

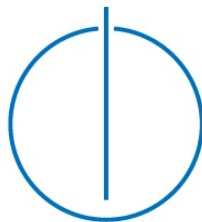
**Technische Universität
München**

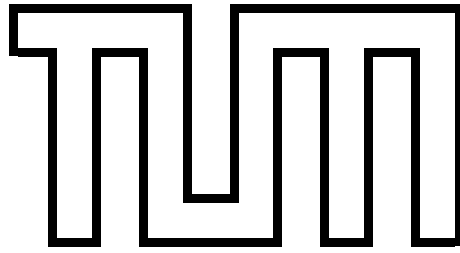
Fakultät für Informatik

Bachelor's Thesis in Wirtschaftsinformatik

Modeling an Information Architecture on Market Data for Data
Governance

Alexander Roschlaub





**Technische Universität
München**

Fakultät für Informatik

Bachelor's Thesis in Wirtschaftsinformatik

Modeling an Information Architecture on Market Data for Data
Governance

Modellierung einer Informationsarchitektur der Marktdaten für
die Data Governance

Author: Alexander Roschlaub

Supervisor: Prof. Dr. Florian Matthes

Advisor: Alexander Schneider

Submission: 15.08.2015

I assure the single handed composition of this bachelor's thesis only supported by declared resources.

München, 15.08.2015

(Alexander Roschlaub)

Abstract

The goal of this bachelor thesis is to develop an information architecture (IA) for a selected company tackling multiple information demands. The information demands range from regulatory requirements by Solvency II or audit requirements to concerns by the system's stakeholder and are collected into the first research artifact, the requirements. The second research artifact is the the IA itself, which manages the inherent complexity of a large-scale system. It offers a preliminary system overview with an exhaustive explanation. The architectural description further comprises of concerns representing information demands, viewpoints addressing these concerns and visualizing information of information models, which define and describe the used information objects. The IA is then evaluated by functional simulation with experts. This study adheres to the design science paradigm and is conducted in (collaboration) with a company offering financial reporting services for which market data is being utilized.

Abbreviations

AD	Architectural Description
AL	Application Landscape
CD	Client Data
CRUD	Create, Read, Update, Delete
CSV	Comma Separated Value
DB	Database
DW	Data Warehouse
EIA	Enterprise Information Architecture
ETL	Extract-Transform-Load
IA	Information Architecture
IAM	Information Architecture Model
ICOFR	Internal Control over Financial Reporting
IM	Information Model
IS	Information System
LOB	Line of business
MD	Market Data
MDP	Market Data Pool
MDS	Market Data System
ODB	Operational database
OE	Organizational Entity
SOX	Sarbanes-Oxley-Act
UML	Unified Modeling Language

Contents

Abstract	7
Abbreviations	9
1 Introduction	5
1.1 Motivation	5
1.2 Problem Statement	6
1.3 Research Approach	6
1.3.1 Natural vs. Design Science	7
1.3.2 Reference Design Science Process	7
1.3.3 Applied Research Approach	8
2 Foundation	11
2.1 Market Data	11
2.2 Theory of Information Architecture	14
2.2.1 Approach 1 – [J C 89]	15
2.2.2 Approach 2 – [IEEE 00]	17
2.2.3 Approach 3 – [S Bu 08]	19
2.2.4 Approach 4 – [M Go 10]	20
2.2.5 Architectural Description Language	24
2.3 Data Warehouse Reference Architecture	25
2.4 Interim Conclusion	28
3 Requirements	29
3.1 Legislation	30
3.2 Regulation	32
3.2.1 License Agreements	32
3.2.2 External Audit Requirements	34
3.3 Company internal stakeholder	37
3.4 Interim Conclusion	44
4 Information Architecture	45
4.1 System Overview	45

4.1.1	Operational Database – Market Data Pool	47
4.1.2	Data Warehouse	49
4.2	System Stakeholder and their key concerns	52
4.3	Viewpoints	55
4.3.1	Structural viewpoint	55
4.3.1.1	Data Storage viewpoint	61
4.3.1.2	Data Flow Viewpoint	64
4.3.2	Infrastructure viewpoint	68
4.3.3	Authorization viewpoint	70
4.3.4	Stakeholder viewpoint	72
4.4	Information Models	75
4.4.1	Structural Information Model	75
4.4.2	Data Flow Information Model	75
4.4.3	Infrastructure Information Model	76
4.4.4	Authorization Information Model	77
4.4.5	Stakeholder Information Model	77
4.5	Data Governance	78
4.5.1	Theory of Data Governance	78
4.5.2	Governance controls	80
4.5.3	Governance viewpoint	81
4.5.4	Governance Information Model	84
4.6	Interim Conclusion	84
5	Evaluation of IA	86
5.1	Theory of Architectural Analysis	86
5.2	Architectural Analysis	87
5.3	Interim Conclusion	93
6	Conclusion	94
	Bibliography	97

List of Figures

1.1	Applied Research Approach, Source: adopted from [P Of 09]	9
2.1	The architecture description life cycle Source: taken from [M La 05]	15
2.2	Conceptual model of architectural description Source: taken from [M La 05]	18
2.3	Data Warehouse Reference Architecture Source: Adopted from [V Ko 12]	26
4.1	System Overview Diagram of Selected company	46
4.2	System Overview of selected company's market data pool	47
4.3	System Overview of selected company's data warehouse	50
4.4	Overview of selected company's Data Standardization Process	51
4.5	Structural Viewpoint on the Market Data Pool	59
4.6	Structural Viewpoint on the Data Warehouse	60
4.7	Data Storage Viewpoint on the Market Data Pool	62
4.8	Data Storage Viewpoint on the DWH	63
4.9	Data Flow Viewpoint on the Market Data Pool	66
4.10	Data Flow Viewpoint on the Data Warehouse	67
4.11	Infrastructure Viewpoint on selected company's System	69
4.12	Authorization Viewpoint on the System	71
4.13	System Stakeholder Viewpoint	74
4.14	Structural Information Model	75
4.15	Data Flow Information Model	76
4.16	Infrastructure Information Model	76
4.17	Authorization Information Model	77
4.18	Communication Information Model	78
4.19	Governance Viewpoint on the Market Data Pool	82
4.20	Governance Viewpoint on the Data Warehouse	83
4.21	Governance Information Model	84

List of Tables

3.1	List of Vendor Audit Requirements for single application	35
3.2	List of External Auditor's Concerns	36
3.3	List of Application Security Audit Requirements	36
3.4	List of relevant Stakeholder	37
3.5	List of Management Board's Concerns	38
3.6	List of Internal Auditor's Concerns	39
3.7	List of Procurement Manager's Concerns	40
3.8	List of Application Owner's Concerns	40
3.9	List of Account Manager's Concerns	41
3.10	List of Database Engineer's (MDP) Concerns	42
3.11	List of Database Engineer's (DW) Concerns	42
3.12	List of Operation Manager's (MDP) Concerns	43
3.13	List of Operation Manager's (DW) Concerns	44

Chapter 1

Introduction

The first chapter elaborates on the the motivation for the research, what the problem is and evaluates different research processes.

1.1 Motivation

There is an increasing need to disclose information about an organizations internal structure. Governmental control, as well as auditing by vendors, clients or parent companies gets specific on organizations to ensure compliance with laws, license agreements and standards. In the financial industry laws such as Basel II were introduced to strengthen company's liquidity after the global financial crisis which started in the middle of 2007 [G Mu 08] and the Sarbanes-Oxley-Act which was passed after several accounting scandals to ensure an appropriate operational risk management being in place [G Mu 08]. The distributor of market data, the vendors, have also been focusing more and more on their license agreements. The selected company could identify an increasing need by the vendor to test their clients for compliance with their market data's terms of use. Therefore new mechanisms need to be implemented within the companies' systems in order to comply to certain requirements. Additionally there need to be mechanisms or ways of proofing the compliance with these requirements. This research solely concentrates on Market Data, due to their major influence on governments, companies and people, which is shown by the high number of existing regulations.

The goal of this bachelor thesis is to create insight into the system via an information architecture (IA) and by using this IA to information for the audit reports on the basis of this insight. The requirements are gathered by analyzing performed audits, analyzing

the literature regarding regulations from laws and other sources of regulations. As second artifact serves the IA to map these requirements to the selected company's system.

1.2 Problem Statement

"IT research studies artificial as opposed to natural phenomena" and "deals with human creations such as organizations and information systems" [S T 95]. In this case, the artificial phenomena and the origin of this study is the need to disclose information about the internal structure of a company in external, as well as internal audits. To successfully conduct an audit, the generation of a report is mandatory. The consecutive problem for the creation of such a report is that companies often operate based on large-scale distributed systems with high complexity, little documentation and no implemented EA. The ensuing result is a black box as a system. A black box has defined inputs and outputs, but everything within the system is unknown. The reasons for the complexity and lack of documentation in the selected company are mostly due to the accumulation process over the years and the company's fast growth. Calling the system a black box is most certainly an exaggeration, but the expression obtains validity, since there is not a single instance or division within the selected company, having the knowledge to describe every process and elements of its system. It is a tedious, interdivisional collection process to gather referenceable information that is accurate, complete and consistent. Accurate, so that the actual and physical instance is being reflected, complete, so that every necessary piece of information is included and consistent, so that anyone creating the report will get the same results. Another problem is that the assessment criteria are very often underspecified and vendors would ask the selected company to proof the adherence to their license agreements. It would be very helpful to already have requirements that have proven to be adequate for the vendors. This would significantly reduce the time it would take to gather the necessary information and the process of evaluating what the vendor actually wants to know would unnecessary. Research Questions:

1.3 Research Approach

"Research in IT must address the design tasks faced by practitioners. Real Problems must be properly conceptualized and represented, appropriate techniques for their solution must be constructed, and solutions must be implemented and evaluated using appropriate criteria" [S T 95].

[S T 95] present two frameworks for research in IT: Natural Science and Design Science.

1.3.1 Natural vs. Design Science

Natural science, as described by [S T 95], "is concerned with explaining how and why things are" and tries to understand facts of existence. The outcome are "sets of concepts", which are the tools to distinguish the phenomena. Those tools are "used in higher order constructions - such as laws, models, and theories - that make claims about the nature of reality".

According to [S T 95] natural science consists of the two activities: discovery and justification. Discovery "is the process of generating or proposing scientific claims. Justification "includes activities by which such claims are tested for validity."

However, IT research is interested in "artificial phenomena operating for a purpose within an environment" [S T 95], which leads to Design Science. Design science, as described by [S T 95], focuses on technology and "attempts to create things that serve human purpose." Its results are evaluated against "criteria of value or utility", as well as achieved improvements. "Progress is achieved in design science when existing technologies are replaced by more effective ones" [S T 95].

According to [S T 95], design science has four types of products: constructs, models, methods and implementations. The first product, the constructs, are used to distinguish phenomena and they are a "basic language of concepts". Secondly, are models used to "describe tasks, situations, or artifacts". The third product, the developed methods, are "ways of performing goal-directed activities". The final product, the "foregoing can be instantiated in specific products, physical implementations intended to perform certain tasks."

How do we get these products? [S T 95] say that design science consists of the two activities build and evaluate. Building "is the process of constructing an artifact for a specific purpose" and evaluation "is the process of determining how well the artifact performs."

1.3.2 Reference Design Science Process

[P Of 09] goes further than [S T 95], but his design science process is based on their results and divides design science into three phases:

1. Problem identification

2. Solution Design

3. Evaluation

In the first phase problems are identified & collected, as well as their relevance evaluated & described subsequently. In the second phase the solution artifacts are designed in an attempt to solve the previously identified problems. The design is based on a literature research where the "existing knowledge base including state-of-the-art has to be taken into account to ensure research rigour" [P Of 09]. The third phase starts after finishing the design of the artifacts representing the solution.

1.3.3 Applied Research Approach

The applied research approach in this bachelor thesis, as seen in 1.1 is based on the design science process in [P Of 09], but modified to fit the needs for the creation of an appropriate IA based on an IS processing market data. The design science approach still consists of three phases: Foundation, Solution Design and Evaluation. The problem identification phase was not replaced by the foundation phase, but rather integrated. This was done due to the need of describing and abstracting the market data and system this bachelor thesis is built upon.

Foundation

The first phase consists itself of two parts. In the beginning the artificial phenomena are distinguished by identifying the problem. After the problem is sufficiently described, its relevance is evaluated by conducting the first literature research. In the second part of the first phase the foundation for the IA is built, which means that all the necessary information to understand and apply the IA is described. Also described is the system, which the IA is based on, as well as market data.

Solution Design

In the second phase the solution to the described problem is being designed. In this case the solution is the IA. The starting point is a requirement analysis & collection of company related artifacts in cooperation with experts. The first product will be a requirements catalogue containing the information demand from audits, laws and license agreements. The audit requirements, as well as the license agreement requirements are collected in

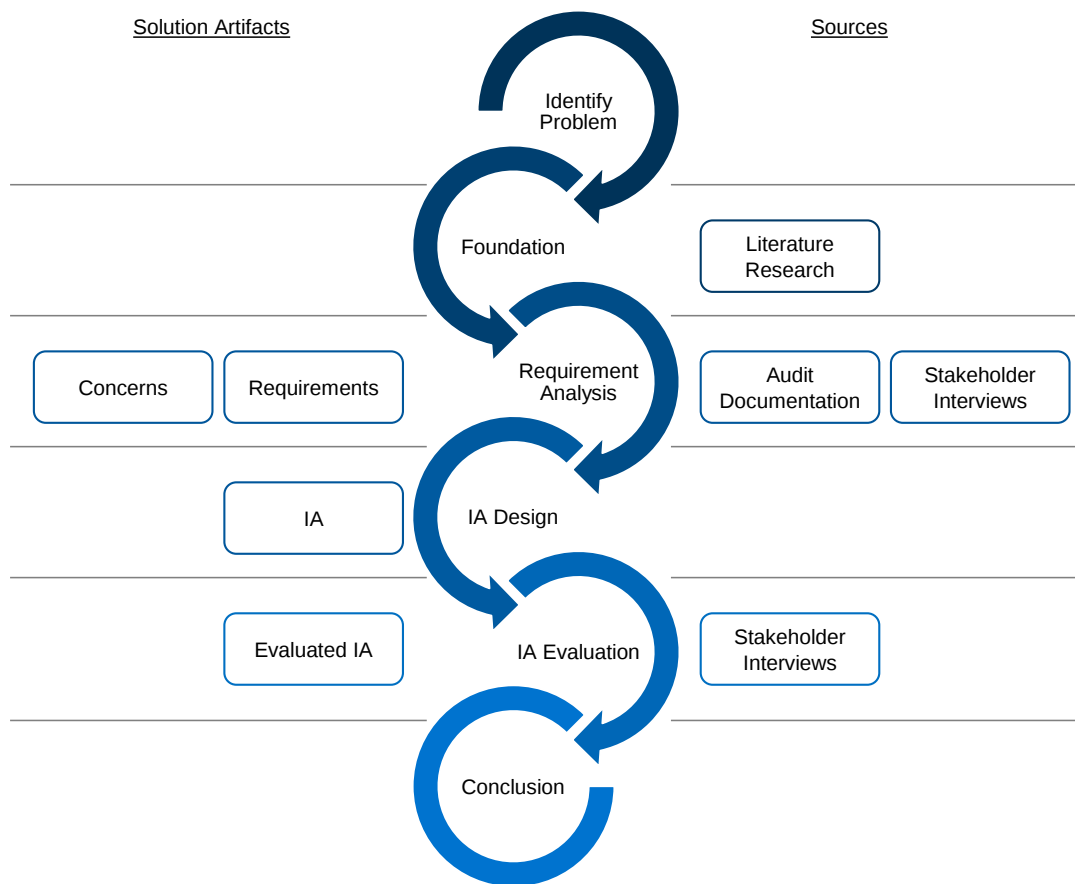


Figure 1.1: Applied Research Approach,
Source: adopted from [P Of 09]

collaboration with the selected company by conducting interviews with experts and the studying of company's literature and documents. The product of this will be a IA. This IA is afterwards enhanced by adding data governance.

