. The ETL processes that support the other three DMs only required functionality to deliver data from the SSSD to the DMs. The module developed to support ETL is discussed further in APPENDIX E, section E.4.

7.4.5. Presentation Layer

7.4.5.1. Design of Presentation Layer

The next step was the development of the Presentation Layer. The main requirement of the Presentation Layer was that it should support access to BI reports from all four BI systems and supporting DMs. A decision was taken to develop a web environment (WE) to provide a single point of access to all BI reports and to support the full application of the Reporting Layer design phase in the MLED_BI design processes. Figure 7-7 shows the architecture of the web environment.

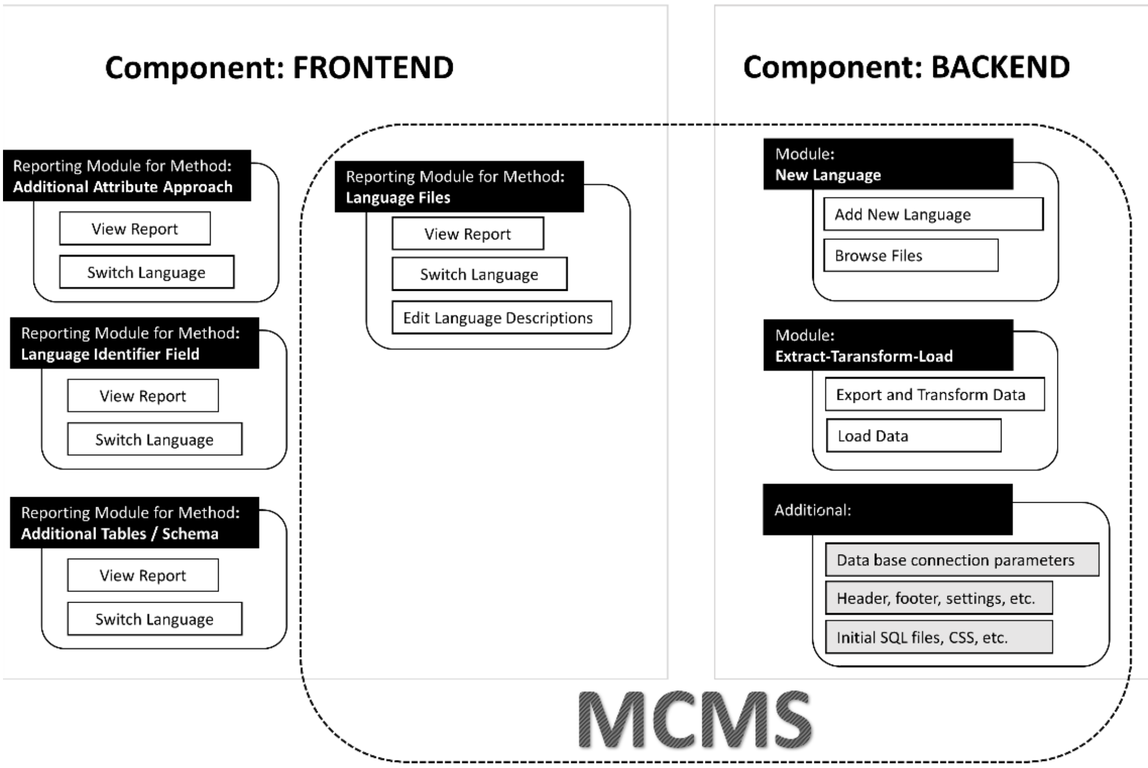


Figure 7-7: Web environment architecture

As shown in Figure 7-7, the reporting front end includes a reporting module for each data mart. The reporting modules support viewing of reports and switching languages for previously executed reports. Existing ML BI/DWH design approaches support the viewing and some manipulation of data in the reporting layer but as discussed in 4.4.2, the "no data change policy" in existing data warehouse design approaches means that content changes are permitted only in the source system. For this reason the BI reporting layer developed for the AA, LIF and ATS approaches provides only visualization of the data stored in the DW and does not include a content management system. The web interfaces retrieve business content (master and transactional data) directly from the supporting data marts.

7.4.5.2. Design of the MCMS Web Environment

Figure 7-7 includes the MCMS component shown in the extended version of MLED_BI (Figure 4-9). There are three elements to the development of the MCMS: design of the MCMS frontend, design of the MCMS backend, and physical design of data join concepts, such as how to assign master data descriptions from language files to master data IDs from data marts. The MLED_BI design approach is independent of the use of a multilingual content management system. However, from the reporting perspective, the fact that MLED_BI makes possible the use of an MCMS, is one of the major benefits of MLED_BI.

Using the MLED_BI BI design approach, Star schema do not use attribute descriptions and hierarchies as the basis for data aggregation, but operate with identifiers. Thus, it is possible to change descriptions without creating unrelated data or additional categories for the same data at reporting level or in the data warehouse, which means that the MLED_BI approach opens the door for the use of a web content management system which extends standard reporting operations with functionality such as editing descriptions for existing languages directly via the web interface and adding new languages and their variations directly by business users, independently of the existing languages in source systems. This was discussed in section 4.6. and the design differences between reporting supported in existing ML BI layers and the content management system supported by MLED_BI are shown in figure 4-8, which is given again here as figure 7-8 for ease of reference.

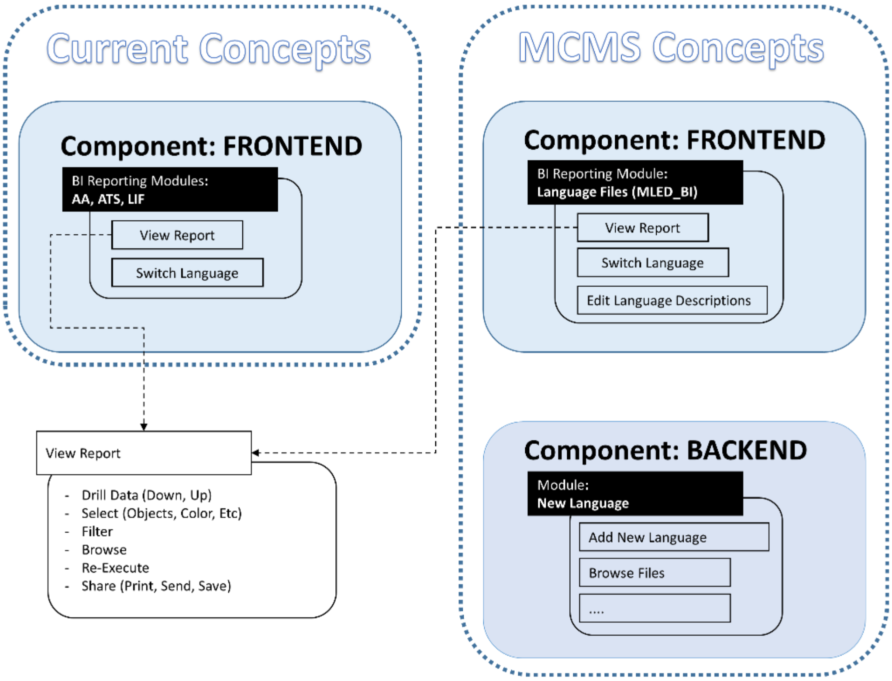


Figure 7-8: Comparison of conventional reporting layer functionality and functionality provided by a MCMS supported by MLED_BI

The MCMS web interface has two components, the frontend element, that in addition to direct editing of master data descriptions, enables execution of the standard reporting activities and the backend element that provides language management functionality. Depending on user requirements, the backend functionality could be extended with additional modules, as shown in Figure 7-8, for example to enable the execution of ETL processes by business users or to edit various aspects of web interface. The MCSM approach allows business users to change erroneous descriptive content directly and as discussed in chapter 4, this would simplify or possibly in some cases eliminate, the ETL processes required to perform language changes.

In existing ML BI design approaches, all business content (master and transactional data) used by web applications in BI reports is retrieved directly from data mart entities. In MLED_BI, as discussed in section 4.5, master data descriptions from languages files are assigned to the result set acquired by the means of querying the data mart on the fly, during execution of the BI report. The result set holds the numerical values of master data identifiers in addition to transactional data; the language file delivers master data descriptions, which are then assigned to master data identifiers in the report to provide meaningful information for the report users. The architecture of the MCMS web environment is given in Appendix G. To demonstrate the functionality made possible by the MLED_BI approach, the MCMS Backend included additional modules supporting ETL operations, allowing the user to add new languages and support for administration functions. Details of the additional functionality are given in appendix H. The additional modules are not required elements but demonstrate how an MCMS based on MLED_BI can provide additional functionality for users.

7.5. Conclusion

This chapter described the design and development of a BI environment to support the comparison of Multilingual Business Intelligence design approaches. The BI environment presented in this chapter is a substantial artefact which simulates a real world BI system and includes an implementation of the three existing ML BI design approaches as well as MLED_BI. The BI environment was developed by following all stages of the Business Intelligence MLED_BI design and development process and consisted of the Sample Source System Database and Data Warehouse Layers, ETL processes and the Presentation Layer, including the development of an MCMS. In addition to validating the feasibility of translating the proposed MLED_BI design approach into a full Business Intelligence system that simulates a real-world environment, the main reason for the development of the artefact described in this chapter was to support a comprehensive validation of the MLED_BI design approach. The implementation was a prerequisite for conducting the next stage of the research, namely, the comparison of conventional design approaches with the MLED_BI design approach. The comparison of the different design approaches required an implementation which supported the use of the measurements and aspects identified as relevant in previous chapter, chapter 7. The following chapter, chapter 8, describes the way in which the BI environment presented in this chapter was used to enable comparison of MLED_BI with existing ML BI design approaches and to support the validation of MLED_BI with business users and domain experts.

Chapter 8: Validation of MLED_BI Design Approach

8.1. Introduction

The previous chapter described the development of a BI environment to support the validation of the MLED_BI design approach. The successful implementation of a large scale BI system based on MLED_BI verified that the MLED_BI approach could be used in a real world context. This chapter discusses the validation of MLED_BI using quantitative and qualitative techniques based on the evaluation tool presented in chapter 6. The data collection and data analysis process is described and the results from the quantitative investigation are presented and discussed. The qualitative element of the validation was carried out with domain experts who commented on some aspects of the technical validation and with business end users who were given the opportunity to test all four design solutions for multilingualism in BI and were then asked to evaluate the strengths and limitations of the approaches. The overall findings from the validation are presented and evaluated.

8.2. Technical Validation

The technical validation was based on the technical functionality cluster of measurements identified in the evaluation tool developed in chapter 6. The tool proposed eleven technical metrics which were summarised in Table 6-5; Table 6-5 is presented in this chapter as Table 8-1 for ease of reference. The metrics evaluate the technical effectiveness of changes to BI systems and based on the validation of the tool, described in section 6.3., are considered to cover the technical measurements relevant in the context of MLED_BI. The tool covers elements such as speed of execution and memory consumption and are labelled TM1 through to TM11 (Table 8-1).

Table 8-1: Technical Functionality Measurements

Code	**Metrics**
TM1	*- Speed of execution time for Initial BI report or dashboard*
TM2	*- Speed of execution time for SQL query*
TM3	*- Speed of re-execution time when changing report language, currency or unit*
TM4	*- Speed of execution time when drilling-down, conditioning, removing or adding columns in reports*

TM5	*- CPU memory usage during execution of initial BI report or dashboard*
TM6	*- CPU memory usage during re-execution of report when changing language, currency or unit*
TM7	*- CPU memory usage during execution of SQL query*
TM8	*- Database memory consumption*
TM9	*- Amount of Time required to change erroneous descriptions of descriptive attributes and hierarchies*
TM10	*- Technical scalability of proposed solution in the existing environment*
TM11	*- Support for possible extension of the system in the future*

8.2.1. Test Environment

Every BI report used for testing included code that measured and provided information about the execution speed of the web application and the relevant SQL query. The fact table holding transactional data in the four DMs had 1,199,989 records, which reflected the transactional data from the source system as discussed in section 7.4.2. The number of records in the dimensional tables reflected the requirements of the respective implementation method. Despite using different structures based on different DM implementation methods, the dimension tables were implemented in a manner which ensured that they provided the same data to the end user via BI reports.

The BI environment developed to support validation implemented a BI system for each of the four design approaches to support ML in BI. The test protocol involved performing the same test on each of the four BI systems. To ensure a fair test, each BI report for each implementation method was executed 20 times in the same environment and provided the same data to the end user. The systems are identified in the following discussion as AA (the additional attributes approach), ATS (The additional table/schema approach), LIF (Language file identifier approach). The MLED_BI design approach (the novel design approach proposed in this thesis) is referred to as MLED_BI or the Language FILES method, depending on context.

8.2.2. Validation with TM1/TM2

TM1 “Speed of execution time for initial BI report or dashboard” and TM2 “Speed of execution time for SQL Query” are related metrics and are discussed here together. Table 8-2 gives the results for these metrics for all four design approaches. Based on the values recorded in Table 8-2, both TM1, “Speed of execution time for initial BI report or dashboard" and TM2 "Speed of execution time for SQL Query", showed improved performance when using BI reports supported by a data mart based on the FILES implementation method, which is part of the MLED_BI design concept.

Table 8-2: Execution speed for initial BI report and underlying SQL Queries

Initial Report Execution								
	AA		ATS		LIF		MLED_BI	
Attempt	Complete Web Report	SQL stat. used in Web Report	Complete Web Report	SQL stat. used in Web Report	Complete Web Report	SQL stat. used in Web Report	Complete Web Report	SQL stat. used in Web Report
1	12,5275	12,5233	12,4374	12,4344	16,2937	16,2902	7,4950	7,4898
2	12,4332	12,4298	12,4122	12,4090	16,0858	16,0823	7,5801	7,5749
3	12,5809	12,5778	12,4608	12,4575	16,1425	16,1392	7,6185	7,6134
4	12,6224	12,6190	12,4000	12,3967	16,2740	16,2706	7,5277	7,5224
5	12,4993	12,4960	12,4355	12,4324	16,3258	16,3224	7,6781	7,6729
6	12,5360	12,5328	12,5702	12,5669	16,2008	16,1975	7,5741	7,5690
7	12,6437	12,6405	12,4289	12,4223	16,3136	16,3105	7,5398	7,5348
8	12,5897	12,5864	12,4190	12,4157	16,2899	16,2864	7,5564	7,5510
9	12,5156	12,5123	12,5165	12,5132	16,1993	16,1885	7,5491	7,5437
10	12,6172	12,6140	12,3446	12,3410	16,4247	16,4214	7,5749	7,5696
11	12,5074	12,5042	12,4005	12,3972	16,2205	16,2174	7,6166	7,6103
12	12,7594	12,7564	12,5111	12,5079	16,1771	16,1588	7,6618	7,6564
13	12,6050	12,6018	12,4352	12,4320	16,4911	16,4876	7,6142	7,6090
14	12,6377	12,6343	12,4342	12,4311	16,1792	16,1760	7,5595	7,5543
15	12,6069	12,6038	12,4079	12,4044	16,4089	16,4057	7,6616	7,6561
16	12,6464	12,6432	12,4766	12,4735	16,3596	16,3562	7,5892	7,5834
17	12,5294	12,5261	12,3616	12,3584	16,2835	16,2805	7,5122	7,5069
18	12,5773	12,5742	12,3611	12,3577	16,2369	16,2337	7,6045	7,5989
19	12,6389	12,6355	12,6371	12,6338	16,2732	16,2580	7,5885	7,5824
20	12,6754	12,6721	12,4772	12,4743	16,1471	16,1440	7,6106	7,6048
Average	**12,5875**	**12,5842**	**12,4464**	**12,4430**	**16,2664**	**16,2613**	**7,5856**	**7,5802**

8.2.3. Validation with TM3

TM3 relates to "Speed of re-execution time when changing report language, currency or unit". Multilingual issues in BI, especially those related to business content descriptions (master data), are the focus of this research, thus the interest is in measuring re-execution time when changing the reporting language. Changing currency or unit descriptions in BI reports reflects the issues involved in changing the reporting language for any other business content. Currency or unit recalculations or transformations on transactional data are not relevant for business information descriptions (master data) and are not considered as part of this research. To evaluate TM3, the report language was changed 20 times in a previously executed BI report in the same environment. The same master and transactional data was used throughout. The results showed that the MLED_BI approach provides a significant advantage (Table 8-3).

Comparison between Table 8-2 and Table 8-3 shows that changing the report language for a BI report based on existing BI design approaches requires as much time as the initial report execution. This is due to the fact that in existing BI design approaches the SQL query must be re-executed to provide business content descriptions in another language. However, this is not the case in the MLED_BI design approach. As shown in Table 8-3, the time required to change the preview language in the MLED_BI design approach of an already executed BI report was less than a hundredth of a second. In MLED_BI, the different understanding of the star schema means that language data is not stored in dimension tables. The decoupling of dimension tables from language storage means that there is no need to re-execute the SQL query, as the new language file was loaded and applied to an already existing SQL result set. For this reason, there are no SQL execution times recorded for MLED_BI in Table 8-3.

Table 8-3: Execution speed for language change in already executed reports

Language Switch in Already Executed Report								
	AA		ATS		LIF		MLED_BI	
Attempt	Complete Web Report	SQL stat. used in Web Report	Complete Web Report	SQL stat. used in Web Report	Complete Web Report	SQL stat. used in Web Report	Complete Web Report	SQL stat. used in Web Report
1	12,6434	12,6399	12,3644	12,3609	16,3424	16,3390	0,0018	0,0000
2	12,5360	12,5326	12,5002	12,5468	16,5138	16,5099	0,0020	0,0000
3	12,6559	12,6526	12,4133	12,4099	16,3648	16,3615	0,0019	0,0000
4	12,6109	12,6058	12,3949	12,3915	16,2968	16,2938	0,0017	0,0000
5	12,6787	12,6757	12,6204	12,6172	16,4638	16,4607	0,0028	0,0000
6	12,5642	12,5608	12,3574	12,3541	16,2970	16,2939	0,0019	0,0000
7	12,6296	12,6264	12,5167	12,5137	16,2809	16,2775	0,0017	0,0000
8	12,6848	12,6814	12,5724	12,5691	16,2914	16,2879	0,0024	0,0000
9	12,7347	12,7313	12,4189	12,4160	16,3695	16,3664	0,0017	0,0000
10	12,5303	12,5269	12,5835	12,5802	16,2624	16,2591	0,0024	0,0000
11	12,5255	12,5221	12,4638	12,4606	16,4954	16,4924	0,0019	0,0000
12	12,7773	12,7740	12,4307	12,4275	16,2686	16,2653	0,0017	0,0000
13	12,5921	12,5746	12,5570	12,5533	16,2960	16,2925	0,0017	0,0000
14	12,6057	12,6025	12,5353	12,5418	16,4996	16,4962	0,0018	0,0000
15	12,5757	12,5727	12,4645	12,4613	16,3162	16,3129	0,0027	0,0000
16	12,6254	12,6219	12,4685	12,4654	16,2748	16,2714	0,0017	0,0000
17	12,5687	12,5654	12,5770	12,5738	16,3962	16,3928	0,0018	0,0000
18	12,7437	12,7406	12,4703	12,4669	16,3274	16,3240	0,0018	0,0000
19	12,5493	12,5456	12,4770	12,4740	16,4062	16,4024	0,0020	0,0000
20	12,5773	12,5741	12,4858	12,4826	16,5049	16,5012	0,0017	0,0000
Average	**12,6205**	**12,6163**	**12,4836**	**12,4833**	**16,3634**	**16,3600**	**0,0020**	**0,0000**

8.2.4. Validation with TM4

TM4 refers to “Speed of execution time when drilling-down, conditioning, removing or adding columns in reports” . This was identified as a relevant technical measurement during the development of the evaluation tool and is particularly relevant in the context of data manipulation at the BI presentation layer. However, the performance of processes such as drilling-down, reflects the performance of the initial BI report execution; thus, it does not require a separate test. Although sometimes visually implemented as a function of an existing report, drilling-down, conditioning, removing or adding new columns is in fact the execution of a report under new criteria, or with different columns at different level of business content. For that reason, TM4, although relevant in the context of the evaluation tool developed in chapter 6, was used in the validation of MLED_BI as the results, in this context, would produce the same results as TM1.

8.2.5. Validation with TM5/TM6

TM5 and TM6 are related measures and are discussed here together. TM5 relates to "CPU memory usage during execution of initial BI report or dashboard". CPU memory usage during the execution of the initial BI report or dashboard, during the execution of an SQL query, and during re-execution of reports when changing language, currency or unit were identified in chapter 6 as a relevant measurement for BI reports. CPU memory usage during the execution of the initial BI report or dashboard was monitored using the built-in functionality of phpMyAdmin, which enables CPU status monitoring for any process executed on the localhost. The overhead of measurement was the same for all systems. During the execution of BI reports based on any method or approach, CPU system usage in the test system was between 20% and 40%. No significant differences are identified for any DM implementation method or for any BI design approach.

TM6 relates to "CPU usage during re-execution of report when changing language". To measure TM6, a language changing process in a previously executed BI report was activated while the CPU status of the web application was simultaneously monitored. The same process was applied for each BI report developed for each design approach. The language change process using the MLED_BI design approach was found to have better resource usage than the same process based on an existing BI design approach. Figure 8-1 presents the results in graph form.

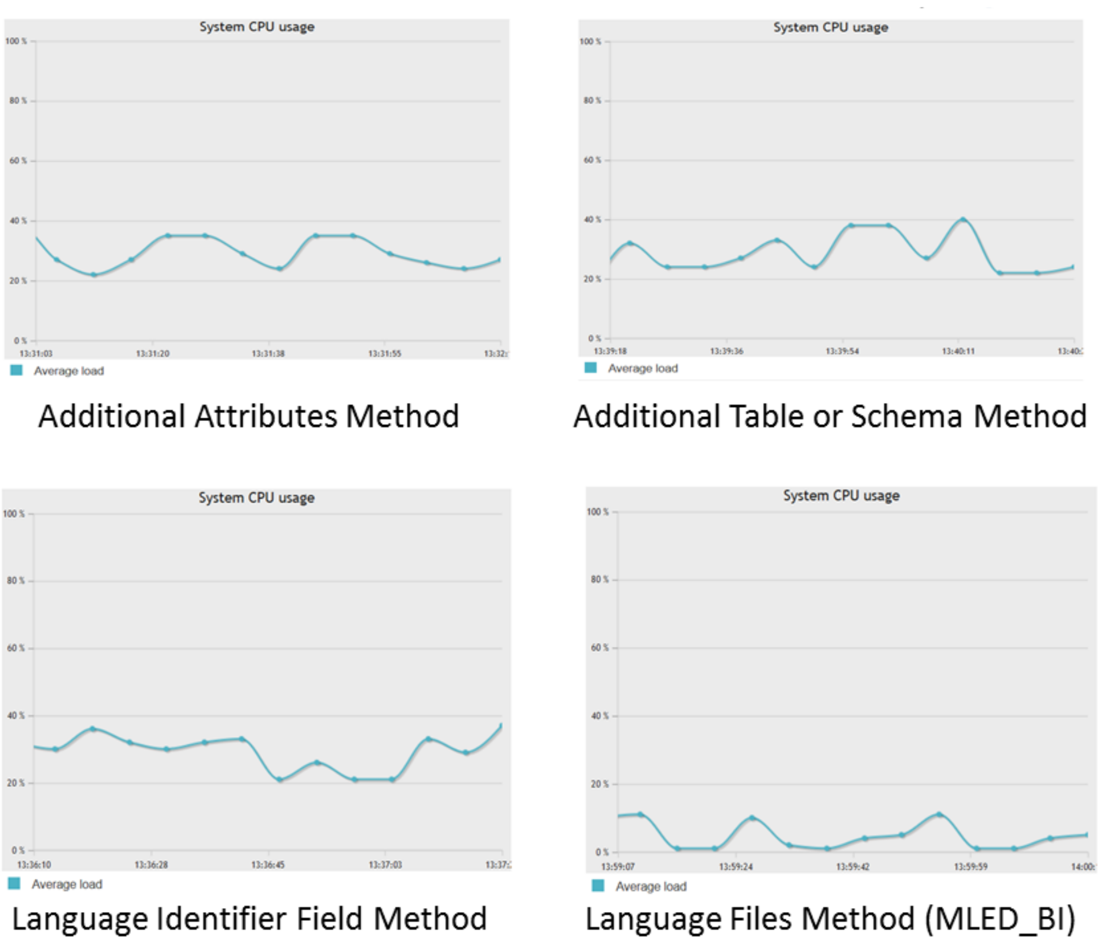


Figure 8-1: CPU usage during re-execution of report when changing language

In existing approaches to support ML (AA, ATS, LIF) , changing the language in a BI report based on existing DM design approaches requires almost the same CPU resources as the initial execution of the report which in the test system is somewhere between 20% and 40%. This result was expected as in these approaches, the SQL query needs to be re-executed when a language is changed, to take business information descriptions from the database in another language. However, this is not the case with BI reports implemented using MLED_BI. The language changing process for a BI report based on this approach had CPU usage of 10% or less in the test system. This is explained by the fact that the use of language files means that descriptive data is not stored in the dimension tables and so there is no requirement to rerun the SQL query to acquire business information descriptions (master data descriptions) in a different language: the CPU was used only to load and apply another language file in the existing web application. This is useful in environments with limited CPU resources as it could enable smoother operations with BI reports for a larger number of users. It could also prevent problems that might be created by excessive use of CPU.

8.2.6. Validation with TM7

TM7 refers to “CPU usage during execution of SQL query only”. To measure the speed of execution of the web application or a part of that application, such as an SQL query, a modular approach can be used, for example, implementing measuring code at appropriate places would be sufficient. This approach was used to measure TM1“Speed of execution time for initial BI report or dashboard” and TM2 “Speed of execution time for SQL Query”. CPU usage could be measured by executing the whole web application. However, to measure and compare TM7 “CPU usage during execution of SQL query only”, an environment independent of previously developed BI reports or the web application was needed.

This is due to the fact that each BI report requires the execution of different code, reflecting the different design approaches. The phpMyAdmin application environment was used for this element. It is important to note that using the phpMyAdmin environment itself requires additional CPU resources to enable the execution of SQL queries. However, this applied to all queries and the purpose of the test was to establish which approach had more optimal CPU usage rather than to establish the actual level of CPU usage for each item. As shown in Figure 8-2, a query on the DM based on the MLED_BI design approach was observed to have the most optimal CPU usage. While other SQL queries had large oscillations in CPU usage rising as high 80% in the test system, this SQL query had linear usage of CPU resources barely exceeding 20% in the test system.

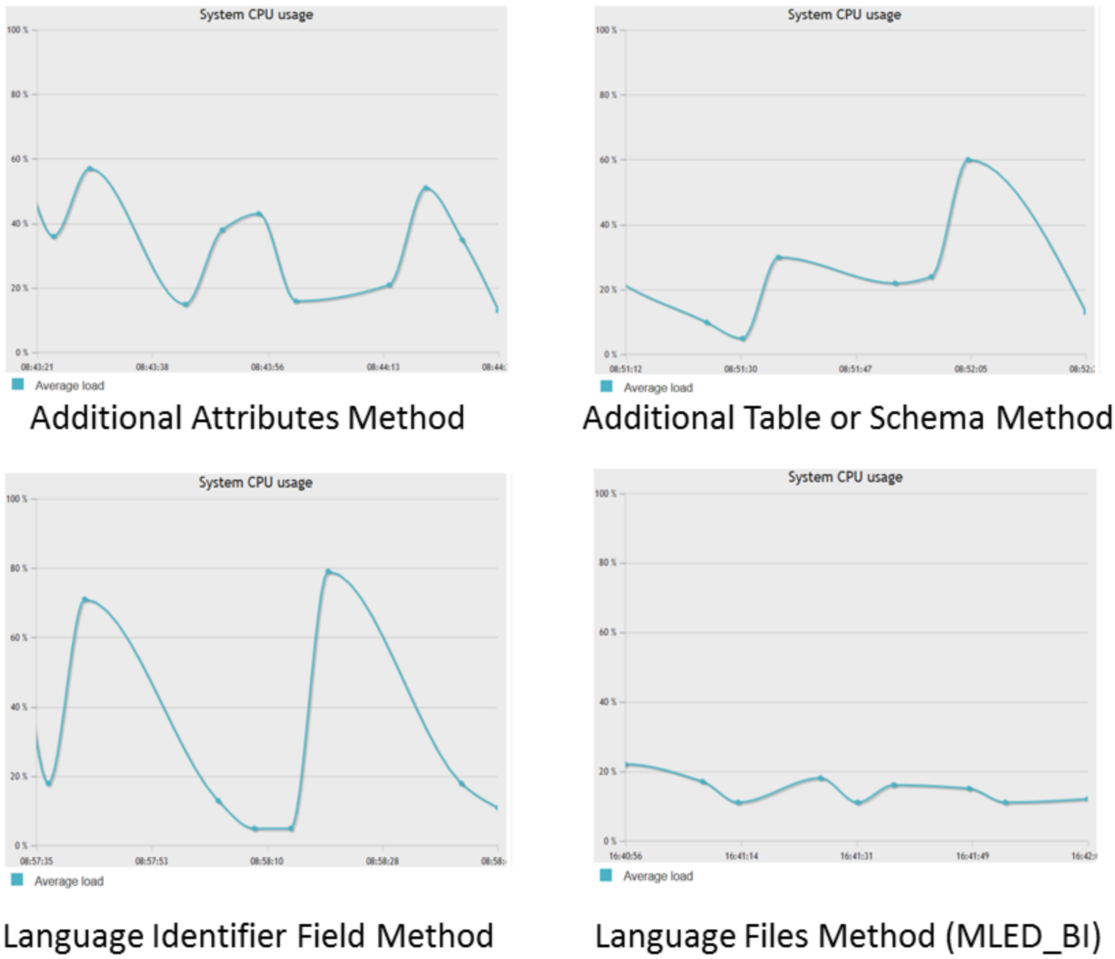


Figure 8-2: CPU usage during execution of SQL query only

8.2.7. Validation with TM8

TM8 refers to “Database memory consumption”. A clustered size sums of database tables provides sufficient information in this case. As previously noted, 1.199.989 records representing transactional data were used for each fact table used in this research. Every fact table in every data mart had the same size and required the same amount of memory, thus, Fact Table size had no influence on cumulative size differences between observed data marts. Dimension data was limited to the amount required to support testing. 216 products, 100 customers, 100 employees, 100 locations, 361 days and 6 units were used in the dimensional tables. In some cases where necessary to meet the requirements of supporting additional language in a particular design approach, data values were duplicated, as for example in the AA approach. As shown in Table 8-4, a data mart developed on the MLED_BI design philosophy uses the smallest amount of database memory in the test system. Differences in memory consumption in this example are not significant due to the small volume of master data. However, differences in memory consumption would be significant if the sample product dimension had 4.000.000 records, which is the current standard Walmart product pallet (Scrapehero.com, 2015). It was anticipated that MLED_BI design approach would reduce database memory consumption in the DM given that business information descriptions are stored outside the database as language files elsewhere on the server. Taking into account the cumulative requirement for memory to store information to the DM, including sever memory requirements for storage of language files, the actual advantage of the MLED_BI approach for this element is arguable and this is not presented as an element which either supports or does not support the MLED_BI approach.

Table 8-4: Database memory consumption comparison

Data Mart	Size in MB
phd_project _aa	85,00
phd_project _ats	85,00
phd_project_files (MLED_BI)	84,90
phd_project_lif	85,10

8.2.8. Validation with TM9

TM 9 refers to the “Amount of time required to change erroneous descriptions of descriptive attributes and hierarchies”: As there are significant structural differences between MLED_BI and the other ML design approaches, a standard measurement and comparison process was not appropriate and would not have provided a fair test. Error changing activities in BI reports based on existing BI design approaches requires external human intervention and communication with other teams. This is not the case with the MLED_BI approach. For this element, validation was through use of an expert panel, composed of six BI domain experts, from three different companies and drawn from three different countries (Germany, Austria and Slovenia). The domain experts had technical and user understanding of BI processes and had more than 50 years of combined BI experience. Table 8-5 provides insight into profiles of BI domain experts.

Table 8-5: Profiles of BI domain experts

Coding	Position	BI Experience in Years	Highest Level of Qualification
DE1	BI Solution Consultant	6	Graduate Diploma
DE2	BI Solution Consultant	6	PhD
DE3	BI Application Engineer	3	Master
DE4	Product Manager BI	8	Bachelor
DE5	Product Manager BI	15	Graduate Diploma
DE6	BI Application Engineer	15	Master

In a simple BI report implemented using the MLED_BI approach supported by a web environment implementation, less than 30 seconds is required to change erroneous business information descriptions. In the web environment, the business user can select an erroneous description. This action leads to a landing page where the user is allowed to change the erroneous content directly in the relevant language file. There is no need to communicate with any other team or to wait for processes to be executed. The process of changing erroneous content in a BI report was summarised previously in 4.7.4 and is discussed here in more detail to illustrate the issues. In an ideal environment where BI reports have been implemented as a part of a BI system based on a existing design approach, the process of changing business information descriptions would take a minimum of two hours. Empirical observation and discussions with the BI domain experts identified a timescale of between 24 and 36 hours as the standard timescale for the application of changes to business information descriptions. The reason for the lengthy timescale is the requirement to communicate with other teams and to wait for processes to be completed. For example, when a business user notices an error in BI report, a typical process requires the following stages: the user must inform the relevant department responsible for maintenance of the master data in the source system. After the error has been corrected in the source system, a member of the data maintenance team or equivalent needs to inform the responsible person in the BI or DWH team to start the ETL process to transfer the amended data from the source system to the DW and to the relevant DM. Immediate execution of ETL processes is rare in a real world environment, especially if re-aggregation of existing data is required; to avoid problems with overload, it is standard to wait until scheduled ETL processes are executed. In a more usual BI environment, which provides BI reports on data for the following day, ETL processes are usually executed every 24 or 36 hours. Despite the fact that most of the ETL processes to load master data might already be scheduled, there is still a need to inform the BI or DWH team if business information descriptions in source systems have been changed. In some cases, there are no scheduled ETL processes for specific master data – those that do not change often. After successful execution of the ETL process, a member of BI or DWH team informs the original business user that the erroneous content has been changed successfully. During the user satisfaction evaluation, discussed in section 8.4., business users identified the delays in changing erroneous content as one of the most frustrating aspects of working with reports in BI systems based on existing design approach. The MLED_BI approach offers a clear benefit in terms of speed and flexibility when changing erroneous content descriptions. It should be noted, however, that companies would need to establish policies and procedures to manage the change process.