Evaluation

In the third phase the solution, the IA, is evaluated by conducting a case study with the stakeholders, where examples are selected and the IA model, the data collection process and possible visualizations will be implemented. On the basis of the case study, the IA is reviewed and modified according to the results. The product of this phase will be an evaluated IA.

Chapter 2

Foundation

The foundation this research consists of an abstraction and description of what market data actually is, as well as in which context it is used. Furthermore is the concept of architecture in principle and IA in particular described. Different approaches are presented and taken into consideration for this research's approach for its IA. The data warehouse (DW) reference architecture as basis for the system overview is described as well.

2.1 Market Data

Market data can be described as data reflecting information about a financial instrument or investment. It can be purchased from data providers, called vendors, and/or stock exchanges. But what is a financial instrument or investment and why do companies need market data? To answer this question and for insight into the subject, an excursion to the insurance industry is made. The following paragraph describes a very simple and abstracted use case of market data within an insurance company.

Example – insurance company

An insurance company assumes the monetary risk for a defined entity of an individual. The individual has to pay a certain price for the transfer of the risk, which is usually a payment over time. Simplified, the insurance company collects that money in an imaginary pot. When they might have to pay for the monetary claims connected to the insured entities, money is taken out of the imaginary pot. But the insurance company doesn't have all

of the "pot's" content lying around in deposits, but rather invest the money in financial instruments or assets. This act is called an investment.

Investment is the purchase of an asset or item with the hope that it will generate income or appreciate in the future and be sold at a higher price. It generally doesn't include deposits with a bank. Description of the security identified by investment identifier and investment type.

- Long-Term outlook: trading or speculation
- Short-Term outlook: higher degree of risk

In such a pot there is enough money to make multiple investments. An investment does not have to be traded publicly. This means that there is no market data available. Only the involved parties have data concerning the investment. Those are collected in a portfolio, which has one or more positions. A position is always associated to an investment. Transactions are movements within a portfolio.

Portfolio Bundle of investments or group of positions. Identified by a unique portfolio code. Attributes are static data like description, currency, benchmark, or time series data like total amount, returns, etc.

Position Single entry within a portfolio, mainly defined by portfolio, investment, validity date and currency. Thus within one portfolio a single investment can occur more than once.

Transaction movements within the portfolio like buy, sell, incomes, as well as withdrawal or payments on portfolio level

Therefore the need for market data originates in the need the invest money into financial instruments. But those investments need to be evaluated, whether their performance is good or bad compared to other factors or indices, calculated in benchmarks. The results are presented in form of reports and delivered to the clients.

Benchmark Reference and comparison value for an investment, which is used to evaluate the investment's performance.

Derivatives a contract that derives its value from the performance of an underlying asset

Equity refers to the buying and holding of shares of stock on a stock market in anticipation of income from dividends and capital gain

Fixed Income refers to any type of investment under which the borrower/issuer is obliged to make payments of a fixed amount on a fixed schedule

Obviously, every financial instrument or investment needs to be identified, so that different people from different countries know whether they speak of the same investment. Due to the global trading of securities and the lack of a global regulation, a lot of different security identifications exist. The three most prevalent security identifications within the selected company:

- ISIN: used "to identify specific securities. They are composed of a 12-digit alphanumeric code and act to unify different ticker symbols which can vary by exchange and currency for the same security"
- SEDOL: primary security identification in the US. Integrated in the ISIN.
- WKNR: German security identification.

Market data is often assigned to sectors. A sector can be a industry, a country, a region or another defined scope. Also of importance is the trading at multiple stock exchanges whereat a primary stock exchange exists. The primary stock exchange is used as the source of market data, if available. This means that even though the investment was purchased at a stock exchange in America, the market price from the primary stock exchange is used for further reports & analyses.

There are also different 'versions' of market data, which are of interest for the requirements. The subsequent definitions are "customary in the market" and in this form used within the selected company, but "have no legal basis" and can be "obsolete through individual license agreements."

Original/Raw data Source data as they are made available by the vendor. This Raw data is fully subject to license.

Manipulated data Data, which was mathematically changed, but can be reckoned back. This manipulated data is fully subject to license.

Derived data Is the original data changed by mathematical calculations, but they cannot be reckoned back, then original data transforms into derived data. Derived data is in most cases not yet subject to license.

Overview about the grouping of market data and some exemplary facts in the selected company:

- Convertibles & Preferreds: The coupon frequency used for computing the security yield, Conversion rate, Convertible start date, Convertible until
- Descriptive: Name, Maturity, Contract Size, Total Shares Issued, Dividend, etc.
- Dividend Record: Dividend Currency, etc.

- Funds: Inception Date, Management Style/fee, Net Asset Value (NAV), Strategy, etc.
- Futures: Conversion factor/size/value, First/Last Delivery/Trade date, etc.
- Exchange Rates: Currency from code, Currency to code, Exchange Rate(London 16:00), etc.
- Security Identifier: ISIN ID, Cusip ID, Sedol ID, Vendor's Company ID, Vendor's Security ID, Vendor's Unique ID, Country's ID, etc.
- Indexes: Constituent list, constituent deflator, constituent market values, constituent vendors, constituents weights
- Options: Contract size, Expire data, Price, Strike price, etc.
- Pricing: Mid/Bid/Low/High price, etc.
- Rating: Long term, short term, watch, etc.
- Warrants: Exercise Date/Price/Type, Put or Call, Expire Date, etc.

2.2 Theory of Information Architecture

It is not easy to describe what an architecture or information architecture (IA) is and all its deliverables, since "there is no single, accepted framework for codifying architectural thinking" [IEEE 00]. This means that different accepted frameworks, thus approaches, need to be taken into consideration and evaluated regarding the appropriateness for the problem in hand.

Architecture Process

Before presenting the different architecture approaches, it is important to understand that "architecture is a process as well as a product." The architecture "serves to guide" stakeholders [M La 05]. The architecture description life cycle, as seen in figure 2.1, can be divided into four phases: idea, design, use and management [M La 05]. The architecture process "consists of the usual steps that take an initial idea through design and implementation phases to an operational system." If it reaches that the operational state, it changes or replaces the system and by doing so closes the loop [M La 05].

This thesis is based on a problem, which represents the idea. It is supposed to deliver

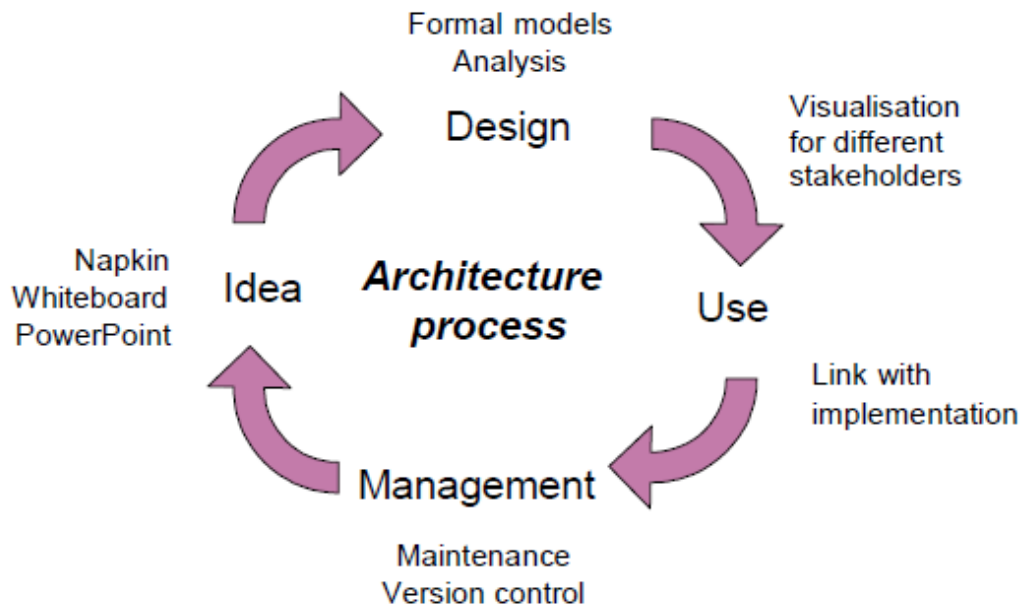


Figure 2.1: The architecture description life cycle
Source: taken from [M La 05]

formal models and analysis techniques from the design phase and is evaluated visualizing the design products for the different stakeholders. However, it does not address the management phase.

2.2.1 Approach 1 – [J C 89]

The first and oldest approach is from J. C. Brancheau and S. T. March from 1989 [J C 89]. It certainly is outdated, but due to its small scope it is easy to understand and serves as a good introduction for the topic. [J C 89] define IA as follows:

"An information architecture is a high-level map of the information requirements of an organization. It is a personnel, organization and technology independent profile of the major information categories used within an enterprise"

IA offers a method "to map the information needs of an organization, relate them to specific business functions and document their interrelationship" [J C 89]. The first

problem is the "broad scope"[J C 89] and "corporate-wide information requirements determination certainly qualifies as having a broad scope"[J C 89]. Additionally does the "lack of structure" also pose a problem, since "the less structured the problem under consideration, the more difficult it is to define and communicate"[J C 89].

[J C 89] states that "three basic inputs" are "required" in order to create an IA: "business functions, organization structure, and existing applications"[J C 89]. Thus the "basic functions of the business must be identified and defined"[J C 89]. Of interest is here "what business the organization is in and determining what functions need to be performed.[J C 89]" Secondly, "the existing organization structure must be mapped to the business functions"[J C 89]. The "information about existing applications must be mapped to business functions"[J C 89]. This is enabled by "gathering information about the functions provided by existing systems and how well they meet the organization's information needs"[J C 89].

The actual IA consists of three components [J C 89]: a business function model, a global data model, and an entity description/definition.

A business function model "takes the form of a BSP-like Matrix." One axis of the matrix specifies business functions and the other specifies data classes. The business functions are important to define, because they "are the fundamental activities necessary for operation of the organization"[J C 89]. Data classes are of interest, since they "are the general categories or types of data needed to perform these business functions"[J C 89]. In the business function model the interaction of both is defined.

The second component, the global data model, is a Logical Data Structure (LDS) diagram containing entities and their relationship [J C 89].

The third component is the entity description/definition and "is used to document additional information about each data class or each entity that is thought to be valuable for future systems [...]"[J C 89]. The entity description contains the data class name, a short definition of the scope of the data class and information "about the ownership of the data"[J C 89]. This entity description can be tailored for a specific purpose or for a specific organization [J C 89].

Important for the IA is the "appropriate level of accuracy." A "100% accuracy is not required", because it would include "too much detail". "Deemphasizing total accuracy [...] shortens the time-frame for architecture development" and "enables development of a first-cut information architecture quickly using a minimum of resources"[J C 89]. Another key success factor is the "cataloging of basic business functions"[J C 89]. According to [J C 89] is it important to not base those on the organization's structure, but rather on the organization's functions. This is founded on the assumption that "business functions are very stable" and "organization structures change over time"[J C 89].

The problem with this IA is, that it is not sufficient enough to solve our problem. However, its three components are of interest and therefore will be integrated into the IA.

2.2.2 Approach 2 – [IEEE 00]

The second approach was developed by the IEEE Computer Society, which approved the IEEE Standard 1471-2000 in 2000. It "builds a solid theoretical base for the definition, analysis, and description of systems architectures" [M La 05]. Their definition of architecture is as follows [IEEE 00]:

"The fundamental organization of a system, embodied in its components, their relationships to each other and the environment, and the principles governing its design and evolution."

An architecture can therefore be described as a blueprint of a system, with a description of its entities and their relationship, as well as principles that need to be adhered to. The question is what information and insight is required and needs to be displayed in order to get a sufficient architecture.

The IEEE Standard 1471-2000 is a recommended practice for the following purposes (only an excerpt from the actual quantity, which actually matches with this thesis' purpose) [IEEE 00]:

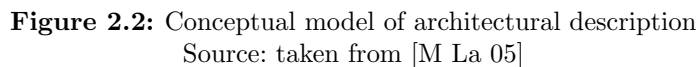
- Expression of the system
- Communication among the system stakeholders
- Evaluation of architectures in a consistent manner

Architectural Description Model

The architectural description (AD) is defined as "a collection of products to document an architecture" [M La 05] and therefore can be seen as the components of the architecture. The IEEE Standard 1471-2000 [IEEE 00] introduces a conceptual framework for architectural description as seen in figure 2.2.

The architecture is for a system, which comprises "aggregations of interest." Those can be individual applications, systems, subsystems, whole enterprises.

The system "inhabits an environment" which can have an impact on the system. The environment "determines the setting and circumstances of developmental, operational, political, and other influences." The environment also specifies "the scope of the system of interest relative to other systems."



The AD identifies one or more system stakeholder, their concerns and selects viewpoints. A viewpoint is "a specification of the conventions for constructing and using a view. A

pattern or template from which to develop individual views by establishing the purposes and audience for a view and the techniques for its creation and analysis”

”Information, which is not contained in any constituent views, may appear in an AD. Examples of such information are the system overview, the system context and the system’s stakeholders and their key concerns”[IEEE 00].

A viewpoint ”establishes the conventions by which a view is created, depicted and analyzed.” Furthermore, a viewpoint ”determines the languages (including notations, model, or product types) to be used to describe the view, and any associated modeling methods or analysis techniques to be applied to these representation of the view.” The AD ”selects one or more viewpoints for use. The selection of viewpoints typically will be based on consideration of the stakeholders to whom the AD is addressed and their concerns.” The viewpoint ”may originate with an AD, or it may have been defined elsewhere.” The source of this viewpoint would be called library viewpoint.

2.2.3 Approach 3 – [S Bu 08]

The third approach was presented in the Enterprise Architecture Management Pattern Catalog [S Bu 08].

The EAM Pattern Catalog consists of four different product elements: Concerns, which are addressed by Methodologies(M) Patterns. Those utilize Viewpoint(V) Patterns for communication. Viewpoint Patterns in turn visualize information of Information Model(I) Patterns.

The definition and use of **Concerns** matches the definition in [M La 05]: ”an interest of a stakeholder with regards to the architecture description of some system, resulting from the stakeholder’s goals, and the present or future role(s) played by the system in relation to these goals”.

M-Patterns define ”steps to be taken in order to address a given concern”[S Bu 08]. With ”steps” being ”activities” that need to be performed by the executing individual or group with the purpose to meet the information demand of the stakeholder. Moreover ”statements about the intended usage context are provided”[S Bu 08]. This M-Patterns help ”to complement activities carried out in an ad-hoc manner” or to rely ”on implicit knowledge with activities carried out more systematically”[S Bu 08]. They should be seen as guideline and represent a base case. There are many special cases, which are not taken into consideration and therefore the methods are not always in applicable in full extent. However, these M-Patterns are not in scope. They include the as-is state and a to-be state, as well as the roadmap of how to get there. This IA does not want to change the architecture of the evaluated system. The purpose of the IA is to generate different

viewpoints for stakeholder to answer their concerns and help to comply to regulations.

V-Patterns provide "languages used by M-Patterns" [S Bu 08]. A V-Pattern "proposes a way to present data stored according to one or more I-Patterns" [S Bu 08]. Furthermore do "industrial users often specify viewpoints by example" and therefore an "exemplary view is provided for the viewpoint, possibly together with some textual explanation" [S Bu 08]. It is to be noted that a V-Pattern only proposes, just like all the other patterns. It does not present the one and only solution. Most often adjustments need to be made in order to fulfill the specific information demand of the stakeholder (concern) or limit the view for external people. Making viewpoints into patterns "offers the advantage to easily add layers to viewpoints" and enables to "visualize applications on one layer and different key performance indicators on additional layers" [S Bu 08].

I-Patterns supply "underlying models (the abstract syntax) for the data visualized in one or more V-Patterns." It "contains an information model fragment including the definitions and descriptions of the used information objects." [S Bu 08] propose using a language adequate to the problem to be addressed, thereby strongly considering UML as the default language. The reasons are that it is "widely understood" and is concluded by them to be "problem-adequate in many situations in the context of EA management information models" [S Bu 08].

The approach of the EAM pattern catalog will be the basis for this IA. It will contain concerns from system stakeholder, viewpoints to help stakeholder generate a certain view onto a system and information models to include the definitions and descriptions the the used entities in the viewpoints.

2.2.4 Approach 4 – [M Go 10]

The fourth approach is by [M Go 10]. This approach was included, because it is the preferred approach within the selected company's company group¹. Furthermore does this approach give insight in how companies create architectures of their system and the difference between Information Architecture (IA) and EA is explained.

Definition of EA according to [M Go 10]:

"An Enterprise Architecture is a tool that links the business mission and strategy of an organization to its IT strategy. It is documented using multiple architectural models that meet the current and future needs of diverse user populations, and it must adapt to changing business requirements and technology."

Definition of IA according to [M Go 10]:

¹It is important to note that the selected company itself does not yet use the concepts of Enterprise Architecture (EA).

“IA provides the foundational information-relevant concepts and frameworks for dealing in a consistent and integrated manner with the technology to guarantee the responsiveness and trusted information insight that the business requires [...]”

The scope of an IA is much smaller than that of an EA. The IA becomes an Enterprise Information Architecture (EIA) by “applying enterprise-wide business context” [M Go 10]. EIA differs again from the EA. “The Enterprise Information Architecture translates the business requirements into informational strategies and defines what data components are needed by whom and when in the information supply chain.” EA looks at the whole enterprise and tries to align different architectures and the enterprise’s strategies with each other as well as coping with changes in the future.

The EA includes heterogeneous architectural domains with different purposes: Business Architecture, Application Architecture, Information Architecture, Infrastructure Architecture, Integration Architecture, Operational Architecture, Security Architecture, Network Architecture and many more. They all “address specific situations or problems to be solved within an enterprise and are thus related in some way to the overall EA” [M Go 10]. This aligns with the approaches by [IEEE 00] and [S Bu 08]: The specific situation or problems can be seen as concerns and the different architectures as viewpoints. The IA is not a complete representation of a company’s architecture like the EA. The IA is only a domain of the EA. A domain is “any subset of a conception (being a set of elements) of the universe that is conceived of as being some ‘part’ or ‘aspect’ of the universe” [M La 05]. An IA is based on the information requirements for a certain artificial problem and tries to meet this information demand in a satisfactory, ordered and timely manner. Satisfactory, so that all the stakeholder’s concerns are covered. Ordered, so that there is some guideline for what steps need to be taken in order to meet information demand and timely, so that a basic, defined procedure rather than on-the-fly procedure exists. In addition does it serve the purpose to eliminate chaos within the system to be consistent.

The IEEE Std 1471-2000 definition “concentrates on design and evolution”, but other issues need to be addressed “such as guidelines and principles for implementation, operations, administration and maintenance” [M Go 10]. Additionally, should the AD “provide a definition and an explanation about the components [...] of the architecture, its properties and collaboration among each other” [M Go 10]. This is also a problem discovered by [S Bu 08]. Their solution was to include the definition and explanation in form of Information Models.

Core principles that guide an IA were defined by [M Go 10]:

- Access and exchange of information: “Information services should provide unconstrained access to the right users at the right time.”

- Service re-use: "Facilitate discovery, selection and re-use of services and whenever possible encourage the use of uniform interfaces."
- Information Governance²: "Adequate information technology should support the efficient execution of an Information Governance strategy."
- Standards: "A set of coherent standards for data and technology should be defined to promote simplification across the Information Infrastructure."

It should be noted that Data Governance is a core principle of an IA and should therefore be included.

Data view

One of the issues that need to be addressed, but are not included in the [IEEE 00] is the data view. "The data-centric view describes the data domains that are derived from classification criteria, the relationship between data domains, and the flow of data through the components of the [...] Architecture"[M Go 10]. The data view is also of importance for the previously described IA principles by [M Go 10] and data governance. Data governance and especially data quality as one of its aspects is only applicable on defined data.

According to [M La 05], the business of a company is data dependent. This data is being collected, enhanced and shared to achieve the business goals. Data can be anything ranging from the type of data to digitalized data which is "data that can be processed by computing systems such as desktops, laptops, [...], servers, and mainframes"[M Go 10]. Data is being used by a variety of applications and by many users. [M Go 10] divides data into two different formats:

Structured Data represents a business object and uses a relational data model. It is identified by a defined number of attributes and is stored, managed and persisted in relational databases

Unstructured Data has no structure and is not based on a data model. Examples are files with data about meetings or tabular list of data, such as Excel or CSV files

Also of interest is the scope of integration, since companies have confidential information and tend to operate on a need to know basis. Need to know means, that only the users, who actually need to know the information about a certain data object have access to

²[M Go 10] uses Information Governance instead of Data Governance to hint at the inclusion of "all the data domains, not just the operational data domain."

that specific piece of information. According to [M Go 10], the scope of integration can be divided into three parts:

Local Scope In this scope the data is only needed by a team, department, or a LOB. An example is the calculation of benchmarks. The data used for the calculation of those benchmarks does not need to be known to the downstream applications.

Enterprise-wide-scope In this scope the data is needed by the whole enterprise. An example would be identification data about an investment. Enterprise-wide data can also be used across enterprises.

Cross-enterprise scope In this scope the data is being used in multiple enterprises, but not necessarily in all LOB's.

This division is very important and useful for market data, due to the licensing of data and can be used as a first classification for the market data. The need to scope is due to the vendor's license agreements, in which it is stated who may get and use the market data. It is of importance that the scope is being adhered to. For example can a vendor have a license regulation as follows: The means that a company may give the data to every affiliated company, if its share is bigger or equal a certain percentage. Therefore you have to scope in such a way to ensure that only the affiliated companies, get the data. Market data serves different purposes. They are used for the generation of reports, risk analyses, benchmark calculations and many more. Therefore it is helpful to divide the data into data domains, so that the purpose of the data is visible at first glance. [M Go 10] divides data into five data domains:

Metadata Domain Additional data which adds information to the underlying data and brings it in a certain context. An instance of metadata could be a timestamp.

Master Data Domain Instances of data which represents the core business data and therefore have many dependencies with other data entities.

Operational Data Domain transactional data which originates from business transactions

Unstructured Data Domain has no structure and is not based on a data model. Examples are text files and Excel/CSV files

Analytical Data Domain is data which is calculated on the basis of operational data to address a specific purpose and gain knowledge. It usually has a time dependent aspect.

The purpose of a data view is to obtain [M Go 10]:

Accuracy is a measure that identifies the difference between the value of the stored data and its actual value. If there is a difference, the data is considered bad data and its quality is bad. To increase the data quality exist several techniques: data standardization, data matching, data deduplication and data validation.

Completeness is a measure to which the data has all relevant attributes and values.

Consistency same data entity in various applications, but still represents the same value.

Timeliness propagation of changes, freshness of data

Relevance degree to which the data satisfies the need of data consumer and is usable

Trust generate and maintain trusted data complies with the above mentioned data principles

Furthermore will the data view of the IA include definitions and descriptions of market data entities, as well as a VM visualizing the data, its specification and transformation within and in between the system's components.

2.2.5 Architectural Description Language

Modeling is crucial when trying to visualize, specify, construct and document the structure and behavior of a system's architecture [Pier 05]. Models are the architects means to answer the analysis questions that the stakeholder have. Modeling is "the act of purposely abstracting a model from (what is conceived to be) a part of the universe" [M La 05] and helps to [Pier 05]:

- Visualize – to show the interaction of a system's elements
- Specify – to aid univocal communication
- Construct – to support consistent design and implementation
- Document – to illustrate the structure and behavior of a system's architecture

Models can be a good or bad representation, but this is not only depending on a single, but rather multiple aspects. In order to ensure that the models are as good as possible the following maxims is being adhered to [M La 05]:

- Maxim of Quantity: As informative as necessary
- Maxim of Quality: Model evidence based and believe to be correct
- Maxim of Relevance: Relevant related to modeling goal

- Maxim of Manner: Avoid ambiguity

The decision what modeling language to use should be based on the knowledge and background of the intended audience (the system stakeholder). If the stakeholder have a technical background (computer science, engineering, etc.), the Unified Modeling Language (UML) is a viable option, as it is de-facto the standard for object oriented programming [Kech 15]. If the stakeholder have a background in other fields, such as finance in this case, UML is not the modeling language to go, because for some it might not be self-explanatory or easily understandable. Additionally it is always of avail to look at what the company uses for its modeling language. The selected company does not use a modeling language, but rather boxes and lines. Boxes and lines means that there is no standard or modeling language is used, but rather the modeler is just models as he pleases. However, this would be very unproductive and not constructive. Therefore each time boxes and lines is being used as modeling language a legend is provided containing map symbols, which are the elements the model consists of and visualization rules describing the interaction of those elements.

However, for the information models UML's definition of class diagrams will be used. The reasoning behind this is, that those IMs serve as definition and description of the architecture's components. Definitions serve a formal purpose and therefore should also have a formal description. Furthermore are class diagrams among the most simple UML diagrams and rather easily understandable and intuitive.

For the particular viewpoints boxes-and-lines will be used in order to create the most value by delivering a model, which is understandable by all the stakeholder. The need for a more specific and detailed model is also not given. The to be delivered architecture and its models only serve the purpose of giving insight and an overview of the system. If the model would be more specific it would get overly complex, because so many special cases and instantiations would need to be considered.

2.3 Data Warehouse Reference Architecture

The system under investigation is based on the principles of a data warehouse (DW). In order to get an adequate understanding of what an DW is, the reference architecture for a DW is described in the following.

Figure 2.3 shows the reference architecture of a DW as described by [V Ko 12].

Data sources The data sources are not part of the DW. They can either be external or internal and are heterogeneous regarding its structure, content and interfaces.

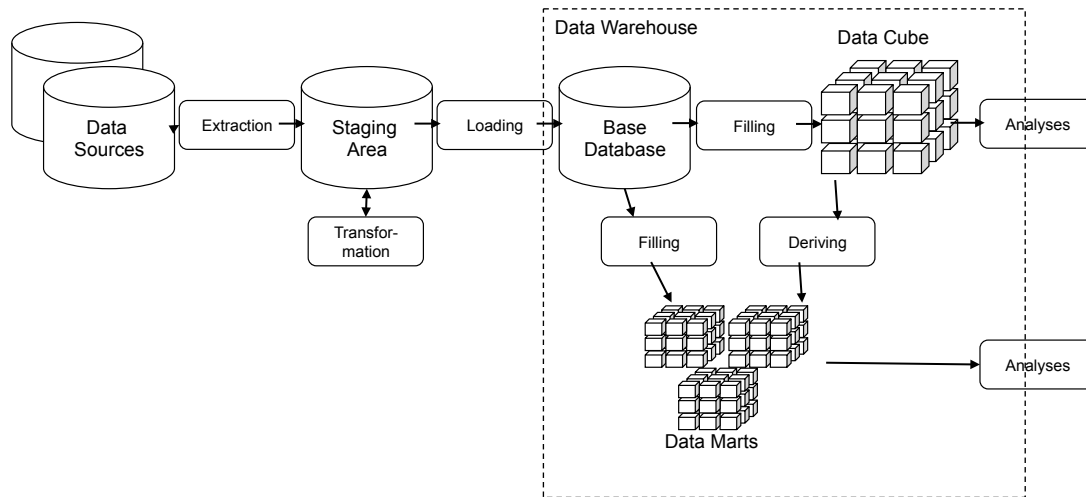


Figure 2.3: Data Warehouse Reference Architecture

Source: Adopted from [V Ko 12]

The classification [V Ko 12] of the sources is important for further processing and analysis:

- Origin: internal, external
- Time: effective, historical
- Use: primary (facts), metadata
- Content: Number, Character string, reference
- Representation: fact type
- Language and character set
- Degree of confidentiality

Staging Area The staging area is a temporary storage for data integration and cleansing. It is being filled with raw data by extraction components. The mainly

use of the staging area are the transformation tasks, which are performed on the data set stored in the staging area. The loading of the data into the base database is only performed after the complete and successful transformation. [V Ko 12]

Extraction Components The extraction components transfer the raw data from the sources to the staging area. Two aspects need to be considered [V Ko 12]:

- What is to be extracted?
- When is to be extracted?

Transformation Components The transformation components work on data stored in the staging area. They extract the data, transform it and then write the data back into the staging area. Types of transformation [V Ko 12]:

1. Homogenization transforms the data into a uniform representation in order to enable integration. Typical homogenization tasks are transformation of data types or standardization of timestamps.[V Ko 12]
2. Integration merges the data from one or more sources. [V Ko 12]
3. Cleansing corrects problems with the data quality:[V Ko 12]
 - Identification and reparation of incorrect data
 - Quantification and remove or complement incomplete data
 - Identification of redundancies
 - Identification and update of outdated data

Loading Components Loading can be understood as the transferring of the adjusted and conditioned data into the base database. For the loading are often special loading tools used in order to efficiently load large amounts of data in a timely manner.[V Ko 12]

Base Database The base database contains all the data of the data warehouse as basis for the analyses. The data is adjusted and historicized. Therefore the Base Database is an Operational Data Store. The main task of the base database is the provisioning of data for the data cube. It needs to be evaluated whether the physical storage for the base database is required.[V Ko 12]

Filling Components Filling is the equivalent to the loading of the adjusted and conditioned data from the base database into the data cube.[V Ko 12]

Data Cube The data cube is the basis for efficient analyses by means of preliminary aggregation and saving data denormalized in order to reduce network operations.[V Ko 12]

Data Marts Data marts are extractions of the complete analysis data set. The extraction can be based on the range of data (vendor, data history) or on the granularity of the data (daily or monthly data). Data marts are materialized views on the data cube. The data marts are filled by derive components. If the data mart get its data from the base database, the transfer of data is called filling. If the data mart gets its data from the data cube, the transfer of data is called deriving.[V Ko 12]

2.4 Interim Conclusion

This chapter gives an overview of what market data is and what it is needed for. The terms investment, portfolio, position, transaction and company should be understood, since this is the basis for the selected company's data model in the DW. Consecutively is the scope for this thesis' IA created by evaluating different approaches. The approach of [J C 89] is relatively simple, but is not sufficient enough to give the required insight into the system and doesn't even give a system overview. Another approach by [IEEE 00] was given, which included the definition of AD and its components. [M La 05] added to the static architecture description the architectural analysis, which promotes the system stakeholder's communication, as well as the component's interaction and is therefore of significance. The approach by [S Bu 08] gives the IA more structure and delivers best practices in form of patterns. The last evaluated approach from [M Go 10] embeds the IA into company specific aspects and requires the creation and implementation of principles, such as Data Governance. The data view provided by [M Go 10] is also presented and is the basis for the described Data Governance. As ADL there will be UML class diagrams used for the IMs and boxes-and-lines in the VMs.