8.2.9 Validation with TM10/TM11

The remaining technical factors identified as relevant when measuring the success of changes to support better BI reporting are TM10 “Technical scalability of proposed solution in the existing environment” and TM11 “Support for possible extension of the system in the future”. These factors cannot be measured using metrics in the same way as, for example, CPU usage. Instead, the domain experts referred to in 8.2.8, were asked for their judgements as to whether the MLED_BI approach would be scalable and extensible. The use of separate language files means that additional languages can be added easily and without needing to amend the Star Schema. The decoupling of descriptions in specific languages from descriptive content in the Star Scheme itself, promotes logical independence, supporting extensibility. Based on the evaluation of the domain experts, MLED_BI was found to support scalability (TM10) and extensibility (TM11).

8.3. Technical Evaluation

8.3.1. Performance Factors

The discussion in 8.2 demonstrates that the MLED_BI design approach provides quantifiable benefits in terms of performance and flexibility. The most significant benefits are considered to be those discussed in 8.2.2 and 8.2.3 and relate to performance speed when executing queries and particularly when re-executing a query and changing the report language. The discussion in 8.2.5 and 8.2.6 showed benefits in terms of CPU usage; this is seen as a secondary benefit but might be significant in some contexts. 8.2.7 showed that MLED_BI led to reduce DB memory consumption. However, as it is still necessary to store language files and given that memory costs are falling, this is not regarded as a key element. The evaluation with domain experts indicated that the ability to use a MCMS to change erroneous content descriptions is a significant benefit.

8.3.2. Implementation Feasibility

One consideration is whether the MLED_BI design approach can in practice be integrated within an existing BI environment. The evaluation with domain experts showed that MLED_BI is regarded as a scalable and extensible system. MLED_BI uses a modular design approach and because it is based on the widely used Star Schema construct, it does not require a complete redesign of existing systems. This is one of the features that contributes to the extensibility and scalability of the approach. MLED_BI can be applied as an additional module within an existing BI system or can be implemented as a new standalone BI system. Implementing MLED_BI in an existing BI system would require the creation of language files, and the addition of columns to dimensional tables. Those columns would hold attribute IDs to reference existing attributes with language files. MLED_BI would also require amendments to existing ETL processes. If an MCMS is created as part of the application of MLED_BI, new BI reports and back end functionality would also be created. It would not be necessary to create new data marts or new dimensional tables and existing BI reports can be retained and used in parallel with new reports based on MLED_BI since the language files approach can be implemented without removing data from the star schema. Extending existing dimensions with additional columns does not require the deletion or modification of any data. Extending existing ETL processes to support MLED_BI would not affect the data content of existing BI reports. This allows the organisation to roll back to its previous approach if this is required for any reason. In addition, previous BI reports could be integrated into the MCSM. The MLED_BI design approach supports full integration with existing BI systems. This compares favorably with existing workarounds to support ML in BI which require the creation of new dimensional tables/modification of dimensional tables. This is due to the fact that every existing BI method that supports ML has a specific architecture for dimensional tables. For example, there is no effective way to integrate the AA (additional attributes) approach for ML with a system which uses the ATS (additional tables/schemas or additional rows) to support ML. Creating new dimensional tables requires new ETL processes, new BI reports, and loading of the new data to support changes made. This is in effect a new implementation of the BI system. Moreover, once is a new BI system based on any existing approach to support ML had been created, it would be very difficult to roll back to the previous system.

8.3.3. Support for Multilingualism

The motivation for developing MLED_BI was to provide better support for multilingualism in BI. The MLED_BI approach supports the use of all languages available in the source system in BI reports and as discussed in 8.3.1, provides better performance. In addition, MLED_BI makes it possible to work with languages which are not in the source system. The number of languages used in BI reports is independent of the number of languages available in source systems. Subject to the necessary consideration of resources to transfer content, to enable additional languages for BI reports based on MLED_BI, it would be sufficient to provide only a language file with content for the new language. As soon as a new language file is available on the server, business users could use BI reports in that language. In the MLED_BI approach, there is no need to implement and enable a language in all source systems or to modify ETL processes to support the new language or to modify dimensional tables to support the new language. This is beneficial where there is a need to support BI reporting in languages or dialects that are generally not available in source systems.

In contrast to the MLED_BI approach, enabling additional languages in BI reports in a system based on traditional ML workarounds, requires the new language to be provided in the source system. The conclusion from the evaluation with domain experts is that enabling a new language in source systems is a challenging and time consuming process, which is resource intensive for both technical and human resources. In addition to the activities related to the translation of the content found with MLED_BI approach as well, additional activities include modification and extension of source systems architectures and data entry of translated content in relevant applications. Enabling an additional language, using existing ML approaches, would require source systems to be modified or extended. For example, in the AA (additional attributes) approach, the existing dimensional tables must be extended with additional columns. In the ATS approach (additional tables or schema), a new schema containing appropriate tables must be implemented for a new language. If there is a requirement for a new dialect in BI reports, even if the difference compared to the standard (received) language is minimal, the whole process used to enable a new language must be applied. This is not a case in MLED_BI based BI system.

8.3.4. Issues and Limitations identified through the Technical Validation

The domain experts described in Table 8-5 were asked to evaluate the strengths and limitations, from a technical perspective, of the MLED_BI approach. One limitation that was identified is that more resources are required for the design and development phases of MLED_BI than in existing BI ML design approaches. The design and development phase requires more resources because it is necessary to establish language files, as against existing ML workarounds which rely on extensions/amendments to the star schema.. The domain experts recognised that the MLED_BI approach produced benefits in terms of reduced processing and greater flexibility further down the data chain. All the domain experts confirmed that the benefits of implementing the MLED_BI design approach, given the anticipated future benefits, outweighed the greater resources required for initial design and implementation compared to existing ML workarounds. Evaluation with domain experts demonstrated support for the MLED_BI approach particularly for larger companies although it was noted that for smaller companies it would be necessary to calculate the break-even point and suitability would depend on the market in which the companies operated. Companies with an existing ML solution, operating with a fixed number of languages and relatively small data volumes, might find the cost of amending their systems with the MLED_BI approach outweighed the benefits.

8.4. Validation with Business / End Users

8.4.1. Design of the Validation Process

User satisfaction is regarded as a key measure in BI (Dedić & Stanier, 2016b; Petter, DeLone & McLean, 2013; Rahman, 2013; Hou, 2012; Dastgir & Mortezaie, 2010; Davidson & Deeks, 2007; DeLone & McLean, 1999, 1992) and the MLED_BI approach was evaluated for user satisfaction as well as technical effectiveness. Where participants have a high level of knowledge and expertise in relation to the research area, four to five participants are seen as a sufficient sample size to achieve data saturation in qualitative interviews (Romney, Weller & Batchelder, 1986). Guest, Bunce & Johnson (2006) propose a range of between 6-12 participants for projects having a narrow research scope focused on an homogenous target audience. Miller (2012) sees a sample size of 6-70 as sufficient taking into account the scope of research and resources available. Bonde (2013) identified that most of the scientists propose a 1+ sample size according to the research scope and type of inquiry as sufficient sample size for data saturation; meaning that the appropriate number of respondents can be between one and any other number depending on scenario and complexity of research field (Back, 2012; Baker & Edwards, 2012; Denzin, 2012).

Based on the literature, six business users who identified themselves as key BI users, coming from three international companies using multilingual BI systems were interviewed, to evaluate MLED_BI from a user perspective,. Table 8-6 presents the profiles of business users who took part in the validation processes, anonymised to preserve confidentiality.

Table 8-6: Profiles of business users who took part in the validation

Code	Position	Years of Experience with BI	Highest Level of Qualification	Relevant Characteristics
BU1	- Business Relationship Manager; - BI Key User	14	Graduate Diploma	- Communicate country level business requirements to BI team; - Currently faced with the issues of ML in BI; - Experienced problems in the context of ML in BI; - Had deep technical understanding of BI and DW; - Excellent understanding of BI from business perspective; - Graduate in Organization, Management and Information Sciences;
BU2	- Team Manager	4	Master	- Lead for Business Processes and Relationship Management; - Behave as interface between business departments and technical users; - Deep understanding of multilingual issues in BI systems: - Excellent understanding of BI and DW; - Faced with multilingual issues in BI through communication with business users; - Delegate activities concerned in this

				research; - Graduate in ERP Systems and Business Process Management;
BU3	- Business Process Manager	18	PhD	- Leader of Business Intelligence Competency Centre; - Communicate enterprise level business requirements to BI team; - BRM between business and technical departments; - Lead BI key user at enterprise level; - Deep understanding of BI from business and technical perspective; - Hold PhD in Psychology;
BU4	-Process Project Manager	1	Bachelor	- Works as Project Leader for diverse business systems, including BI; - Works as BI key user for SAP Business Warehouse (SAP BI); - Understand BI and related processes very well; - Familiar with ML issues in BI systems; - Graduate in International Business;
BU5	- Expert Associate	4	Graduate Diploma	- Business user in BI domain; - Communicate country level business requirements to BI team; - Involved in evaluation of BI reports; - Actually faced with multilingual issues in BI reports; - Country level BRM between local business and enterprise BI team; - Graduate in Economics, Organization an Management;
BU6	- Senior Expert Associate	11	Graduate Diploma	- Country level BRM between local business and enterprise BI team; - Communicate country level business requirements to BI team; - Actually faces with multilingual issues in BI reports; - Frustrated with multilingual process in current BI systems; - Use BI to support everyday activities; - Graduate in Geodesy;

Face to face interviews, on a 1:1 basis, were held in three different countries (Austria, Slovenia and Croatia). The validation process consisted of a presentation, a demonstration, hands on use of the BI environment by the interviewee and completion of an evaluation questionnaire. At the start of the interview, the researcher gave a presentation to the business user, explaining MLED_BI and the differences compared to existing ML BI design approaches. Next, the artefacts developed to validate MLED_BI, including the MCMS were demonstrated. The demonstration covered the three existing approaches to support ML in BI (AA, LIF, ATS) and the MLED_BI approach. The business users were then able to use the BI systems and experience for themselves the functionality and differences between the four approaches. This was followed by completion of an evaluation questionnaire (APPENDIX I), which was based on the user satisfaction cluster of measurements extracted from evaluation tool developed in chapter 6 and presented in table 6-5. For ease of reference, the user satisfaction measurements are given in Table 8-7.

Table 8-7: User satisfaction measurements

Code	***User Satisfaction***
BM1	*- Information content meets your needs?*
BM2	*- The information provided in the reports is accurate?*
BM3	*- Output is presented in a format that you find useful?*
BM4	*- The system and associated reports are easy for you to use?*
BM5	*- Information in the reports is up to date?*
BM6	*- Reports have the functionality that you require?*
BM7	*- The BI system is flexible enough to support easy change of "descriptive content"?*
BM8	*- Is the change of "descriptive content"* fast enough to fulfil business requirement?*
BM9	*- Exporting and sharing content functionalities meet your needs?*

As an introduction to the evaluation questionnaire, users were provided with a product sales scenario and asked to test the four approaches used to support multilingualism against this scenario and to give their comments. As all the BI reports provided the same content and the scenario, for the purposes of validation, assumes that the information content in BI reports meet the needs of business users, the first question (BM1) from Table 8-7, namely "Information content meets your needs?" was not used in the MLED_BI evaluation process.

8.4.2. Validation Process

All business users answered "Yes" to all the following questions for all BI reports regardless of the ML BI design approach used: *(BM2)* "*The information provided in the reports is accurate?*", *(BM3) "Output is presented in a format that you find useful?*", *(BM4)* "*The system and associated reports are easy for you to use?*", *(BM5)* "*Information in the reports is up to date?*", and *(BM9)* "*Exporting and sharing content functionalities meet your needs?*". Due to the nature of the scenario, the application of ML in BI, and the output of the BI reports presented in the demonstration this answers was expected since to ensure a fair test, all the reports were based on the same data source and provided the same information. A conclusion would be that every ML BI design approach supported by any DM implementation method has the capability to provide BI reports that meet user needs and to provide a BI system that delivers accurate information presented in useful format, reports that are easy to use, are up to date, and have appropriate content sharing functionalities. In this context, we found no advantage of MLED_BI over existing BI design approaches, or DM implementation method.

However, based on the scenario, only BI reports developed on the MLED_BI design approach which supports a MCMS received "Yes" from all business users as an answer to the following questions: *(BM6)* "*Reports have the functionality that you require?*", *(BM7)* "*The BI system is flexible enough to support easy change of "descriptive content"?*", and *(BM8)* "*Is the change of descriptive content fast enough to fulfil business requirement?*". This confirmed that one of the end user advantages of MLED_BI, compared to existing BI design approaches, is that the greater immunity to change and data independence supported by the MLED_BI approach, enables the user to carry out activities such as changing the language of previously executed reports, making corrections to erroneous content and enabling new languages for reports.

Business users were able to provide free text comments on the MLED_BI approach. Most of these comments related to the additional functionality made possible by the MCMS as this was an area where the MLED_BI design approach provided a different end user experience to existing ML BI approaches. The additional comments given by users are shown here in Table 8-8.

Table 8-8: Additional comments provided by users during evaluation of business/end user satisfaction

Comment	User ID
"As we have similar problems every day, proposed solution is interesting and will bring improvements"	BU1
"I would like to have it (proposed solution) in all relevant ERP systems" *"Simple, fast, flexible and uncomplicated for the end users."*	BU2
"The proposed MLED approach is very helpful in regard to performance, usability and business requirements"	BU3
"In my opinion this approach is an improvement to the existing approaches. If we would have the possibility to implement this at our company, I would vote YES."	BU4

"Report by MLED approach is much faster than other approaches. Easy usability. Users can define their own content (descriptions of the data). No frustration." *"I like it (MLED approach) a lot. It would be great to be implemented in our multilingual system."*	BU5
A Report is much faster. Language change can be made with just one click without a need to start a report (again) or even whole SAP BW system." *"I would apply it (MLED approach) immediately, not only in SAP BW, but in our IMAge system as well."*	BU6

8.4.3. Issues and Limitations identified through End User Validation

One issue that was identified during the validation with business users is highlighted by the comment that *"Authorization is very important."* (BU3). The flexibility provided by the MCMS gives end users control over their data but makes changes to master data possible without the checks and balances provided by traditional approaches to changing master data. This is an implementation and management issue for the companies that implement MLED_BI but existing data security policies would need to be modified to reflect the change in functionality. This point was also noted in 8.2.8. The MLED_BI approach gives end users more flexibility and control and it was expected that for this reason, the MCMS would be welcomed by end users. However, the other side of the increased flexibility for end users, is that a strict change management policy would be required as implementing MLED_BI might have implications for corporate data governance.

One user suggested an extension to provide additional flexibility. This was to extend the functionality of the MCMS to include automatic translation: *"Automatic translation of already used variables in another report of the same language. Possibility to translate a whole variable package of a language at once in the frontend."* (BU4). This element is outside the current scope of MLED_BI.

8.5. Summary

The development of the large scale implementation of MLED_BI, described in chapter 7, demonstrated that it was possible to translate MLED_BI into a fully functionally real-world artefact. Technical functionality and business/end users satisfaction were assessed using the measures identified in the evaluation tool developed in Chapter 6.

The evaluation of technical functionality showed that MLED_BI compared favorably to existing BI design approaches in terms of:

- *Speed of execution time for Initial BI report or dashboard;*
- *Speed of execution time for SQL query;*
- *Speed of re-execution time when changing report language, currency or unit;*
- *CPU memory usage during execution of initial BI report or dashboard;*
- *CPU memory usage during execution of SQL query;*
- *CPU memory usage during re-execution of report when changing language, currency or unit;*
- *Amount of Time required to change erroneous descriptions of descriptive attributes and hierarchies;*

The technical functionality measurement "*Database memory consumption*" also showed some advantage when using MLED_BI but as discussed in section 8.2. and 8.3.1., the benefits are arguable if other factors such as data volumes are taken into consideration. Evaluation with domain experts indicated that MLED_BI is more scalable and more easily integrated into existing BI environments than existing approaches.

An important limitation of the MLED_BI approach is that the initial design and implementation requires more resources for the design and development phases than existing BI design approaches. For larger organisations, this initial increased resource demand would be outweighed by benefits, such as increased performance and flexibility in data management, following implementation. For smaller companies, however, and particularly those that have limited ML requirements, the benefits of MLED_BI would be questionable.

The evaluation of *business/user satisfaction* confirmed the benefits of MLED_BI, including the multilingual content management system, compared to existing ML BI design approaches in respect of activities such as changing language of already executed report, making corrections to erroneous content, or enabling new languages for reports. However, no advantage was identified, compared to existing approaches, in terms of provision of BI reports suggesting that for non technical users, one of the main benefits of the MLED_BI approach is the greater flexibility and ease of data manipulation that

MLED_BI provides. It was noted, however, that this flexibility would have implications for BI data management in companies.

8.6. Conclusion

This chapter presented the validation of the MLED_BI design approach and evaluated the findings from the validation. The measures identified in the evaluation tool developed in Chapter 6 were used to validate the technical performance, extensibility and scalability of the system and end user satisfaction. The results of the validation showed that MLED_BI provides technical advantages in terms of performance, particularly when changing the language of reports and that the MCMS, which is made possible by the MLED_BI approach, provides users with greater flexibility and control of BI processes. The process identified some issues and limitations in that the increased upfront design and development costs of MLED_BI make the approach most suitable for larger companies and the increased control and flexibility for end users would need to be balanced by data governance policies and procedures. The following chapter, chapter 9, presents the overall evaluation and conclusions from the research and gives suggestions for future work.

Chapter 9: Conclusions and Future Work

9.1. Introduction

This chapter provides an overview of the research, summarising the investigation carried out, the findings from the investigation and the validation of the research. The chapter evaluates the outcomes from the research and the research as a whole and presents the conclusions from the research. The first section outlines the content of each chapter of the thesis. The chapter next summarises the relationship between the research objectives and the methods of investigation and discusses the contribution to knowledge. The research limitations are discussed and areas for future work are identified.

9.2 Research Overview

The main aim of this research was to develop a novel design solution to the problem of supporting multilingualism in Business Information as a contribution to knowledge. To support this aim a number of objectives were developed. The study began by critically reviewing the existing literature about BI and ML, current DW/BI theories, tools and techniques and DW/BI approaches to support ML in BI. The literature review identified a number of issues and challenges when considering ML from the BI perspective. From the limitations identified during the literature review, two additional objectives were identified as minor contributions to knowledge, the development of a novel, holistic BI framework (HBIF) to support understanding of the BI environment and the development of an evaluation tool to support measurement of the success of changes to the BI reporting environment. A novel design approach, MLED_BI was developed and validated initially through a proof-of-concept artefact and then through a full implementation that simulated the real world environment. The MLED_BI approach was validated by business users and technical domain experts and was found to make a significant contribution to the issue of ML in BI although some limitations were also identified. The following sections give an overview of each chapter in the thesis, identifying some key issues.

- **Chapter 1 (Introduction)**

This chapter provides an overview of the research and gives the background and motivation for the research. The aim and objectives of the research are explained and the contribution to knowledge and ethical issues are discussed. The research approach is also discussed and justified. It was initially intended to adopt a positivist approach. However, as it was identified that acceptance and usability are also key elements in evaluating the effectiveness of the MLED_BI approach, the research also reflects the philosophy of interpretivism. Thus, this research uses a mixed methods research, which combines both quantitative and qualitative research approaches. The principal limitation of adopting a mixed method approach in this research was the significantly greater resources needed for the validation and evaluation of the research compared to a positivist approach. However, the mixed methods approach was helpful because it supported a more in depth validation and evaluation of MLED_BI, complementing the limitations of quantitative and qualitative approaches when used individually.

- **Chapter 2 (Literature Review)**

Chapter 2 presented the literature review which provided a comprehensive discussion of BI, including the definition of BI and current trends in BI. Multilingualism was defined and the issues and challenges of ML in BI were discussed. The underpinning concepts of BI which provided the theoretical basis for this research were reviewed, including data independence and immunity from changes, DW design and development approaches, DW modelling concepts, ETL processes, and data presentation and visualisation. Existing approaches to supporting ML in BI were discussed and evaluated. The conclusion from the literature review was that existing solutions to support ML in BI are not optimal and that a new solution, based on a redefinition of the Star Schema and the concept of immunity from changes, was required.

- **Chapter 3 (A Holistic Framework for Business Intelligence)**

To support the development of the new solution identified as necessary in Chapter 2, it was in turn necessary to identify the components of BI systems that would be affected by support for ML. Chapter 3 included a further literature review which examined business intelligence frameworks to support requirements analysis for the development of MLED_BI. Twelve existing BI frameworks and data warehousing approaches were discussed and evaluated and it was shown that none of the existing BI framework satisfied the requirements of the research. For this reason, a new framework, HBIF, was developed. Using an iterative approach, all components from each of the evaluated frameworks were analysed and grouped into appropriate categories and allocated to the appropriate data layer to provide the basis for the first version of HBIF. The HBIF was then developed further based on input from domain experts, and was validated by means of a pilot survey. HBIF was iterated based on the feedback from the pilot survey and a second version was developed. The second version of HBIF was validated by means of a larger survey and was modified based on the feedback received. The final version of HBIF is presented at the end of Chapter 3. The HBIF is one of the minor contributions to knowledge in the thesis.

- **Chapter 4 (MLED_BI: A New BI Design Approach**

Chapter 4 discussed the design and development of MLED_BI. Based on the findings of the literature review in chapter 2 and the analysis of BI systems supported by the HBIF developed in chapter 3, chapter 4 identified the requirements for a new approach to support multilingualism, including the need to support immunity from changes and the requirement to be compatible with existing BI environments. The underpinning concepts for MLED_BI were presented and the chapter explained how the MLED_BI approach was compatible with existing BI approaches based on Inmon and Kimball. The chapter demonstrated that the revised Star Schema approach used in MLED_BI supports immunity to changes at different levels of the BI environment and that the MLED_BI approach makes possible the use of a Multilingual Content Management System to support improved reporting and flexibility in the management of languages.

- **Chapter 5 (MLED_BI: Initial Validation and Technical Feasibility)**

The validation of MLED_BI consisted of three phases; an initial validation using a Proof of Concept to investigate technical feasibility, a full implementation to examine feasibility and performance in more detail and a qualitative evaluation with end users and technical domain experts. Chapter 5 presented the first phase of the validation, the PoC. The development of the PoC demonstrated that the MLED_BI design approach could be translated into implementation and was compatible with both the Inmon and Kimball approaches. The PoC was evaluated against one of the existing approaches used to support ML, the Language Identifier Approach (LIF). The LIF approach was chosen as experimentation had shown that LIF was the fastest of the existing ML design approaches. The results of the tests described in chapter 5 showed that MLED_BI met the requirement to provide improved performance. However, it was acknowledged that the PoC used a trivial data set and was evaluated on only one metric. Having established technical feasibility, it was necessary to use a wider range of measures and a full implementation to validate and evaluate MLED_BI in detail.

- **Chapter 6 (Development of an Evaluation Tool)**

Chapter 6 discussed the development of an evaluation tool to support the further validation of the MLED_BI design approach. Technical feasibility was demonstrated through the PoC described in Chapter 5 but it was also necessary to evaluate the design approach in terms of whether it provided a better technical and reporting experience for technical domain experts and business end users. A review of the literature on evaluation tools in Information Systems established that a suitable tool did not exist. However, the literature review provided the basis for the development of an appropriate evaluation tool. The review identified user satisfaction and technical functionality as the most important clusters to be considered when measuring success of BI improvements in the context of BI reporting. Appropriate metrics for each of these two clusters were identified from the literature as were BI user groups, roles and user activities in BI. The evaluation tool was tested by means of a pilot survey which led to a number of revisions. The revised version of the tool was then evaluated with users working in the BI field and minor revisions were made. The final version of the tool includes the core elements used to evaluate MLED_BI and optional elements which can be used to extend the evaluation tool, depending on the requirements of the user. The evaluation tool is one of the minor contributions to knowledge in the thesis.

- **Chapter 7 (Implementation of BI Design Approaches)**

Chapter 7 describes the implementation of MLED_BI which was developed to support the full validation of the novel design approach. The MLED_BI approach enables the use of a MCMS to support the management of multilingual elements and this is one of the benefits of the MLED_BI design approach. For this reason a Multilingual Content Management System (MCMS) was also developed as part of the implementation. To allow a full evaluation of MLED_BI against existing approaches to support ML in BI, the three existing ML design approaches were also implemented, the Additional Attributes approach (AA), the Language Identifier approach (LIF) and the Additional Tables/Schema approach (ATS).

- **Chapter 8 (Validation of the MLED_BI Design Approach)**

Chapter 8 describes the quantitative and qualitative approaches used in the validation of MLED_BI. The quantitative validation was based on the use of technical metrics from the evaluation tool described in chapter 6. To perform the metrics based evaluation, data was collected from processes executed in a controlled environment, the results were recorded and then compared. The same tests were applied to the MLED_BI environment and to the AA, LIF and ATS environments. The results of the technical evaluation showed that the use of MLED_BI led to improved performance a number of areas particularly when re-executing a query in a different language and in terms of CPU usage. For the metric, use of database memory, MLED_BI was not found to offer any significant advantage. Qualitative validation was carried out with domain experts who were able to test the MLED_BI implementation and compare this with implementations based on existing ML BI design approaches. The results of the qualitative validation showed that the MLED_BI was found to provide a satisfactory solution to the challenges of ML in BI and that the MCMS provided clear benefits for end users. It was identified, however, that the greater upfront design and development costs of the MLED_BI approach meant that this solution might not be suitable for smaller companies and that organisations adopting MLED_BI would need to develop policies to deal with authorisation and change management.

9.3 Research Summary

This section presents the objectives set for the research, the method of investigation used for each objective and shows the chapter in which each objective was addressed.

Table 9-1: Objectives summary

	Objective	*Method(s) of investigation*	*Chapter*
1	To critically review the literature covering issues involved in ML in BI, current DW/BI theories, tools and techniques and relevant data design concepts such as data independence, BI approaches used to support BI in multilingual context, and validation and evaluation of BI systems	- Secondary research through review of existing literature	2
2	To develop a novel Multilingual Enabled Design solution (MLED_BI) to the problem of supporting multilingualism in BI	MLED_BI was developed based on a synthesis of the findings from the secondary research and information from HBIF and novel redefinition of the Star Schema	4
3	To initially validate that MLED_BI translates into functional implementation by establishing technical feasibility through a proof of concept implementation before considering other issues	- validation of technical feasibility though a PoC implementation - experimental validation by metric	5
4	To further validate that MLED_BI translates into full-functional implementation by establishing technical feasibility through a large-scale system that simulates the full real world environment to support comprehensive validation of approach	- Implementing BI environment by applying inputs from chapter 4, and considering findings from PoC artefact and relevant literature	7
5	To conduct comprehensive validation of MLED_BI design approach	- Comparison of performance metrics achieved in a multilingual BI system based on MLED_BI and on conventional BI design approach - semi-structured interviews with business users working with multilingual BI system on daily basis - semi-structured interviews with technical BI/DWH experts	8
6	To critically evaluate the outcomes of the research	- Synthesising the findings of the thesis	9
7	To develop and validate a novel BI Framework to support the analysis stage of MLED_BI	- Identification of key elements from the review of existing literature and investigation with BI and DWH domain experts - pilot validation through survey of 29 users Bi users - Final validation and evaluation through survey with 109 users	3
8	To develop an evaluation tool to provide evaluation criteria to measure the success of changes to existing BI solutions to support overall validation and evaluation of MLED_BI	- The evaluation tool was developed based on synthesis of elements identified through a review of the existing literature and discussions with BI/DW team members from eight European companies using BI - pilot validation through a survey of 10 BI domain experts/report users - Final validation and evaluation through a survey of 30 (domain experts) working in BI field	6

9.4 Research Contribution

This research makes a number of contributions to knowledge. The main research contribution of this thesis is MLED_BI, a novel BI design approach to support ML in a BI environment. MLED_BI is based on a revised approach to the star schema which reintroduces data independence and immunity from changes and enables extensible support for ML in BI by making possible the use of a multilingual content management system to provide greater flexibility in ML data reporting and ML data manipulation.

Minor contributions of the thesis are the development of the HBIF (Holistic Business Intelligence Framework), the development of the Evaluation Tool and the contribution to the body of knowledge represented by the review of BI and ML in BI.

The HBIF is a novel framework which uses the 3 layer approach to identify the five perspectives (concepts, users, applications (software), types of data, and hardware) which describe the BI environment. In this research, the HBIF was used to support analysis and identify the elements of the BI environment which might be affected when considering changes to the BI system. However, the HBIF is generic and is also customisable and extensible and represents a contribution to the understanding of the BI environment.

The evaluation tool also addresses a gap in the literature as the review identified that a comprehensive tool to measure the success of changes to the reporting layer in BI environment did not exist. Like the HBIF, the evaluation tool is extensible and customisable and represents a contribution to knowledge and to the evaluation of BI reporting.

Multilingualism in BI is an understudied element although as discussed in chapter 2, ML is increasingly important in BI applications. The thesis presents a comprehensive review of the issues, challenges and existing approaches and this also represents a contribution to the body of knowledge.

9.5 Research Journey

In addition to supporting the further development and advancement of knowledge in the scope of technical and conceptual competence in the field of the research, this research provided a structured and solid framework to support the researcher in progressing from an industry professional to a fully formed researcher. The research journey began by extending the researcher's competence and understanding of research philosophies and relevant concepts. Once the researcher had acquired the ability to evaluate existing research approaches and to identify and apply those appropriate for this research, the next stage of progression included conducting a real-world research that further refined and extended his research skills. Real-world research was carried out several times and encompassed different research methods from diametrically opposite research philosophies, thus extending the actual practical research experience of the researcher by including different perspectives, strategies, design and methods. The research process initially provided insight into a structured, organised and theoretically-based approach to problem solving, which researcher sharpened and successfully applied throughout the course of the research. Reflecting on the research from a personal point of view, one of the biggest benefit of the research was found to be the way in which critical thinking became embedded in the mindset of the researcher. There was a shift from the simple and ad-hoc view of reality to the view where the reality is to be seen through a complex network of relevant and mutually intertwined perspectives and phenomena.

9.6 Research Limitations

This research recognises some limitations and restrictions. The investigation considered only business information descriptions, known as master data. Other aspects of BI internationalisation related to multilingualism are not considered, such as different types of script, the direction of writing of specific language, or currency and unit conversions for different countries. The MLED_BI files approach supports all known languages but further work would be required to address some of these issues at the presentation level.

Due to European data protection legislation, it was not possible to validate MLED_BI using a live system. The implementation developed to validate MLED_BI simulated a production system but did not use production data volumes and was not used on an enterprise network. However, the data used was a simulation of real world data, data volumes were sufficient for testing and tests were conducted on all four design approaches under the same conditions.

The evaluation by technical experts identified the greater design and development effort of MLED_BI as a possible limitation since for smaller companies, or companies with limited multilingual requirements, the greater upfront cost might outweigh the benefits of implementing MLED_BI. However for larger companies and companies working in a true multilingual environment, the benefits of MLED_BI were clear.

The MLED_BI design approach makes the use of a multilingual content management system possible and this in turns offers end users much greater flexibility in multilingual data manipulation. This is seen as a strength of MLED_BI not as a limitation but it is necessary to recognise that implementing a MCMS would require companies to develop policies to regulate data changes.

9.7 Areas for Further Work

Based on the discussion, the following areas have been identified for future work

- The design and development effort required by the MLED_BI approach was identified as one of the limitations of the research as it indicates that MLED_BI would be challenging to apply in the context of smaller organisations. One area identified for future work is the development of a tool to support the implementation of MLED_BI and the development of MCMS
- The evaluation of MLED_BI identified that further work is required to address ML presentation issues at reporting level. This suggest an area for future work, linked to the development of a tool to support the implementation of MLED_BI and a MCMS.
- ETL processes were implemented to support the validation of MLED_BI but ETL itself was outside the scope of the research. An area for further work is to investigate ETL in the context of the MLED_BI design approach and to examine whether the MLED_BI design approach should be extended to include a specific ML ETL element
- The focus of MLED_BI is on the use of structured data in a multilingual environment. Extending the MLED_BI approach into fields such as Big Data,

where the focus is on unstructured and semi-structured data, is identified as an area for further research to support the delivery of multilingual content to end users.

References

Alexander, A., 2014. Case Studies: Business intelligence. *Accounting Today*, (June), p. 32.

American Institute of CPAs (2015). Business Intelligence. *AICPA Website*. Available at: http://www.aicpa.org/INTERESTAREAS/INFORMATIONTECHNOLOGY/RESOURCES/BUSINESSINTELLIGENCE/Pages/default.aspx [Accessed 27 March, 2015].

Anadiotis, G., 2013. Agile Business Intelligence : Reshaping the Landscape., pp.1–15.