Chapter 3

Requirements

The selected company assures compliance with risks and regulations in three steps. The first step being "operational controls within processes and performed in a structured, consistent and timely manner." The second step being "regular assessment of both the controls and the people performing those controls". The third step contains the audits. Those are required by laws, defined by regulations and conducted by auditors.

The information requirements for said audits are not always clear and if they were, no single, central point within the selected company was able to deliver the required information. This also means that the information gathering process is not defined and had to be constructed anew each time a audit was performed. In a first step said requirements were gathered in collaboration with the selected company. This was done by conducting interviews with the responsible employees in different business units in combination with the studying of the companies documentation, as well as academic research papers and laws.

Collection of Information

The collection process for the information to gather the requirements for the IA was done from the top down, since this is the same approach externals are given an overview and introduction to the system.

In the beginning interviews were conducted with the company's experts, who were involved in the creation and evolution of the system. The purpose of these interviews was to create an understanding of the system, its purpose and its structure. Additionally, the selected company's processes. This led to a system overview draft. This resulting informal system

overview and the selected company's department organization helped divide the system into components to create a black box on the big components of the system, as well as identifying corresponding departments. To be noted is here that system components and departments were selected that are processing or using market data and client data.

After conducting an interview for the confirmation of these results and selection of key employees of each department roughly the same process for the interviews was used with more details. An initial interview to create an understanding of the system's component and the key employee's tasks was conducted and resulted in a component overview. The second interview was to confirm the results of the first and key employee's were renamed to stakeholder in order to depict the affiliation to the system architecture. Furthermore was evaluated whether any legislation or regulation affected the interviewee. If appropriate these requirements were collected. Additionally, the stakeholder's concerns of the system regarding market data were collected, wherein always two key employees were interviewed for one stakeholder role. The requirements were gathered in the scope of the following principles [M Go 10]:

- Compliance with all information security requirements
- Compliance with all relevant regulations
- Deliver information with the appropriate data quality

3.1 Legislation

The system of the selected company is being utilized for the financial reporting and is thus a financial system. Financial systems need to be "stable and reliable" and legislators are "prompted to react by passing regulations preventing further crises" [B Wa 14]. Those legislation request companies to implement operational risk controls, which needs to be adhered to by the IT, since it "has become central within the financial system" [B Wa 14]. The business of financial institutions includes "evaluating risks, coping with uncertainty and managing changes" [B Wa 14]. Operational risk is the "risk of loss resulting from inadequate or failed processes, people and systems or from external events." However, "the derivation of concrete measures for existing systems is not trivial" [B Wa 14]. The following three laws are regulating operational risk, affecting financial reporting and finance data:

Sarbanes-Oxley Act (SOX) The SOX was established by the US government "in order to protect investors and improve the accuracy and reliability of corporate

disclosures”[G Mu 08]. This law applies to all the companies ”that are listed in US stock exchanges”, including subsidiary companies [G Mu 08]. *the nature and characteristics of the use of IT systems affect a company’s internal control over financial reporting. Data about asset values [...] may contain errors that compromise the accounting systems. In this case, the valuation of a company would be affected and thus the financial reporting*[G Mu 08]. ”The second influence of IT systems on the financial report is the value of IT systems itself and the security risks that are associated with their employment.” Dervining from this ”companies have to execute IT audits that require risk management”[G Mu 08].

Basel II & III is another ”regulatory requirement that makes the management of security risks for banks necessary” [G Mu 08]. It requires the affected companies (banks) to identify security risks and treasure enough solvency to assure the stability of the company, for the case these risks occur. However, the selected company is not yet affiliated with the banking sector and thus is not affected by Basel’s regulatory requirements.

Solvency II The selected company is affiliated in the insurance sector and thus affected by Solvency’s regulatory requirements. It was introduced by the EU and serves to purpose to strengthen the insurance protection and uniform competition standards in order to ensure a uniform control practice within the EU [Bund 14]. According to [Bund 14], can the control system be divided into three ”pillars”: Quantitative regulation, governance instructions and reporting & disclosure duties. The quantitative regulation comprise rules for the assessment of financial assets to ensure that enough equity is treasured. The governance instructions stipulate a risk control function, a compliance function, insurance mathematical function and internal revision. The reporting & disclosure duties consist of the Solvency and Financial Condition Report (SCFR) and Regular Supervisory Report (RSR) and contain mostly qualitative information about the company. The reports are of no further importance for the market data and are thus not further explained, but mentioned to demonstrate the elements that were evaluated as requirements for market data. For the selected company are the quantitative regulations, in particular the rules for the assessment of financial assets, of high importance, as well as the risk control function, the compliance function and the internal revision. This thesis is in fact a concept, which originates in the need for a compliance function and internal revision in form of audits.

There exist many regulatory requirements, but they are mostly not directly linkable to market data. The only requirements that can be derived from those laws is that

the appropriate data needs to be used. Everything else is functions that need to be implemented within the company, but are not specific enough regarding data.

In Germany exists the MaRisk, which is a legal document containing the "minimum requirements for risk management." It has "direct impact on the technical-organizational equipment" [B Wa 14]. MaRisk defines four security criteria, which IT systems need to follow: Integrity, availability, authenticity and confidentiality. A big problem with MaRisk is, that "few general principles [...] are mentioned", but it "lacks of detailed requirements and concrete counter measures for risks in the context of IT" [B Wa 14]. This is why [B Wa 14] suggest "a method to derive the requirements from audits."

These laws focus a lot on risk management, which is not in scope of this thesis, but would be a highly interesting topic for further research and is certainly of high importance for companies operating in the financial sector. The reason for risk management not being in scope is the focus on market data. Market data certainly needs to be considered for risk management, but constitutes only a small portion of risk management.

3.2 Regulation

After an unsatisfactory requirement gathering from laws, a more practical approach was made. The regulations were derived from the requirements from audits, the method suggested by [B Wa 14]. The majority of the performed audits were internal audits and were strongly focused on the ICOFR [Brau 10] concept. But the only identifiable influence the ICOFR audits had regarding market data was, that processes had to be defined. Those processes only depict the selected company's system only unsatisfactory and is not being utilized. They were performed on a yearly basis. Other audits included audits from consulting companies, such as a application security audit. The vendors performed only very few audits in the past, but the selected company expects the numbers of vendor audits to increase significantly in the future. Due to the small number of past vendor audits, the license agreements from the vendors were studied, in order to derive requirements with sufficient accuracy.

3.2.1 License Agreements

License Agreements are mainly used for the issue of use of patents, brands, know-how or software. It is not defined in the German civil code, the BGB, as its own agreement category, but rather as an agreement of its own kind. With this agreement the owner of an protected right issues a defined right of use to the licensee.

Via license agreements single or exclusive rights can be granted. Central points are the description of the license artifacts, the definition of use and spatial validity, the duration, the fee, the terms of use and if applicable contract penalty for noncompliance. Legal consequences can include:

- Sustainable data exchange cannot be guaranteed without appropriate license agreements
- By noncompliance with the terms of a license, the vendor is authorized to instantaneously stop the data delivery
- In almost all cases is the vendor authorized to conduct an audit in order to verify the compliance with license agreements
- Most of the license agreements with vendors are based on American law, therefore court of jurisdiction and place of fulfillment are usually in the USA
- By noncompliance is the licensee liable for the resulting damage for the vendor

Possible license requirements can be categorized by content, by use as well as by source and can include:

1. By content:

- Singular products
- Indexlevel and/or Constituents
- Special Data such as sectors, countries. GICS, ICB)
- Frequency: daily, monthly, realtime, EOD, Midday

2. By use:

- Data-, Reporting-, Vendor-, Distributionslicense
- Within reports, within systems, in raw data format

3. By source:

- Supplied by third-party supplier
- Supplied by different systems

These requirements vary greatly depending on the incoming data and the vendor. This breakdown of the license requirements was created by studying the company's license agreements and by interviewing three key employees, who are working in the procurement

management.

Of interest is what information needs to be available to procurement management for the formal acquisition of market data from the data vendor:

- Client Information
- Data Vendor
- Data Type
- Format: Delivery to DWH or by file
- Delivery Timing
- Request Process
- Cost
- Usage of Data
- Licensing
- Additional Information

Without this information, no formal acquisition is initiated by the procurement management. If the initial information is available, collected and forwarded to the procurement management, the market data is being ordered from the vendor. The vendors mainly have two different ways of delivering the data. The **Back Office Data** is a standard package pre-defined by the vendor. There might also be the option for an extended back office data subscription available. Overall is the scope of the data packages narrowed down by the vendor. If additional market data is required or wanted, then the data needs to be separately requested. This is called **Per-request Data**. This differentiation is needed, because different license requirements can exist.

3.2.2 External Audit Requirements

The vendor's audit requirements as seen in 3.1 were collected by studying the documentation of performed vendor audits in the past.

Table 3.1: List of Vendor Audit Requirements for single application

#	Requirement	Description
R-1	Application name	What is the name of the Application
R-2	Application server ID	What is the ID for the Application Server
R-3	Owner & Supporter	Who supports/owns the Application
R-4	Location of applications server	List address where the application servers reside
R-5	Security measures	What security measures are in place to ensure only authorized access
R-6	Data sourcing	Is the data source a datafeed, flat file, database, other application, spreadsheet, other source
R-7	Data access	Can the data be accessed or viewed?
R-8	Data type	Data type meaning Equities, FX, etc.
R-9	Data frequency	Frequency as end of day, delayed or intraday data
R-10	Data storing	Duration & type
R-11	Display of data	Website: open or intranet
R-12	Export of data	To other application
R-13	User authorization	View or access
R-14	User list	For each application user: Application ID, User Name: First and last name of the end user, Location
R-15	External users	Yes/no
R-16	User location	Location where the user resides
R-17	RIC list	Provision of a list of all unique instruments used by the application

Table 3.2: List of External Auditor's Concerns

#	Concern	Description
C-2	Market Data Flow	Where does the market data end up and how many instances do exist?
C-3	Compliance to requirements	Requirements as described in table 3.1
C-4	Information source	Whom to contact for certain problems or questions regarding market data for completion of audits

Table 3.3: List of Application Security Audit Requirements

#	Requirement	Description
R-18	Data input & change	Needs to be logged, see R-20
R-19	Deletion of data	Only allowed for some needs to be logged, see R-20
R-20	Logging	Each entry needs to include the user ID, System Component, Timestamp, Event type and result

Vendor Auditor

The stakeholder vendor auditor is an external auditor, who demands insight into the company's system to test for compliance to the license agreements. Those audits are not yet performed on a regular basis, but have increased and might be performed regularly in the future.

The Vendor Auditor's Concerns are collected in table 3.2.

Application Security Auditor

Within the company there were also application security audits performed, which are included for the sake of completeness. The result of those is a catalogue of requirements, which need to be adhered to for every new application. The requirements regarding the use of market data exist in the application security audit requirements as seen in table 3.3. There are a lot more requirements, but the requirements presented are the only ones regarding market data.

Table 3.4: List of relevant Stakeholder

#	Name	Description
S-1	Management Board Member	Duty of legality, duty of care in a narrow sense and monitoring duty
S-2	Internal Auditor	Compliance to existing regulations & standards
S-3	Procurement Manager	Interface between vendor & company; Formal acquisition of market data
S-4	Application Owner	Front-end or back-end applications; Essentially creating reports or analyses
S-5	Account Manager	Interface between customer & company; Managing sales and building customer relationship
S-6	Database Engineer MDP	Change of MDP processes & exports
S-7	Operation Manager MDP	Run of MDP processes & support in MDP
S-8	Database Engineer DW	Change of DW processes & exports
S-9	Operation Manager DW	Run of DW processes & support in DW

3.3 Company internal stakeholder

The List of Stakeholder was created on the basis of the described system, as well as employee tasks and affiliation of the companies divisions. In table 3.4 are all the stakeholders listed, who are somehow involved in the lifecycle of the market data and are organizational roles.

Those stakeholder have different concerns regarding the use of market data within the system. A concern represents a goal "which is to be achieved" for certain stakeholders [S Bu 08]. The management board and auditor roles serve as control instances.

Management Board Member

The Management Board Members are the topmost people responsible. In a survey [C Ge 13] for the European Commission, the Management Board Members' duties and liabilities were listed. For Market Data are three duties can be distinguished:

Duty of legality "requires the directors to abide by the law while acting for the company" [C Ge 13]. These requirements have an "internal aspect" requiring the

Table 3.5: List of Management Board's Concerns

#	Concern	Description
C-1	System Understanding	Creation of an understanding of the system for multiple purposes
C-2	Market Data Flow	Traceability of the flow of market data
C-3	Requirements adherence	Adherence to legislation & regulation.

directors to abide by the Stock Corporation Act, the articles of association and the by-laws and "external aspects" from "various statutes other than the Stock Corporation Act"[C Ge 13]. The market data is affected by statutes such as Basel III, SOX and

Duty of care in a narrow sense can be divided in four aspects [C Ge 13]:

1. Duty to plan and control
2. Organizational responsibilities
3. Financial responsibilities of the directors
4. Informational responsibilities"

Market data is involved in the second and fourth aspect. The second aspect, the organizational responsibilities, require the directors "to establish an organizational structure that is in compliance with the applicable laws and articles of association"[C Ge 13]. The fourth aspect, informational responsibilities, "require them to set up structures that ensure the flow of information"[C Ge 13]. This thesis originates from this aspect.

Monitoring duty "have a horizontal and vertical aspect." The horizontal aspect "refers to the overall responsibility of the board members to monitor all business transactions." To perform the monitoring duty "relevant information" needs to be obtained and accessed [C Ge 13]. The vertical aspect "refers to the delegation of duties." This means that the board member is "required to choose the person to whom the duties are delegated carefully, instruct and monitor the person"[C Ge 13]. The solution product of this thesis helps the management board members to fulfill the monitoring duty by providing the means and models to extract information about market data from the system.

The Management Board Member's Concerns are collected in table 3.5.

Table 3.6: List of Internal Auditor's Concerns

#	Concern	Description
C-4	Process & Role Uniqueness	Are processes & roles defined and well defined from one another?
C-5	Risk Prevention	Are risks prevented through controls?
C-6	Control Execution	Are the controls carried out as defined?

Internal Auditor

The third stakeholder, the internal auditor, tests whether the company abides to existing regulations and standards. This internal auditing is performed by an Internal Audit Function, which administrates a key function within the control system.

Monitoring and audit processes: *audit processes, in contrast, may have very different frequencies that range from an event-driven instantiation to several instantiations per month or year to a continuous (since automated) execution. Also, audit processes differ from business processes in that they are specifically designed to run “outside” of a firm’s regular operations with the intention to control a particular “audit object” such as a business process, a record of transactions etc.*[S St 10] Risk assessment areas of IT architectures in legal audits from the selected company:

1. IT strategy
2. IT risk management
3. IT revision
4. IT outsourcing
5. Emergency management
6. Application development
7. User authorization

The Internal Auditor's Concerns are collected in table 3.6.

Procurement Manager

The Procurement Manager's duty is the formal acquisition of market data. The role serves as the interface between the company and the data vendor, which negotiates the license

Table 3.7: List of Procurement Manager's Concerns

#	Concern	Description
C-7	Utilization of Market Data	Is all the market data actually needed?
C-8	Data Payer equals Data Obtainee	Is the OE paying for the market data also the one receiving the market data for correct cost application?
C-9	User authorization	Do appropriate measures exist to prevent misuse of market data?

Table 3.8: List of Application Owner's Concerns

#	Concern	Description
C-10	Support inquiry	Whom to contact for support inquiries?
C-11	Data quality	Is the data being used correct?
C-12	Impact Mitigation measures	What measures are in place for for delayed or missing data and to prevent server outages?

agreements.

The Procurement Manager's Concerns are collected in table 3.7.

Application Owner

The fifth stakeholder, the Application Owner, is the role responsible for the front-end or back-end applications in the DW. Those applications are used to create the reports or analyses for the customers in the applications. Front-end applications are directly accessible by the user, unlike back-end applications, which have no graphical interface for the users. They are mostly utilized by technical users for automated processes.

The Application Owner's Concerns are collected in table 3.8.

Account Manager

The sixth stakeholder, the Account Manager, is the interface between the company and the customer. This role manages the sales, as well as builds and maintains a customer relationship. It is the contact person for information about the customer and

Table 3.9: List of Account Manager's Concerns

#	Concern	Description
C-13	Data Quality checks	Are exported data packages in good data quality. Is data complete? Is data retrieved from vendor without any business errors? What data quality checks are performed?
C-14	Timely delivery of requested data	How long does an export take? If data is erroneous, how long does it take to restore the data? How is information flow and communication towards back end processes?
C-15	Transparency of Market Data System	Are there any monopolies on knowledge concerning the system set up? Is there good governance and due diligence? How are errors communicated? How are changes communicated which affect back end systems?

its acquisition of market data or service products.

The Account Manager's Concerns are collected in table 3.9. Additionally was of interest the stability of the system. Outage of server platforms (How often do outages occur? How fast is the reaction time if an outage occurs? What are critical dependencies?

Database Engineer – MDP

The Database Engineer performs changes in the MDP-processes. This includes the development of new applications, adjustments to existing applications for optimization or quality reasons, inclusion of new vendors or deliveries, as well as exclusion and changes in the processing of market data. Other infrequent tasks regarding the software or hardware such as migration to a new platform is also in the scope.

The Database Engineer's (MDP) Concerns are collected in table 3.10.

Table 3.10: List of Database Engineer's (MDP) Concerns

#	Concern	Description
C-16	Process Performance	Is the performance of the processes good enough. Good enough meaning whether the processes finish on time.
C-17	Regulation adherence	When developing a new application/tool, does it adhere to company's user authorization policy, license agreements and other restrictions?
C-18	System Understanding	Means to create a system understanding for other non-technical users.

Table 3.11: List of Database Engineer's (DW) Concerns

#	Concern	Description
C-19	Market Data Interface Change	Does a change in the market data's content or logic (e.g. new calculation method for index weightings), need to be considered in the existing Interface?
C-20	New requirement MD Interface	Does new market data need to be provided by the system (business-driven). New types, as well as new vendors. How does the specification look like?
C-21	User authorization	How is the user authorization implemented in the data model? Do restrictions/license requirements exist and need to be adhered to during implementation?

Database Engineer – DW

The Database Engineer performs changes in the DW-processes. The role doesn't differ much in the scope and tasks from the Database Engineer in the MDP. The only difference is the system each role is active in. Additionally is the DW more closely connected to the clients than the vendor.

The Database Engineer's (DW) Concerns are collected in table 3.11. Some more information demands of the Database Engineer (DW): What is the performance of Market Data Processes? How long are you allowed to keep the history of the data? How are data correction deliveries supposed to be handled?

Table 3.12: List of Operation Manager's (MDP) Concerns

#	Concern	Description
C-22	Data deliveries	Is the data delivery erroneous, too late or missing? Behaviour of the processes during holidays
C-23	Interface maintenance	Market data interface needs to continuously cope with changing market data (Type of delivery, scope of the delivery or market specific adaptations)
C-24	Transparency for the client	Transparency for the client in the provision of data. Where does a value originate and what is the meaning of it?

Operation Manager – MDP

The Operation Manager monitors whether the MDP processes were successful for each valuation date. Their main concern is that the market data is delivered up-to-date, correct, complete and in time by the vendor. Very important for them is to understand how the data is processed, as well as how it is conditioned, but the data is delivered of vendor's own choice delivered and it is difficult to standardize.

The Operation Manager's (MDP) Concerns are collected in table 3.12. Additional information demands: Information policy of the vendors: Every vendor communicates differently about its deliveries. Whether changes occur, when they are performed and how they are communicated differs from vendor to vendor. Storage of data: Storage of mass data and possibility to easily access this data on a daily basis. Auditability assured by saving of historic data, which leads to the need of time/storage efficient archiving for huge amounts of market data.

Operation Manager – DW

The operation manager in the DW monitors whether the processes in the DW were successful for each valuation date. This includes the main

The Operation Manager's (DW) Concerns are collected in table 3.13.

Table 3.13: List of Operation Manager's (DW) Concerns

#	Concern	Description
C-25	Process stability	How stable are my market data processes? Does a process interruption cause the risk, that depending services are supplied not at all or delayed?
C-26	Data quality	Where is market data missing or incorrect? What is the cause?
C-27	Scalability	Does the process still run in a timely manner, even though the volume increased?

3.4 Interim Conclusion

The requirements should be applicable to most of the companies in the financial reporting and analysis field using market data and therefore having reference character. The collection is based on compliance reviews of some of the biggest data vendors worldwide, on internal standards within a dominating corporate group and on laws and regulations in the United States of America, as well as Germany.

The requirements are very specific. Information about each application with its hardware and user, information about the data and information about the users. Due to the specific information requirements some kind of method is needed to gather all the information in a timely manner and as little effort as possible. This is made possible by the use of an information architecture.

The concerns of the company's internal stakeholder are very specific as well and creating an IA on the basis of these concerns is often not possible. Therefore a selection of the most important and solvable concerns the IA will solve is presented in the next section.

Chapter 4

Information Architecture

An information architecture (IA) always has a goal that needs to be achieved. The goal of this IA is to get the needed information regarding the processes handling market data to enable the report generation for audits. The requirements represent the information demands of the auditors and other system stakeholder who ask for insight into a company's system. The major problem is that the architecture is either one, big black box from a top down view or one small white box with many black boxes for the individual workers in the departments. This IA tries to give a guideline for those demanding insight into the system's use of market data. It is supposed to have a reference character and should be usable for companies that offer analysis and reporting services in the financial industry. In [S Bu 08] Methodologies are allocated into EA topics. However, this thesis does not have any allocation, mainly due to its small scope as an IA. This IA covers only the information needs for the legislation and regulations, as well as its system stakeholder, regarding market data and not of the whole enterprise and its systems. Furthermore were no EA management frameworks, like Zachman or TOGAF," used for the creation of the IA, since they "are usually either too abstract and therefore not implementable, or too extensive to be used in real world" [S Bu 08], which is the case for this IA.

4.1 System Overview

Using the System Overview figure 4.1 as a basis, the selected company's system can be divided into three large system components: The market data pool (MDP), the data warehouse (DW) and the subsequent application landscape (AL) using the data. Due to the size, complexity and independency of these system components the market data pool

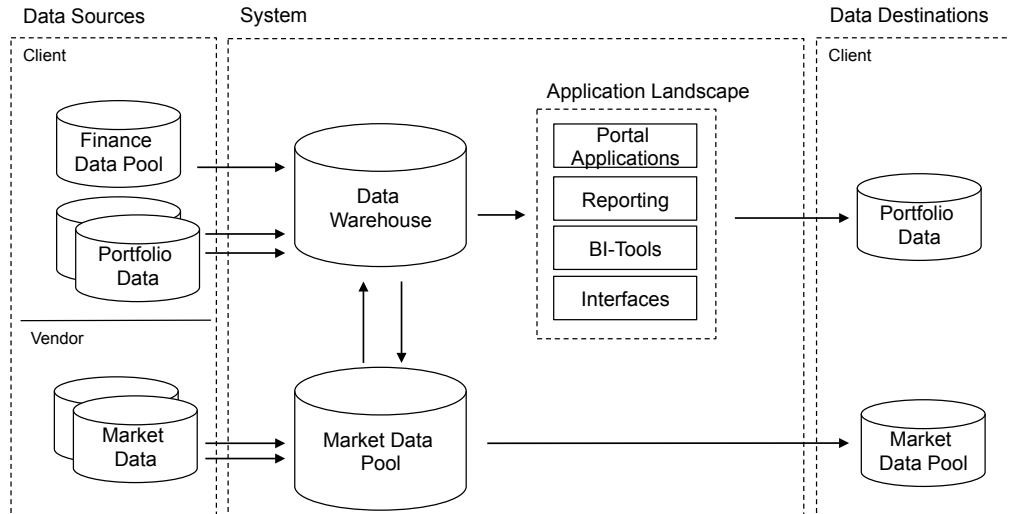


Figure 4.1: System Overview Diagram of Selected company

and the data warehouse receive their own system overview. The subsequent application landscape does not receive its own system overview, because its heterogeneity it can hardly be abstracted and the applications are not really of interest for the requirements regarding market data as presented in the previous chapter 3.

The market data pool primarily collects the market data from the data vendors and stores the data into a time series DB. After some post-processes the data is being exported to clients. The DW can be seen as a client, because the same conditions apply as to other clients. If a client receives the raw market data from the market data pool, the client is performing his own reporting and does not use the provided reporting service by the selected company.

The DW serves as data basis for the subsequent analytical tools. It combines and enhances the client's data with market data, if publicly available and creates a trusted data basis. The system stakeholder database engineer in the DW is responsible for the creation of applications and implementation of changes. The system stakeholder operational manager

in the DW is responsible that all the processes were running correctly for each valuation date and solve occurring problems.

The applications in the application landscape use the data from the data warehouse and are mostly mapped to a DM and a department. The system stakeholder application owner is responsible for the application. This includes the initial creation, changes, migrations and the daily processing.

In the following are the system components described in more detail.

4.1.1 Operational Database – Market Data Pool

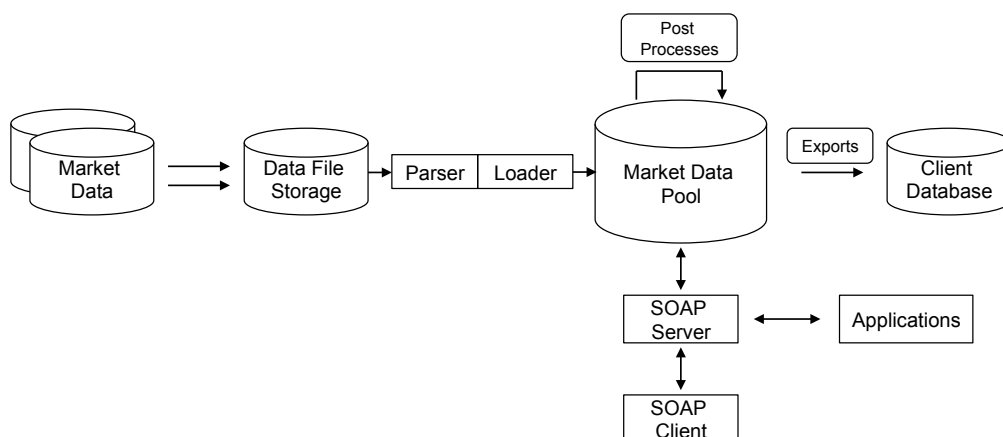


Figure 4.2: System Overview of selected company's market data pool

The day-to-day business for the market data pool consists of loading raw data files into the database, several post-processes and the export of its data. To do this in the most efficient way possible an online transaction processing (OLTP) database application is used. OLTP database applications are characterized by only having to process a limited

amount of data for the execution of a transaction [A Ke 06]. The transactions in this case are the daily market data files, that need to be loaded into the market data pool. OLTP applications realize the day-to-day business [A Ke 06], among which these transactions can be counted.

As in figure 4.2 displayed, are the data sources for the market data market data pool in the selected company consists of multiple file based DBs, each DB representing a certain type of market data, which can also be divided into smaller parts and is separated by vendor. For example is the equity data from a Back-office package divided into five regions (euro, namr, lamr, asia1, asia2) and therefore loaded into five different DBs. This division serves mostly the purpose to be able to process data parallel and therefore faster.

The operational database in the selected company is a time-series database, which has two data types:

Time-series data is data which is updated periodically, mostly daily or monthly. So even though the value at a certain time is the same as the value on the consecutive period, the value is still updated. Examples for time-series data is price data for equities, where it is necessary to know the price for each day, month or some other time period.

Static data is data which is only updated, if the value changes. This is mostly descriptive data, such as name, amount issued or industry sector.

Whether a fact is time-series or static data is decided by the developer, when a new file is being set up. The operational DB contains all market data the selected company gets from bulk or per-request data.

Data Sources

The market data is collected from data vendors or from stock exchanges. The data collection process is initiated by uploading a data request file onto the data vendor's server, which triggers the generation of the requested file. After the requested file is generated the content on the vendor's server is mirrored with the content on the selected company's server. This mirroring only occurs directional from vendor to company and not the other way around. Therefore mirroring means that if new content on the vendor's side is available, it will be copied into the system. The incoming new or updated file is saved in an initial directory, for incoming, raw vendor data files. Ensuing, the file is decoded if necessary, zipped and the name is normalized. After completion of these tasks it is archived and copied into another directory, where it is prepared to be loaded into the

system.

The interesting part for this topic is, what data is requested and what data is the company allowed to request, most of all for which customer. Data vendors offer two ways to order market data:

Back-office data is a standard package, where the vendor delivers you a set of market data from its universe. The set is determined by the vendor and is final.

Per-request data is data which is not included in the standard package and needs to be requested individually. Requesting the data doesn't refer to the previously described requesting in the data collection process. Per-request data is much more expensive than bulk data.

This data comes in a variety of many different file types, which is solely determined by the vendor:

CSV In Comma-Separated-Value-Files every horizontal line represents a data object. The data object can have 1 to n attributes represented in its line, as defined in the header and identical for each data object. Each attribute of an data object has a value, which is separated by a delimiter. A delimiter can be a comma(,), a semicolon(;) or a vertical bar/pipe(—). The header is always at the beginning of the file.

Initial Fundamentally a CSV-file containing the complete amount of data objects and its attributes. The Initial-File is used for the initial loading process.

Difference Fundamentally a CSV-file, but not containing the complete amount of data objects. Only the data objects, which have changed since the last delivery are included. E.g. on one day the file contains pricing information about asset A and asset B. On the next valuation Date, only the price for asset A changes, but not for asset B. Therefore the file contains only the pricing information for asset A for this valuation date. However, the file always contains the vendor's identifier.

The actual file format of initial or difference file is company specific. The biggest data vendor **Bloomberg** uses .out as initial-file and .diff for the difference-file.