Anderson, C., 2010. Presenting and evaluating qualitative research. *American Journal of Pharmaceutical Education*, *74*(8). Available at: http://doi.org/10.5688/aj7408141

Androutsopoulos, J., 2015. Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism*, *19*(2), pp. 185–205. Available at: http://doi.org/10.1177/1367006913489198

Arefin, M. S., Morimoto, Y., & Yasmin, A., 2011. Multilingual Content Management in Web Environment. In *2011 International Conference on Information Science and Applications*, pp. 1–9. Available at: http://doi.org/10.1109/ICISA.2011.5772328

Atzeni, P., Jensen, C. S., Orsi, G., Ram, S., Tanca, L., & Torlone, R., 2013. The relational model is dead, SQL is dead, and I don't feel so good myself. *ACM SIGMOD Record*, *42*(1), pp. 64. Available at: http://doi.org/10.1145/2503792.2503808

Aufaure, M.-A., Chiky, R., Curé, O., Khrouf, H., & Kepeklian, G., 2016. From Business Intelligence to semantic data stream management. *Future Generation Computer Systems*, *63*, pp. 100–107. Available at: http://doi.org/10.1016/j.future.2015.11.015

Awad, M. M. I., Abdullah, M. S., & Ali, A. B. M., 2011. Extending ETL framework using service oriented architecture. *Procedia Computer Science*, *3*, pp. 110–114. Available at: http://doi.org/10.1016/j.procs.2010.12.019

Baars, H., & Kemper, H.-G., 2008. Management Support with Structured and Unstructured Data—An Integrated Business Intelligence Framework. *Information Systems Management*, *25*(2), pp. 132–148. Available at: http://doi.org/10.1080/10580530801941058

Back, L., 2012. Expert voices. Expert voices. In S.E. Baker & R. Edwards. *How many qualitative interviews is enough.* National Centre for Research Methods Review Discussion Paper, pp. 12–14.

Bailey, J. E., & Pearson, S. W., 1983. Development of a Tool for Measuring and Analyzing Computer User Satisfaction. *Institute of Management Sciences, 29*(5), pp. 530–545.

Baker, S. E., & Edwards, R., 2012. *How many qualitative interviews is enough? Expert voices and early career reflections on sampling and cases in qualitative research*. Southampton, United Kingdom: National Centre for Research Methods.

Bansal, S. K., & Kagemann, S., 2015. Integrating Big Data: A Semantic Extract-Transform-Load Framework. *Computer*, *48*(3), pp. 42–50. Available at: http://doi.org/10.1109/MC.2015.76

Beattie, V., & Pratt, K., 2003. Issues concerning web-based business reporting: An analysis of the views of interested parties. *British Accounting Review*, *35*(2), pp. 155–187. Available at: http://doi.org/10.1016/S0890-8389(03)00016-7

Boakye, E. A., 2012. From Design and Build to Implementation and Validation. In C. McKinney (Ed.), *Implementing Business Intelligence in Your Healthcare Organization*, pp. 73–86. Chicago: HIMSS.

Bonde, D., 2013. Qualitative Interviews: When Enough is Enough. *Qualitative Market Research*, pp. 1–10.

Boone, H. N. J., & Boone, D. A., 2012. Analyzing Likert data. *Journal of Extension*, 50(2), p. 30. http://doi.org/10.1111/j.1365-2929.2004.02012.x

Borrego, M., Douglas, E. P., & Amelink, C. T., 2009. Quantitative, qualitative, and mixed research methods in engineering education. *... of Engineering Education*, *98*(1), pp. 53–66. Available at: http://doi.org/10.1002/j.2168-9830.2009.tb01005.x

Bouzeghoub, M., & Kedad, Z., 2000. A Logical Model for Data Warehouse Design and Evolution. *Data Warehousing and Knowledge Discovery SE - 18, 1874*, pp. 178–188. Available at: http://doi.org/10.1007/3-540-44466-1_18

Boyer, J., Frank, B., Green, B., & Harris, T., 2010. *A Practical Guide for Achieving BI Excellence*. Ketchum, US: MC Press Online, LLC.

Brannon, N., 2010. Business Intelligence and E-Discovery. *Intellectual Property & Technology Law Journal*, 22(7), pp. 1–5.

Breslin, M., 2004. Data Warehousing Battle of the Giants : Comparing the Basics of the Kimball and Inmon Models. *Business Intelligence Journal*, *9*(1), pp. 6–20.

Brichni, M., Dupuy-Chessa, S., Gzara, L., Mandran, N., & Jeannet, C., 2017. BI4BI: A continuous evaluation system for Business Intelligence systems. *Expert Systems with Applications*, *76*, pp. 97–112. Available at: http://doi.org/10.1016/j.eswa.2017.01.018

Bryman, A., & NetLibrary Inc., 1988. *Quantity and Quality in Social Research*. *Contemporary social research series 18*. London, UK: Routledge. Available at: http://doi.org/10.4324/9780203410028

Cenoz, J., 2009. *Towards Multilingual Education: Basque Educational Research from an International Perspective*, Bristol, UK: Multilingual Matters.

Chamberlin, D. D., 1976. Relational Data-Base Management Systems. *ACM Comput. Surv.*, *8*(1), pp. 43–66. Available at: http://doi.org/10.1145/356662.356665

Chaudhuri, S., Dayal, U., & Narasayya, V., 2011. An Overview of Business Intelligence Technology. *Communications of the ACM*, *54*(8), p. 88. Available at: http://doi.org/10.1145/1978542.1978562

Chen, X., Li, J., Zhang, W., Zou, Z., Ding, F., Liu, Y., & Li, Q., 2006. The Development and Application of Data Warehouse and Data Mining in Aluminum Electrolysis Control Systems. *TMS Light Metals*, *2006*, pp. 515–519.

Chu, T.-H., 2013. A Framework for BI Systems Implementation in Manufacturing. *International Journal of Electronic Business Management*, *11*(2), pp. 113–120.

Chu, Y. S., Tseng, S. S., Tsai, Y. J., & Luo, R. J., 2009. An Intelligent Questionnaire Analysis Expert System. *Expert Systems with Applications*, *36*(2 PART 2), pp. 2699–2710. Available at: http://doi.org/10.1016/j.eswa.2008.01.076

Chung, W., Zhang, Y., Huang, Z., Wang, G., Ong, T. H., & Chen, H., 2004. Internet searching and browsing in a multilingual world: An experiment on the Chinese business intelligence portal (CBizPort). *Journal of the American Society for Information Science and Technology*, *55*(9), pp. 818–831. Available at: http://doi.org/10.1002/asi.20025

Cios, K., Pedrycz, W., Swiniarski, R., & Kurgan, L., 2007. *A Knowledge Discovery Approach*. New York, USA: Springer Science+Business Media, LLC.

Clason, D. L., & Dormody, T. J., 1994. Analyzing Data Measured By Individual Likert-Type Items. *Journal of Agricultural Education*, *35*(4), pp. 31–35. Available at: http://doi.org/10.5032/jae.1994.04031

Codd, E. F., 1970. A relational model of data for large shared data banks. *Communications of the ACM*, *13*(6), pp. 377–387. Available at: http://doi.org/10.1145/362384.362685

Collins, R. W., 2002. Software localization for internet software: Issues and methods. *IEEE Software*, *19*(2), pp. 74–80. Available at: http://doi.org/10.1109/52.991367

Corr, L., & Stagnitto, J., 2011. *Agile Data Warehouse Design: Collaborative Dimensional Modeling, from Whiteboard to Star Schema*. Leeds: DecisionOne Press.

Cravero, A., & Sepulveda, S., 2015. Using GORE in data warehouse: A systematic mapping study. *IEEE Latin America Transactions*, *13*(5), pp. 1654–1660. Available at: http://doi.org/10.1109/TLA.2015.7112028

Creswell, J. W., 2003. *Research design Qualitative quantitative and mixed methods approaches*. *Research design Qualitative quantitative and mixed methods approaches* (2nd ed., Vol. 7). London, UK: SAGE Publications, Inc. Available at: http://doi.org/10.3109/08941939.2012.723954

Croatian Constitution, 2014. Constitution of the Republic of Croatia, Article 12.1.

Curino, C. A., Difallah, D. E., Pavlo, A., & Cudre-Mauroux, P., 2012. Benchmarking OLTP/web databases in the cloud. *Proceedings of the Fourth International Workshop on Cloud Data Management - CloudDB '12*, p. 17. Available at: http://doi.org/10.1145/2390021.2390025

Dang, Y., Zhang, Y., Hu, P. J. H., Brown, S. A., Ku, Y., Wang, J. H., & Chen, H., 2014. An integrated framework for analyzing multilingual content in Web 2.0 social media. *Decision Support Systems*, *61*(1), pp. 126–135. Available at: http://doi.org/10.1016/j.dss.2014.02.004

Darwen, H., 2012. The relational model: Beginning of an era. *IEEE Annals of the History of Computing*, *34*(4), pp. 30–37. Available at: http://doi.org/10.1109/MAHC.2012.50

Dastgir, M., & Mortezaie, A. S., 2012. Factors Affecting the End-User Computing. *Business Intelligence Journal*, *5*(2), pp. 292–298.

Date, C. J., 1975). *An Introduction to Database Systems*. *An Introduction to Database Systems* (Vol. 1). Pearson/Addison Wesley. Available at: http://doi.org/10.3145/epi.2009.jul.14

Davis, G. B., & Olson, M. H., 1985. *Management information systems: conceptual foundations, structure, and development* (2nd ed.). New York City, US: McGraw-Hill, Inc.

Davison, J., & Deeks, D., 2007. Measuring the potential success of information system implementation. *Measuring Business Excellence*, 11(4), pp. 75–81.

De Groot, A., 1969. *Methodology*. *Foundations of inference in the behavioral sciences*. The Hague: Mouton.

Dedić, N., & Stanier, C., 2016a. An Evaluation of the Challenges of Multilingualism in Data Warehouse Development. In *Proceedings of the 18th International Conference on Enterprise Information Systems* (Vol. 1, pp. 196–206). Rome: SCITEPRESS - Science and Technology Publications, Lda. Available at: http://doi.org/10.5220/0005858401960206

Dedić N., & Stanier, C., 2016b. Measuring the Success of Changes to Existing Business Intelligence Solutions to Improve Business Intelligence Reporting. In *Research and Practical Issues of Enterprise Information Systems, CONFENIS 2016. Lecture Notes in Business Information Processing* (Vol. 268, pp. 225–236). Vienna, Austria: Springer International Publishing AG 2017. Available at: http://doi.org/10.1007/978-3-319-49944-4_17

Dedić, N., & Stanier, C., 2017a. Towards Differentiating Business Intelligence, Big Data, Data Analytics and Knowledge Discovery. In *Innovations in Enterprise Information Systems Management and Engineering: 5th International Conference, ERP Future 2016 – Research. Lecture Notes in Business Information Processing* (Vol. 285, pp. 114–122). Hagenberg, Austria: Springer International Publishing AG 2017. Available at: http://doi.org/10.1007/978-3-319-58801-8_10

Dedić, N., & Stanier, C., 2017b. "Measuring the success of changes to Business Intelligence solutions to improve Business Intelligence reporting", *Journal of Management Analytics*, Volume 4, Issue 2. pp. 130 – 144. Available at: http://dx.doi.org/10.1080/23270012.2017.1299048

Dekkers, J., Versendaal, J., & Batenburg, R., 2007. Organising for Business Intelligence: A framework for aligning the use and development of information. In *BLED 2007 Proceedings* (pp. 625–636). Bled, Slovenia.

DeLone, W. H., & McLean, E. R., 2003. The DeLone and McLean Model of Information Systems Success. *Journal of Management Information Systems*, *19*(4), pp. 9–30.

DeLone, W. H., & McLean, E. R., 1992. Information systems success: The quest for the dependent variable. *Information Systems Research*, *3*(1), pp. 60–95. Available at: http://doi.org/10.1287/isre.3.1.60

Denzin, N., 2012. Expert voices. In S.E. Baker & R. Edwards. *How many qualitative interviews is enough.* National Centre for Research Methods Review Discussion Paper., pp. 23-24.

Devens, R. M., 1865. *Cyclopædia of Commercial and Business Anecdotes: Comprising Interesting Reminiscences and Facts, Remarkable Traits and Humors ... of Merchants, Traders, Bankers ... Etc. in All Ages and Countries*. Oakland: D. Appleton.

Devlin, B. A., & Murphy, P. T., 1988. An architecture for a business and information system. *IBM Systems Journal*, *27*(1), pp. 60–80. Available at: http://doi.org/10.1147/sj.271.0060

Dhote, V., Mishra, S., Parajapati, G. S., & Shukla, J. P., 2015. Challenges in Big Data Application: A Review. *International Journal of Computer Applications*, *121*(19), pp. 42–46. Available at: http://doi.org/10.5120/21651-4962

Di Sano, M., 2014. Business intelligence as a service: A new approach to manage business processes in the cloud. In *Proceedings of the Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, WETICE* (pp. 155–160). Parma. Available at: http://doi.org/10.1109/WETICE.2014.95

Di Tria, F., Lefons, E., & Tangorra, F., 2017. Cost-benefit analysis of data warehouse design methodologies. *Information Systems*, *63*, pp. 47–62. Available at: http://doi.org/10.1016/j.is.2016.06.006

Didau, D., 2015. *What If Everything You Knew about Education Was Wrong?* Carmarthen, UK: Crown House Publishing.

DiScala, M., & Abadi, D. J., 2016. Automatic Generation of Normalized Relational Schemas from Nested Key-Value Data. In *Proceedings of the 2016 International Conference on Management of Data - SIGMOD '16* (pp. 295–310). Available at: http://doi.org/10.1145/2882903.2882924

Dokeroglu, T., Sert, S. A., & Cinar, M. S., 2014. Evolutionary multiobjective query workload optimization of cloud data warehouses. *Scientific World Journal*, *2014*(14), pp. 1–16. Available at: http://doi.org/10.1155/2014/435254

Doll, W. J., & Torkzadeh, G., 1991. The Measurement of End-User Computing Satisfaction: Theoretical and Methodological Issues. *MIS Quarterly*, *15*(1), p. 5. Available at: http://doi.org/10.2307/249429

El-Sappagh, S. H. A., Hendawi, A. M. A., & El Bastawissy, A. H., 2011. A proposed model for data warehouse ETL processes. *Journal of King Saud University - Computer and Information Sciences*, *23*(2), pp. 91–104. Available at: http://doi.org/10.1016/j.jksuci.2011.05.005

Engel, R. J., & Schutt, R. K., 2014. *Fundamentals of Social Work Research* (2nd ed.). Thousand Oaks, CA, US: SAGE Publications, Inc.

European Union, 2015. Topics of European Union - Multilingualism. Retrieved February 2, 2016, from http://europa.eu/pol/mult/index_en.htm

Fairweather, P., 2003. The World Wide Web: another English-speaking colony? In *International Conference on Information Technology: Research and Education* (pp. 517–520).

Fegaras, L., & Maier, D., 1995. An Algebraic Framework for Physical OODB Design. In *Dbpl* (pp. 1–19). Gubbio, Italy.

Fink, L., Yogev, N., & Even, A., 2017. Business intelligence and organizational learning: An empirical investigation of value creation processes. *Information & Management*, *54*(1), pp. 38–56. Available at: http://doi.org/10.1016/j.im.2016.03.009

Firican, G., 2017. Best Practices for Powerful Dashboards. *Business Intelligence Journal*, *22*(2), pp. 33–39.

Forbes, 2013. The World's Biggest Public Companies. Retrieved November 24, 2015, from http://www.forbes.com/global2000/list/#tab:overall

Garani, G., & Helmer, S., 2012. Integrating Star and Snowflake Schemas in Data Warehouses . *International Journal of Data Warehousing and Mining*, *8*(4), pp. 22–40. Available at: http://doi.org/10.4018/jdwm.2012100102

Gartner, C. P., 2011. *More Than 2,000 CIOs Identifies Cloud Computing as Top Technology Priority for CIOs in 2011*. Stamford, Connecticut, US: Gartner Research. Retrieved from http://www.gartner.com/newsroom/id/1526414

Gatian, A., 1994. Is user satisfaction a valid measure of systems effectiveness? *Information & Management*, *26*(3), pp. 119–131. Retrieved from http://dx.doi.org/10.1016/0378-7206(94)90036-1

Gluchowski, P., & Kemper, H.-G., 2006. Quo Vadis Business Intelligence? Aktuelle Konzepte und Entwicklungstrends. *BI-Spektrum*, *1*(1), pp. 12–19.

Golfarelli, M., Rizzi, S. & Cella, I., 2004. Beyond Data Warehousing: What ' s Next in Business Intelligence ? In *DOLAP '04 Proceedings of the 7th ACM international workshop on Data warehousing and OLAP*. New York: ACM Press, pp. 1–6.

Gouvernement de la République Française, 1958. Constitution of France, Article 2.1.

Government of Federation of Bosnia and Herzegovina, 1995. Federation Constitution, Article 6.1.

Gracia, J., Montiel-Ponsoda, E., Cimiano, P., Gómez-Pérez, A., Buitelaar, P., & McCrae, J., 2012. Challenges for the multilingual Web of Data. *Journal of Web Semantics*, *11*, pp. 63–71. Available at: http://doi.org/10.1016/j.websem.2011.09.001

Graefe, G., Nica, A., Stolze, K., Neumann, T., Eavis, T., & Petrov, I., 2013. Elasticity in cloud databases and their query processing. *International Journal of Data Warehousing and Mining*, *9*(2), 1. Available at: http://doi.org/10.4018/jdwm.2013040101

Grin, F., 1998. Language Policy in Multilingual Switzerland: Overview and Recent Developments. *Cicle de Confèrencies Sobre Política Lingüística: Direcció General de Política Lingüística*. Barelona.

Guarda, T., Santos, M. F., Augusto, M. F., Silva, C., & Pinto, F., 2013. Process Mining: A framework proposal for Pervasive Business Intelligence. In *Information Systems and Technologies (CISTI), 2013 8th Iberian Conference on* (pp. 1–4). Lisaboa: IEEE.

Guest, G., 2006. How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*, *18*(1), pp. 59–82. Available at: http://doi.org/10.1177/1525822X05279903

Hainaut, J.-L., Hick, J.-M., Henrard, J., Roland, D., & Englebert, V., 1997. Knowledge transfer in database reverse engineering: a supporting\ncase study. In *Proceedings of the*

Fourth Working Conference on Reverse Engineering (pp. 194–203). Available at: http://doi.org/10.1109/WCRE.1997.624590

Hair, J. F. J., Celsi, M. W., Money, A. H., Samouel, P., & Page, M. J., 2011. *Essentials of business research methods*. New York, USA: Routledge.

Hannula, M., & Pirttimäki, V., 2003. Business Intelligence Empirical Study on the Top 50 Finnish Companies. *Journal of American Academy of Business*. Retrieved from http://www.umsl.edu/~lacitym/bi1.pdf

Harzing, A. W., Köster, K., & Magner, U., 2011. Babel in business: The language barrier and its solutions in the HQ-subsidiary relationship. *Journal of World Business*, *46*(3), pp. 279–287. Available at: http://doi.org/10.1016/j.jwb.2010.07.005

Hau, E., & Aparício, M., 2008. Software internationalization and localization in web based ERP. In *Proceedings of the 26th annual ACM international conference on Design of communication - SIGDOC '08* (p. 175). New York: ACM. Available at: http://doi.org/10.1145/1456536.1456570

Hensch, K., 2005. IBM history of far eastern languages in computing, part 1: Requirements and initial phonetic product solutions in the 1960s. *IEEE Annals of the History of Computing*, *27*(1), pp. 17–26. Available at: http://doi.org/10.1109/MAHC.2005.9

Henze, N., Shrazi, A. S., Schmidt, A., Pielot, M., & Michahelles, F., 2013. Empirical research through ubiquitous data collection. *Computer*, *46*(6), pp. 74–76. Available at: http://doi.org/10.1109/MC.2013.202

Herschel, R. T., Clements, N., 2017. The Importance of Storytelling in Business Intelligence. *International Journal of Business Intelligence Research, 8*(1), pp. 26-39.

Hillier, M., 2003. The role of cultural context in multilingual website usability. *Electronic Commerce Research and Applications*, *2*(1), pp. 2–14. Available at: http://doi.org/10.1016/S1567-4223(03)00005-X

Hooda, R., & Gill, N. S., 2012. Evolution of Data Warehousing and Data Mining. *International Journal of Management, IT and Engineering*, *2*(5), pp. 96–104.

Hossain, S., Islam, F., Karim, R., & Siddique, K. N., 2014. A Critical Comparison Between Distributed Database Approach and Data Warehousing Approach. *Paper*, *5*(1).

Hou, C.-K., 2012. Examining the effect of user satisfaction on system usage and individual performance with business intelligence systems: An empirical study of Taiwan's electronics industry. *International Journal of Information Management*, 32(6), pp. 560–573. Available at: http://doi.org/10.1016/j.ijinfomgt.2012.03.001

Huang, S., & Tilley, S., 2001. Issues of content and structure for a multilingual web site. In *Proceedings of the 19th annual international conference on computer documentation* (pp. 103–110). ACM. Available at: http://doi.org/10.1145/501516.501537

Huff, A., 201. Big Data II: Business Intelligence - Fleets Analyze Information from several sources to improve overall performance. *Commercial Carrier Journal*, (April), 53.

Imhoff, C., Galemmo, N., & Geiger, J. G., 2003. *Mastering Data Warehouse Design: Relational and Dimensional Techniques*. *Wiley Publishing, Inc.* Indianapolis: Wiley Publishing, Inc.

Inmon, B.W., 1995. *Building the Data Warehouse*, New York: Wiley.

Inmon, B.W., 2005. *Building the Data Warehouse* (4th ed.), Indianapolis: John Wiley & Sons.

Işik, Ö., Jones, M. C., & Sidorova, A., 2013. Business intelligence success: The roles of BI capabilities and decision environments. *Information and Management*, *50*(1), pp. 13–23. Available at: http://doi.org/10.1016/j.im.2012.12.001

Jain, A., Garg, S., & Sharma, N., 2015. The Management of Conformed ETL Architecture, *118*(10), pp. 20–22.

Jamaludin, I. A., & Mansor, Z., 2011. Review on Business Intelligence “BI” Success Determinants in Project Implementation. *International Journal of Computer Applications*, *33*(8), p. 24.

Jensen, C. S., Pedersen, T. B., & Thomsen, C., 2010. *Multidimensional Databases and Data Warehousing*. *Synthesis Lectures on Data Management* (1st ed., Vol. 2). Morgan & Claypool Publishers. Available at: http://doi.org/10.2200/S00299ED1V01Y201009DTM009

Jourdan, Z., Rainer, R. K., & Marshall, T. E., 2008. Business Intelligence: An Analysis of the Literature [1]. *Information Systems Management*, *25*(2), pp. 121–131. Available at: http://doi.org/10.1080/10580530801941512

Jovanović, V., Subotić, D., & Mrdalj, S., 2014. Data modeling styles in data warehousing. In *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2014 - Proceedings* (pp. 1458–1463). Opatija, Croatia: MIPRO. Available at: http://doi.org/10.1109/MIPRO.2014.6859796

Kappelman, L., Mclean, E., Vess, J., & Gerhart, N., 2014. The 2014 SIM IT Key Issues and Trends Study. *MIS Quarterly Executive*, *2014* (December), pp. 237–263.

Khouri, S., Bellatreche, L., Boukhari, I., & Bouarar, S., 2012. More investment in conceptual designers: Think about it! In *Proceedings - 15th IEEE International Conference on Computational Science and Engineering, CSE 2012 and 10th IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, EUC 2012* (pp. 88–93). Available at: http://doi.org/10.1109/ICCSE.2012.22

Khouri, S., Berkani, N., & Bellatreche, L., 2017. Tracing data warehouse design lifecycle semantically. *Computer Standards and Interfaces, 51*, pp. 132–151. Available at: http://doi.org/10.1016/j.csi.2016.12.002

Kianoff, L., 2010. Dashboards: Business Intelligence At A Glance. *CPA Technology Advisor (Ft. Atkinson, Wis. 2004), 20*(7), p. 17.

Kimball, R., 1998. *The Data Warehouse Lifecycle Toolkit* (2nd ed.). Indianapolis: John Wiley & Sons.

Kimball, R., 2001. Design Tip #24: Designing Dimensional In A Multinational Data Warehouse. Retrieved December 1, 2015, from http://www.kimballgroup.com/2001/06/design-tip-24-designing-dimensional-in-a-multinational-data-warehouse/

Kimball, R., & Ross, M., 2011. *The data warehouse toolkit: the complete guide to dimensional modelling* (2nd ed.). John Wiley & Sons. Available at: http://doi.org/10.1145/945721.945741

Kimball, R., & Ross, M., 2013. *The data warehouse toolkit, third edition: the definitive guide to dimensional modeling* (3rd ed.). John Wiley & Sons.

Kimball, R., Ross, M., Thorthwaite, W., Becker, B., & Mundy, J., 2008. *The Data Warehouse Lifecycle Toolkit*. Wiley India Pvt. Limited.

Köhler, H., & Link, S., 2016. SQL Schema Design: Foundations, Normal Forms, and Normalization. In *SIGMOD '16 Proceedings of the 2016 International Conference on Management of Data* (pp. 267–279). San Francisco, USA.

Krzysztof, C., Witold, P., Swiniarski, R., & Kurgan, L., 2007. *A Knowledge Discovery Approach* (1st ed.). New York, USA: Springer Science+Business Media, LLC.

Kumaran, A., Chowdary, P. K., & Haritsa, J. R., 2006. On pushing multilingual query operators into relational engines. In *Proceedings - International Conference on Data Engineering* (Vol. 2006, p. 98). IEEE. Available at: http://doi.org/10.1109/ICDE.2006.107

Kumhyr, D., 2001. Method and system for generating and serving multilingual web pages. United States.

Kurbel, K. E., 2013. *Enterprise Resource Planning and Supply Chain Management*. Springer Science & Business Media. Available at: http://doi.org/10.1007/978-3-642-31573-2

Kurniawan, Y., Gunawan, A., & Gregorius Kurnia, S., 2014. Application of business intelligence to support marketing strategies: A case study approach. *Journal of Theoretical and Applied Information Technology*, *64*(1), pp. 240–248.

Laha, A., 2008. Rap. In *Proceedings of the 1st Bangalore annual Compute conference on - Compute '08* (p. 1). New York, NY, USA: ACM. Available at: http://doi.org/10.1145/1341771.1341796

Larson, D., & Chang, V., 2016. A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management*, *36*(5), pp. 700–710. Available at: http://doi.org/10.1016/j.ijinfomgt.2016.04.013

Law No. 482., 1999. Norme in materia di tutela delle minoranze linguistiche in Italia. *Gazzetta Ufficiale*, *297*(Article 1.1). Retrieved from http://www.camera.it/parlam/leggi/99482l.htm

Lepouras, G., & Vassilakis, C., 2000. Multilingual Web Site Construction and Maintenance. In *Proceedings of the SCI 2000 conference* (pp. 56–61). Orlando, Florida.

Likert, R., 1932. A Technique for the Measurement of Attitudes. *Archives of Psychology*, *22 140*(140), p. 55. Available at: http://doi.org/2731047

Linstedt, D., Graziano, K., & Hultgren, H., 2010. *The New Business Supermodel, The Business of Data Vault modeling*. *Lulu. com* (2nd ed.). Lulu.com.

Lokhande, S., & Khare, N., 2015. An Outlook on Big Data and Big Data Analytics. *International Journal of Computer Applications*, *124*(11), pp. 37–41. Available at: http://doi.org/10.5120/ijca2015905658

Lönnqvist, A., & Pirttimäki, V., 2006. The Measurement of Business Intelligence. *Information Systems Management*. Available at: http://doi.org/10.1201/1078.10580530/45769.23.1.20061201/91770.4

Lord-Castillo, B. K., Mate, B. R., Wright, D. J., & Follett, T., 2009. A customization of the arc marine data model to support whale tracking via satellite telemetry. In

Transactions in GIS (Vol. 13, pp. 63–83). Available at: http://doi.org/10.1111/j.1467-9671.2009.01159.x

Luftman, J. , Zadeh, H. S., Derksen, B., Santana, M., Rigoni, E. H., & Huang, Z., 2012. Key information technology and management issues 2011-2012: An international study. *Journal of Information Technology*, *27*(3), pp. 198–212. Available at: http://doi.org/10.1057/jit.2013.22

Luftman, J., & Ben-Zvi, T., 2010. Key Issues for IT executives 2009: Difficult Economy's Impact on IT. *MIS Quarterly Executive*, *9*(1), pp. 203–213.

Luhn, H. P., 1958. A Business Intelligence System. *IBM Journal of Research and Development*, *2*(4), pp. 314–319. Available at: http://doi.org/10.1147/rd.24.0314

Lymer, A., Debreceny, R., Gray, G. L., & Rahman, A., 1999. *Business Reporting On The Internet*. *Business*. London, UK: International Accounting Standards Committe. Retrieved from http://www.iasc.org.uk

Marchand, M., & Raymond, L., 2008. Researching performance measurement systems An information systems perspective. *International Journal of Operations & Production Management*, *28*(7), pp. 663–686. Available at: http://doi.org/10.1108/01443570810881802

Maxwell, C. J., 2008. Building and operating a web-based reporting system: A case study. *New Directions for Institutional Research*, *2008*(137), pp. 41–56. Available at: http://doi.org/10.1002/ir.237

Menzel, K., Scherer, R., Schapke, S., & Eisenblätter, K., 2002. Potentials of Data Warehouse Technology to Support Case Sensitive Information Representation on Mobile Devices. In *Concurrent Engineering – Research and Applications: Proceedings of the 9th ISPE International Conference on Concurrent Engineering, Cranfield, U.K.*

Miller, D., 2012. Expert voices. In S.E. Baker & R. Edwards. *How many qualitative interviews is enough.* National Centre for Research Methods Review Discussion Paper., pp 31).

Miller, G. J., Brautigam, D., & Gerlach, S. V., 2006. *Business Intelligence Competency Centers: a team approach to maximizing competitive advantage* (1st ed., Vol. 2). Hoboken, US: John Wiley & Sons. Available at: http://doi.org/10.4018/jbir.2011070102

Morgan, T., Luttrell, C., & Liu, Y., 2001. Designing Multilingual Web Sites: Applied Authoring Techniques. In *SIGDOC '01 Proceedings of the 19th annual international conference on Computer documentation* (pp. 230–231). Santa Fe, US.

Moss, L. T., & Shaku, A., 2003. *Business Intellgence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*. Addison-Wesley Professional.

Mühl, J. K., 2014. *Organizational trust: Measurement, impact, and the role of management accountants. Organizational trust: Measurement, impact, and the role of management accountants.* Edinburgh, UK: Springer.

Muñoz, L., Mazón, J. N., & Trujillo, J., 2011. ETL Process Modeling Conceptual for Data Warehouses: A Systematic Mapping Study. *IEEE Latin America Transactions, 9*(3), pp. 360–365. Available at: http://doi.org/10.1109/TLA.2011.5893784

Muriithi, G. M., & Kotzé, J. E., 2013. A conceptual framework for delivering cost effective business intelligence solutions as a service. In *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference on - SAICSIT '13* (p. 96). East London, South Africa: AC. Available at: http://doi.org/10.1145/2513456.2513502

Mykityshyn, M. G., 2007. *Assessing the Maturity of Information Architectures for Complex Dynamic Enterprise Systems*. Georgia Institute of Technology.

Nebot, V., & Berlanga, R., 2016. Statistically-driven generation of multidimensional analytical schemas from linked data. *Knowledge-Based Systems, 110*, pp. 15–29. Available at: http://doi.org/10.1016/j.knosys.2016.07.010

Niglas, K., 2010. The multidimensional model of research methodology. In A. Tashakkori & C. Teddlie (Eds.), *SAGE Handbook of Mixed Methods in Social & Behavioral Research* (2nd ed., pp. 215–36). Thousand Oaks, CA, US: SAGE.

O'Neill, D., 2011. Business Intelligence Competency Centers. *International Journal of Business Intelligence Research*, *2*(3), pp. 21–35. Available at: http://doi.org/10.4018/jbir.2011070102

Obeidat, M., North, M., Richardson, R., Rattanak, V., & North, S., 2015. Business Intelligence Technology, Applications, and Trends. *International Management Review*, *11*(2), pp. 47–56.

Odysseas G. Tsatalos, Marvin H. Solomon, Y. E. I., 1996. The GMAP: a versatile tool for physical data independence. *Vldbj*, *5*(3), pp. 101–118. Available at: http://doi.org/10.1007/s007780050018

Olaru, C., 2014. Business Intelligence in Telecommunications Industry. *International Journal of Economic Practices and Theories*, *4*(1), pp. 89–100.

Olszak, C. M., 2014. Towards an Understanding Business Intelligence. A Dynamic Capability-Based Framework for Business Intelligence. In *Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on* (pp. 1103–1110). IEEE. Available at: http://doi.org/10.15439/2014F68

Olszak, C. M., & Ziemba, E., 2006. Business Intelligence Systems in the holistic infrastructure development supporting decision-making in organisations. *Interdisciplinary Journal of Information, Knowledge, and Management*, *1*, pp. 47–58.

Onwuegbuzie, A. J., & Johnson, R. B., 2004. Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, *33*(7), pp. 14–26. Available at: http://doi.org/10.3102/0013189X033007014

Oracle, 2012. The Business Intelligence Competency Center : Enabling Continuous Improvement in Performance Management. Oracle Corporation.

Orlikowski, W. J., & Baroudi, J. J., 1991. Studying information technology in organizations: Research approaches and assumptions. *Information Systems Research*, *2*(1), pp. 1–28. Available at: http://doi.org/10.1287/isre.2.1.1

Orlov, V., 2014. Data Warehouse Architecture: Inmon CIF, Kimball Dimensional or Linstedt Data Vault? Retrieved February 20, 2015, from http://blog.westmonroepartners.com/data-warehouse-architecture-inmon-cif-kimball-dimensional-or-linstedt-data-vault/

Paltridge, B. & Phakiti, A., 2015. *Research Methods in Applied Linguistics: A Practical Resource* (2nd ed.), London, UK: Bloomsbury Academic.

Perry, M. P., 2016. *Business Driven Project Portfolio Management : Conquering the Top 10 Risks That Threaten Success*. Fort Lauderdale, FL, US: J. Ross Publishing.

Peters, T., Işik, Ö., Tona, O., & Popovič, A., 2016. How system quality influences mobile BI use: The mediating role of engagement. *International Journal of Information Management, 36*(5), pp. 773–783. Available at: http://doi.org/10.1016/j.ijinfomgt.2016.05.003

Petter, S., DeLone, W., & McLean, E., 2013. Information Systems Success: The Quest for the Independent Variables. *Journal of Management Information Systems, 29*(4), pp. 7–61. Available at: http://doi.org/10.2753/MIS0742-1222290401

Pierini, F., & English., 2016. Multilingualism and Englishization in international business communication: the case of some Italian wine producers. *English Today*, *32*(1), pp. 48–55. Available at: http://doi.org/10.1017/

Ponniah, P., 2001. *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals*. *Data Warehousing Fundamentals - A Comprehensive Guide for IT Professionals* (Vol. 6). New Yor: John Wiley & Sons.