4.1.2 Data Warehouse

"initial loading and periodical refreshment of the DW" [A Ke 06] But the system's purpose is not only to store data. The data is solely the basis on which analytical and reporting processes build upon. Since a OLTP DB's focus is on the transaction of

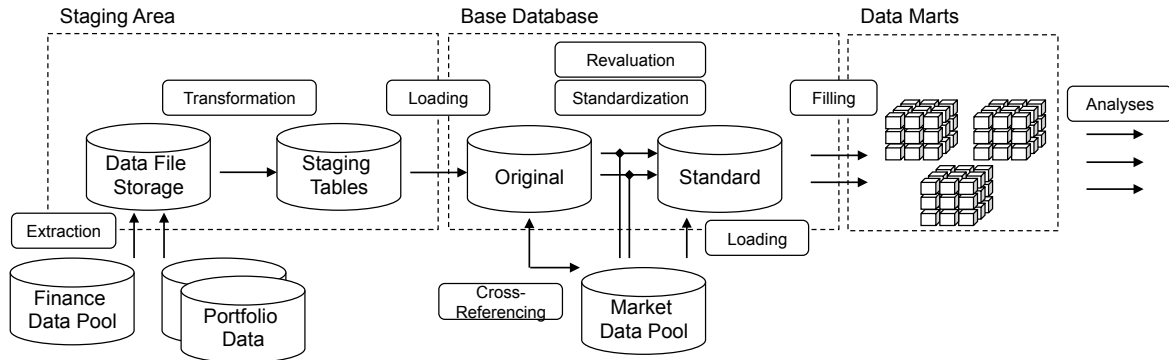


Figure 4.3: System Overview of selected company's data warehouse

data or throughput and not analysis, the need for an OLAP DB arises. OLAP is optimized for the analysis of its data content, but is not as efficient when it comes to transactions [A Ke 06].

A Data Warehouse (DW) "is based on a normalized data model in a relational database. It supports a diverse set of analytical requirements" [Krcm 10] and [M Go 10] and can be divided into two parts: The DW itself and multiple Data marts. The Data marts are "optimized for specific analytical purpose often using data mining and other techniques" [Krcm 10]. They contain a selection of the data from the whole universe of data within the data warehouse.

In the selected company the DW's data content roughly represents the customer's portfolio data and the data needed for analysis and report generation. It is explicitly talked about data generically, since the DW contains not only market data. This is where the definition of market data is not clear enough.

The selected company defines market data as "all data, which are purchased from data

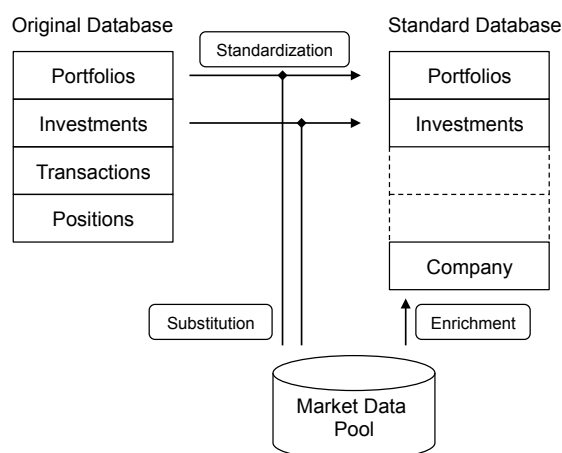


Figure 4.4: Overview of selected company’s Data Standardization Process

providers (Vendors) and/or stock exchanges (third-parties).” This is certainly correct, but the data coming from the customer, might also contain market data by definition. This classification is significant and will be important, when the data from vendors and data from the customers are merged.

The reporting and analyses for customer is grouped into divisions. For the most part it can be said, that they have one or more data marts or extract the data directly from the DW. They are also the one’s actually requesting the market data.

Reasons for the use of a DW, according to [Krcm 10]:

- Data from different data storage and computing systems can be analyzed with good performance as well as problems occurring due to the different storing of same data types can be solved.
- Data from outside sources can be embedded.
- Performance of the operative systems is not affected by the data analyses. The

transfer of the data from the MDP to the DW can occur during the times there is a small load processed by the hardware.

- The DW is stable compared the changes and extensions of the MDPs. If applicable, interfaces are to be adjusted.

In contrast there are risks as well [Krcm 10]:

- Quality of data is a major risk factor, since on the basis of the DW data important decision might be made.
- The establishment of a DW requires capital expenditures, whose benefit are hard to predict.

Data Distribution

The system has two endpoints/purposes:

1. delivery of market data
2. creation of reports & analyses

A data delivery, is defined as "the amount of data from a delivery system for one validity date" and executed in a periodical, defined time for a distinct customer.

4.2 System Stakeholder and their key concerns

The previously collected concerns in the subsection 3.2.1 License Agreements and section 3.3 Company Internal Stakeholder are not always answerable by an IA and some of them are not in the scope of this IA. That is why only a selection from these requirements were considered and were taken as concerns in the IA. Of the ten interviewed stakeholder, only six were identified as having an appropriate information demand regarding market data and its processes. Included in the IA as system stakeholder were:

External(Vendor) Auditor The external auditor from vendors as the main addressee of this IA was included, simply due to his importance and focus. The information demands gathered in form of requirements (table 3.1 in subsection 3.2.2) were previously collected and need to be answerable by the IA. Additionally, can the concerns of a vendor auditor be summarized into the following three concerns. For an description of the concerns refer to table 3.2 in subsection 3.2.2:

- C-2: Market data flow – is addressed by the data flow viewpoint.
- C-3: Compliance to vendor requirements – is addressed by all viewpoints.
- C-4: Information source – is addressed by the stakeholder viewpoint.

Procurement Manager The procurement manager was included since the role is the interface between the vendor and the selected company. Furthermore, are his concerns very important for the selected company due to their impact on costs and contract adherence.

- C-7: Market data utilization – is addressed by the structural viewpoint, data flow viewpoint and authorization viewpoint.
- C-8: Payer equals obtainee – is addressed by the structural viewpoint, data flow viewpoint and authorization viewpoint.
- C-9: User authorization – is addressed by the structural viewpoint, data flow viewpoint and authorization viewpoint.

Account Manager The account manager was included since the role is the interface between the client and the selected company. His concerns are highly influenced by the client and can be interpreted as in agreement with the client. The client's focus lies in the data quality, but also that he receives the data on time. Furthermore, is the transparency for the selected company of importance for reasons such as data governance.

- C-13: Data Quality Checks – is addressed by the governance viewpoint.
- C-14: Timely Delivery of requested data – is addressed by the data flow viewpoint.
- C-15: Transparency of MDS – is addressed by all the viewpoints since transparency is very generic. For different transparency intentions, different viewpoints can be selected.

Database Engineer (MDP) The database engineer (MDP) was included due to the role's great extent of knowledge about the system. Of utmost importance is this stakeholder for the evaluation of the IA, since this roles possess the most knowledge about the system. The regulation adherence is of utmost importance for the selected company and a system understanding could solve a lot of problems.

- C-16: Process Performance – is addressed by the data flow viewpoint.
- C-17: Regulation adherence – is addressed by all viewpoints.

- C-18: System understanding – is addressed by all the viewpoints since an understanding is very generic. For the different purposes of the understanding, different viewpoints can be selected.

Database Engineer (DW) The database engineer (DW) was included due to the same reasoning as the database engineer in the MDP.

- C-19: Market data interface change – is addressed by the structural, authorization viewpoint and governance viewpoint.
- C-20: New requirement market data interface – is addressed by the structural, authorization viewpoint and governance viewpoint.
- C-21: User authorization – is addressed by the structural viewpoint, data flow viewpoint and authorization viewpoint.

Operations Manager (MDP) The operation manager (MDP) was included to have the requirements of someone dealing with the daily processing of market data and therefore including this knowledge is essential.

- C-22: Data deliveries – is addressed by the structural viewpoint and data flow viewpoint.
- C-23: Interface maintenance – is addressed by the structural viewpoint, authorization viewpoint and governance viewpoint.
- C-24: Transparency for the client – is addressed by all the viewpoints since transparency is very generic. For different transparency intentions, different viewpoints can be selected.

Not included in the IA were the following system stakeholder:

Management Board Member The management board member as stakeholder was included since this stakeholder is the one liable and ultimately assuring the compliance to legal requirements and regulations. However, the management board member delegates this responsibility down to his employees working in a specific system and component. Furthermore, are his concerns very broadly formulated and can in general be answered by the viewpoints.

Internal Auditor The internal auditor's focus was on risk prevention, processes uniqueness and control execution. Even though the risk prevention measures would be interesting, they were excluded due to their little direct impact on the market data itself other than their timely delivery and thus defined as out of scope. Control

execution was not added due to its broad formulation and identifying controls other than the presented governance controls were hard to identify. The actually defined processes of the selected company could not be added due to being under closure.

Application Owner The application owner's concerns were focused on the existence and the quality of the data, as well as the measures in place to prevent delayed data deliveries. These concerns do not differ much from the concerns of the operations managers in the MDP and the DW. Since they are the core systems receiving and providing the data it is better to focus on those. If their concerns are answered and can be satisfied by the system, they will also be so for the application owner, since in most cases the application owner only has an view on the data.

Operations Manager (DW) The operations manager concerns such as process stability and scalability of the processes can not easily be answered by an IA.

Application Security Auditor The application security auditor with his requirements (table 3.3 in subsection 3.2.2) was solely included to highlight the importance of logging all the necessary information.

4.3 Viewpoints

An "exemplary view is provided for the viewpoint,[...] together with some textual explanation." The viewpoint "proposes a way to present data stored according to"[S Bu 08] information models.

4.3.1 Structural viewpoint

The structural viewpoint identifies the static components of a system. A component is a "individual unit of computation"[IEEE 00] and can be anything from a database, an application or a simple script. Those components have connections in some form to other components. A connection is usually some kind of information flow, being data sent from one component to the other or a component triggering actions in other components.

The structural viewpoint is divided into two parts the MDP and the DW.

Description of structural viewpoint components in the MDP in figure 4.5:

Data Source The data source is a provided server by the vendor to enable the retrieval of market data.

Request Engine To be able to retrieve data from the vendor, requests need to be uploaded. The request engine generates those requests.

Raw Data Storage The raw data storage stores all the data files received from the vendor. It also prepares the files for loading by standardizing filenames, decrypting and zipping them. The data storage is simply a directory tree in the shared filesystem.

File Archive The files are archived after they are retrieved from the vendor's server. The file archive is simply another directory in the shared filesystem.

Loading Unit The loading unit processes the raw files and writes the file's data objects into the designated database.

Data Storage The data storage is the central component of the MDP and the actual database containing the complete, received MD with history. Due to the nature of MD the database is a time-series based database.

Database Archive The database archive is a backup of all the file-based databases, which is updated weekly.

Data Catalog The data catalog serves several purposes. It serves as component, logging all the processing steps. Furthermore, it serves in principle as a hash table for fast processing and accessing of data. It is also a channel that enables data viewing and access.

Post Processing To keep the ordered MD from the vendor at a minimum, a lot of post-processes need to be performed to derive this difference. Post processing includes the loading and logging of the necessary data into the data catalog, the cross-referencing with the client data and the building of the indices' constituents. Furthermore is the calculation of custom benchmarks and indices performed.

Data Provision The data provision is a component consisting of several interfaces for applications from the application landscape or even clients' applications to extract the data directly from the databases via back-end applications.

Data Client The data client uses the same interface as the data provision component itself and is an application that enables users or clients to look up MD in the MDP via Microsoft Excel.

Export System The export system generates exports for the clients, mostly in form of files.

Data Target The data target is a clients' desired destination for the data. It can be a file-storage or an application, depending on the communication channel.

Description of structural viewpoint components in the DW in figure 4.6:

Data Source The data source are the clients' servers, where the clients data is being retrieved from.

Raw Data Storage The data storage is a directory in the filed-shared system and contains the files, which were received or retrieved from the client.

File Archive The clients' data files are being archived, for reasons such as having a multiple layer backup, as well as a reference point for wrong data.

Transformation Unit The transformation unit takes the raw client data and transforms it into client data objects as defined by the target system.

Staging Tables The staging tables contain the transformed client data.

Original Database The original database is essentially mirroring the staging tables and represents the complete amount of client data the selected company receives. The clients' data consists of portfolio, position, investment and transaction data.

Standard Database The standard database contains the investment and portfolio data from the client. However, this data is being standardized with the market data from the MDP. Additionally is company data from the MDP added.

MDP-interface The Market Data Pool (MDP) interface contains all the information the DW receives from the MDP.

Cross-Referencing Cross-referencing is performed to create a connection from the clients' data to the vendors' data by assigning each a unique, custom identifier.

Standardization Standardization is the process of evaluating the clients' data and use the standard data, if it is not already used. For example is the clients' data being overwritten, if the investment trading data of another stock exchange other than the main stock exchange was used.

Data Enrichment Data enrichment is the process of adding market data to missing facts of the clients' data.

Revaluation Revaluation is the process of overwriting the client data facts with market data, if it is a better fit.

Performance Calculation Performance calculation is the process of calculating the performance of portfolios compared to a comparative value.

Data Mart The data mart contains a smaller portion of the client and market data. Often an application from the application landscape does not directly access the core tables in the base database of the DW, but rather its own data mart.

Data Target The data target is an application in the application landscape performing analyses, creating reports and offering services.

It is to be noted that the Standardization, Data Enrichment and Revaluation need to consider the quantitative regulations from Solvency II need to be adhered to, which states rules for the appropriate assessment of financial assets. VP visualizes information of the Structural IM.

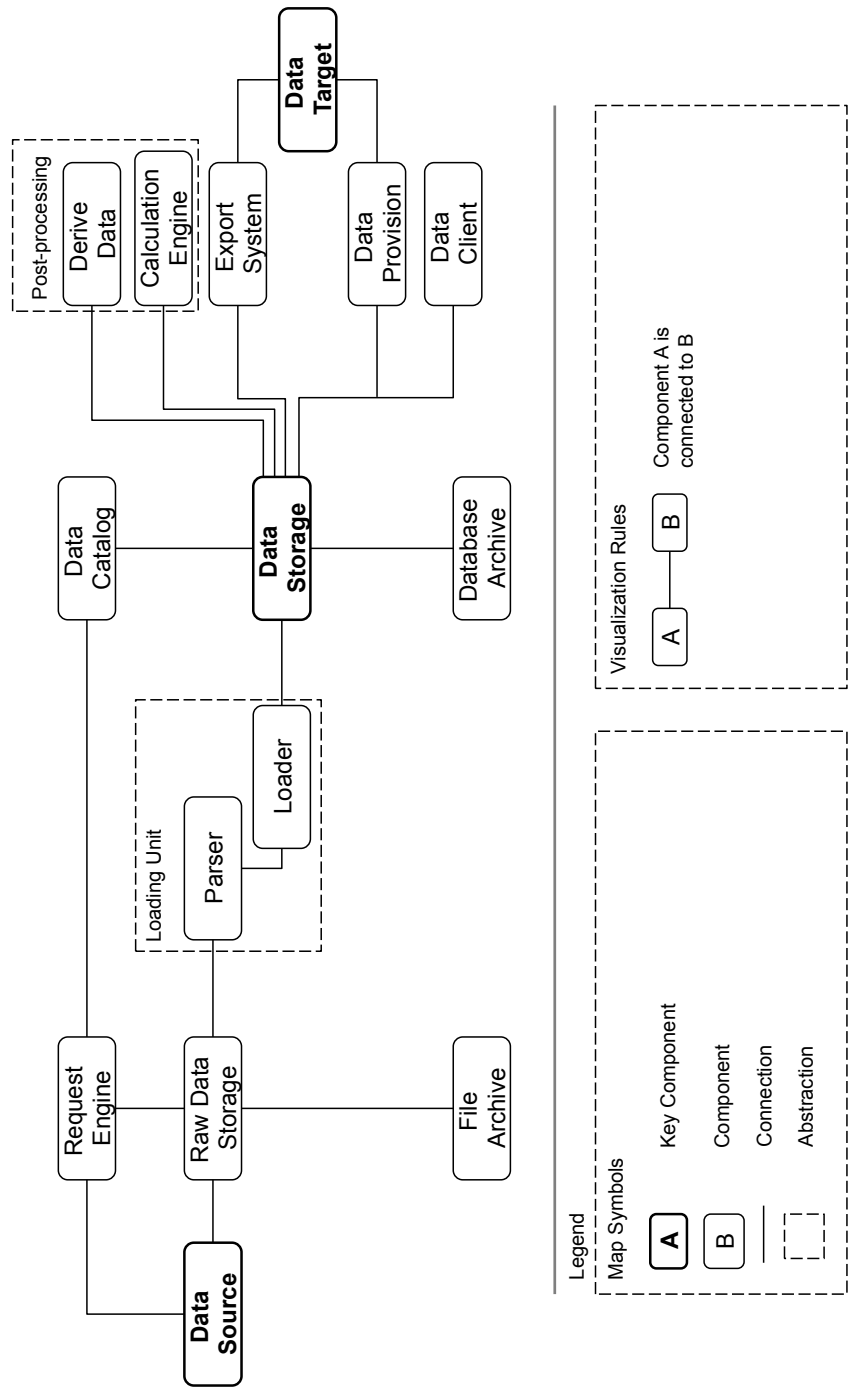


Figure 4.5: Structural Viewpoint on the Market Data Pool

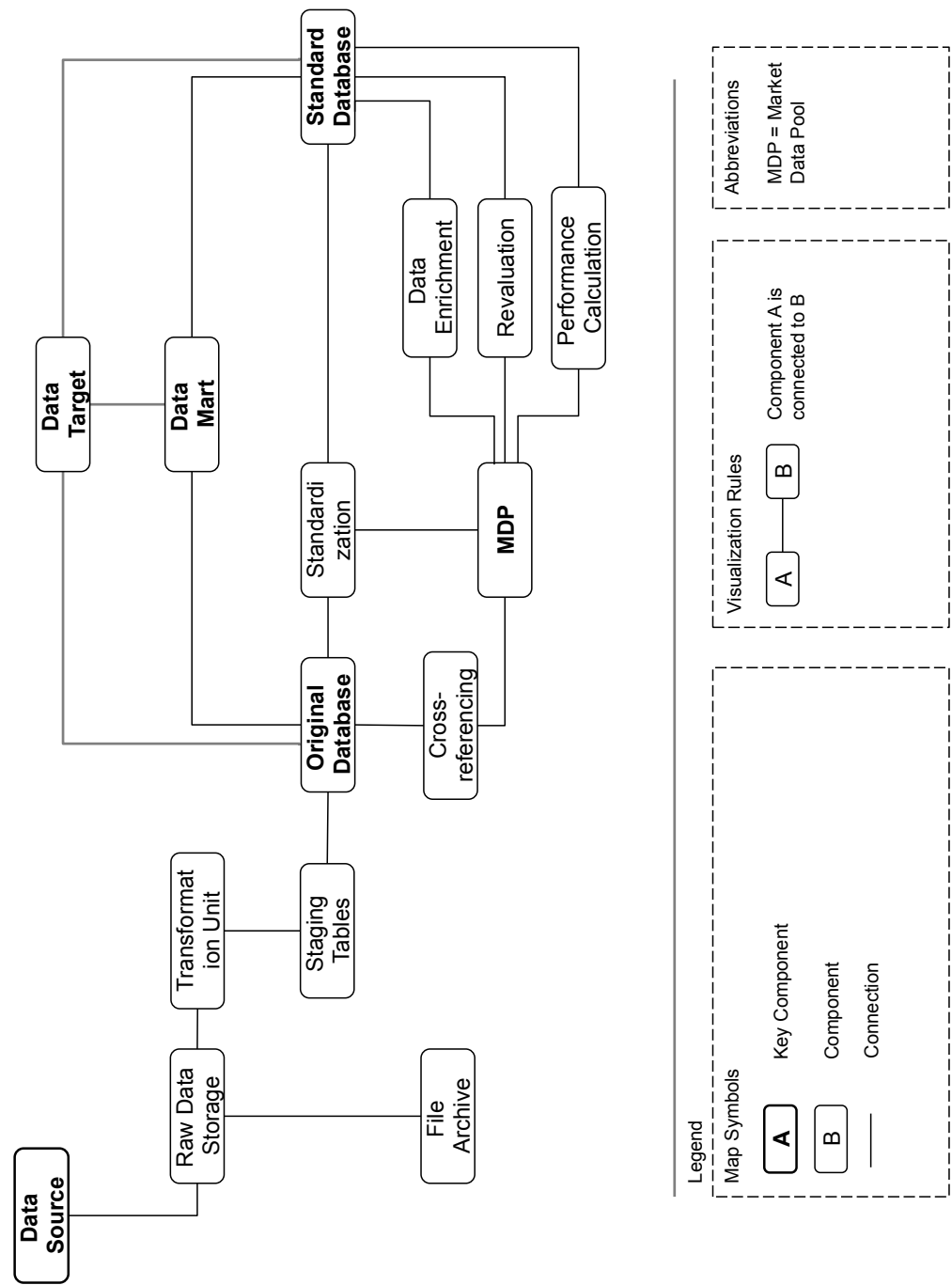


Figure 4.6: Structural Viewpoint on the Data Warehouse

4.3.1.1 Data Storage viewpoint

The data storage viewpoint is based on the structure viewpoint and depicts the structure components that permanently store client data or market data. This viewpoint enables the differentiating between components processing data and components storing data. An exception to the previous statement are the data source and the data target. Depending on the data source and the data target they might or might not fulfill both roles depending on the individual instantiation of the component.

VP visualizes information of the Structural IM.

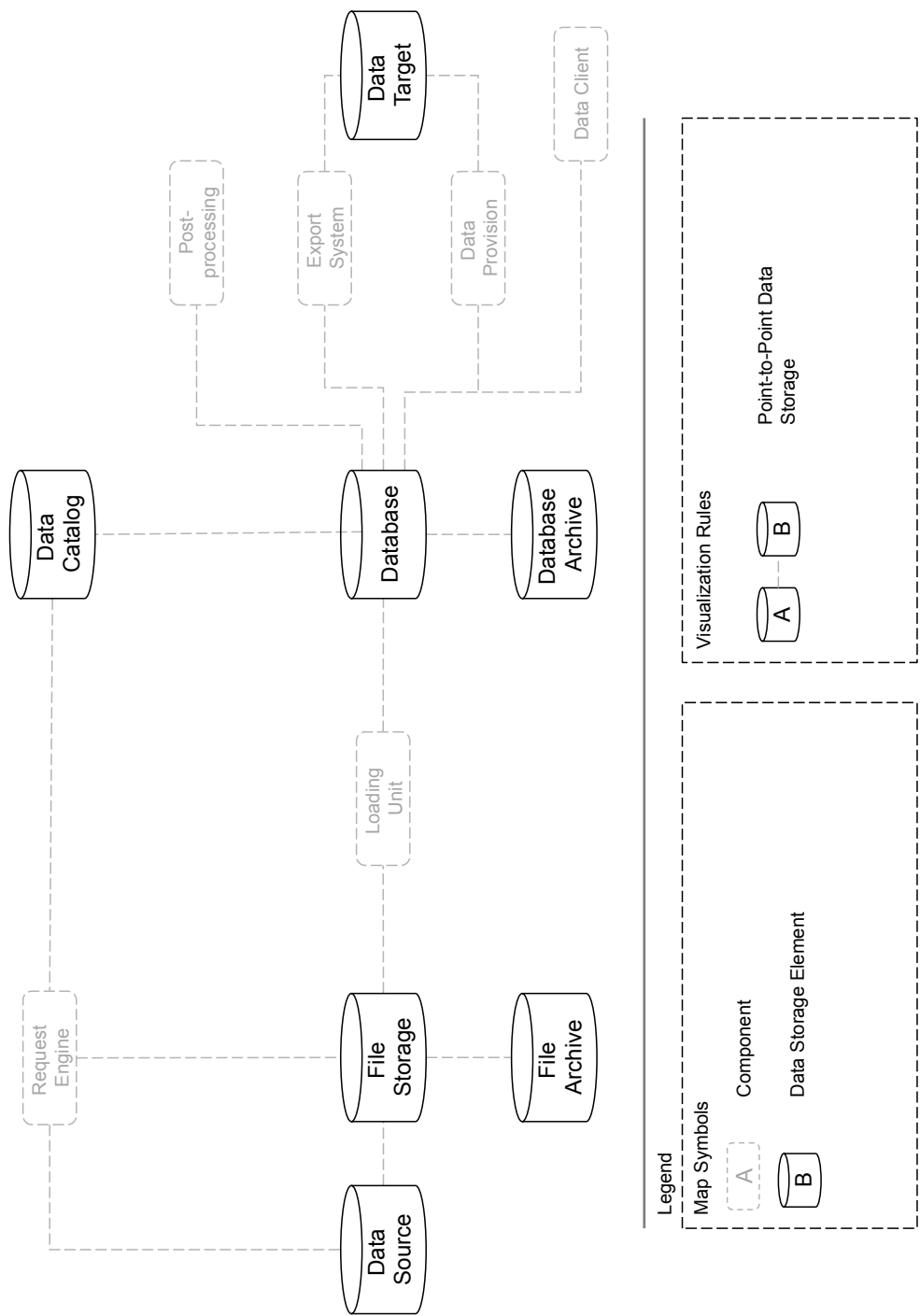


Figure 4.7: Data Storage Viewpoint on the Market Data Pool

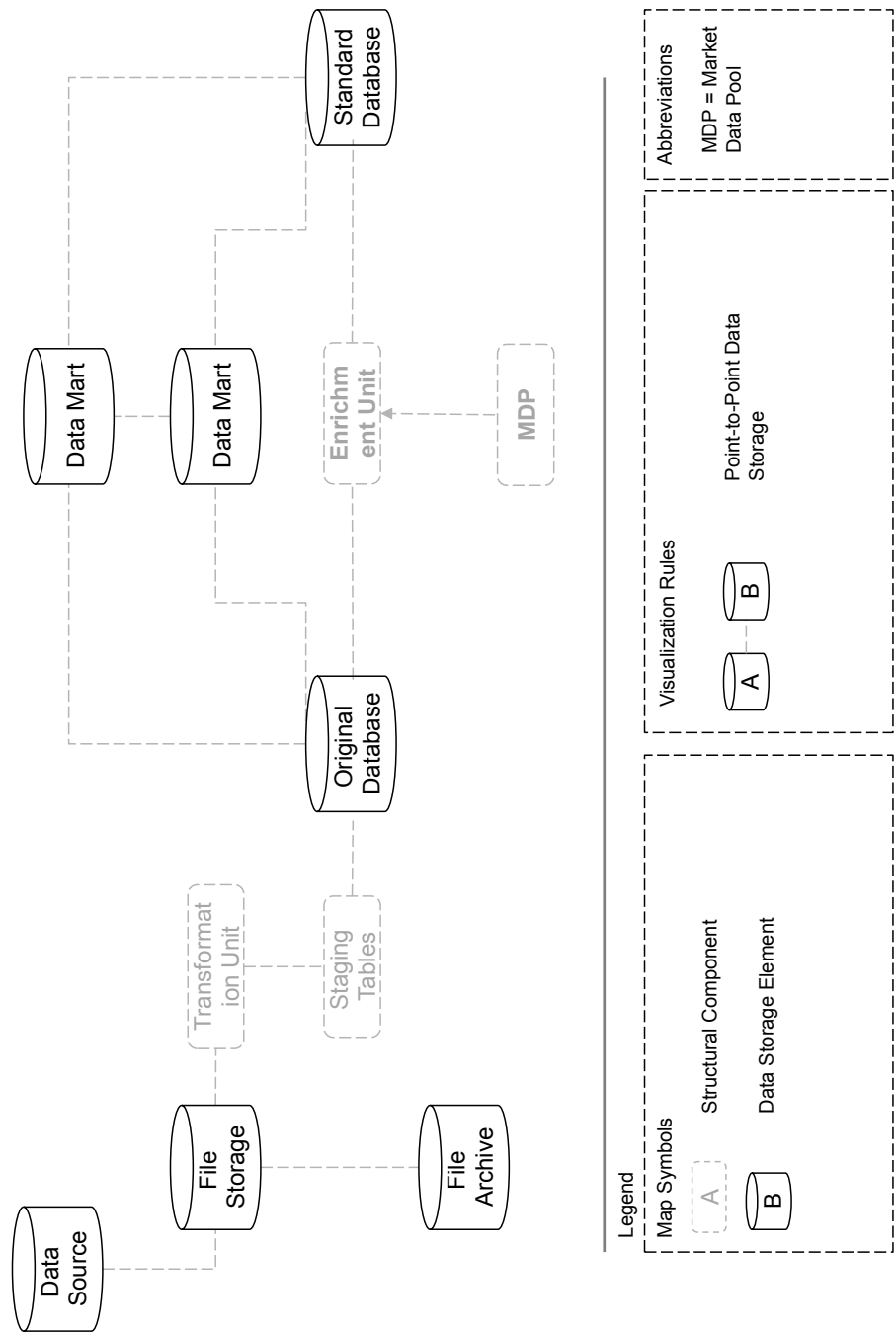


Figure 4.8: Data Storage Viewpoint on the DWH

4.3.1.2 Data Flow Viewpoint

The data flow viewpoint is based on the structure viewpoint and describes the data flow from the structural viewpoints' components and enables to create a point-to-point data flow report. The viewpoints contain a core data flow, which is performed every valuation date, as well as special event-triggered flows, which are performed independent from the core data flow. Corrections of market data or client data are not explicitly included in the data flow, but run through the same process as the initial data.

Description of data flow in the MDP in figure 4.9 (is based on figure 4.5):

- 1.Collecting** The first process starting the data flow is the collection of the necessary information by the request engine from the raw data storage and the data catalog.
- 2.Requesting** The requesting of the market data is performed by uploading the previously created requests from the request engine and uploading them to the data source to trigger the generation of the request market data.
- 3.Retrieving** The generated market data is retrieved from the vendor's data storage.
- 4.Archiving** After the retrieval, the market data is being archived.
- 5.Reading** The loading unit opens the files and reads them.
- 6.Loading** The loading unit loads the files' content into the data storage by creating market data objects.
- 7.Filling** The data catalog is then filled with information about the market data stored in the data storage.
- 8.Enhancing** Data Enhancing is performed by post-processing components to derive and calculate additional data from the original market data.
- (9.)Accumulating** The export system accumulates the portion of the market data, which is exported to clients after the enhancing of the market data. The data provision component accumulates the market data for the data target, if requested to do so. It is not dependent on the core data flow chain.
- 10.Exporting** Exporting is the process of sending the previously accumulated market data to the data target.

Archiving Archiving of the data storage is performed on a weekly basis on a specific date and time.

Providing The provision of market data is essentially the result of an users' query from the data client or the data target.

Description of data flow in the DW in figure 4.10 (is based on figure 4.6):

1.Receiving The first process starting the data flow is the receival of client data from the client and saving the client data files into the raw data storage.