Poolet, M. A., 2008. Data Warehousing: Rapidly Changing Monster Dimensions. Retrieved November 24, 2015, from http://sqlmag.com/database-administration/data-warehousing-rapidly-changing-monster-dimensions

Popovič, A., Turk, T., & Jaklič, J., 2010. Conceptual Model of Business Value of Business Intelligence Systems. *Management: Journal of Contemporary Management*, *15*(1), pp. 5–29.

Porter, J., & Rome, J., 1995. Lessons from a Successful Data Warehouse Implementation. *Cause/Effect*, *18*(4), pp. 43–50.

Power, D. J., 2002. *No Title Decision Support Systems: Concepts and Resources for Managers*. Westport: Greenwood Publishing Group.

Purba, S., 1999. *Handbook of Data Management* (3rd ed.). Boca Raton, FL, US: Auerbach, CRC Press.

Rahman, N., 2013. Measuring Performance for Data Warehouses-A Balanced Scorecard Approach. *International Journal of Computer and Information …*, *4*(1). Retrieved from http://ijcit.org/ijcit_papers/vol4no1/IJCIT-130701.pdf

Rajan, M. N., & Makani, B., 2016. Methodological Considerations in Cross-Cultural Research: A Discussion of the Translation Issue, *28*(2), 2016.

Rajesh, R., 2010. *Supply Chain Management for Retailing*, Kalkota: Tata McGraw-Hill Education.

Raniner, R. K., & Cegielski, C. G., 2010. *Introduction to Information Systems: Enabling and Transforming Business* (3rd ed.). John Wiley & Sons.

Ranjan, J., 2009. Business Intelligence: Concepts, Components, Techniques and Benefits. *Journal of Theoretical and Applied Information Technology*, 9(1), pp. 60 – 70.

Richards, G., Yeoh, W., Chong, A. Y.-L., & Popovič, A., 2014. An empirical study of business intelligence impact on corporate performance management. In *PACIS 2014 Proceedings* (p. 341).

Romney, A. K., Weller, S. C., & Batchelder, W. H., 1986. Culture as Consensus: A Theory of Culture and Informant Accuracy. *American Anthropologist*, *88*(2), pp. 313–338. Available at: http://doi.org/10.1525/aa.1986.88.2.02a00020

Ruffle, P., 2001. Managing multilingual Web sites. *The Seybold Report*, *1*(13), p. 12.

Runeson, P., Rainer, A., Höst, M., & Regnell, B., 2012. *Case Study Research in Software Engineering: Guidelines and Examples* (1 edition). Haboken, New Jersey, US: Wiley-Blackwell.

Rylander, D. H., & Provost, T., 2006. Improving the Odds: Combining Six Sigma and Online Market Research for Better Customer Service. *SAM Advanced Management Journal*, *71*(1), p. 13.

Sabherwal, R., & Becerra-Fernandez, I., 2011. *Business intelligence: Practices, technologies, and management* (1st ed.). John Wiley & Sons.

Sabherwal, R., & Chowa, C., 2006. SCIENCE Information Success : Individual Determinants Organizational System. *Management*, *52*(12), pp. 1849–1864. Available at: http://doi.org/10.1287/mnsc.l060.0583

Sandu, D. I., 2010. Operational and real-time Business Intelligence. *Informatica Economic*, *3*(3), pp. 33–36.

SAP AG., 2015. SAP Library - XML: BW - Data Warehousing - Modeling. Retrieved March 2, 2015, from http://help.sap.com/saphelp_snc70/helpdata/en/4c/89dc37c7f2d67ae10000009b38f889/frameset.htm

Saporito, P. L., 2014. *Applied Insurance Analytics: A Framework for Driving More Value from Data Assets, Technologies, and Tools*. Upper Saddle River, US: FT Press.

Sauders, M. Lewis, P & Thornhill, A., 2009. *Research Methods for business Students* (6th ed.). Harlow, UK: Pearson Education Ltd. .

Saunders, M., & Tosey, P., 2013. The_Layers_of_Research_Design. *Rapport*, (Winter 2012/2013), pp. 58–59.

Schutt, K. R., 2012. *Investigating the Social World: The Process and Practice of Research*. Thousand Oaks, CA, US: SAGE Publications, Inc.

Scrapehero.com, 2015. How many products does Walmart.com sell in comparison to Amazon.com? Retrieved November 24, 2015, from http://learn.scrapehero.com/how-many-products-does-walmart-com-sell-vs-amazon-com/

Sedera, D., & Tan, F. T. C., 2005. User Satisfaction: An Overarching Measure of Enterprise System Success. *PACIS 2005 Proceedings*, *2*, pp. 963–976.

Sell, D., da Silva, D. C., Beppler, F. D., Napoli, M., Ghisi, F. B., Pacheco, R. C. S., & Todesco, J. L., 2008. SBI: A Semantic Framework to Support Business Intelligence. In *Proceedings of the first international workshop on Ontology-supported business intelligence - OBI '08* (pp. 1–11). Karlsruhe, Germany: ACM. Available at: http://doi.org/10.1145/1452567.1452578

Sen, A., & Sinha, A. P., 2005. A Comparison of Data Warehousing Methodologies. *Communications of the ACM*, *48*(3), pp. 79–84. Available at: http://doi.org/10.1145/1047671.1047673

Singh, S. kumar., 2011. *Database Systems: Concepts, Design and Applications* (2nd ed.). New Delhi: Dorling Kindersley (India) Pvt. Ltd.

Sommerville, I., & Sawyer, P., 1997. *No Title* (1st ed.). New York: John Wiley & Sons, Inc.

Song, X., Yan, X., & Yang, L., 2009. Design ETL metamodel based on UML profile. *2009 2nd International Symposium on Knowledge Acquisition and Modeling, KAM 2009*, *3*, pp. 69–72. Available at: http://doi.org/10.1109/KAM.2009.112

Spanish Constitution, 1978. *The Spanish Constitution*, p. 132.

Srinivasan, S., & Kamalakannan, T., 2017. Multi Criteria Decision Making in Financial Risk Management with a Multi-objective Genetic Algorithm. *Computational Economics*, pp. 1–15. Available at: http://doi.org/10.1007/s10614-017-9683-7

Starr, J., 2005. Design considerations for multilingual web sites. *Information Technology and Libraries*, *24*(3), pp. 107–116. Available at: http://doi.org/10.6017/ital.v24i3.3370

Sun, H., 2001. Building A Culturally-Competent Corporate Web Site : An Exploratory Study of Cultural Markers In Multilingual Web Design. In *Language* (Vol. 10/01, pp. 95–102). Santa Fe, US: ACM. Available at: http://doi.org/10.1145/501516.501536

Talburt, J., & Zhou, Y., 2015. *Entity Information Life Cycle for Big Data: Master Data Management and Information Integration*. Morgan Kaufmann.

Thamir, A. & Poulis, E., 2015. Business Intelligence Capabilities and Implementation Strategies. *International Journal of Global Business*, 8(1), pp. 34–45.

The European Commission., 2008. *Final Report: Commission of the European Communities High Level Group on Multilingualism*. Luxembourg.

Tilling, B., 2003. European union local language requirements and how labelling systems may provide a solution. *Pharmaceutical Technology Europe*, *15*(5), pp. 58–61.

Toumi, L., Moussaoui, A., & Ugur, A., 2014. Particle swarm optimization for bitmap join indexes selection problem in data warehouses. *Journal of Supercomputing*, *68*(2), pp. 672–708. Available at: http://doi.org/10.1007/s11227-013-1058-9

Turban, E., Sharda, R., Delen, D., & King, D., 2010. *Business Intelligence : A Managerial Approach* (2nd ed.). Prentice Hall.

Vázquez, S. R., 2013. Localizing Accessibility of Text Alternatives for Visual Content in Multilingual Websites. In *ACM SIGACCESS Accessibility and Computing* (pp. 34–37). ACM. Available at: http://doi.org/10.1145/2444800.2444807

Wang, R. Y., & Strong, D. M., 1996. "Beyond Accuracy: What Data Quality Means to Data Consumers," *Source J. Manag. Inf. Syst.*, 12(4), pp. 5–33.

Warren, J., Benbow, H. M., 2008. *Multilingual Europe : Reflections on Language and Identity*. *Cambridge Scholars Publishing*. Newcastle upon Tyne, UK: Cambridge Scholars Publishing. Retrieved from http://site.ebrary.com/lib/uon/docDetail.action?docID=10655385

Watson, H. J., & Wixom, B. H., 2007. The Current State of Business Intelligence. Computer, 40(9), pp. 96–99.

Wixom, B., & Goul, M., 2014. The Current State of Business Intelligence in Academia : The Arrival of Big Data The Current State of Business Intelligence in Academia : The Arrival of Big Data. *Communications of the Association for Information Systems*, *34*(January), pp. 1–13. Available at: http://doi.org/10.1016/j.cub.2013.07.021

Wixom, B., Ariyachandra, T., Goul, M., Gray, P., Kulkarni, U., & Phillips-Wren, G., 2011. The current state of Business Intelligence in academia. *Communications of the Association for Information Systems*, *29*(1), pp. 299–312. Available at: http://doi.org/10.1109/MC.2007.331

Wu, X., Zhu, X., Wu, G. Q., & Ding, W., 2014. Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, *26*(1), pp. 97–107. Available at: http://doi.org/10.1109/TKDE.2013.109

Yeung, A. K. W., & Hall, G. B., 2007. *Spatial Database Systems: Design, Implementation and Project Management*. Dordrecht, The Nethelands: Springer Science & Business Media.

Yrjö-Koskinen, P., 1973. Johto – Tiedontarve – Tiedonhankinta. Miten informaatiopalvelu voi palvel la yrityksen johtoa? *Publications of Insinöörijärjestö, No. 77-73*.

APPENDIX A. Existing BI Frameworks

In this APPENDIX Dxisting BI frameworks are reviewed to identify which elements of BI system, that might be relevant in multilingual context, are included in the frameworks and which are excluded. The frameworks are categorised into three groups: *high level* approaches which provide an overview or conceptual level view of BI but do not consider implementation or data management details, *data oriented* approaches which typically use the concept of layers to describe the data journey from data source to presentation, and *business oriented* approaches which discuss BI from a business perspective, but without considering data management or data processing details.

High Level and Conceptual Level Approaches

Humphrey (1997) used the term high-level conceptual approach to define software environments described at a high level of abstraction. The BI frameworks and approaches evaluated in this group are focused mostly on defining functional abstractions of the BI environment while offering simplified representation of components. This section discusses four BI frameworks and DWH approaches proposed by Inmon and Kimball.

- **Watson & Wixom BI Framework (2007)**

This framework has only two major components/functions: “getting data in” and “getting data out” (Figure A-1). For this reason, this framework is classified as high level and process focused. It offers only superficial understanding of BI concepts, and identification of few applications (e.g. data warehouse and data marts) and types of data (Metadata). The strength of the framework is the capability to explain the function of BI in a readily comprehensible and non-technical way to different categories of users. However, the limitation of this framework is that it provides only an abstracted, high level view of input and outputs, and does not provide information about other aspects or components of a BI system. Information about further applications or types of data relevant in BI, hardware or user groups is not provided. In a scenario where it is proposed to develop a BI system or to extend or modify an existing BI environment to provide new capabilities such as Multilingualism, the framework does not support the identification of all relevant elements of the BI environment.

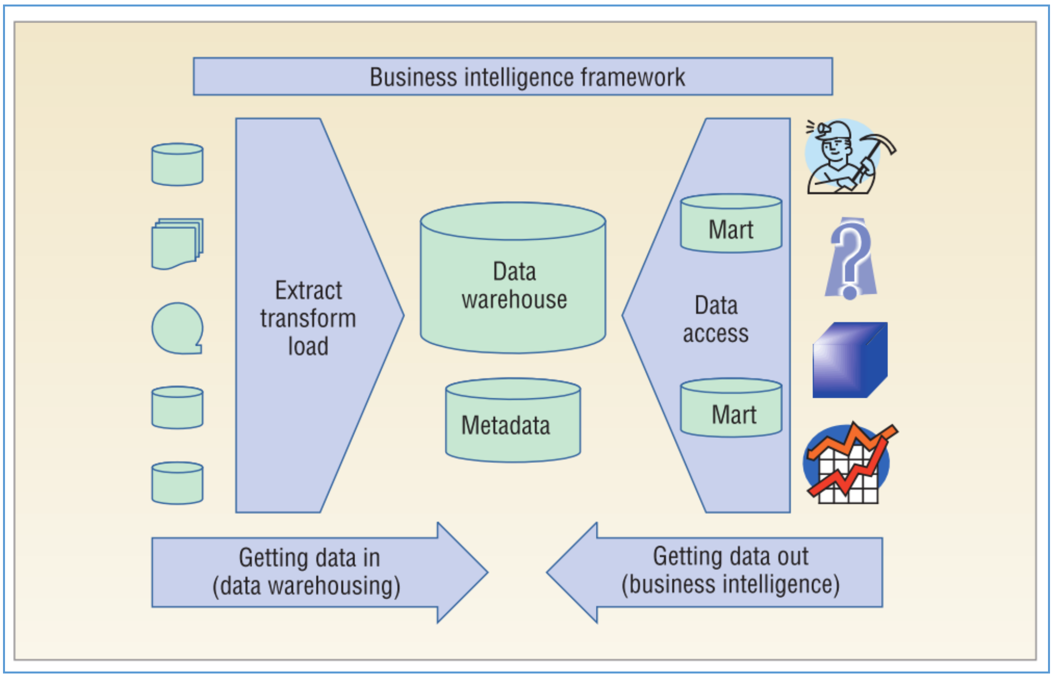


Figure A-1: Watson & Wixom BI Framework (Source: Watson & Wixom, 2007, p. 97)

- **RAP: A Conceptual Business Intelligence Framework**

Reference-Activity-Projection (RAP), a conceptual BI framework, was developed by Laha (2008). This framework has three layers as shown in Figure A-2: archived data and information elements belong to the *Reference* layer mostly covering activities in source systems; computational and processing activities relate to DW system and are building blocks of the *Activity* layer; an overall view of the future business conditions, comprising estimated values of various Key Performance Indicators (KPI) along with their interrelationship is represented in the *Projection* layer, which can be understood as the presentation level. According to Laha (2008), the strength of this framework is support for decision-making processes based on organisational experience and accessed through systematically organised mechanisms. Laha (2008) himself stresses that this is a conceptual framework and not representational, while divisions identified in the RAP framework are not intended to be translated to physical or logical DW design. The framework has a very limited discussion of types of data and does not include any application, hardware or user group elements. Some of the elements in the framework, particularly in the reference section, are outside the BI scope of this research. Evaluated against the requirement to support the development or extension of a BI environment, the framework does not support identification of any applications, hardware or user groups and it is difficult to identify types of data.

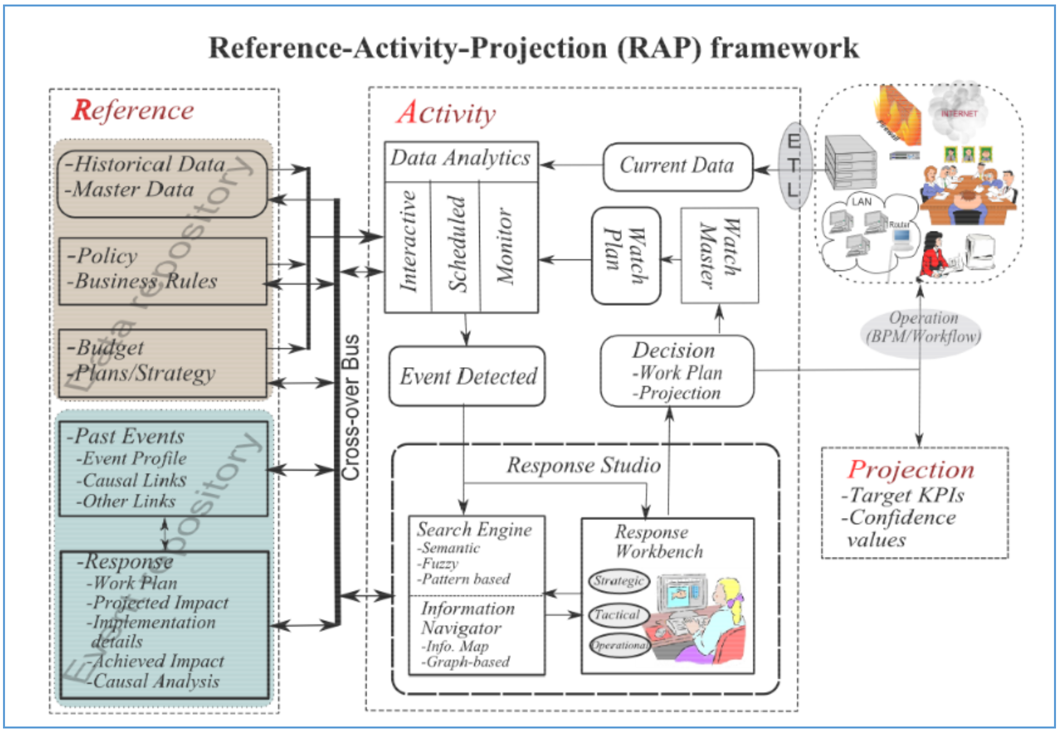


Figure A-2: RAP BI Framework (Source: Laha, 2008, p. 2)

- **SBI: A Semantic Framework to Support Business Intelligence**

In the same year as Laha, Sell *et al.* (2008) presented a semantic framework to support Business Intelligence (SBI) and to enable developers to customize BI solutions according to business needs – Figure A-3. According to Sell *et al.* (2008), the SBI framework develops ontologies from the description of business rules and concepts in order to support semantic-analytical functionalities that extend traditional OLAP operations. The focus of the framework is on presentation. It is interested in how semantic inference is supported by using batch and on-the-fly based strategies, and how such semantic infrastructure makes access to heterogeneous data sources transparent. The approach proposed by Sell *et al.* (2008) refers to the typical three-layered BI architecture that contains DW, an ETL tool, and an analytical tool. The strengths of this approach are its flexibility and the possibility of integrating heterogeneous data sources, analytical tools and business semantics for the purpose of more optimal decision making (Sell *et al.*, 2008). In addition, the benefits of the framework are illustrated exclusively through Extracta software, which makes it difficult to assess the generalisability of the approach. In the context of support for the development of a new BI environment or extending or modifying an existing BI environment to enable additional features or functionality, the identification of relevant aspects and components would not be supported by SBI. This framework does not include implementation and user issues and does not cover elements such as applications (software), hardware, types of data and user requirements.

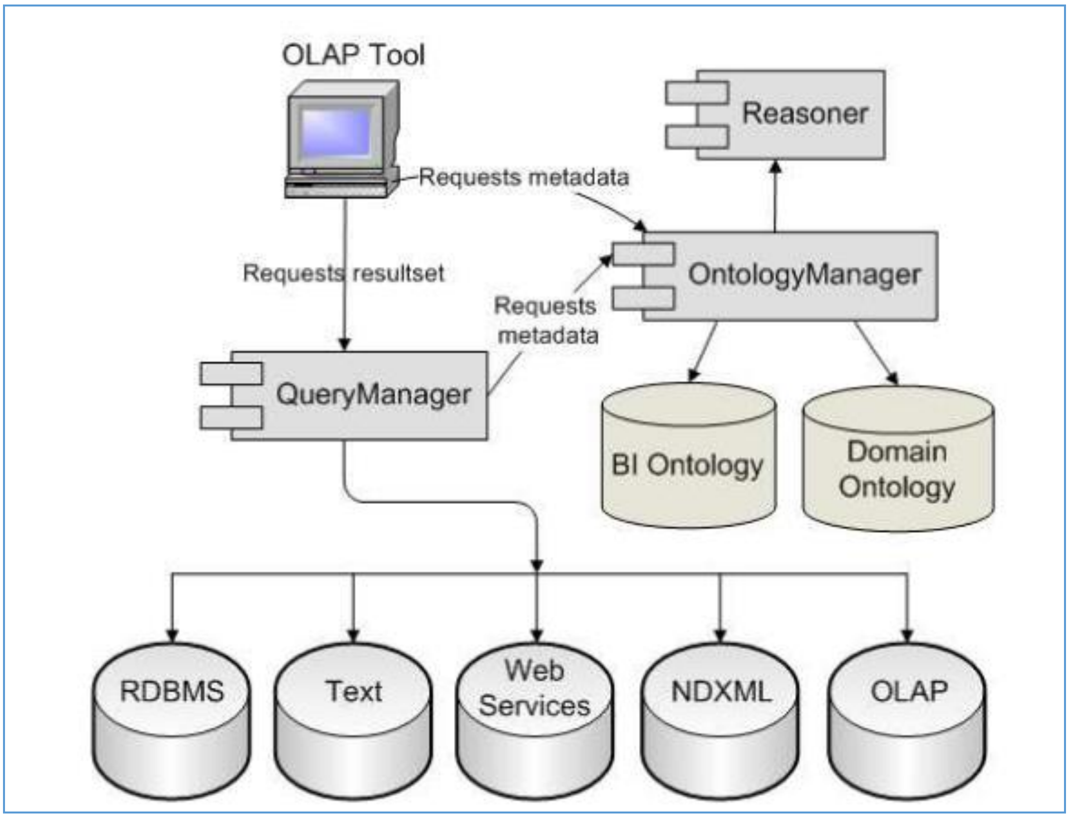


Figure A-3: Illustration of SBI components (Source: Sell et al, 2008, p. 3.)

- **A Conceptual Framework for Delivering Cost Effective BI Solutions as a Service**

Muriithi & Kotzé (2013) proposed a conceptual framework primarily intended to support the adoption of cloud-based BI (Figure A-4). The strength of this framework is the focus on leveraging transactional data through cloud solutions, thus enabling smaller companies suffering from resource constraints, to get an insight into how to use BI. The framework offers an additional perspective, presenting BI as a service over the Internet, which could, because of its lower costs, lead to faster acceleration and adoption of BI in the company. The focus is on enabling outsourcing of some part of BI into cloud solutions through componentising BI. In the context of this discussion, the strength of the Muriithi & Kotzé (2013) framework is also the biggest limitation. Since the framework focuses on Cloud BI, it cannot be considered as a holistic or generic solution. The framework lacks information about BI applications, hardware, types of data, user groups, possible layers and their concepts, thus, in the given scenario of developing a new BI environment or extending or modifying existing BI environments it would not be sufficient.

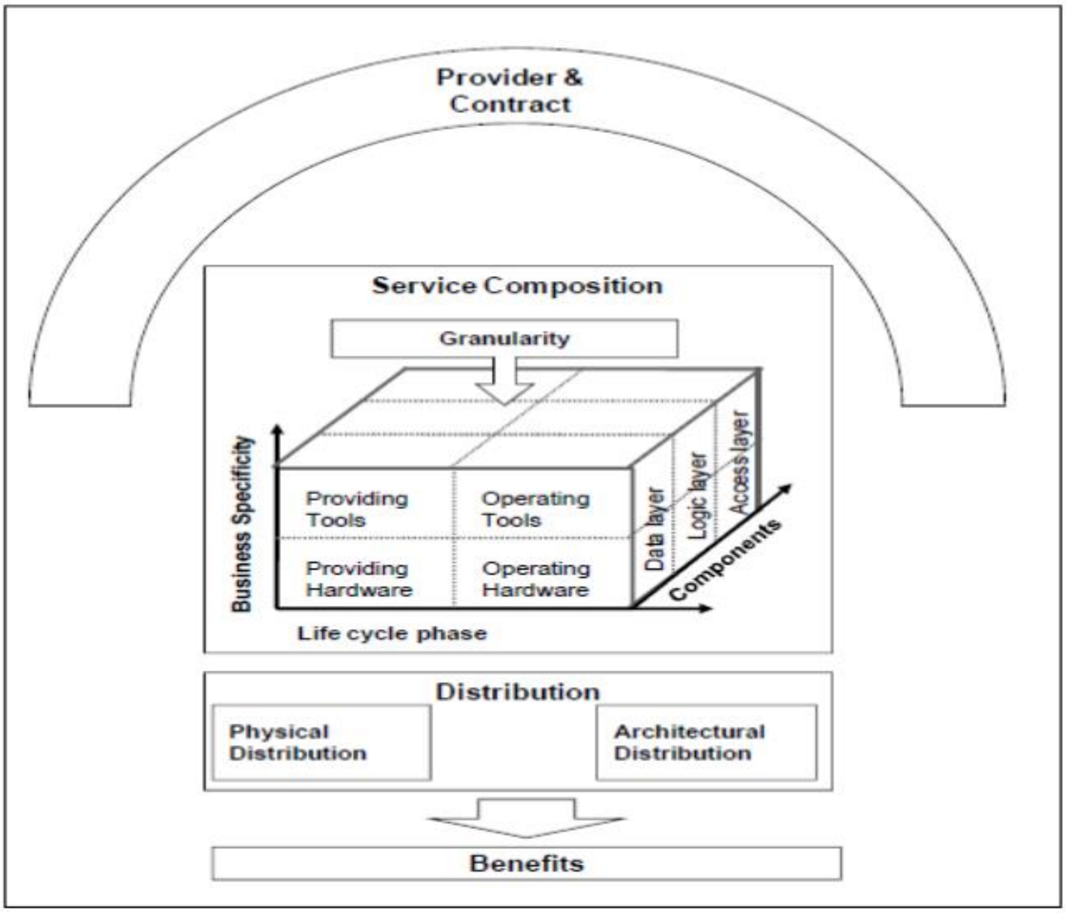


Figure A-4: Figure Cloud BI Framework (Source: Muriithi & Kotzé, 2013, p. 97)

- **Inmon's approach: A BI framework for enterprise data**

A seminal work by Inmon (2005) introduces the Corporate Information Factory (CIF) which is a top-down approach to the implementation of a DW and adopts a holistic view of enterprise data. Breslin (2004) defines Inmon's philosophy as evolutionary where a warehouse is an integral part of the CIF. In this case, DW, reporting applications (such as reports, queries or dashboards), data marts and operational database are the building parts of a "larger block". Inmon does not explicitly define this "larger block" as BI; however, it is a holistic view of processes and applications in the BI environment.

The simplified view of Inmon's CIF framework supports the identification of components that could be relevant to enable the development, modification and extension of a BI environment. As shown in Figure 2-3, the CIF approach supports the concept of a three-layered BI framework: (i) data sources layer, (ii) DWH layer that contains the staging area, DW itself, data marts holding information for reporting and (iii) reporting and querying layer.

Inmon (2005) proposed data marts to hold information directly used by the "reporting and querying" component. This component could, for example, help to identify the necessary components required by multilingualism for reporting data marts and querying. Inmon's approach is not domain specific and can therefore be considered as a generic solution. The enterprise-wide application element is one of the biggest strengths of this approach as it covers most relevant aspects of the BI environment, namely data sources, DW and data marts, and the presentation aspect. The other strengths include supporting easy understanding of components and the overall view of the BI environment.

The discussion by Inmon (2005) covers a large number of issues including environment, design, granularity, technology, internationalization, external data, database models, costs and other elements. The range of the discussion is also a limitation as it is difficult to identify functional relationships between relevant users, hardware and applications in the context of a specific BI environment. For example, it is not easy to identify which users, applications and hardware are related to which layer in the framework shown in Figure 2-3. Data relationships cannot easily be linked to users and tools. It would be possible to identify the nodes to be modified to optimise or improve the application of, for example, multilingualism in an existing BI environment, but it would be difficult to identify the relationships between these nodes and other components.

- **Kimball's approach: A BI framework with the focus on business needs**

An alternative to Inmon's approach and also a highly influential approach is that of Kimball. In the Kimball *et al.* (2008) approach to BI Architecture, the DW is not implemented separately as an additional storage element which holds all organisational information as in Inmon's approach. Thus, there is no additional physical database representing DW. In the Kimball approach, a DW is only a conceptual idea that encompasses data marts and relevant functionalities. The data marts are tightly integrated to enable efficient data retrieval, using a common set of conformed and standardised dimensions and facts (Poolet, 2007). A sketch of the BI framework, which is based on the Kimball approach, is shown in Figure 2-4.

The strengths of the framework based on the Kimball approach are similar to those of the framework derived from the Inmon approach. The BI framework provides an overall view of the components included in the BI environment, and supports identification of most components relevant to the development of a new BI environment, and optimisation and improvement of an existing BI environment.

However, the Kimball *et al.* (2008) approach is focused on explaining and defining the DW lifecycle rather than developing a holistic framework for BI. As shown in Figure 2-4, the framework extracted from the Kimball's approach focuses on high level, not lower level implementation concepts. As with the Inmon approach, the relevant components can be identified, at a high level of detail, but it is not possible to identify interrelationships and interconnectivity at a lower level of detail, for example, functional relations between relevant users, hardware and applications. Although not explicitly stated, the framework extracted from the Kimball approach also suggests the idea of the three-layered BI framework which includes (i) data sources,(ii) DW based on conformed dimensions, and (iii) reporting and querying layer.

In summary, most of the conceptual approaches with respect to the definition of BI and BI frameworks provide high level representations of BI and a useful overview of the BI environment. However, none of these frameworks and approaches can be considered as providing a holistic view of BI because they do not map to lower level representations of the components and relationships between the components which together compose the BI environment. The Inmon and Kimball approaches, although not formally defined as BI frameworks, seem to offer the most useful overview of the generic BI environment. It is important to note that both Inmon and Kimball promote the idea of the three-layered approach to BI.

Data Oriented Approaches

Data oriented approaches typically rely on the concept of layers to describe the data journey from data source to presentation layer. Their focus is mostly on usability of different types of data at different levels rather than on components of BI environment. We distinguish data oriented approaches from conceptual approaches, which also make reference to data issues, because in the data oriented approach, the focus is primarily on the data journey rather than on architectural or other elements. This section discusses three data oriented approaches.

- **Three-layer framework (Baars & Kemper)**

The three-layered framework developed by Baars and Kemper (2008) describes BI in terms of (i) an access layer which allows users to access information, (ii) a logical layer which handles data analysis and supports knowledge distribution and data analysis, and (iii) a data layer which handles data storage and content management (Figure A-5). The data layer receives input from data generation operational systems and external data sources.

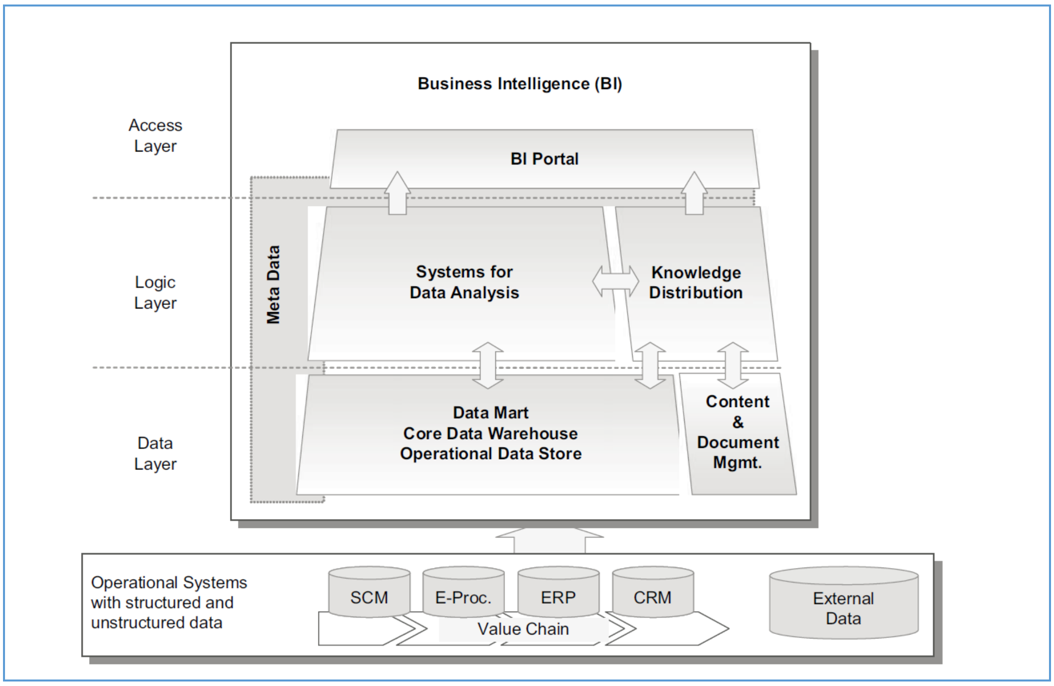


Figure A-5: A three-layered Business Intelligence Framework (Source: Baars and Kemper, 2008, p.137)

The three-layer framework provides a logical high level view of the BI architecture and maps the logical components of BI and their relations (Dod & Sharma, 2012). The strength of this framework is the clarity of the representation of relevant data layers in the BI environment, and its wider focus which goes beyond transactional and master data only. It covers other aspects of the data, such as content and document management, knowledge distribution and metadata.

However, the framework has been criticised for supporting only one way data flow to the BI portal and for its weak handling of metadata (Ong *et al.*, 2011). A further limitation of this framework is the fact that it considers the BI framework from the perspective of the data only. The framework does not consider relationships to and between applications, hardware, users or concepts. Source systems are considered as external to the framework although the input from source systems is recognised. It is argued here that as source systems are a requirement in the BI environment, given that without them there is no data, it is questionable whether source systems can appropriately be seen as an element outside the BI environment. The focus of this framework on data means that it would not support the identification of all the components and interrelationships between and within components required to develop or extend a BI system

- **BI architecture (Ranjan)**

Ranjan (2009) developed a BI framework which covers data and some additional technical aspects of BI. One of the strengths of this approach is the well described reporting layer. Ranjan (2009) separated BI into the following elements: (i) raw data and databases, (ii) DW and relevant applications, and (iii) BI tools (Figure A-6).

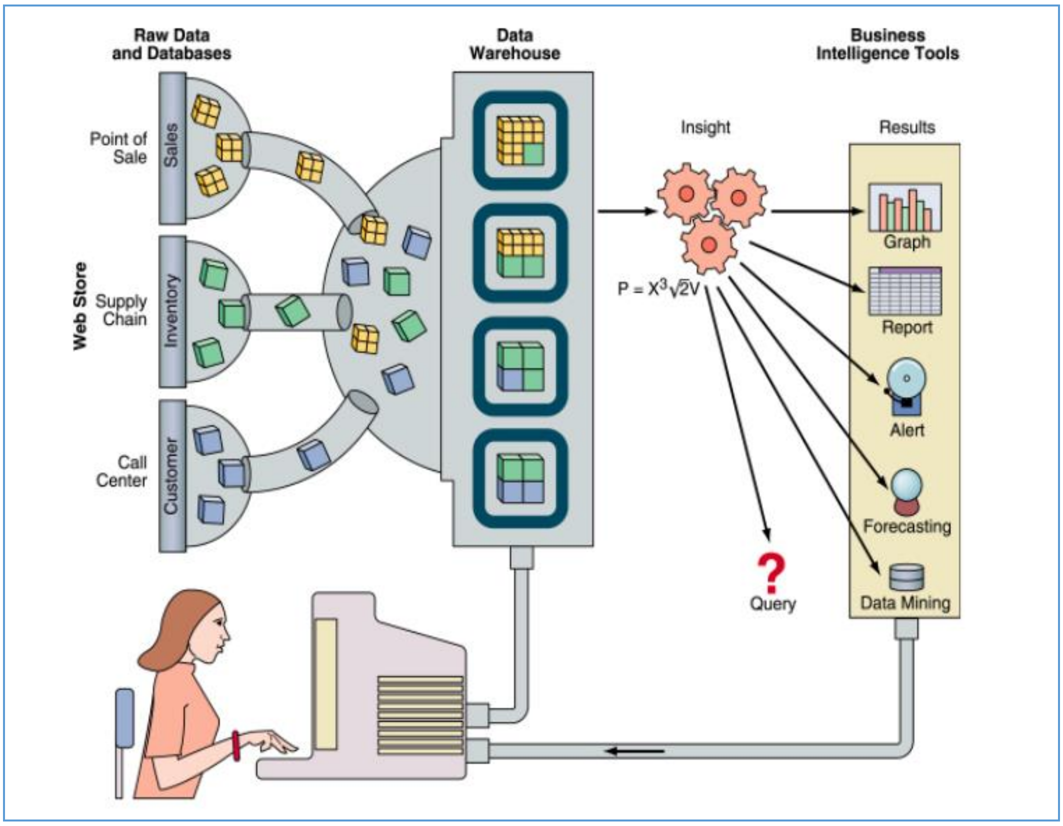


Figure A-6: A BI Framework (Source: Ranjan, 2009, p. 64)

The strengths of this framework are the clarity and simplicity of the presentation of the BI architecture while also considering possible layers, concepts, some applications and some categories of source data.

The Ranjan approach used the term "Business Intelligence Tools" to describe the tasks involved in reporting, analytics and querying. This is potentially misleading as the more usual usage (Kimball *et al.* 2008) is to include DWH and BI in the same category. None of the frameworks presented here define BI exclusively as a reporting category or as a set of the reporting tools only. The framework, as shown in Figure A-6, also includes an outlier object called "Query" and the relationship between query and results is not clear. As the calculation happens during report/query execution or in some cases during transformation process in data warehouse itself, the second outlier, named "Insight", depicted in Figure A-6, should be appended either to the data warehouse category, or to the reporting and querying category. The three-layered approach has a number of limitations including the omission of applications, types of data, relevant information about users, hardware and concepts. Such an approach makes it difficult to identify all relevant aspects and components and thus to extend or modify existing BI environment.

- **Business Intelligence Layers Architecture (Gluchowski & Kemper)**

Gluchowski & Kemper (2006) defined a BI architecture which included all system components that help the gathering and processing of data, their preparation and permanent storage, and their analysis and presentation in appropriate form (Figure A-7). Although not officially defined as such, their definition of architectural layers represents another example of a BI framework. The architecture of the BI environment is separated into three levels: (i) data source level comprising operational systems and external data, (ii) storage and preparation layer comprising memory and ETL, and (iii) presentation and analysis layer comprising various reports, management cockpits, dashboard and related elements.

The strength of the Gluchowski & Kemper (2006) architecture is the clarity of the presentation of relevant layers in the BI environment. However, the difficulty of clearly identifying and separating the relevant horizontal components of BI environment (hardware, concepts, users, applications, types of data) at each layer is a significant limitation.

In the scenario where we need to change or extend the BI environment, this approach provides some support for the identification of the nodes to be changed or application(s) to be used. We can easily identify relevant layers and to some extent relevant applications. However, the approach does not support identification of other elements, such as users, hardware, data types and concepts.

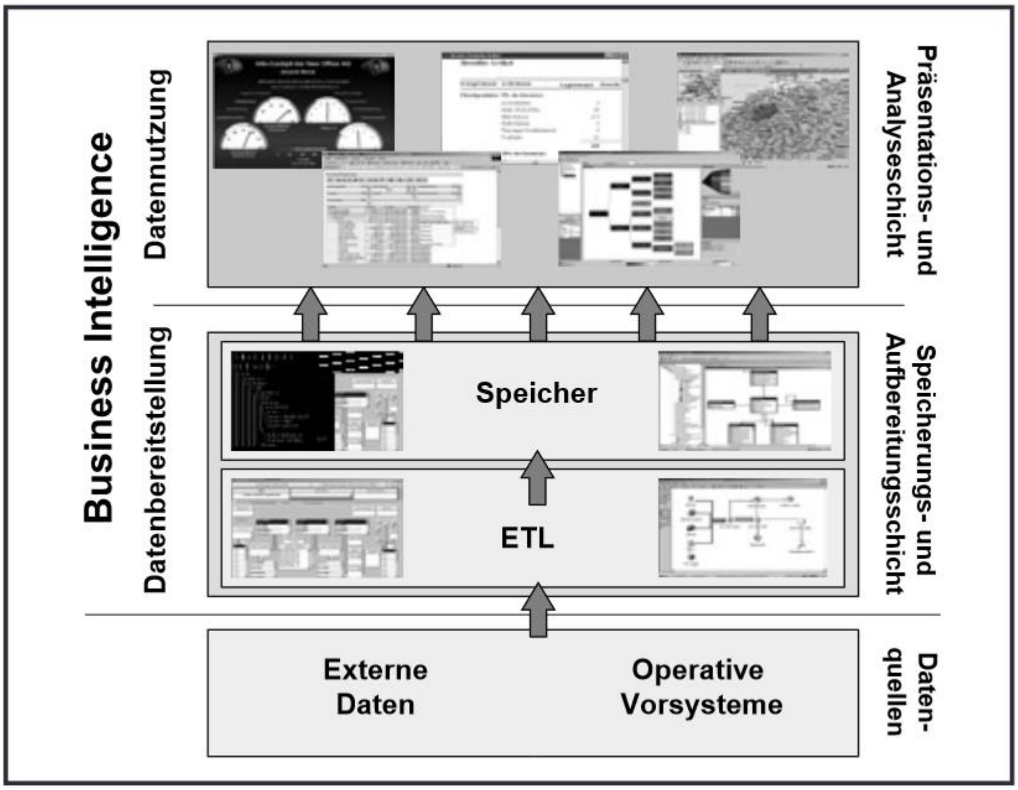


Figure A-7: Business Intelligence Layers Architecture (Gluchowski & Kemper, 2006, p. 14)

As the discussion demonstrates, data oriented approaches support clear descriptions of the BI environment and relevant layers. A further strength of data oriented approaches is the capability to provide simple visual insight into the data journey from source to presentation layer. However, BI frameworks based on data oriented approaches cannot be regarded as holistic, in the sense discussed here of identifying core components and the relationships, dependencies and connectivity between elements at different data layers, since only the data perspective is considered. Relationships to applications, hardware, users groups or concepts are either not considered or are considered only superficially. Data oriented approaches do not support the identification and separation of horizontal components of the BI environment, such as hardware, concepts, user groups and applications relevant for every separate layer. However, all the data oriented approaches can be used to support the idea of a three-layered BI framework when considering this as a vertical perspective.

Business Oriented Approaches

This section describes BI frameworks which may be layer-based or conceptual in their nature and which focus on a specific business category or interest. Three different frameworks are discussed in this section.

- **Process Mining: A framework proposal for Pervasive Business Intelligence**

Guarda *et al.* (2013) proposed a framework for process mining BI, consisting of four layers: i) objectives definition ii) collection iii) analysis and iv) dissemination (Figure A-8). If we disregard the first layer in the proposed framework, which is explicitly process mining based, the approach is very similar to the conventional three-layered BI framework which includes (i) collection (data source) layer, (ii) analysis (DWH) layer and (iii) dissemination layer. The use of the three-layered structure in a business focussed context supports the view that a three-layered based framework is the most generally used BI approach.

The biggest strength of this framework, which is the focus on process mining through pervasive BI, is also its limitation. The discussion of the business usage of BI frameworks is outside the scope of this paper. However, considering the requirement for support for multiple perspectives in BI, it can be noted that this framework does not provide sufficient information about relevant applications (software), hardware, types of data, user groups, possible layers and their concepts. This approach does not support the identification of those aspects and so lacks support for the development or modification of a BI environment.

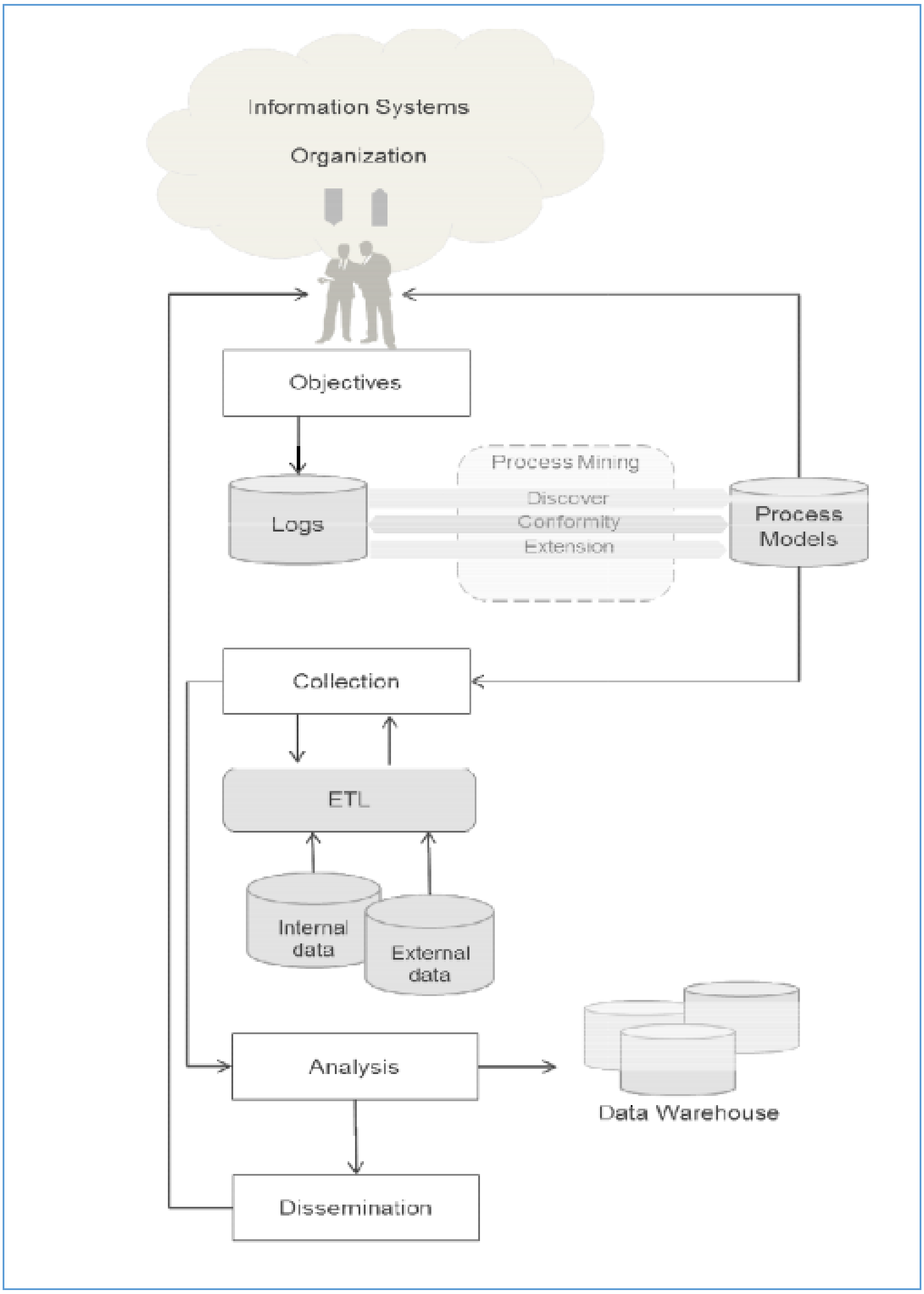


Figure A-8: Process mining framework for PBI (Source: Guarda et al, 2013, p.3)

- **BI Systems Implementation in Manufacturing**

Chu (2013) proposes a conceptual framework for BI systems implementation in manufacturing. The BI infrastructure contains three layers as shown in Figure A-9, thus supporting the idea of the three-layered framework presented by Baars and Kempers

(2008) & Ranjan (2009). The infrastructure includes components for data transformation (ETL); data storage (DW and data marts); and operational data. It can be used to support our argument that the BI architecture consisting of three layers (data source, DW and reporting) is the most widely used and understood approach and is suitable for use in a generic BI framework. The framework provides an overview of a possible BI environment in manufacturing companies, which can be understood easily by non-technical users and identifies a number of the concepts and applications used in the BI environment such as metadata and analysis. Other key perspectives such as stakeholders, types of data and the majority of BI applications are not identifiable. Because of its focus on one segment of industry this approach is not readily generalisable.

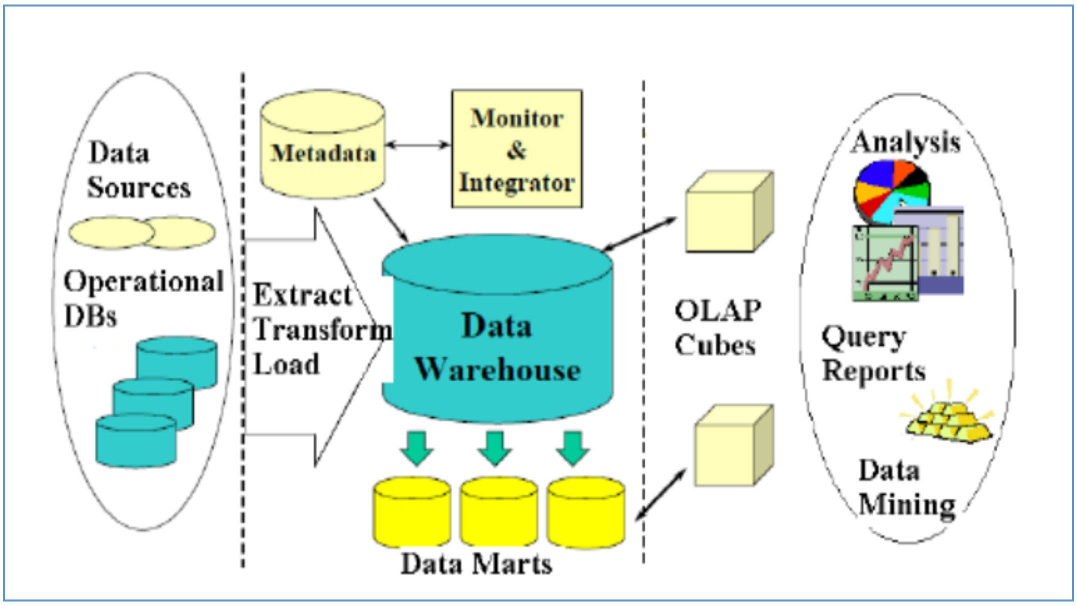


Figure A-9: Conventional BI Infrastructure according to Chu (2013, p. 114)

- **A Dynamic Capability-Based Framework for Business Intelligence**

An alternative to the three-layered architecture is the capability approach developed by Olszak (2014) who identifies six capabilities covering governance, culture, technology, people, processes, and change management & creativity (Figure A-10). This framework illustrates the complexity of BI since the emphasis is on the wider BI environment. The approach is high level and capability-based and does not provide sufficient detail to identify relevant BI components.

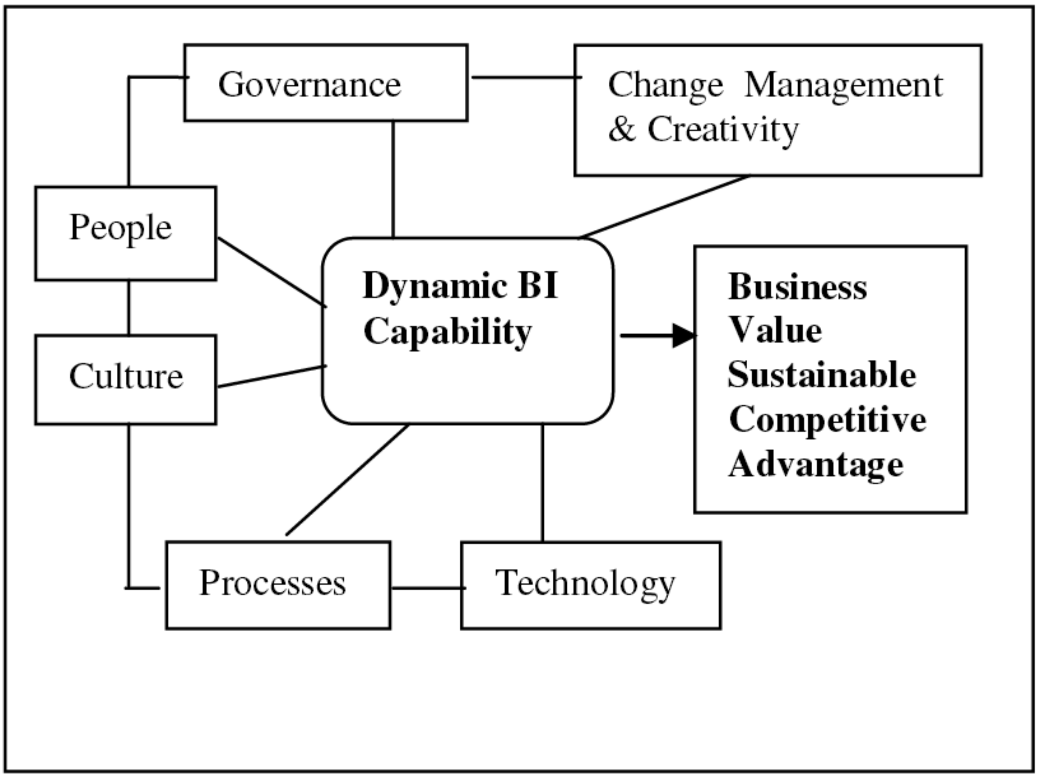


Figure A-10: Framework for BI capabilities (Source: Olszak, 2014, p.1106)

In summary, the strengths of business oriented frameworks are also their biggest limitations as their focus tends to be limited to specific business areas. Some of the frameworks partially identify relevant components and aspects, such as data layers, applications and types of data in the BI environment, but because of their focus on specific business categories or interest, these frameworks cannot be considered as holistic or generic, in the sense defined in section 1, solutions.

APPENDIX B. Business Intelligence Framework Evaluation (Pilot Survey)

The Figure 1 below shows a Business Intelligence Framework that is proposed a holistic solution. This framework is a product of scientific research conducted in 2015.

It is intended to be used by **business users**, **management** and **technical users** for easier understanding of all relevant components involved in one Business Intelligence project.

It should enable immediate identification of all relevant objects and understanding or relevant aspects when considering changes in existing Business Intelligence environment, such as development of the new report, modifying existing one, etc.

As you're probably belong to the category of the users that has works with Business Intelligence solutions (reports, data warehouse, source systems, etc.), we would like to ask you to participate in this short survey.

All the questions in the survey are related to the framework below and it takes **1 to 2 minutes** to complete the survey.

All results are anonym!

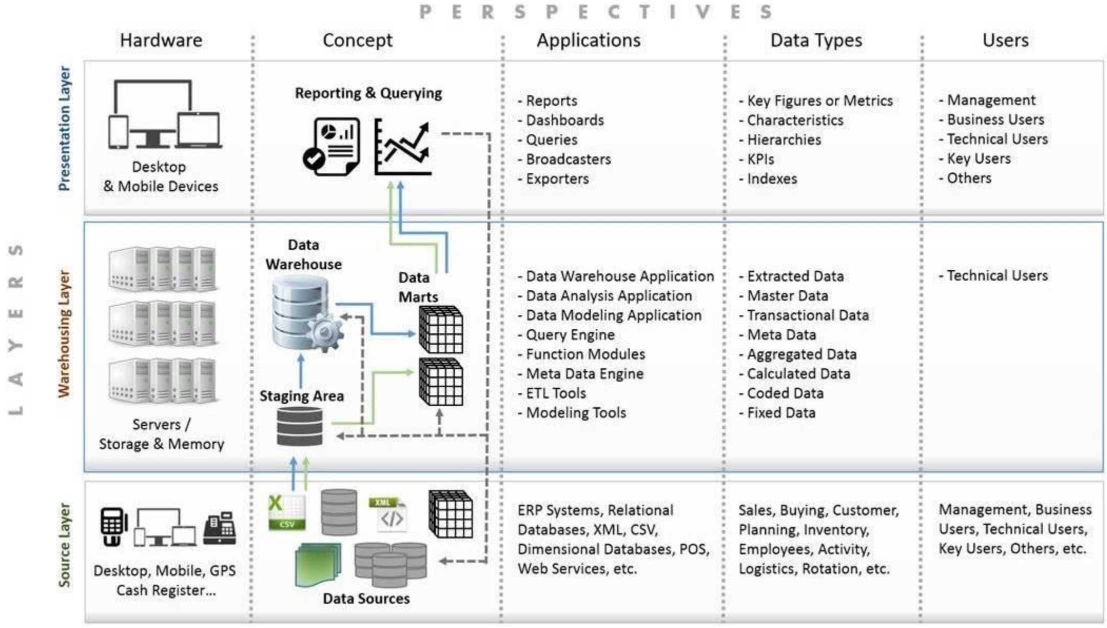


Figure B-1: Proposed Framework for Business Intelligence

There are 8 questions in this survey

1. Please select your age group: *

Please choose **only one** of the following:

- ○ 0 - 29
- ○ 30 - 49
- ○ 50 or more

2. Please select your gender: *

Please choose **only one** of the following:

- ○ Female
- ○ Male

3. What type of Business Intelligence user you are? *

Please choose **only one** of the following:

- ○ Business user
- ○ Management user
- ○ Technical user
- ○ Other

Business user (uses reports for every day activities); **Management** (uses reports to make decisions); **Technical user** (developers reports, data warehouse, etc); **Other** (all other users);

4. How long are you working in or you had some activities related to Business Intelligence? *

Please choose **only one** of the following:

- ○ 0 to 2 years
- ○ 3 to 5 years
- ○ 6 or more years

5. From the figure 1 above, how easy it would be for you to identify relevant PERSPECTIVE components (such as hardware**,** software**,** applications**, etc.) that might be involved in respective Business Intelligence project? ***

Please choose the appropriate response for each item:

	Impossible	*Very hard*	*Hard*	*Undecided*	*Easy - additional help might be needed*	*Easy - without additional help*	*Very easy*
Hardware	○	○	○	○	○	○	○
Concept	○	○	○	○	○	○	○
Applications	○	○	○	○	○	○	○
Data Type	○	○	○	○	○	○	○
Users	○	○	○	○	○	○	○

6. From the figure above, how easy it would be for you to identify LAYERS components (source layer**,** data warehousing **and** presentation layer**), etc. that might be involved in respective Business Intelligence project? ***

Please choose the appropriate response for each item:

	Impossible	*Very hard*	*Hard*	*Undecided*	*Easy - additional help might be needed*	*Easy - without additional help*	*Very easy*
Presentation layer components	○	○	○	○	○	○	○
Data Warehousing layer components	○	○	○	○	○	○	○
Data Source layer components	○	○	○	○	○	○	○

7. Do you find the figure above useful for understanding of Business Intelligence project and components that might be involved in respective project? *

Please choose **only one** of the following:

1. ○ Yes
2. ○ No

8. Generally speaking, do you find the concept of the framework from the picture above easy to understand? *

Please choose **only one** of the following:

- ○ Yes
- ○ No

Thank your very much!

Nedim

APPENDIX C. Business Intelligence Framework Evaluation Survey

The diagram below (Figure 1) shows a Business Intelligence Framework that is proposed as a holistic representation of the components involved in Business Intelligence (BI) processes.

This framework was developed based on research conducted in 2015/16. It is intended to be used by technical, business, management and other Business Intelligence users to provide a high level overview and easier understanding of the components that may be involved in a Business Intelligence project.

The aim of the Framework is to support immediate identification of all relevant components, and understanding of the interactions between components, when developing a new BI project or considering changes in existing Business Intelligence environments, such as development of a new report or modification of an existing report.

As you belong to the category of the users that work with Business Intelligence solutions (reports, data warehouse, source systems, etc.), we ask you to be kind enough to participate in this short survey.

All the questions in the survey are related to the framework in the Figure 1 below and it takes **2** to **5 minutes** to complete the survey. All results are anonymous. If you would like further information about the Framework or the research project on which it is based, please contact Nedim Dedić [email: nedim.dedic@research.staffs.ac.uk].

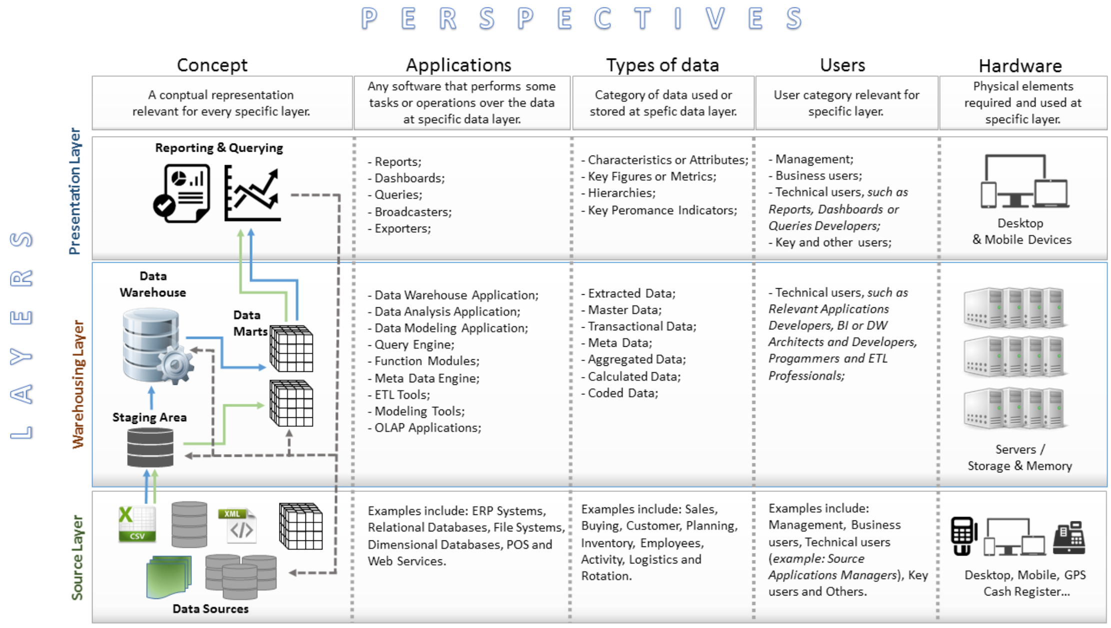


Figure C-1: Business Intelligence Framework

There are 9 questions in this survey

1. How long have you been working in, or had some involvement with, Business Intelligence projects? *

Please choose **only one** of the following:

- ○ 0 – 2 years
- ○ 3 – 5 years
- ○ 6 or more years

2. What type of Business Intelligence user you are? *

Please choose **only one** of the following:

- ○ Technical user
- ○ Data-centric user
- ○ Business user
- ○ Management user
- ○ Other user

Explanation:

Technical users *– Examples include: Report or Data Warehouse developers, BI Architects or Solutions Designers, Programmers and Source Systems Application Managers. Any user that perform technical activities in respective Business Intelligence project;*

Data-centric user *– Examples include: Statisticians or Mathematicians, Data Scientists or Data Miners. Users that create and define adequate formulas and standards to discover patterns in large data sets, or to extract knowledge or insights from data in various forms.*

Business users *- Includes people from various areas, such as controlling, finance, human resources, sales and logistics, which use Business Intelligence reports to perform their daily work;*

Management users *– Examples include: Company CEO, Owner, Department or Team Manager. This category uses Business Intelligence reports to make decisions;*

Other users *– All other users not belonging to the first three categories;*

If you work in more than one category, please pick the category which most reflects your area of expertise

3. From the Business Intelligence Framework diagram above, how easy it would be for you to identify relevant PERSPECTIVE components (Concept, Applications, Type of data, Users or Hardware) that might be involved in a Business Intelligence project? *

Please choose the appropriate response for each item:

	Impossible	*Very Hard*	*Hard*	*Undecided*	*Easy with help*	*Easy without help*	*Very easy*
Concept	○	○	○	○	○	○	○
Applications	○	○	○	○	○	○	○
Types of data	○	○	○	○	○	○	○
Users	○	○	○	○	○	○	○
Hardware	○	○	○	○	○	○	○

4. From the Business Intelligence Framework diagram above, how easy it would be for you to identify LAYER components (Source, Data Warehousing or Presentation Layer) that might be involved in a Business Intelligence project? *

Please choose the appropriate response for each item:

	Impossible	*Very hard*	*Hard*	*Undecided*	*Easy with help*	*Easy without help*	*Very easy*
Presentation Layer	○	○	○	○	○	○	○
Warehousing Layer	○	○	○	○	○	○	○
Data Source Layer	○	○	○	○	○	○	○

5. Altogether, do you find the proposed Business Intelligence framework useful for the understanding of Business Intelligence activities? For example, would it be useful in identifying components that might be involved in a BI project? *

Please choose **only one** of the following:

- ○ Yes
- ○ No

6. Generally speaking, do you find the concept of the framework as shown in the Framework diagram easy to understand? *

Please choose **only one** of the following:

- ○ Yes
- ○ No

7. Please list any additional components which you feel should be included in the Business Intelligence framework

Please write your answer here:

8. Are there any components which you feel should not be included in the Business Intelligence Framework?

Please write your answer here:

9. Any additional comments?

Please write your answer here:

Thank you very much!

APPENDIX D. Differences between Design Approaches to implement BI / DW System

Table D-1: Differences between Design Approaches to implement BI / DW System

	Inmon, Data Vault	Kimball	MLED_BI
Business Intelligence System / Environment Concept	- Reporting Applications - Data Warehouse + Data Marts - Source System	- Reporting Applications - Data Marts - Source System	- Reporting Applications + CMS - Data Warehouse + Data Marts + Language Files or Data Marts only + Language Files - Source System
Reporting Layer	Contains Web interfaces to select, browse, filter, drill, re-execute and share reports. Uses information stored in Data Marts for reporting.	Contains Web interfaces to select, browse, filter, drill, re-execute and share reports. Uses information stored in Data Marts for reporting.	Contains Web interfaces to select, browse, filter, drill, re-execute and share reports. Web interfaces are extended with Content Management System (CMS) to manage, add and remove descriptive content, including language manipulation. Uses information stored in Data Marts for reporting. Uses information stored in Language Files for reporting.
Data Marts Concept (based on Star Schema)	Have both: fact tables and dimension tables. Fact tables hold transactional data. Dimension tables hold master data, including IDs and relevant descriptive information. Dimensional data are redundant.	Have both: fact tables and dimension tables. Fact tables hold transactional data. Dimension tables hold master data, including IDs and relevant descriptive information. Dimensional data are redundant.	Has both: fact tables and dimension tables. Fact tables hold transactional data. Dimension tables hold master data, however, only IDs. Relevant descriptive information are stored outside dimensional tables as language files. Dimensional data are NOT redundant.
Data Warehouse Concept	All data from source systems are replicated and saved	Data Warehouse is a concept only that consist of	MLED_BI design approach to BI is conformed to both

	into actual Data Warehouse. Additional Data Marts holding information required by business are used for reporting.	Data Marts connected using conformed dimensions and holding only data needed by business.	approaches compared (Inmon/Data Vault and Kimball). It can have separate Data Warehouse with additional Data Marts, or only Data Marts needed by business. However, as Data Marts hold only numerical (ID) values, descriptive information, although belonging conceptually to Data Warehouse, are stored as Language Files on server.
Multilingualism enabled by:	1) Additional attributes in dimensional tables. 2) Language identifier via additional field. 3) Additional Tables / Schemas.	1) Additional attributes in dimensional tables. 2) Language identifier via additional field. 3) Additional Tables / Schemas.	1) Language files elsewhere on server.

APPENDIX E. Implementation of MCMS

This appendix provides additional information about the implementation of the MCMS.

E.1 Context

Figure E-1, given in Chapter 7, shows the architecture of the MCMS and is reproduced here to give the context of the discussion.

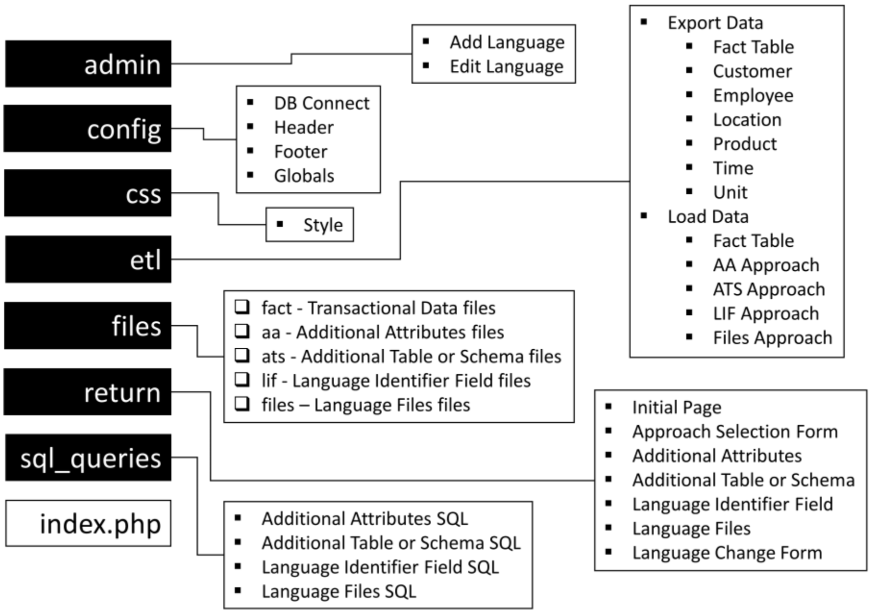


Figure E-1: MCMS Web Environment Architecture

The black boxes shown in Figure E-1 represent the physical structure of a folder, while listings in white boxes represent files in PHP and MYSQL, or in the case of the subfolders of folder Files, the underlying data structures. languages and administration functions.

E.2 Data Mart Implementation

As there are no structural changes to the transactional data, every fact table (sales_fact_table) in every data mart holds same amount data. However, while the DMs based on AA and MLED_BI approaches have same amount of data and same tables, they do not have same table structure. The DM based on the ATS approach has one fact table but double the amount of dimensional tables, representing the two different languages. The DM based on the LIF approach has the same number of tables as those based on AA and MLED_BI approaches; however, it has double amount of data. This can be seen in the figures D-2, D-3, D-4 and D-5.

Server: Local Databases » Database: phd_project_aa

Structure | SQL | Search | Query | Export | Import | Operations | Privileges | Routines | Events

Table	Action	Rows	Type	Collation	Size	Overhead
customer_dimension	Browse Structure Search Insert Empty Drop	100	MyISAM	latin1_general_ci	19.2 KiB	-
employee_dimension	Browse Structure Search Insert Empty Drop	100	MyISAM	latin1_general_ci	12.2 KiB	-
location_dimension	Browse Structure Search Insert Empty Drop	100	MyISAM	latin1_general_ci	13.1 KiB	-
product_dimension	Browse Structure Search Insert Empty Drop	216	MyISAM	latin1_general_ci	37.9 KiB	-
sales_fact_table	Browse Structure Search Insert Empty Drop	1,199,989	MyISAM	latin1_general_ci	90.8 MiB	-
time_dimension	Browse Structure Search Insert Empty Drop	366	MyISAM	latin1_general_ci	23.3 KiB	-
unit_dimension	Browse Structure Search Insert Empty Drop	6	MyISAM	latin1_general_ci	2.2 KiB	-
7 tables	Sum	1,200,877	MyISAM	latin1_general_ci	90.9 MiB	0 B

Figure E-2: Tables of data mart based on AA implementation approach after ETL processes

Server: Local Databases » Database: phd_project_ats

Structure | SQL | Search | Query | Export | Import | Operations | Privileges | Routines | Events | Tri

Table	Action	Rows	Type	Collation	Size	Overhead
de_customer_dimension	Browse Structure Search Insert Empty Drop	100	MyISAM	latin1_general_ci	11.4 KiB	-
de_employee_dimension	Browse Structure Search Insert Empty Drop	100	MyISAM	latin1_general_ci	7.8 KiB	-
de_location_dimension	Browse Structure Search Insert Empty Drop	100	MyISAM	latin1_general_ci	8.1 KiB	-
de_product_dimension	Browse Structure Search Insert Empty Drop	216	MyISAM	latin1_general_ci	23.3 KiB	-
de_time_dimension	Browse Structure Search Insert Empty Drop	366	MyISAM	latin1_general_ci	17.8 KiB	-
de_unit_dimension	Browse Structure Search Insert Empty Drop	6	MyISAM	latin1_general_ci	2.1 KiB	-
en_customer_dimension	Browse Structure Search Insert Empty Drop	100	MyISAM	latin1_general_ci	10.7 KiB	-
en_employee_dimension	Browse Structure Search Insert Empty Drop	100	MyISAM	latin1_general_ci	7.3 KiB	-
en_location_dimension	Browse Structure Search Insert Empty Drop	100	MyISAM	latin1_general_ci	7.8 KiB	-
en_product_dimension	Browse Structure Search Insert Empty Drop	216	MyISAM	latin1_general_ci	21.5 KiB	-
en_time_dimension	Browse Structure Search Insert Empty Drop	366	MyISAM	latin1_general_ci	18.7 KiB	-
en_unit_dimension	Browse Structure Search Insert Empty Drop	6	MyISAM	latin1_general_ci	2.1 KiB	-
sales_fact_table	Browse Structure Search Insert Empty Drop	1,199,989	MyISAM	latin1_general_ci	90.7 MiB	-
13 tables	Sum	1,201,765	MyISAM	latin1_general_ci	90.9 MiB	0 B

Figure E-3: Tables of data mart based on ATS implementation approach after ETL processes

Server: Local Databases » Database: phd_project_lif

Structure | SQL | Search | Query | Export | Import | Operations | Privileges | Routines | Events

Table	Action	Rows	Type	Collation	Size	Overhead
customer_dimension	Browse Structure Search Insert Empty Drop	200	MyISAM	latin1_general_ci	25.6 KiB	-
employee_dimension	Browse Structure Search Insert Empty Drop	200	MyISAM	latin1_general_ci	19.6 KiB	-
location_dimension	Browse Structure Search Insert Empty Drop	200	MyISAM	latin1_general_ci	19.3 KiB	-
product_dimension	Browse Structure Search Insert Empty Drop	432	MyISAM	latin1_general_ci	51 KiB	-
sales_fact_table	Browse Structure Search Insert Empty Drop	1,199,989	MyISAM	latin1_general_ci	90.7 MiB	-
time_dimension	Browse Structure Search Insert Empty Drop	732	MyISAM	latin1_general_ci	43.9 KiB	-
unit_dimension	Browse Structure Search Insert Empty Drop	12	MyISAM	latin1_general_ci	2.3 KiB	-
7 tables	**Sum**	1,201,765	**MyISAM**	**latin1_general_ci**	90.9 MiB	0 B

Figure E-4: Tables of data mart based on LIF implementation approach after ETL processes

Server: Local Databases » Database: phd_project_files

Structure | SQL | Search | Query | Export | Import | Operations | Privileges | Routines | Events

Table	Action	Rows	Type	Collation	Size	Overhead
customer_dimension	Browse Structure Search Insert Empty Drop	100	MyISAM	latin1_general_ci	6 KiB	-
employee_dimension	Browse Structure Search Insert Empty Drop	100	MyISAM	latin1_general_ci	4.1 KiB	-
location_dimension	Browse Structure Search Insert Empty Drop	100	MyISAM	latin1_general_ci	4.1 KiB	-
product_dimension	Browse Structure Search Insert Empty Drop	216	MyISAM	latin1_general_ci	11.1 KiB	-
sales_fact_table	Browse Structure Search Insert Empty Drop	1,199,989	MyISAM	latin1_general_ci	90.8 MiB	-
time_dimension	Browse Structure Search Insert Empty Drop	366	MyISAM	latin1_general_ci	12.1 KiB	-
unit_dimension	Browse Structure Search Insert Empty Drop	6	MyISAM	latin1_general_ci	2 KiB	-
7 tables	**Sum**	1,200,877	**MyISAM**	**latin1_general_ci**	90.8 MiB	0 B

Figure E-5: Tables of data mart based on MLED_BI implementation approach after ETL processes

E.3 Implementation of Reports in the Web Environment

The initial WE access page has a menu offering appropriate navigation possibilities. In addition to the menu, the initial homepage, shown in Figure E-6, allowed the user to select the data mart implementation to be used as a basis for initial execution of BI report. As explained in section 7.3, there were four possible data mart implementations:

- Additional Attributes (based on conventional ML BI design approach)
- Language Identifier Field (based on conventional ML BI design approach)
- Additional Tables or Schema (based on conventional ML BI design approach)
- Language Files (based on MLED_BI design approach)

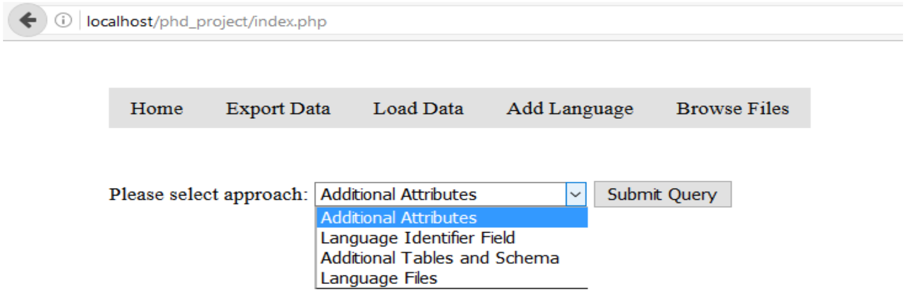


Figure E-6: Homepage of Web Environment

Reports returned the same information to the user, irrespective of the implementation approach selected. However, in column sorting operations, different approaches used different types of data as a basis for sorting, resulting in some minor differences. Figure E-7, Figure E-8, Figure E-9, and Figure E-10 show screenshots of four BI reports based on different DM implementation approaches. Each report returned the same information, which was expected given that all four approaches used data acquired from same source system and enabled like for like comparison between the approaches.

As seen in top right corner of Figure X-3, Figure X-4, Figure X-5, and Figure X-6, the WE provided a menu to enable selection of the language to be used to preview business content descriptions (master data). In this case, end users were only able to switch between the German and English language, as only those two language were available in the sample source system database.

English German

Home Export Data Load Data Add Language Browse Files

Please select approach: Additional Attributes Submit Query
You selected:

Additional Attributes Approach

Products Sales Overview

Year	Area	Category	Subcategory	Gross Sales	Net Sales	Profit
2015	Food	Beverages	Coffee	234,760.66	217,149.61	17,611.05
2015	Food	Beverages	Other Beverages	271,004.20	250,690.72	20,313.48
2015	Food	Beverages	Water	688,013.23	636,421.07	51,592.16
2015	Food	Stable Dairy Products	Shelf Stable Dairy Produc	621,596.83	574,979.87	46,616.96
2015	Food	Stable Food Products	Cereals	62,392.50	57,714.10	4,678.40
2015	Food	Stable Food Products	Culinary Products	127,905.14	118,316.00	9,589.14
2015	Food	Stable Food Products	Frozen Foods	129,133.62	119,453.99	9,679.63
2015	Food	Stable Food Products	Ice Cream	258,410.42	239,038.16	19,372.26
2015	Food	Stable Food Products	Infant Foods	533,165.17	493,168.45	39,996.72
2015	Food	Stable Food Products	Nutrition	125,776.80	116,343.54	9,433.26
2015	Food	Stable Food Products	Refridgerated Products	126,364.27	116,890.20	9,474.07
2015	Food	Stable Food Products	Sweets	488,282.01	451,680.66	36,601.35
2015	Food II	Stable Pet Products	Pets	587,361.36	513,991.44	73,369.92
2015	Non Food	Other Stable Products	Cosmetics	599,662.26	524,730.79	74,931.47
2015	Non Food	Other Stable Products	Pharmacy	101,751.93	89,035.57	12,716.36
2015	Non Food I	Clothing	Clothing	14,840,484.00	9,943,130.98	4,897,353.02
2015	Non Food II	Electronics	Smartphone	584,224,792.50	391,430,736.04	192,794,056.46
2015	Non Food II	Electronics	Smartphones Android	355,256,006.50	238,021,601.14	117,234,405.36
2015	Non Food II	Electronics	Smartphones Windows	30,336,996.50	20,325,794.69	10,011,201.81
2015	Non Food II	Electronics	Tablet	78,195,639.50	52,391,099.37	25,804,540.13
2015	Services	Services	Others	131,722.48	121,851.17	9,871.31
2016	Food	Beverages	Coffee	234,823.98	217,209.12	17,614.86
2016	Food	Beverages	Other Beverages	274,105.73	253,559.88	20,545.85
2016	Food	Beverages	Water	686,898.42	635,388.40	51,510.02
2016	Food	Stable Dairy Products	Shelf Stable Dairy Produc	611,563.92	565,700.18	45,863.74
2016	Food	Stable Food Products	Cereals	62,681.60	57,981.19	4,700.41
2016	Food	Stable Food Products	Culinary Products	128,547.00	118,909.77	9,637.23
2016	Food	Stable Food Products	Frozen Foods	129,288.50	119,597.45	9,691.05
2016	Food	Stable Food Products	Ice Cream	258,262.92	238,901.96	19,360.96
2016	Food	Stable Food Products	Infant Foods	535,587.07	495,409.03	40,178.04
2016	Food	Stable Food Products	Nutrition	127,327.20	117,777.66	9,549.54
2016	Food	Stable Food Products	Refridgerated Products	130,209.75	120,448.12	9,761.63
2016	Food	Stable Food Products	Sweets	486,443.56	449,979.34	36,464.22
2016	Food II	Stable Pet Products	Pets	597,681.71	523,022.49	74,659.22
2016	Non Food	Other Stable Products	Cosmetics	604,204.35	528,705.71	75,498.64
2016	Non Food	Other Stable Products	Pharmacy	100,757.79	88,165.73	12,592.06
2016	Non Food I	Clothing	Clothing	15,435,091.50	10,341,518.37	5,093,573.13
2016	Non Food II	Electronics	Smartphone	583,118,971.00	390,689,835.91	192,429,135.09
2016	Non Food II	Electronics	Smartphones Android	353,354,596.00	236,747,655.21	116,606,940.79
2016	Non Food II	Electronics	Smartphones Windows	30,598,696.50	20,501,133.13	10,097,563.37
2016	Non Food II	Electronics	Tablet	78,323,796.50	52,476,964.76	25,846,831.74
2016	Services	Services	Others	133,398.19	123,400.40	9,997.79

SQL Query Execution Time for additional_attributes is 12.692947149277
WEB Application Execution Time for additional_attributes is 12.696374893188

Figure E-7: Initial BI report based on AA approach of data mart implementation

English German

Home Export Data Load Data Add Language Browse Files

Please select approach: Language Files Submit Query
You selected:

Additional Tables and Schema Approach

Products Sales Overview

Year	Area	Category	Subcategory	Gross Sales	Net Sales	Profit
2015	Food	Beverages	Coffee	234,760.66	217,149.61	17,611.05
2015	Food	Beverages	Other Beverages	271,004.20	250,690.72	20,313.48
2015	Food	Beverages	Water	688,013.23	636,421.07	51,592.16
2015	Food	Stable Dairy Products	Shelf Stable Dairy Produc	621,596.83	574,979.87	46,616.96
2015	Food	Stable Food Products	Cereals	62,392.50	57,714.10	4,678.40
2015	Food	Stable Food Products	Culinary Products	127,905.14	118,316.00	9,589.14
2015	Food	Stable Food Products	Frozen Foods	129,133.62	119,453.99	9,679.63
2015	Food	Stable Food Products	Ice Cream	258,410.42	239,038.16	19,372.26
2015	Food	Stable Food Products	Infant Foods	533,165.17	493,168.45	39,996.72
2015	Food	Stable Food Products	Nutrition	125,776.80	116,343.54	9,433.26
2015	Food	Stable Food Products	Refridgerated Products	126,364.27	116,890.20	9,474.07
2015	Food	Stable Food Products	Sweets	488,282.01	451,680.66	36,601.35
2015	Food II	Stable Pet Products	Pets	587,361.36	513,991.44	73,369.92
2015	Non Food	Other Stable Products	Cosmetics	599,662.26	524,730.79	74,931.47
2015	Non Food	Other Stable Products	Pharmacy	101,751.93	89,035.57	12,716.36
2015	Non Food I	Clothing	Clothing	14,840,484.00	9,943,130.98	4,897,353.02
2015	Non Food II	Electronics	Smartphone	584,224,792.50	391,430,736.04	192,794,056.46
2015	Non Food II	Electronics	Smartphones Android	355,256,006.50	238,021,601.14	117,234,405.36
2015	Non Food II	Electronics	Smartphones Windows	30,336,996.50	20,325,794.69	10,011,201.81
2015	Non Food II	Electronics	Tablet	78,195,639.50	52,391,099.37	25,804,540.13
2015	Services	Services	Others	131,722.48	121,851.17	9,871.31
2016	Food	Beverages	Coffee	234,823.98	217,209.12	17,614.86
2016	Food	Beverages	Other Beverages	274,105.73	253,559.88	20,545.85
2016	Food	Beverages	Water	686,898.42	635,388.40	51,510.02
2016	Food	Stable Dairy Products	Shelf Stable Dairy Produc	611,563.92	565,700.18	45,863.74
2016	Food	Stable Food Products	Cereals	62,681.60	57,981.19	4,700.41
2016	Food	Stable Food Products	Culinary Products	128,547.00	118,909.77	9,637.23
2016	Food	Stable Food Products	Frozen Foods	129,288.50	119,597.45	9,691.05
2016	Food	Stable Food Products	Ice Cream	258,262.92	238,901.96	19,360.96
2016	Food	Stable Food Products	Infant Foods	535,587.07	495,409.03	40,178.04
2016	Food	Stable Food Products	Nutrition	127,327.20	117,777.66	9,549.54
2016	Food	Stable Food Products	Refridgerated Products	130,209.75	120,448.12	9,761.63
2016	Food	Stable Food Products	Sweets	486,443.56	449,979.34	36,464.22
2016	Food II	Stable Pet Products	Pets	597,681.71	523,022.49	74,659.22
2016	Non Food	Other Stable Products	Cosmetics	604,204.35	528,705.71	75,498.64
2016	Non Food	Other Stable Products	Pharmacy	100,757.79	88,165.73	12,592.06
2016	Non Food I	Clothing	Clothing	15,435,091.50	10,341,518.37	5,093,573.13
2016	Non Food II	Electronics	Smartphone	583,118,971.00	390,689,835.91	192,429,135.09
2016	Non Food II	Electronics	Smartphones Android	353,354,596.00	236,747,655.21	116,606,940.79
2016	Non Food II	Electronics	Smartphones Windows	30,598,696.50	20,501,133.13	10,097,563.37
2016	Non Food II	Electronics	Tablet	78,323,796.50	52,476,964.76	25,846,831.74
2016	Services	Services	Others	133,398.19	123,400.40	9,997.79

SQL Query Execution Time for additional_table_or_schema is 12.446967124939
WEB Application Execution Time for additional_table_or_schema is 12.450474977493

Figure E-8: Initial BI report based on ATS approach of data mart implementation

English German

Home Export Data Load Data Add Language Browse Files

Please select approach: Language Identifier Field Submit Query
You selected:

Language Identifier Field Approach

Products Sales Overview

Year	Area	Category	Subcategory	Gross Sales	Net Sales	Profit
2015	Food	Beverages	Coffee	234,760.66	217,149.61	17,611.05
2015	Food	Beverages	Other Beverages	271,004.20	250,690.72	20,313.48
2015	Food	Beverages	Water	688,013.23	636,421.07	51,592.16
2015	Food	Stable Dairy Products	Shelf Stable Dairy Produc	621,596.83	574,979.87	46,616.96
2015	Food	Stable Food Products	Cereals	62,392.50	57,714.10	4,678.40
2015	Food	Stable Food Products	Culinary Products	127,905.14	118,316.00	9,589.14
2015	Food	Stable Food Products	Frozen Foods	129,133.62	119,453.99	9,679.63
2015	Food	Stable Food Products	Ice Cream	258,410.42	239,038.16	19,372.26
2015	Food	Stable Food Products	Infant Foods	533,165.17	493,168.45	39,996.72
2015	Food	Stable Food Products	Nutrition	125,776.80	116,343.54	9,433.26
2015	Food	Stable Food Products	Refridgerated Products	126,364.27	116,890.20	9,474.07
2015	Food	Stable Food Products	Sweets	488,282.01	451,680.66	36,601.35
2015	Food II	Stable Pet Products	Pets	587,361.36	513,991.44	73,369.92
2015	Non Food	Other Stable Products	Cosmetics	599,662.26	524,730.79	74,931.47
2015	Non Food	Other Stable Products	Pharmacy	101,751.93	89,035.57	12,716.36
2015	Non Food I	Clothing	Clothing	14,840,484.00	9,943,130.98	4,897,353.02
2015	Non Food II	Electronics	Smartphone	584,224,792.50	391,430,736.04	192,794,056.46
2015	Non Food II	Electronics	Smartphones Android	355,256,006.50	238,021,601.14	117,234,405.36
2015	Non Food II	Electronics	Smartphones Windows	30,336,996.50	20,325,794.69	10,011,201.81
2015	Non Food II	Electronics	Tablet	78,195,639.50	52,391,099.37	25,804,540.13
2015	Services	Services	Others	131,722.48	121,851.17	9,871.31
2016	Food	Beverages	Coffee	234,823.98	217,209.12	17,614.86
2016	Food	Beverages	Other Beverages	274,105.73	253,559.88	20,545.85
2016	Food	Beverages	Water	686,898.42	635,388.40	51,510.02
2016	Food	Stable Dairy Products	Shelf Stable Dairy Produc	611,563.92	565,700.18	45,863.74
2016	Food	Stable Food Products	Cereals	62,681.60	57,981.19	4,700.41
2016	Food	Stable Food Products	Culinary Products	128,547.00	118,909.77	9,637.23
2016	Food	Stable Food Products	Frozen Foods	129,288.50	119,597.45	9,691.05
2016	Food	Stable Food Products	Ice Cream	258,262.92	238,901.96	19,360.96
2016	Food	Stable Food Products	Infant Foods	535,587.07	495,409.03	40,178.04
2016	Food	Stable Food Products	Nutrition	127,327.20	117,777.66	9,549.54
2016	Food	Stable Food Products	Refridgerated Products	130,209.75	120,448.12	9,761.63
2016	Food	Stable Food Products	Sweets	486,443.56	449,979.34	36,464.22
2016	Food II	Stable Pet Products	Pets	597,681.71	523,022.49	74,659.22
2016	Non Food	Other Stable Products	Cosmetics	604,204.35	528,705.71	75,498.64
2016	Non Food	Other Stable Products	Pharmacy	100,757.79	88,165.73	12,592.06
2016	Non Food I	Clothing	Clothing	15,435,091.50	10,341,518.37	5,093,573.13
2016	Non Food II	Electronics	Smartphone	583,118,971.00	390,689,835.91	192,429,135.09
2016	Non Food II	Electronics	Smartphones Android	353,354,596.00	236,747,655.21	116,606,940.79
2016	Non Food II	Electronics	Smartphones Windows	30,598,696.50	20,501,133.13	10,097,563.37
2016	Non Food II	Electronics	Tablet	78,323,796.50	52,476,964.76	25,846,831.74
2016	Services	Services	Others	133,398.19	123,400.40	9,997.79

SQL Query Execution Time for language_identifier_field is 16.71648812294
WEB Application Execution Time for language_identifier_field is 16.719608068466

Figure E-9: Initial BI report based on LIF approach of data mart implementation

English German

Home Export Data Load Data Add Language Browse Files

Please select approach: Language Files Submit Query

You selected:

Language Files

Products Sales Overview

Year	Area	Category	Subcategory	Gross Sales	Net Sales	Profit
2015	Food	Beverages	Coffee	234,760.66	217,149.61	17,611.05
2015	Food	Beverages	Water	688,013.23	636,421.07	51,592.16
2015	Food	Beverages	Other Beverages	271,004.20	250,690.72	20,313.48
2015	Food	Stable Dairy Products	Shelf Stable Dairy Products	621,596.83	574,979.87	46,616.96
2015	Food	Stable Food Products	Cereals	62,392.50	57,714.10	4,678.40
2015	Food	Stable Food Products	Infant Foods	533,165.17	493,168.45	39,996.72
2015	Food	Stable Food Products	Nutrition	125,776.80	116,343.54	9,433.26
2015	Food	Stable Food Products	Culinary Products	127,905.14	118,316.00	9,589.14
2015	Food	Stable Food Products	Frozen Foods	129,133.62	119,453.99	9,679.63
2015	Food	Stable Food Products	Ice Cream	258,410.42	239,038.16	19,372.26
2015	Food	Stable Food Products	Refridgerated Products	126,364.27	116,890.20	9,474.07
2015	Food	Stable Food Products	Sweets	488,282.01	451,680.66	36,601.35
2015	Services	Services	Others	131,722.48	121,851.17	9,871.31
2015	Food II	Stable Pet Products	Pets	587,361.36	513,991.44	73,369.92
2015	Non Food	Other Stable Products	Pharmacy	101,751.93	89,035.57	12,716.36
2015	Non Food	Other Stable Products	Cosmetics	599,662.26	524,730.79	74,931.47
2015	Non Food I	Clothing	Clothing	14,840,484.00	9,943,130.98	4,897,353.02
2015	Non Food II	Electronics	Smartphones Windows	30,336,996.50	20,325,794.69	10,011,201.81
2015	Non Food II	Electronics	Smartphones Android	355,256,006.50	238,021,601.14	117,234,405.36
2015	Non Food II	Electronics	Tablet	78,195,639.50	52,391,099.37	25,804,540.13
2015	Non Food II	Electronics	Smartphone	584,224,792.50	391,430,736.04	192,794,056.46
2016	Food	Beverages	Coffee	234,823.98	217,209.12	17,614.86
2016	Food	Beverages	Water	686,898.42	635,388.40	51,510.02
2016	Food	Beverages	Other Beverages	274,105.73	253,559.88	20,545.85
2016	Food	Stable Dairy Products	Shelf Stable Dairy Products	611,563.92	565,700.18	45,863.74
2016	Food	Stable Food Products	Cereals	62,681.60	57,981.19	4,700.41
2016	Food	Stable Food Products	Infant Foods	535,587.07	495,409.03	40,178.04
2016	Food	Stable Food Products	Nutrition	127,327.20	117,777.66	9,549.54
2016	Food	Stable Food Products	Culinary Products	128,547.00	118,909.77	9,637.23
2016	Food	Stable Food Products	Frozen Foods	129,288.50	119,597.45	9,691.05
2016	Food	Stable Food Products	Ice Cream	258,262.92	238,901.96	19,360.96
2016	Food	Stable Food Products	Refridgerated Products	130,209.75	120,448.12	9,761.63
2016	Food	Stable Food Products	Sweets	486,443.56	449,979.34	36,464.22
2016	Services	Services	Others	133,398.19	123,400.40	9,997.79
2016	Food II	Stable Pet Products	Pets	597,681.71	523,022.49	74,659.22
2016	Non Food	Other Stable Products	Pharmacy	100,757.79	88,165.73	12,592.06
2016	Non Food	Other Stable Products	Cosmetics	604,204.35	528,705.71	75,498.64
2016	Non Food I	Clothing	Clothing	15,435,091.50	10,341,518.37	5,093,573.13
2016	Non Food II	Electronics	Smartphones Windows	30,598,696.50	20,501,133.13	10,097,563.37
2016	Non Food II	Electronics	Smartphones Android	353,354,596.00	236,747,655.21	116,606,940.79
2016	Non Food II	Electronics	Tablet	78,323,796.50	52,476,964.76	25,846,831.74
2016	Non Food II	Electronics	Smartphone	583,118,971.00	390,689,835.91	192,429,135.09

SQL Query Execution Time for language_files is 7.6936609745026
WEB Application Execution Time for language_files is 7.6990039348602

Figure E-10: Initial BI report based on FILES approach of data mart implementation

E.4 Implementation of ETL module in the MCMS

Implementation of the ETL module was not a required element of the MLED_BI validation but was developed to illustrate the type of flexibility required by end users which can be provided by a content management system. The ETL module supported ETL processes based on all four data mart implementation approaches. After clicking the "Export Data" link in the main menu, the user can access a simple interface enabling extract and transform activities as a part of ETL. (Figure E-11).

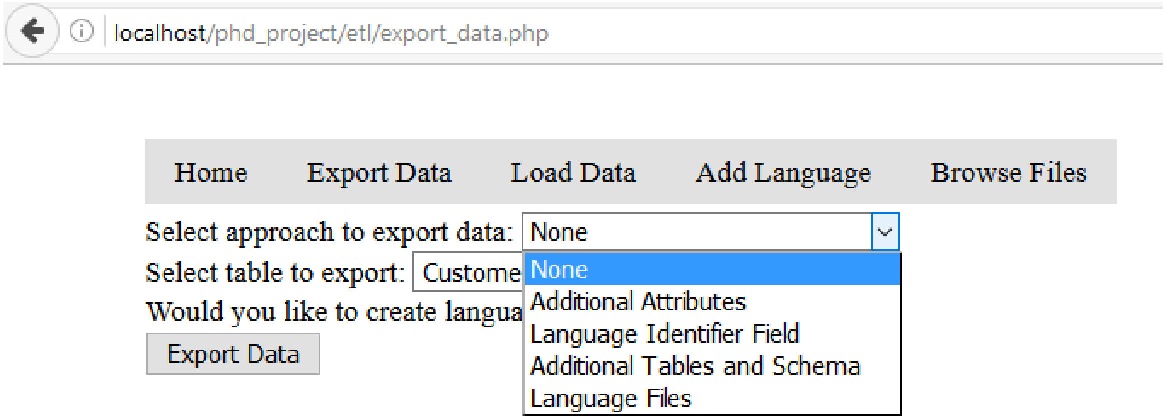


Figure E-11: A part of ETL Backend module that enables extract and transform activities

The business user is able to select appropriate approach for extract and transform, which extract and transform the data from source system (Figure E-12); to select a table (dimension) if needed (Figure E-13); and to select possibility to create/recreate language files (Figure E-14) if a data mart based on MLED_BI star schema was intended as the final destination of extracted and transformed data.

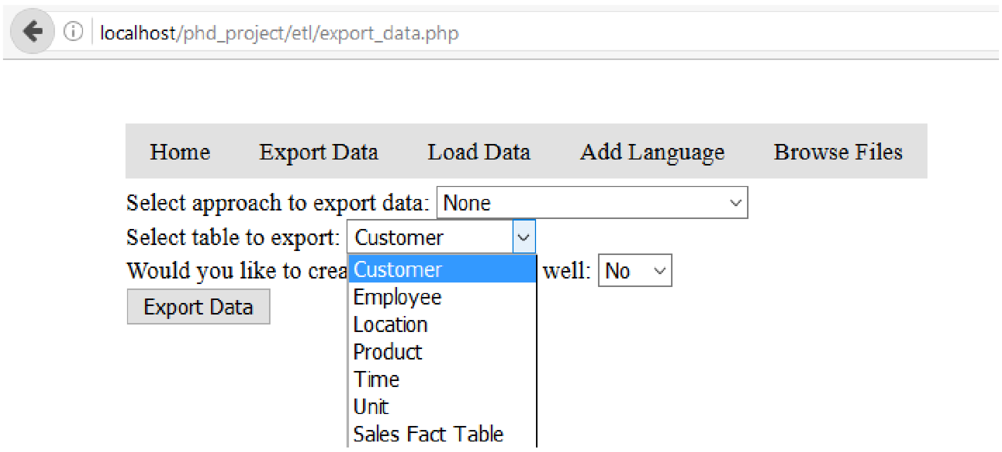


Figure E-12: A dimension selection possiblity

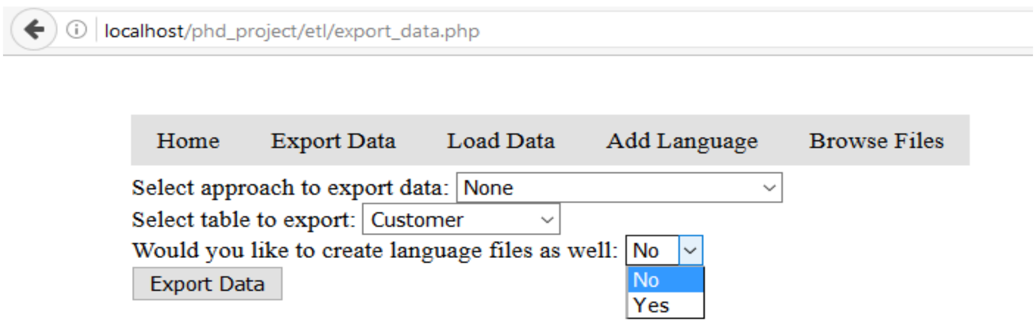


Figure E-13: Creating the languages files possibility

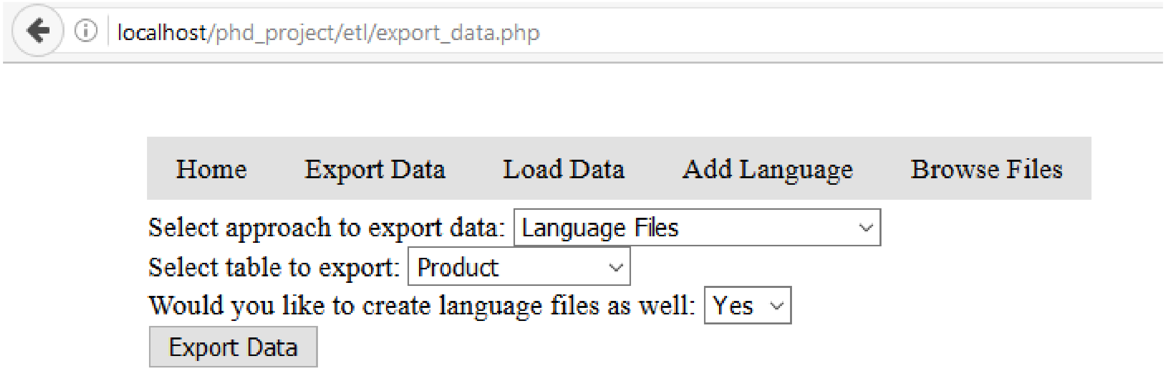


Figure E-14: An example of selection to extract and transform the data

After successful execution (Figure E-15), the data extraction processes creates appropriate files to support further operations (Figure E-16). A file contain dimension identifiers would be loaded into appropriate table, and language files would be moved to the language file folder.

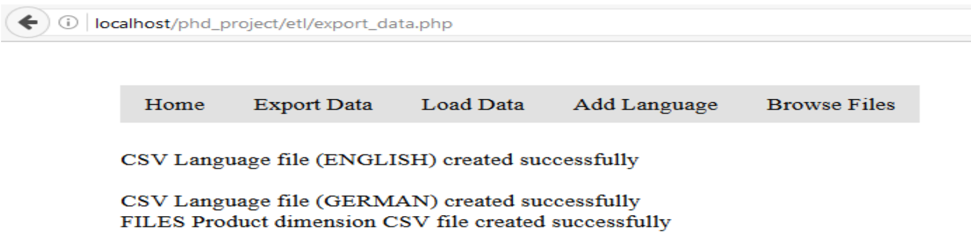


Figure E-15: Information about successful execution of the extract and transform process

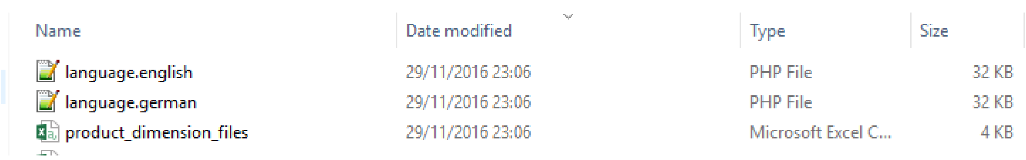
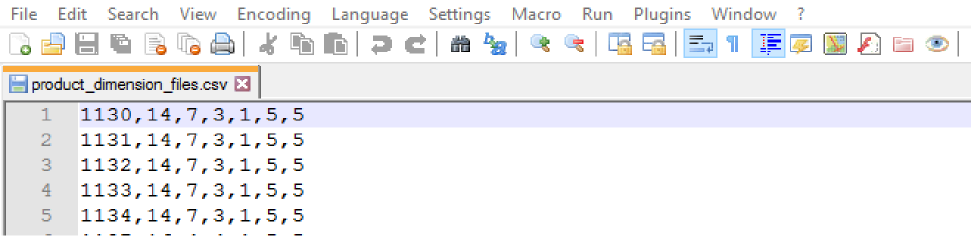


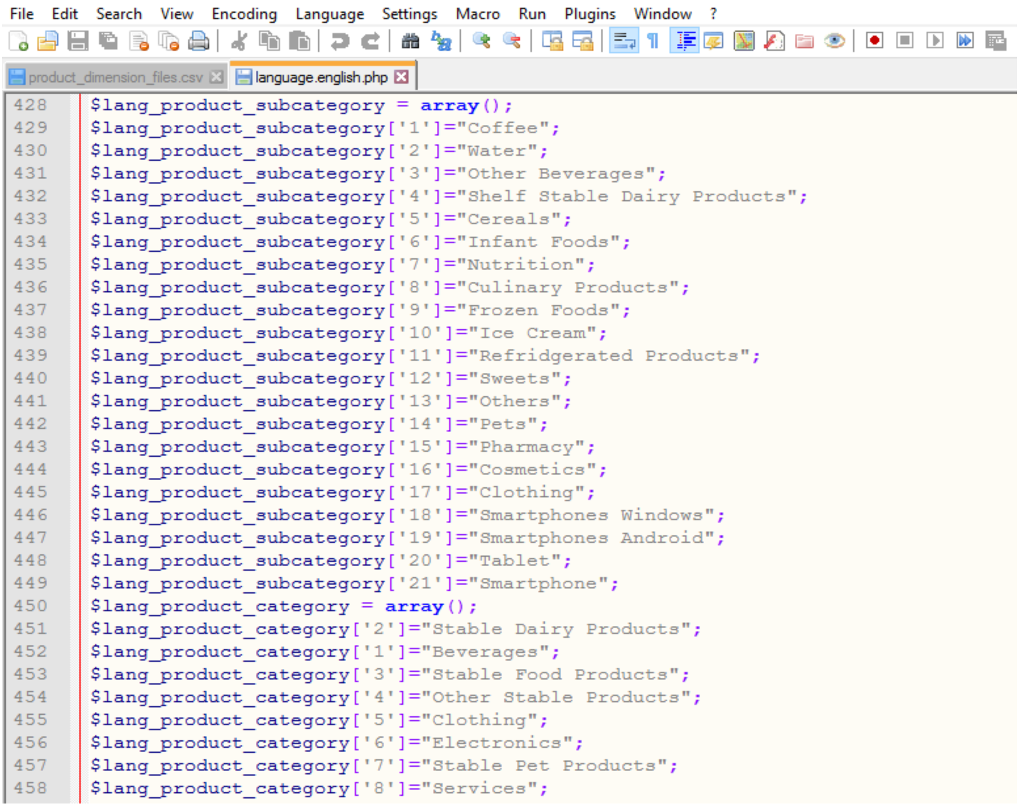
Figure E-16: Files created during executing sample process of extraction and transformation

Figure E-17, Figure E-18, and Figure E-19 shows the actual structure of exported files. The Product dimension files (Figure E-17) contains only identifiers for dimension, while English (Figure E-18) and German (Figure E-19) language files contain actual descriptions.



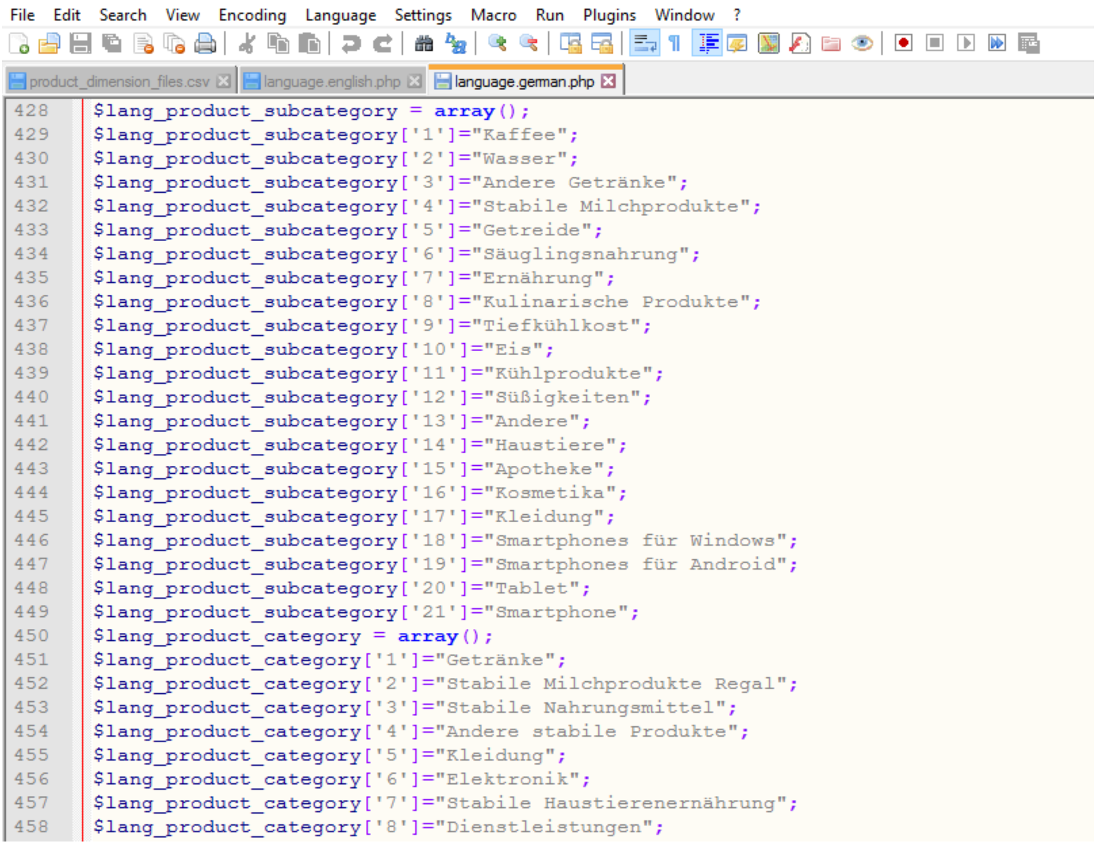

```
1130,14,7,3,1,5,5
1131,14,7,3,1,5,5
1132,14,7,3,1,5,5
1133,14,7,3,1,5,5
1134,14,7,3,1,5,5
```

Figure E-17: Dimension file created as a product of sample extraction and transformation




```
$lang_product_subcategory = array();
$lang_product_subcategory['1']="Coffee";
$lang_product_subcategory['2']="Water";
$lang_product_subcategory['3']="Other Beverages";
$lang_product_subcategory['4']="Shelf Stable Dairy Products";
$lang_product_subcategory['5']="Cereals";
$lang_product_subcategory['6']="Infant Foods";
$lang_product_subcategory['7']="Nutrition";
$lang_product_subcategory['8']="Culinary Products";
$lang_product_subcategory['9']="Frozen Foods";
$lang_product_subcategory['10']="Ice Cream";
$lang_product_subcategory['11']="Refridgerated Products";
$lang_product_subcategory['12']="Sweets";
$lang_product_subcategory['13']="Others";
$lang_product_subcategory['14']="Pets";
$lang_product_subcategory['15']="Pharmacy";
$lang_product_subcategory['16']="Cosmetics";
$lang_product_subcategory['17']="Clothing";
$lang_product_subcategory['18']="Smartphones Windows";
$lang_product_subcategory['19']="Smartphones Android";
$lang_product_subcategory['20']="Tablet";
$lang_product_subcategory['21']="Smartphone";
$lang_product_category = array();
$lang_product_category['2']="Stable Dairy Products";
$lang_product_category['1']="Beverages";
$lang_product_category['3']="Stable Food Products";
$lang_product_category['4']="Other Stable Products";
$lang_product_category['5']="Clothing";
$lang_product_category['6']="Electronics";
$lang_product_category['7']="Stable Pet Products";
$lang_product_category['8']="Services";
```

Figure E-18: English language file created as a product of sample extraction and transformation




```
$lang_product_subcategory = array();
$lang_product_subcategory['1']="Kaffee";
$lang_product_subcategory['2']="Wasser";
$lang_product_subcategory['3']="Andere Getränke";
$lang_product_subcategory['4']="Stabile Milchprodukte";
$lang_product_subcategory['5']="Getreide";
$lang_product_subcategory['6']="Säuglingsnahrung";
$lang_product_subcategory['7']="Ernährung";
$lang_product_subcategory['8']="Kulinarische Produkte";
$lang_product_subcategory['9']="Tiefkühlkost";
$lang_product_subcategory['10']="Eis";
$lang_product_subcategory['11']="Kühlprodukte";
$lang_product_subcategory['12']="Süßigkeiten";
$lang_product_subcategory['13']="Andere";
$lang_product_subcategory['14']="Haustiere";
$lang_product_subcategory['15']="Apotheke";
$lang_product_subcategory['16']="Kosmetika";
$lang_product_subcategory['17']="Kleidung";
$lang_product_subcategory['18']="Smartphones für Windows";
$lang_product_subcategory['19']="Smartphones für Android";
$lang_product_subcategory['20']="Tablet";
$lang_product_subcategory['21']="Smartphone";
$lang_product_category = array();
$lang_product_category['1']="Getränke";
$lang_product_category['2']="Stabile Milchprodukte Regal";
$lang_product_category['3']="Stabile Nahrungsmittel";
$lang_product_category['4']="Andere stabile Produkte";
$lang_product_category['5']="Kleidung";
$lang_product_category['6']="Elektronik";
$lang_product_category['7']="Stabile Haustierenernährung";
$lang_product_category['8']="Dienstleistungen";
```

Figure E-19: German language file created as a product of sample extraction and transformation

In addition to extract and transform functionality, the MCSM ETL Backend module supported data loading into data marts (Figure E-20). When “Load Data” is selected in the main menu, an appropriate interface to perform load activities appears. Any previously source system extracted file that holds any kind of data (master of transactional) can be selected, as can any named DM based on any type of implementation approach, and any table (dimension or fact). An example of selecting a table of Customer dimension, based on AA DM implementation approach to load data in, was shown in Figure E-21. A message after successful loading process is shown in Figure E-22.

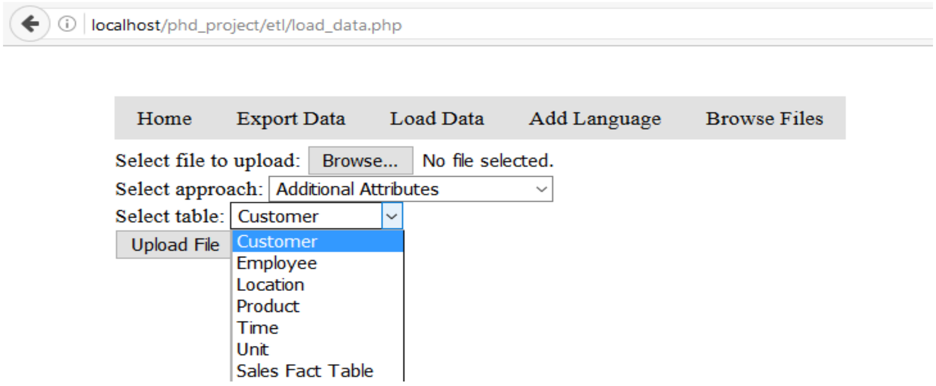


Figure E-20: Interface to perform data load activities as a part of ETL Backend module

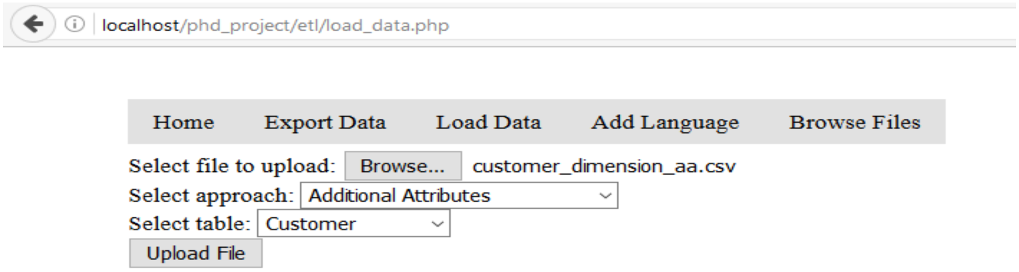


Figure E-21: An example of selecting customer dimension based on AA approach to load data in

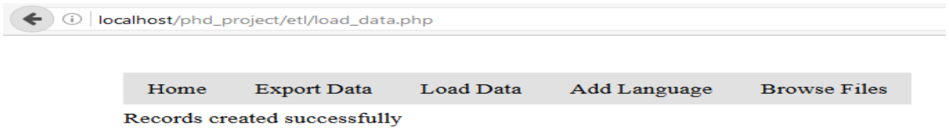


Figure E-22: Message after successful load process

E.4 Implementation of New Language module in the MCMS

The New Language module was implemented as a part of MCSM Reporting Layer based on MLED_BI in WE. The idea behind this module was to enable end business users to create new languages to be used in BI reports themselves. As explained in chapter 1, this functionality for end users is not supported by existing ML BI design approaches.

As shown in Figure E-23, there is an "Add Language" link in main menu which provides the interface to add new languages to be used for BI reports getting data from DM based on FILES approach. An existing language can be used as a template for the new language (for example, to support dialects). The example used here, is that the German language is employed as a template to create a fictitious Austrian language (Figure E-23 and Figure E-24), which is later modified according to the needs of end users. Once created (Figure E-25), the new language file, in this case for an Austrian language, was added to the same folder with existing files (Figure E-26). Except the different name, its content and structure was completely the same as German language file (Figure E-27).

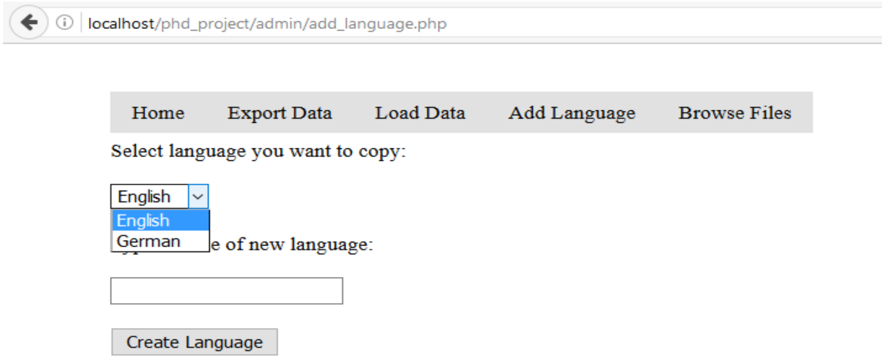


Figure E-23: Initial interface enabling adding of a new language for BI reports

Figure E-24: Creating Austrian from German language

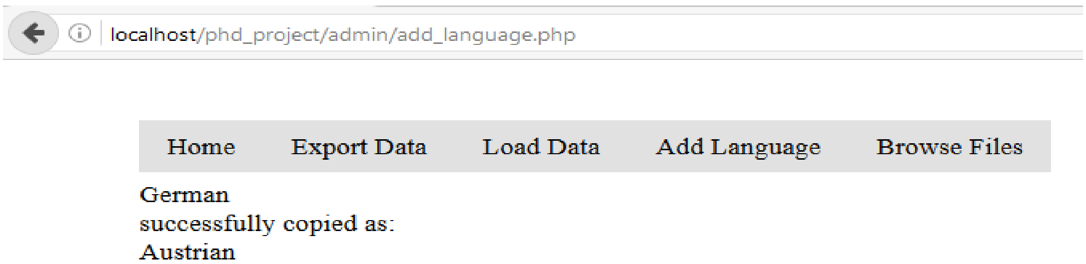


Figure E-25: Message about successful creation of Austrian from German Language

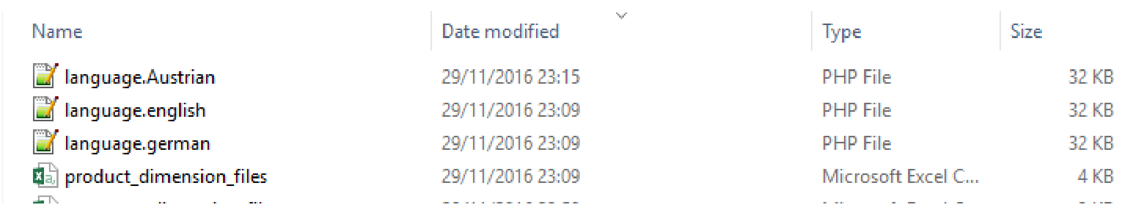
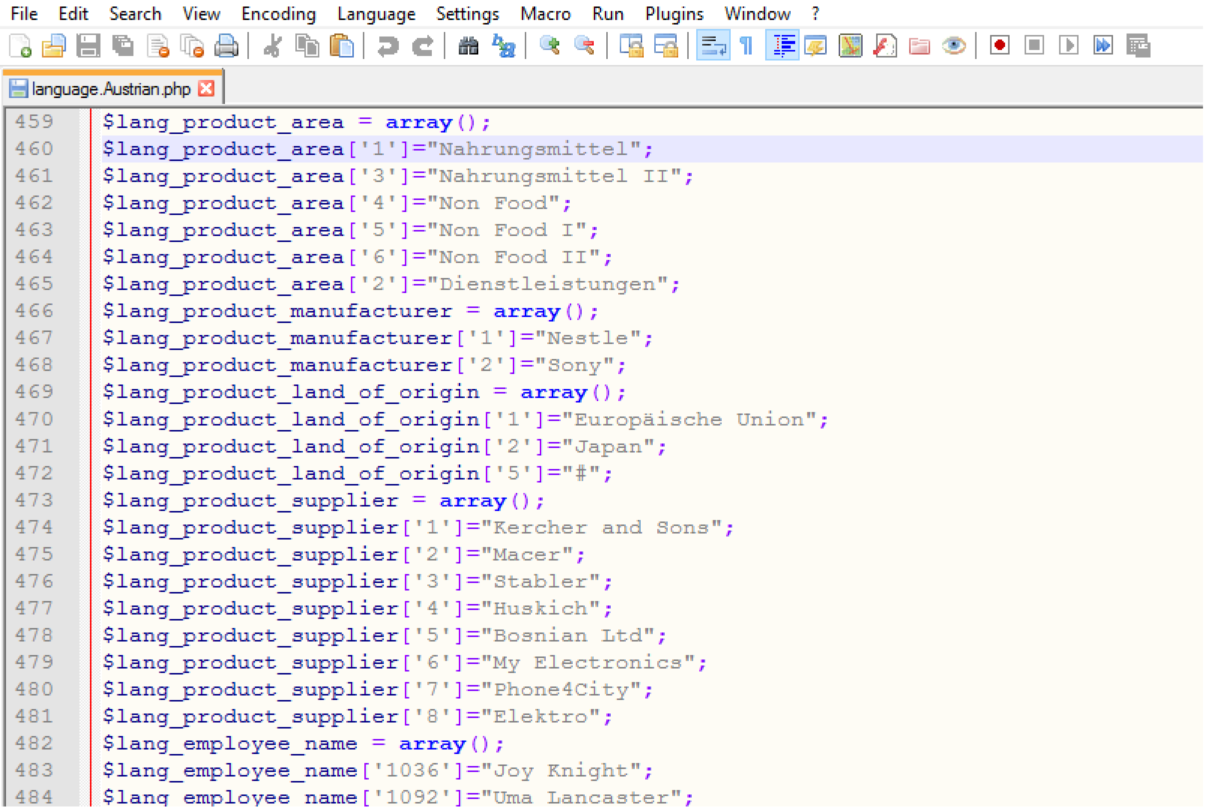


Figure E-26: The new language file was created for Austrian language




```php
$lang_product_area = array();
$lang_product_area['1']="Nahrungsmittel";
$lang_product_area['3']="Nahrungsmittel II";
$lang_product_area['4']="Non Food";
$lang_product_area['5']="Non Food I";
$lang_product_area['6']="Non Food II";
$lang_product_area['2']="Dienstleistungen";
$lang_product_manufacturer = array();
$lang_product_manufacturer['1']="Nestle";
$lang_product_manufacturer['2']="Sony";
$lang_product_land_of_origin = array();
$lang_product_land_of_origin['1']="Europäische Union";
$lang_product_land_of_origin['2']="Japan";
$lang_product_land_of_origin['5']="#";
$lang_product_supplier = array();
$lang_product_supplier['1']="Kercher and Sons";
$lang_product_supplier['2']="Macer";
$lang_product_supplier['3']="Stabler";
$lang_product_supplier['4']="Huskich";
$lang_product_supplier['5']="Bosnian Ltd";
$lang_product_supplier['6']="My Electronics";
$lang_product_supplier['7']="Phone4City";
$lang_product_supplier['8']="Elektro";
$lang_employee_name = array();
$lang_employee_name['1036']="Joy Knight";
$lang_employee_name['1092']="Uma Lancaster";
```

Figure E-27: Actual screenshot of the language file holding descriptions in Austrian

A new menu link for Austrian language was provided in BI report and it could be used immediately (Top right corner of Figure E-28). However, as the fictitious language was copied from German it provided same descriptions.

Austrian English German

Home Export Data Load Data Add Language Browse Files

Please select approach: Language Files Submit Query
You selected:

Language Files

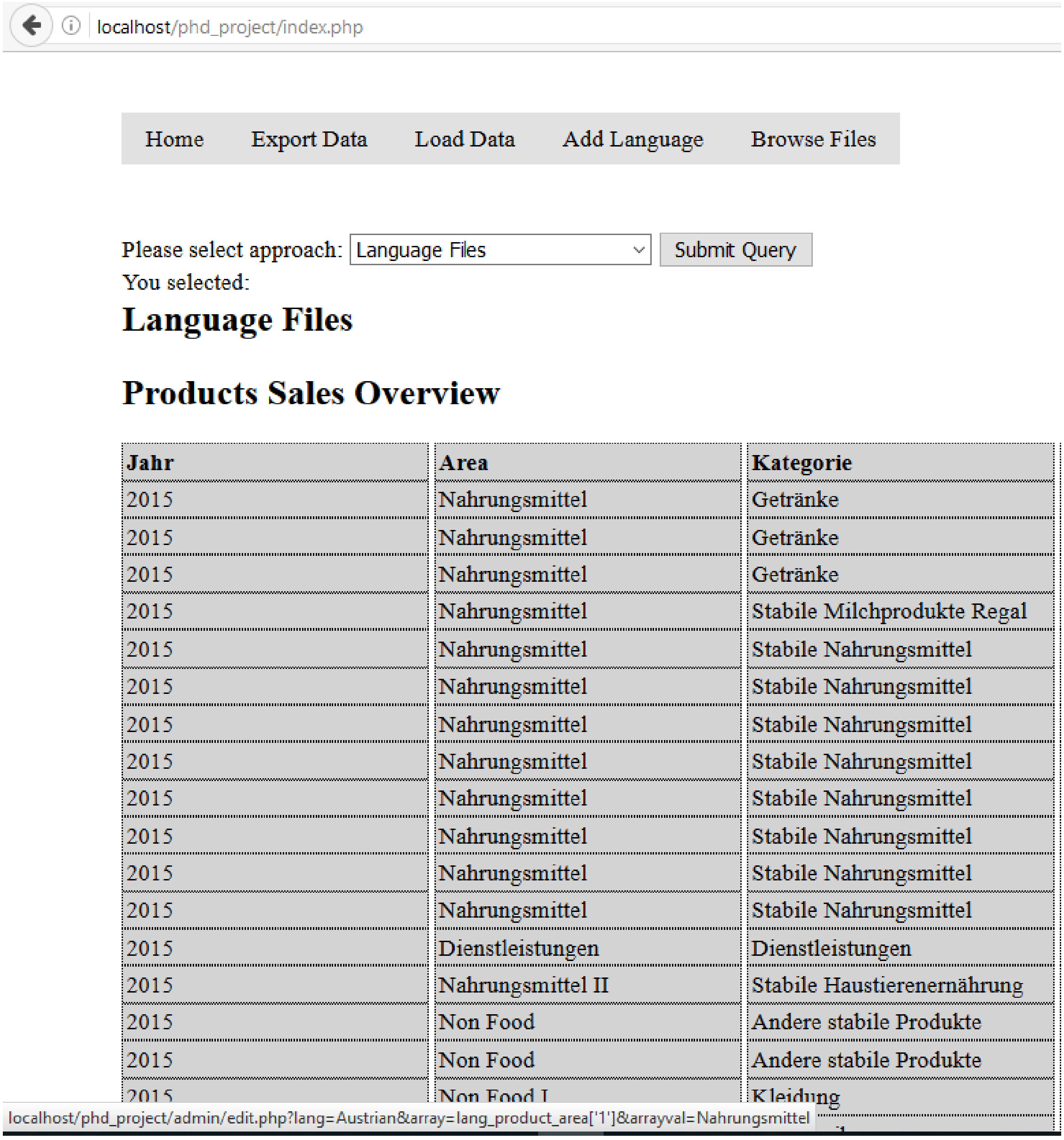
Products Sales Overview

Jahr	Area	Kategorie	Sub-Kategorie	Umsatz Brutto	Umatz Netto	Profit
2015	Nahrungsmittel	Getränke	Kaffee	234,760.66	217,149.61	17,611.05
2015	Nahrungsmittel	Getränke	Wasser	688,013.23	636,421.07	51,592.16
2015	Nahrungsmittel	Getränke	Andere Getränke	271,004.20	250,690.72	20,313.48
2015	Nahrungsmittel	Stabile Milchprodukte Regal	Stabile Milchprodukte	621,596.83	574,979.87	46,616.96
2015	Nahrungsmittel	Stabile Nahrungsmittel	Getreide	62,392.50	57,714.10	4,678.40
2015	Nahrungsmittel	Stabile Nahrungsmittel	Säuglingsnahrung	533,165.17	493,168.45	39,996.72
2015	Nahrungsmittel	Stabile Nahrungsmittel	Ernährung	125,776.80	116,343.54	9,433.26
2015	Nahrungsmittel	Stabile Nahrungsmittel	Kulinarische Produkte	127,905.14	118,316.00	9,589.14
2015	Nahrungsmittel	Stabile Nahrungsmittel	Tiefkühlkost	129,133.62	119,453.99	9,679.63
2015	Nahrungsmittel	Stabile Nahrungsmittel	Eis	258,410.42	239,038.16	19,372.26
2015	Nahrungsmittel	Stabile Nahrungsmittel	Kühlprodukte	126,364.27	116,890.20	9,474.07
2015	Nahrungsmittel	Stabile Nahrungsmittel	Süßigkeiten	488,282.01	451,680.66	36,601.35
2015	Dienstleistungen	Dienstleistungen	Andere	131,722.48	121,851.17	9,871.31
2015	Nahrungsmittel II	Stabile Haustierenernährung	Haustiere	587,361.36	513,991.44	73,369.92
2015	Non Food	Andere stabile Produkte	Apotheke	101,751.93	89,035.57	12,716.36
2015	Non Food	Andere stabile Produkte	Kosmetika	599,662.26	524,730.79	74,931.47
2015	Non Food I	Kleidung	Kleidung	14,840,484.00	9,943,130.98	4,897,353.02
2015	Non Food II	Elektronik	Smartphones für Windows	30,336,996.50	20,325,794.69	10,011,201.81
2015	Non Food II	Elektronik	Smartphones für Android	355,256,006.50	238,021,601.14	117,234,405.36
2015	Non Food II	Elektronik	Tablet	78,195,639.50	52,391,099.37	25,804,540.13
2015	Non Food II	Elektronik	Smartphone	584,224,792.50	391,430,736.04	192,794,056.46
2016	Nahrungsmittel	Getränke	Kaffee	234,823.98	217,209.12	17,614.86
2016	Nahrungsmittel	Getränke	Wasser	686,898.42	635,388.40	51,510.02
2016	Nahrungsmittel	Getränke	Andere Getränke	274,105.73	253,559.88	20,545.85
2016	Nahrungsmittel	Stabile Milchprodukte Regal	Stabile Milchprodukte	611,563.92	565,700.18	45,863.74
2016	Nahrungsmittel	Stabile Nahrungsmittel	Getreide	62,681.60	57,981.19	4,700.41
2016	Nahrungsmittel	Stabile Nahrungsmittel	Säuglingsnahrung	535,587.07	495,409.03	40,178.04
2016	Nahrungsmittel	Stabile Nahrungsmittel	Ernährung	127,327.20	117,777.66	9,549.54
2016	Nahrungsmittel	Stabile Nahrungsmittel	Kulinarische Produkte	128,547.00	118,909.77	9,637.23
2016	Nahrungsmittel	Stabile Nahrungsmittel	Tiefkühlkost	129,288.50	119,597.45	9,691.05
2016	Nahrungsmittel	Stabile Nahrungsmittel	Eis	258,262.92	238,901.96	19,360.96
2016	Nahrungsmittel	Stabile Nahrungsmittel	Kühlprodukte	130,209.75	120,448.12	9,761.63
2016	Nahrungsmittel	Stabile Nahrungsmittel	Süßigkeiten	486,443.56	449,979.34	36,464.22
2016	Dienstleistungen	Dienstleistungen	Andere	133,398.19	123,400.40	9,997.79
2016	Nahrungsmittel II	Stabile Haustierenernährung	Haustiere	597,681.71	523,022.49	74,659.22
2016	Non Food	Andere stabile Produkte	Apotheke	100,757.79	88,165.73	12,592.06
2016	Non Food	Andere stabile Produkte	Kosmetika	604,204.35	528,705.71	75,498.64
2016	Non Food I	Kleidung	Kleidung	15,435,091.50	10,341,518.37	5,093,573.13
2016	Non Food II	Elektronik	Smartphones für Windows	30,598,696.50	20,501,133.13	10,097,563.37
2016	Non Food II	Elektronik	Smartphones für Android	353,354,596.00	236,747,655.21	116,606,940.79
2016	Non Food II	Elektronik	Tablet	78,323,796.50	52,476,964.76	25,846,831.74
2016	Non Food II	Elektronik	Smartphone	583,118,971.00	390,689,835.91	192,429,135.09

WEB Application Execution Time for language_files is 0.0019330978393555

Figure E-28: Example of the automatically generated menu link for Austrian language

Having a fictional Austrian language created by copying German language made it immediately possible to use the same BI report in the newly created language. To enable different descriptions for business content in BI report for Austrian in regard to German language, a business user could perform translations in two ways: direct change via BI report or a by changing language file for Austrian language at local server. Figure E-29 provides a screenshot of a part of actual BI report browsed in Austrian language. This BI report offered clickable descriptions of business content, which when clicked lead to the page that enables its change (Figure E-30). As soon as a new value for respective description of a business content was provided in appropriate text field, "Change Value" link was clicked, and WE returned a message about successful change (Figure E-31), a new translation or content change was visible in exiting BI report (Figure E-32). There was no need to re-execute underlying query for the existing report to load the new language. As the WE in MLED_BI design approach loads only the content of the language file, it would be sufficient to click on the same language once again and the change would be immediately visible. Change is also immediately visible in language file having Austrian business information descriptions (Figure E-33). Changes to descriptions of business content could be done directly by modifying this file as well.


localhost/phd_project/index.php

Home Export Data Load Data Add Language Browse Files

Please select approach: Language Files Submit Query

You selected:

Language Files

Products Sales Overview

Jahr	Area	Kategorie
2015	Nahrungsmittel	Getränke
2015	Nahrungsmittel	Getränke
2015	Nahrungsmittel	Getränke
2015	Nahrungsmittel	Stabile Milchprodukte Regal
2015	Nahrungsmittel	Stabile Nahrungsmittel
2015	Nahrungsmittel	Stabile Nahrungsmittel
2015	Nahrungsmittel	Stabile Nahrungsmittel
2015	Nahrungsmittel	Stabile Nahrungsmittel
2015	Nahrungsmittel	Stabile Nahrungsmittel
2015	Nahrungsmittel	Stabile Nahrungsmittel
2015	Nahrungsmittel	Stabile Nahrungsmittel
2015	Nahrungsmittel	Stabile Nahrungsmittel
2015	Dienstleistungen	Dienstleistungen
2015	Nahrungsmittel II	Stabile Haustierenernährung
2015	Non Food	Andere stabile Produkte
2015	Non Food	Andere stabile Produkte
2015	Non Food I	Kleidung

localhost/phd_project/admin/edit.php?lang=Austrian&array=lang_product_area['1']&arrayval=Nahrungsmittel

Figure E-29: A part of the BI report with clickable descriptions leading to editing page

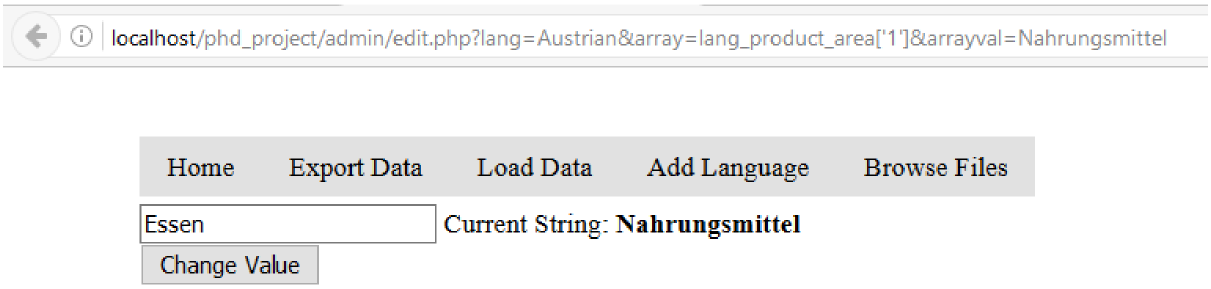


Figure E-30: Description editing interface

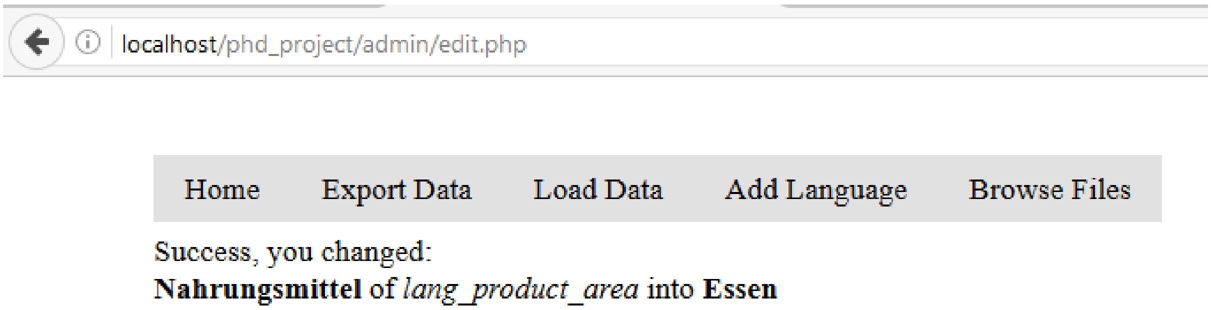


Figure E-31: Message after successful change of business description

Austrian English German

Home Export Data Load Data Add Language Browse Files

Please select approach: Language Files Submit Query

You selected:

Language Files

Products Sales Overview

Jahr	Area	Kategorie	Sub-Kategorie	Umsatz Brutto	Umatz Netto	Profit
2015	Essen	Getränke	Kaffee	234,760.66	217,149.61	17,611.05
2015	Essen	Getränke	Wasser	688,013.23	636,421.07	51,592.16
2015	Essen	Getränke	Andere Getränke	271,004.20	250,690.72	20,313.48
2015	Essen	Stabile Milchprodukte Regal	Stabile Milchprodukte	621,596.83	574,979.87	46,616.96
2015	Essen	Stabile Nahrungsmittel	Getreide	62,392.50	57,714.10	4,678.40
2015	Essen	Stabile Nahrungsmittel	Säuglingsnahrung	533,165.17	493,168.45	39,996.72
2015	Essen	Stabile Nahrungsmittel	Ernährung	125,776.80	116,343.54	9,433.26
2015	Essen	Stabile Nahrungsmittel	Kulinarische Produkte	127,905.14	118,316.00	9,589.14
2015	Essen	Stabile Nahrungsmittel	Tiefkühlkost	129,133.62	119,453.99	9,679.63
2015	Essen	Stabile Nahrungsmittel	Eis	258,410.42	239,038.16	19,372.26
2015	Essen	Stabile Nahrungsmittel	Kühlprodukte	126,364.27	116,890.20	9,474.07
2015	Essen	Stabile Nahrungsmittel	Süßigkeiten	488,282.01	451,680.66	36,601.35
2015	Dienstleistungen	Dienstleistungen	Andere	131,722.48	121,851.17	9,871.31
2015	Nahrungsmittel II	Stabile Haustierenernährung	Haustiere	587,361.36	513,991.44	73,369.92
2015	Non Food	Andere stabile Produkte	Apotheke	101,751.93	89,035.57	12,716.36
2015	Non Food	Andere stabile Produkte	Kosmetika	599,662.26	524,730.79	74,931.47
2015	Non Food I	Kleidung	Kleidung	14,840,484.00	9,943,130.98	4,897,353.02
2015	Non Food II	Elektronik	Smartphones für Windows	30,336,996.50	20,325,794.69	10,011,201.81
2015	Non Food II	Elektronik	Smartphones für Android	355,256,006.50	238,021,601.14	117,234,405.36
2015	Non Food II	Elektronik	Tablet	78,195,639.50	52,391,099.37	25,804,540.13
2015	Non Food II	Elektronik	Smartphone	584,224,792.50	391,430,736.04	192,794,056.46
2016	Essen	Getränke	Kaffee	234,823.98	217,209.12	17,614.86
2016	Essen	Getränke	Wasser	686,898.42	635,388.40	51,510.02
2016	Essen	Getränke	Andere Getränke	274,105.73	253,559.88	20,545.85
2016	Essen	Stabile Milchprodukte Regal	Stabile Milchprodukte	611,563.92	565,700.18	45,863.74
2016	Essen	Stabile Nahrungsmittel	Getreide	62,681.60	57,981.19	4,700.41
2016	Essen	Stabile Nahrungsmittel	Säuglingsnahrung	535,587.07	495,409.03	40,178.04
2016	Essen	Stabile Nahrungsmittel	Ernährung	127,327.20	117,777.66	9,549.54
2016	Essen	Stabile Nahrungsmittel	Kulinarische Produkte	128,547.00	118,909.77	9,637.23
2016	Essen	Stabile Nahrungsmittel	Tiefkühlkost	129,288.50	119,597.45	9,691.05
2016	Essen	Stabile Nahrungsmittel	Eis	258,262.92	238,901.96	19,360.96
2016	Essen	Stabile Nahrungsmittel	Kühlprodukte	130,209.75	120,448.12	9,761.63
2016	Essen	Stabile Nahrungsmittel	Süßigkeiten	486,443.56	449,979.34	36,464.22
2016	Dienstleistungen	Dienstleistungen	Andere	133,398.19	123,400.40	9,997.79
2016	Nahrungsmittel II	Stabile Haustierenernährung	Haustiere	597,681.71	523,022.49	74,659.22
2016	Non Food	Andere stabile Produkte	Apotheke	100,757.79	88,165.73	12,592.06
2016	Non Food	Andere stabile Produkte	Kosmetika	604,204.35	528,705.71	75,498.64
2016	Non Food I	Kleidung	Kleidung	15,435,091.50	10,341,518.37	5,093,573.13
2016	Non Food II	Elektronik	Smartphones für Windows	30,598,696.50	20,501,133.13	10,097,563.37
2016	Non Food II	Elektronik	Smartphones für Android	353,354,596.00	236,747,655.21	116,606,940.79
2016	Non Food II	Elektronik	Tablet	78,323,796.50	52,476,964.76	25,846,831.74
2016	Non Food II	Elektronik	Smartphone	583,118,971.00	390,689,835.91	192,429,135.09

WEB Application Execution Time for language_files is 0.001939058303833

Figure E-32: BI report with immediately changed business descriptions

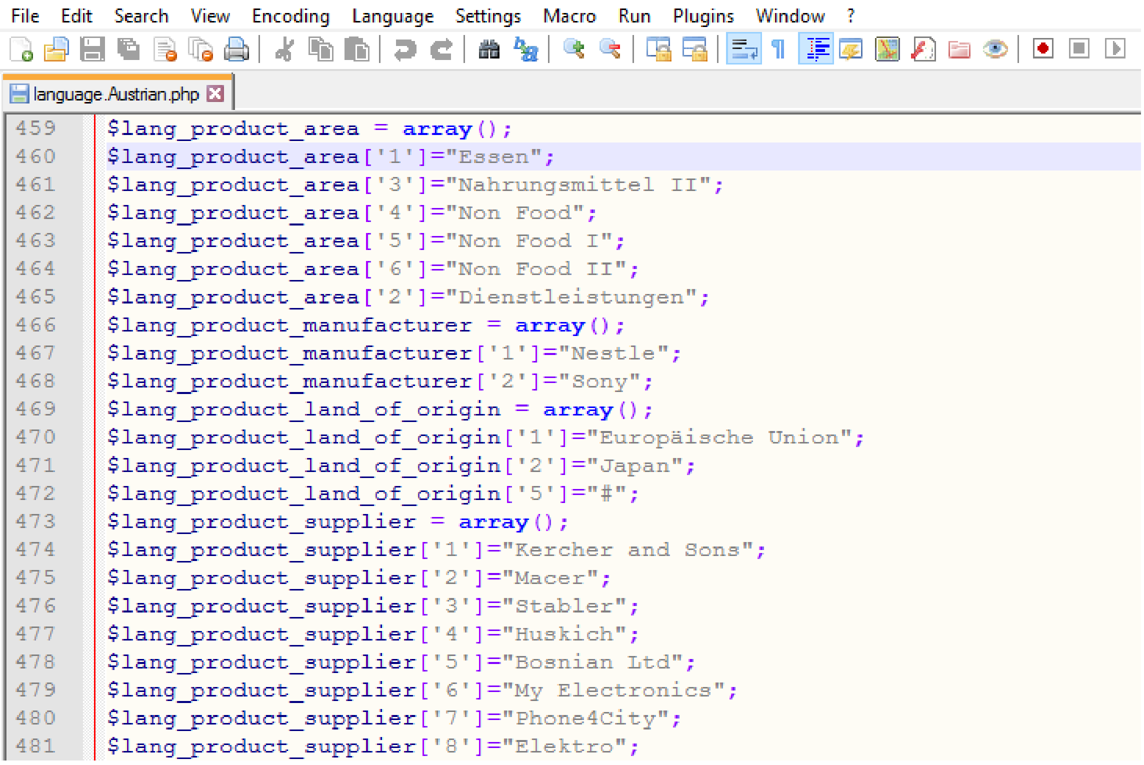


Figure E-33: A screenshot of actual Austrian language file with changed description

APPENDIX F. Larger versions of diagrams

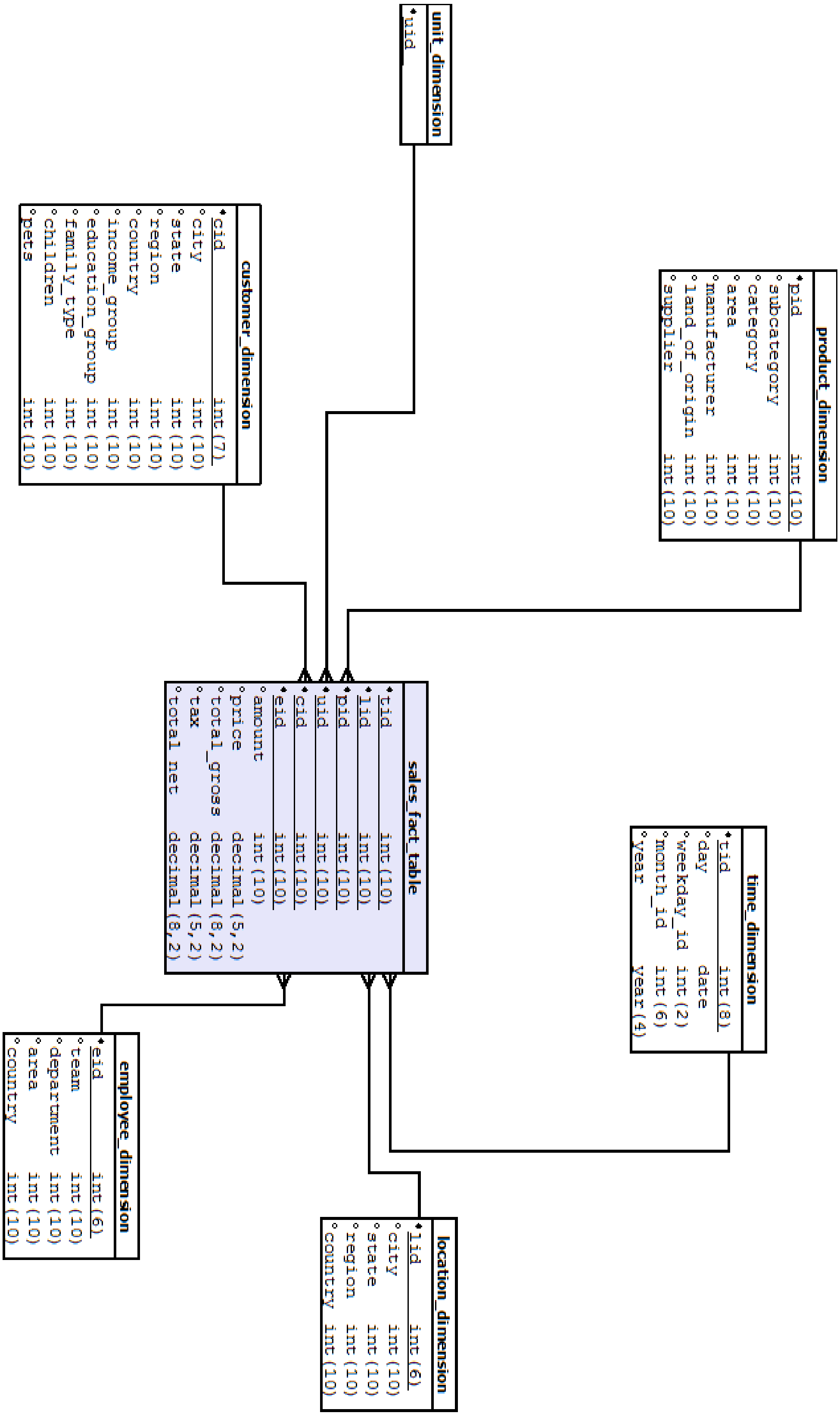


Figure F-1: DM Star Schema based on MLED_BI approach

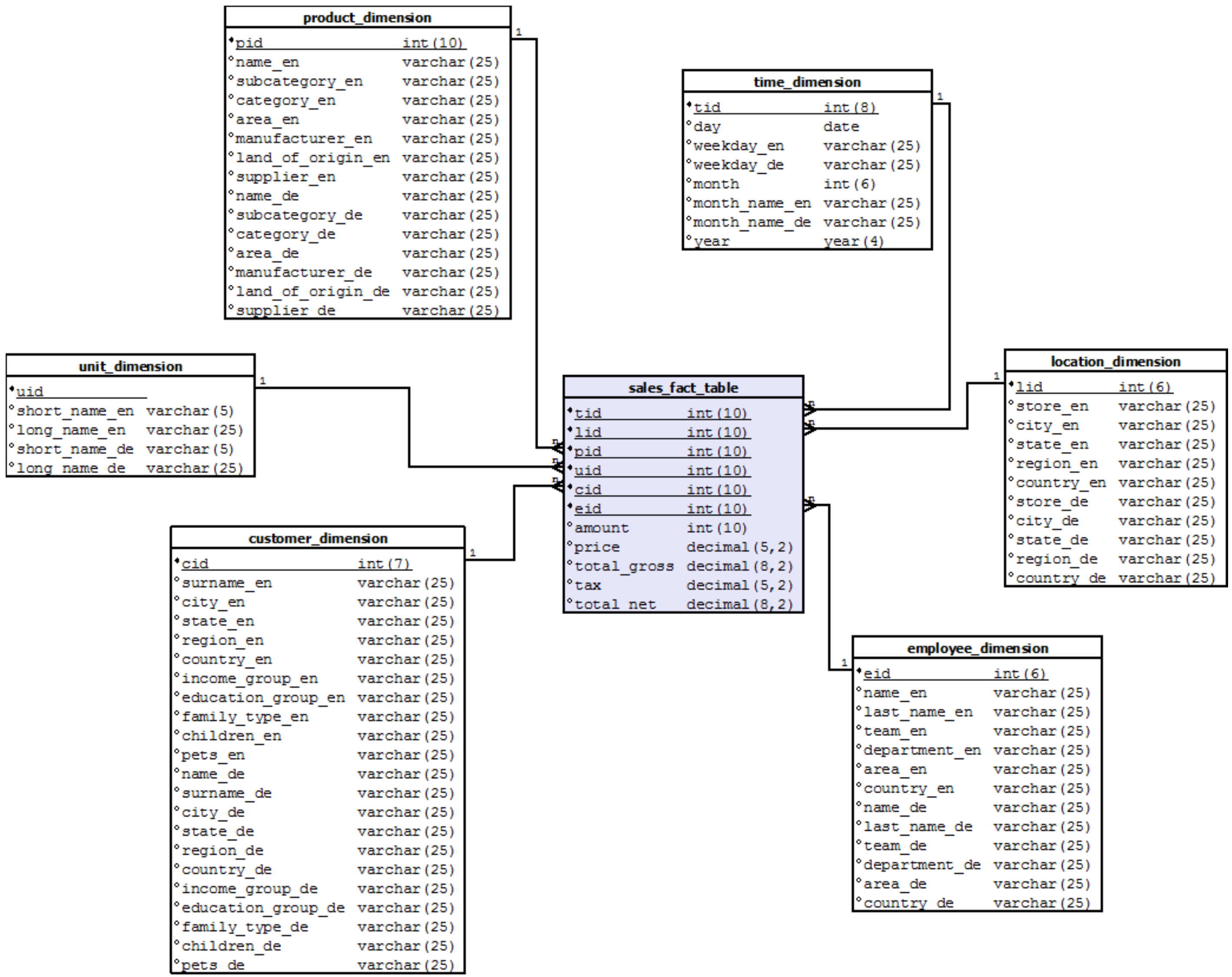


Figure F-2: DM star schema based on the AA approach

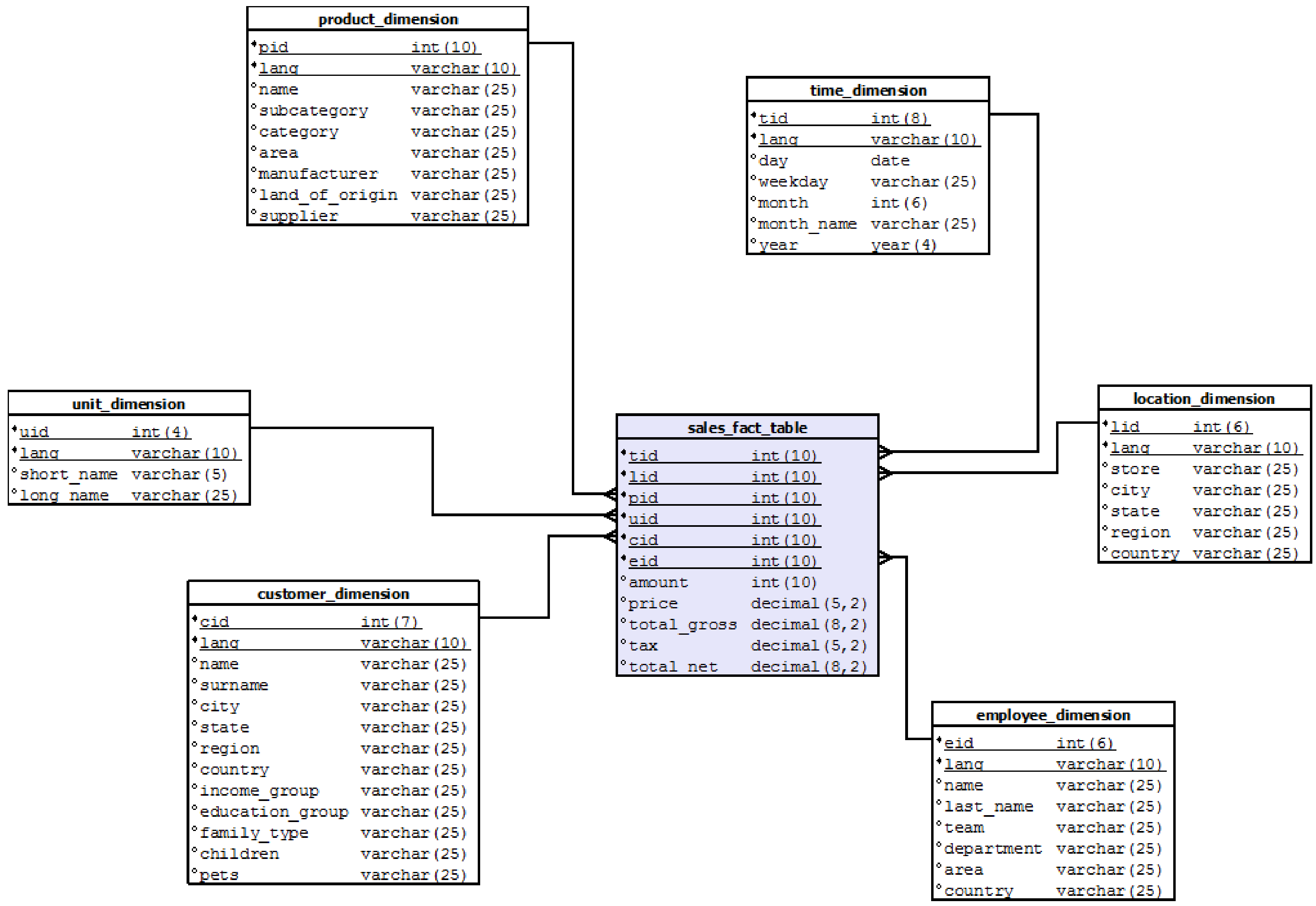


Figure F-3: DM Star Schema based on the LIF approach

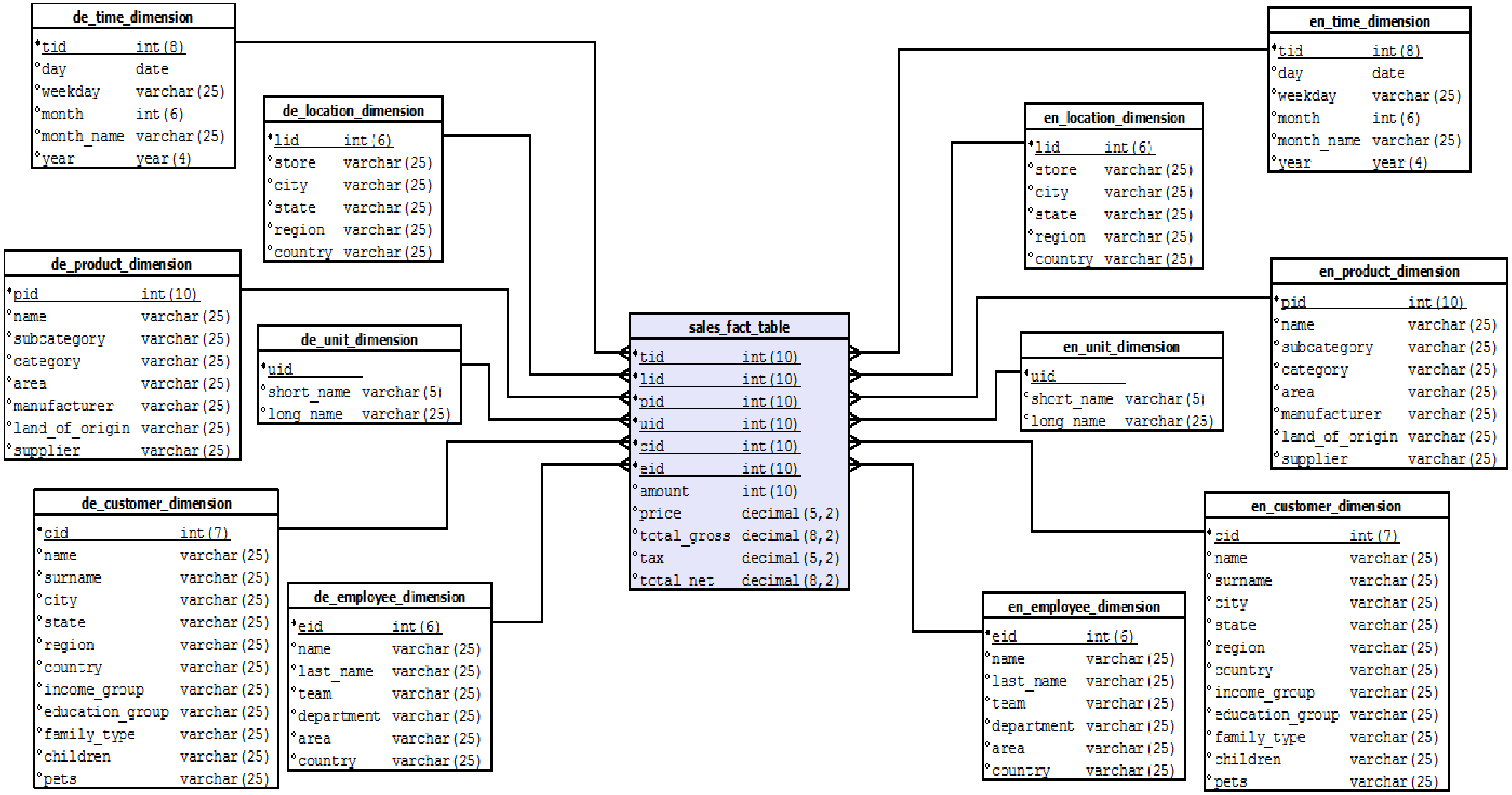


Figure F-4: DM Star Schema based on ATS approach

APPENDIX G. Architecture of the MCMS web environment

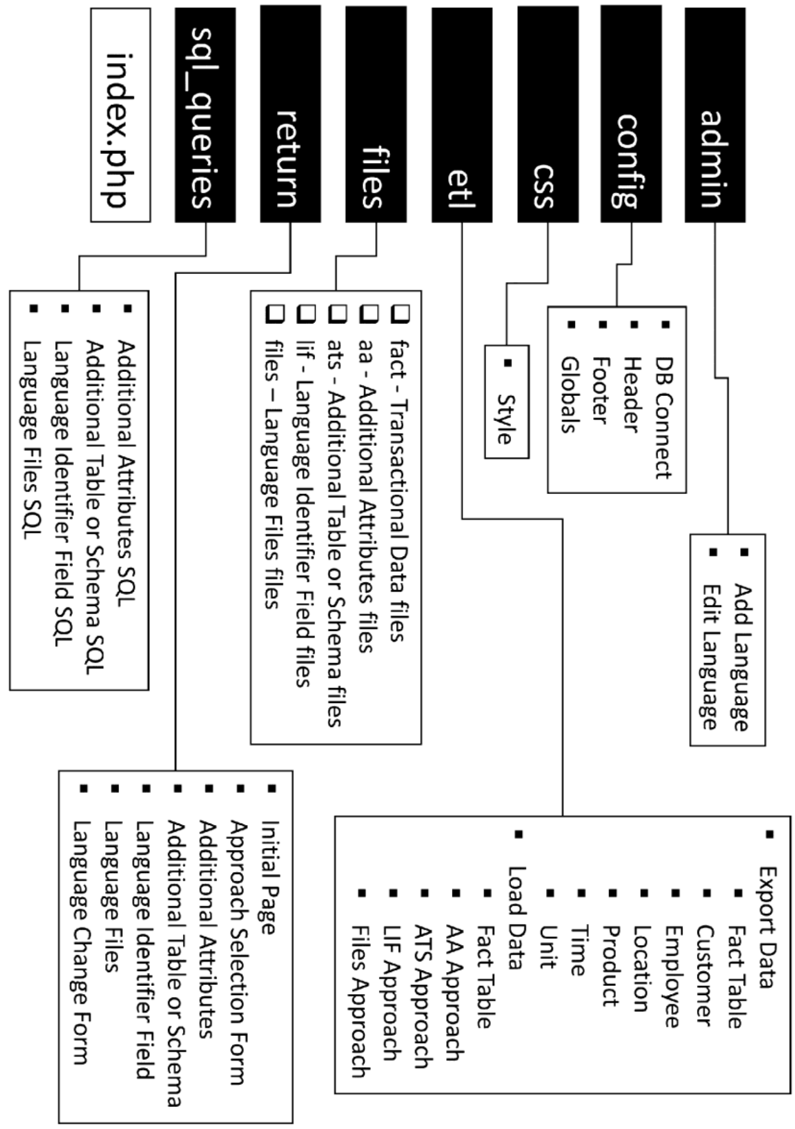


Figure G-1: Architecture of the MCMS web environment

APPENDIX H. Additional modules demonstrating optional functionalities of MCMS based on MLED_BI

- **ETL additional module**

Using the MCMS ETL module interfaces, data were extracted for each entity or file and loaded into appropriate DMs, or in the case of language files moved to appropriate folder at local server. As there are no structural changes within transactional data, every fact entity (sales_fact_entity) in every data mart holds same amount data. Further details of the ETL additional module are given in APPENDIX E, section E.4.

- **New Language additional module**

The rationale for the New Language module was to enable business end users to create new languages to be used in BI reports by themselves. As discussed in section 7.4.5., allowing end users to add new languages is not supported in existing reporting layers for ML approaches and for this reason the Additional Language module was implemented only for the MLED_BI design approach. To demonstrate the approach, the German language was used as a template to create a fictional Austrian language which could then be modified as required by the end user. As the WE in the MCMS loads only the content of the language file, it would be sufficient to click on the language added and the change would be immediately visible. Changes to descriptions of business content could also be implemented directly by modifying this file. This approach means that it is possible to immediate use the same BI report in the newly created language and there was no requirement even to re-execute the underlying query for the existing report to load the new language. Further details of the New Language modules are given in APPENDIX E, section E.4.

APPENDIX I. Evaluation Questionnaire

Staffordshire University, Faculty of Computing, Engineering and Sciences
College Rd, Stoke-on-Trent ST4 2DE, United Kingdom

Evaluation of application of Multilingualism in Business Intelligence

Aim: The main aim of this demonstration is to evaluate application of multilingualism in Business Intelligence (BI) reports based on different BI design approaches. Under phrase "application of multilingualism", we understand activities such as changing language of already executed report, making corrections to erroneous content or enabling new languages for reports.

Demonstration considers two different BI design approaches, which cover four different methods to enable multilingualism in Business Intelligence environment:

1) Conventional Business Intelligence Design approach
    a. Design approach based on additional attributes in data marts dimensional tables
    b. Design approach based on language identifier field in data marts dimensional tables
    c. Design approach based on additional schema/tables for dimensions
2) Newly proposed Multilingual Business Intelligence Design approach (MLED_BI)
    a. Design approach based on language files as a part of data mart concept, plus addition of MCMS (multilingual content management system)

Scenario: As a business key user, you want to browse a Business Intelligence based report that provides products sales overview. This report should provide overview of sales per *year*, *product area*, *category* and *subcategory* and include *gross sales*, *net sales* and *profit* as appropriate metrics. All reports, you visit as a part of this demonstration, should provide same data based on same source systems; however, their implementations are based on different design approach philosophy. The first three approaches have different design philosophies only in regard to data marts, while fourth applies different design approach to whole Business Intelligence concept.

Your task is to test every approach concerning "application of multilingualism" according to moderatos instructions and give your opinion by filling in questionnaire and giving your feedback.

Please check appropriate box under each question:

1. The information provided in the reports is accurate?

1.a.	**1.b.**	**1.c.**	**2.a.**
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No

2. Output is presented in a format that you find useful?

1.a.	**1.b.**	**1.c.**	**2.a.**
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No

3. The system and associated reports are easy for you to use?

1.a.	1.b.	1.c.	2.a.
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No

4. Information in the reports is up to date?

1.a.	1.b.	1.c.	2.a.
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No

5. Reports have the functionality that you require?

1.a.	1.b.	1.c.	2.a.
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No

6. The BI system is flexible enough to support easy change of „descriptive content"?

1.a.	1.b.	1.c.	2.a.
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No

7. Is the change of „descriptive content" fast enough to fulfil business requirement?

1.a.	1.b.	1.c.	2.a.
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No

8. *Please give your evaluation of the proposed MLED_BI approach with regard to functionality, performance, usability, fit to business requirements*:

9. *Please provide any other comments on the MLED_BI approach*:

Thank You

Participant consent form

STAFFORDSHIRE UNIVERSITY

Consent form for Evaluation of the MLED_BI (MultiLingual Enabled Design) approach to the application of Multilingualism in Business Intelligence

My name is Nedim Dedić and I am investigating issues which affect support for Multilingualism in Business intelligence as part of my PhD research at Staffordshire University.

Thank you for agreeing to take part in this evaluation and for supporting my research. Your participation is voluntary and you can withdraw at any time. No participants will be identified in this research and your personal details will not be disclosed.
If you would like any further information about the research or would like to be kept informed of the outcome of the research, please contact me at [nedim.dedic@research.staffs.ac.uk].

Please sign this consent form to confirm that you understand what is involved in the evaluation and that you are happy to take part

Please tick box

- ☐ I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions.
- ☐ I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason.
- ☐ I agree to take part in the study.
- ☐ I give my permission for the researchers to contact me again about other research opportunities.

Name of Participant	Date	Signature
______________________	______________	__________________
Name of Researcher	Date	Signature