2.Archiving After the initial receival of the client data files it is being archived for audit reasons and as backup.

3.Reading The transformation unit opens the client data files.

4.Writing The transformation unit writes the transformed client data objects into the designed staging tables.

5.Loading From the staging tables the client data objects are then loaded into the original database.

6.Evaluating The client data objects are evaluated by enrichment units for enrichment and correction of its facts. A fact being an attribute of the client data object.

7.Utilizing If applicable, the enrichment unit utilizes market data from the MDP-interface.

8.Improving If applicable, the enrichment unit improves the facts of the clients' data object by utilizing market data facts. It is to be noted that usually not whole market data objects are being utilized, but rather only a certain fact of an market data object. A client data object's facts can also be overwritten by market data object facts from different market data objects.

9.Accumulating The data mart accumulates its data portion from the original database and the standard database.

10.Analyzing The data target in the application landscape performs analyses for the creation of reports and offered services.

VP visualizes information of the Data Flow IM.

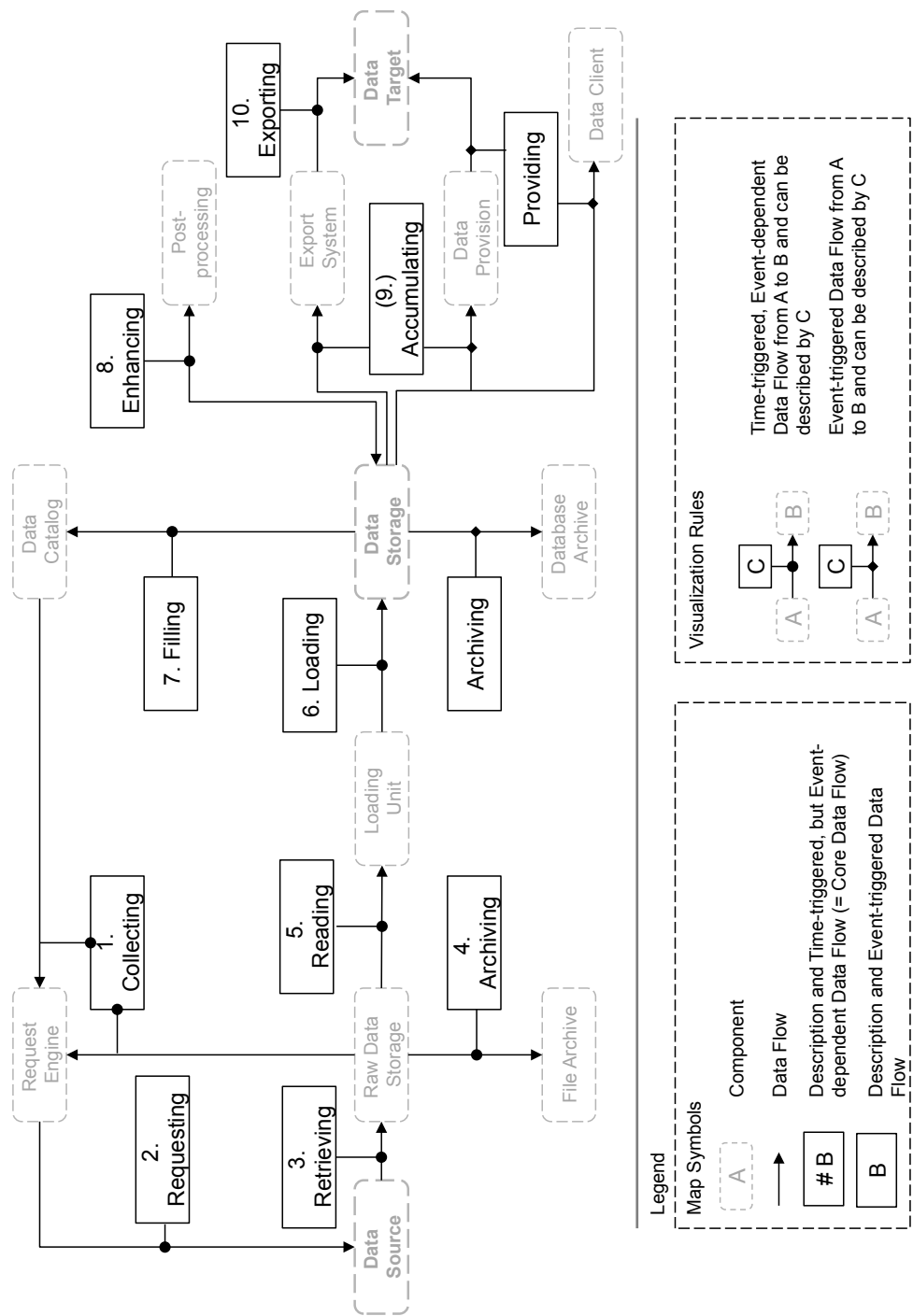


Figure 4.9: Data Flow Viewpoint on the Market Data Pool

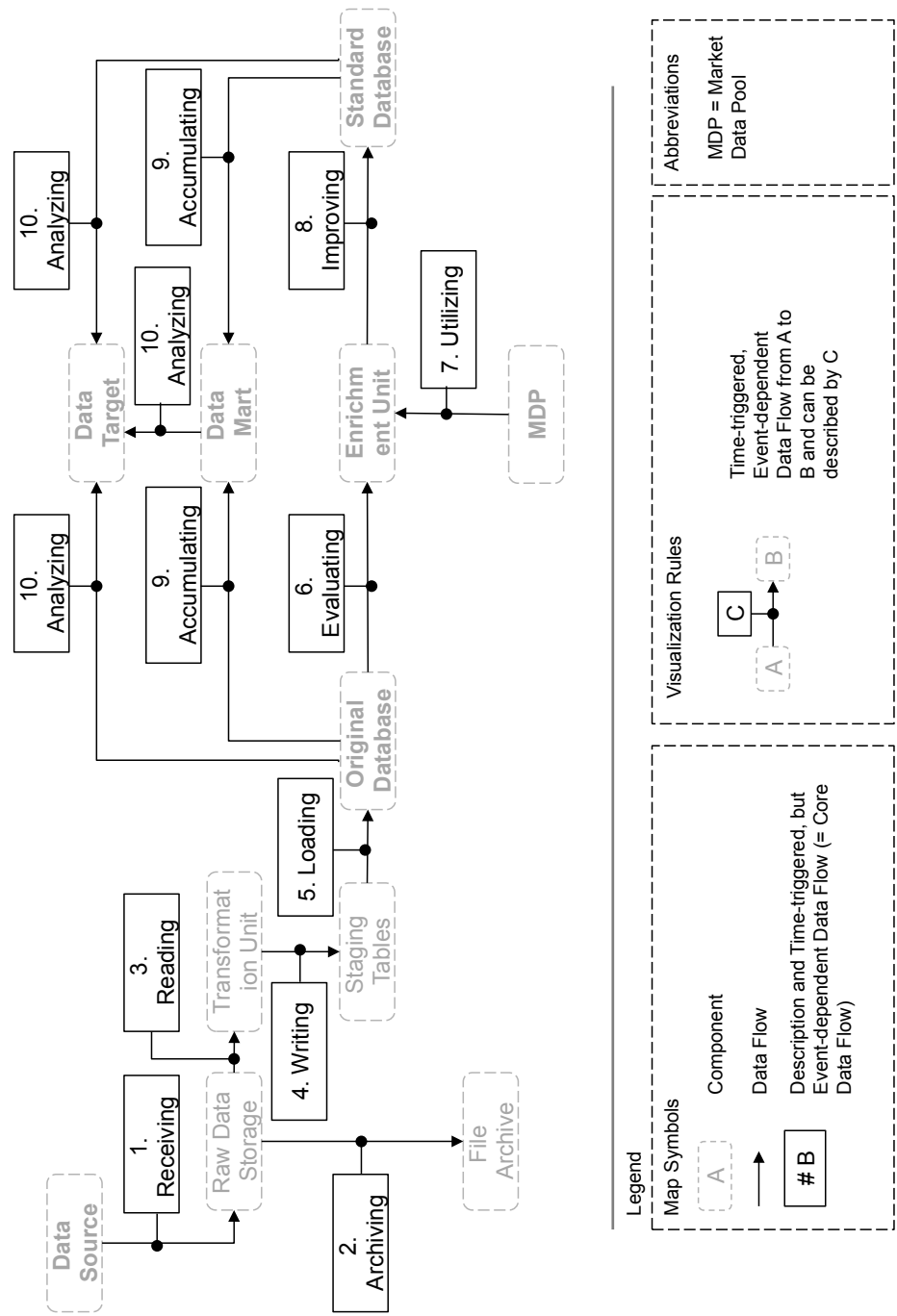


Figure 4.10: Data Flow Viewpoint on the Data Warehouse

4.3.2 Infrastructure viewpoint

The infrastructure viewpoint illustrates the hardware utilization of the MDP, DW and Application Landscape. In the selected company the MDP has its own environment and the DW and the Application Landscape share an environment, but is respectively larger. Generally, it can be said that every system has three different system environments:

Production The production environment is the core processing environment, where all the daily processes run in.

Test The test environment contains all the component's logic with changes in the process of testing. The data is mirrored regularly with the data from the production environment in a defined time frame.

Development The development environment contains all the component's logic with changes in the process of implementation. The data in the development environment is not up to date and is only the amount of data, which is copied from the production or test environment by the database engineer.

Changes may not be performed directly in the production environment, but rather run through a change process originating in the implementation of changes in the development environment, which are being tested in the test environment and after a thorough analysis, being deployed to the production environment.

Each system environment contains two hardware components:

Server Pool The server pool consists of a defined number of servers that access the shared filesystem to run a list of processes. The server pool could have been broken down into single servers, but jobs representing the processes, usually don't run on a defined server, but are rather dynamically allocated to a certain server by a balancer to ensure a maximal processing saturation.

Shared Filesystem The shared filesystem is the physical storage containing all the files, databases and applications' logic. A shared filesystem has different volumes with different purposes, which are not explicitly mentioned due to being out of scope and the lack in information gain.

It is to be noted that the physically separation for processing and storing of market data and client data is desired by clients, a server or a portion of the server pool can designated to this certain client only. The same applies to the shared filesystem, where the client can receive its own physically detached shared filesystem.

VP visualizes information of the Infrastructure IM.

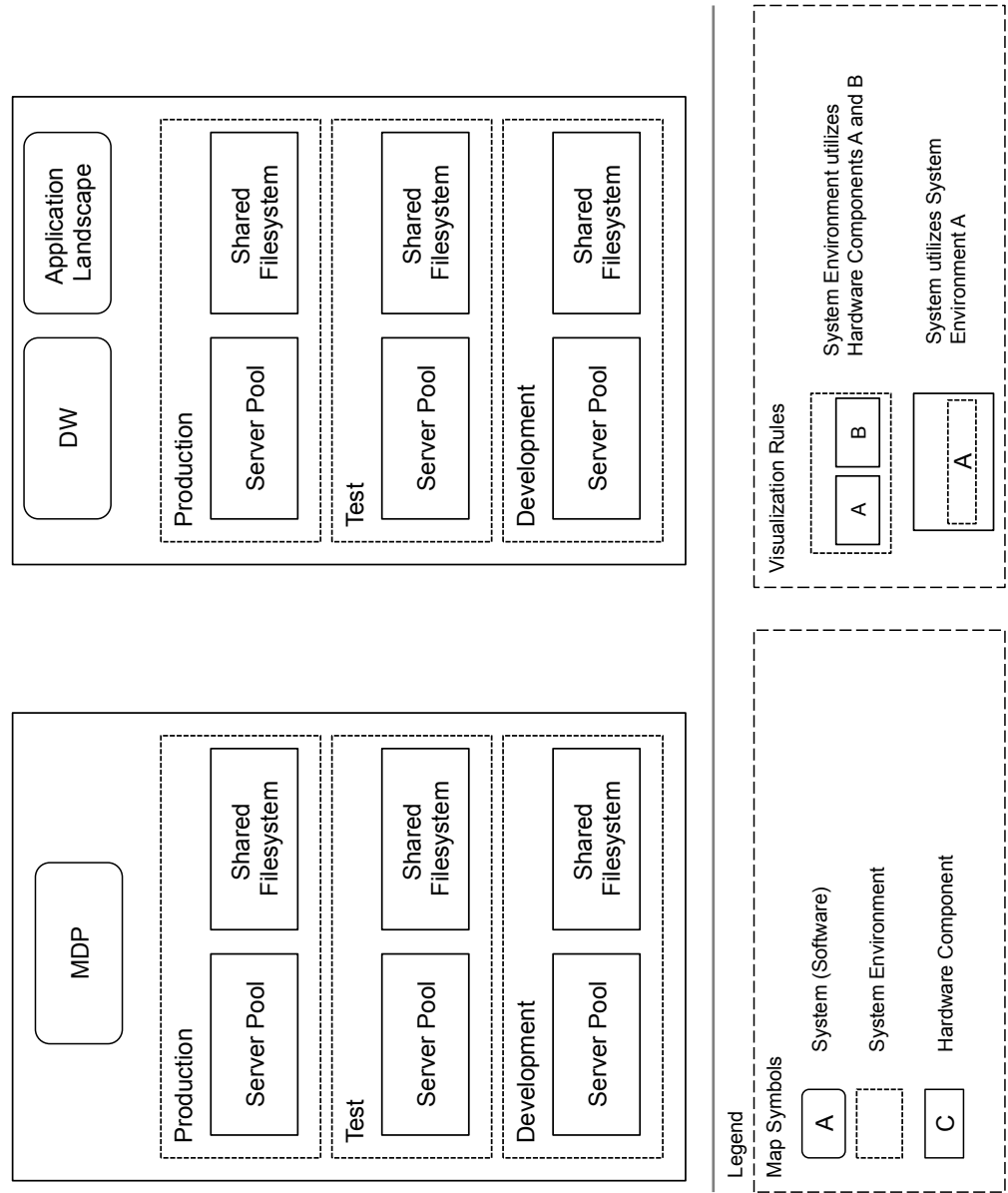


Figure 4.11: Infrastructure Viewpoint on selected company's System

4.3.3 Authorization viewpoint

The authorization viewpoint enables the creation of a mapping of the data, the client has sent, ordered and received and of the data, the vendor had sent by visualizing identifiable objects at certain structural components. The mapping can be performed manually to control whether the user was authorized for the use of the data or automatically by an application, which could provide real-time authorization queries for users.

MD Package The MD data packages are back-office data provided by the vendor. A vendor contract can contain multiple MD packages.

MD Request The MD requests are per-request data provided by the vendor. A vendor contract only includes the right to request MD from a certain pool of MD, but not explicitly lists all the requestable MD objects. The pricing is respectively fitted.

MD Object A MD object is affiliated with a MD package or request. An MD object contains multiple facts.

Enhanced MD Object An enhanced MD object is affiliated with a MD object.

MD Fact The facts of a MD Object containing values such as prices.

Enhanced MD Fact The facts of an enhanced MD Object.

CD Service A CD Service is performed for a certain client and stipulated.

CD Object A CD objects is affiliated with a CD service.

Enhanced CD Object An enhanced CD object is affiliated with a CD object and if respective MD object facts have been utilized, it is affiliated with one or more MD objects and one or more enhanced MD objects.

User The user is affiliated with a CD service and uses The might use (enhanced) MD or CD objects.

VP visualizes information of the Authorization IM.

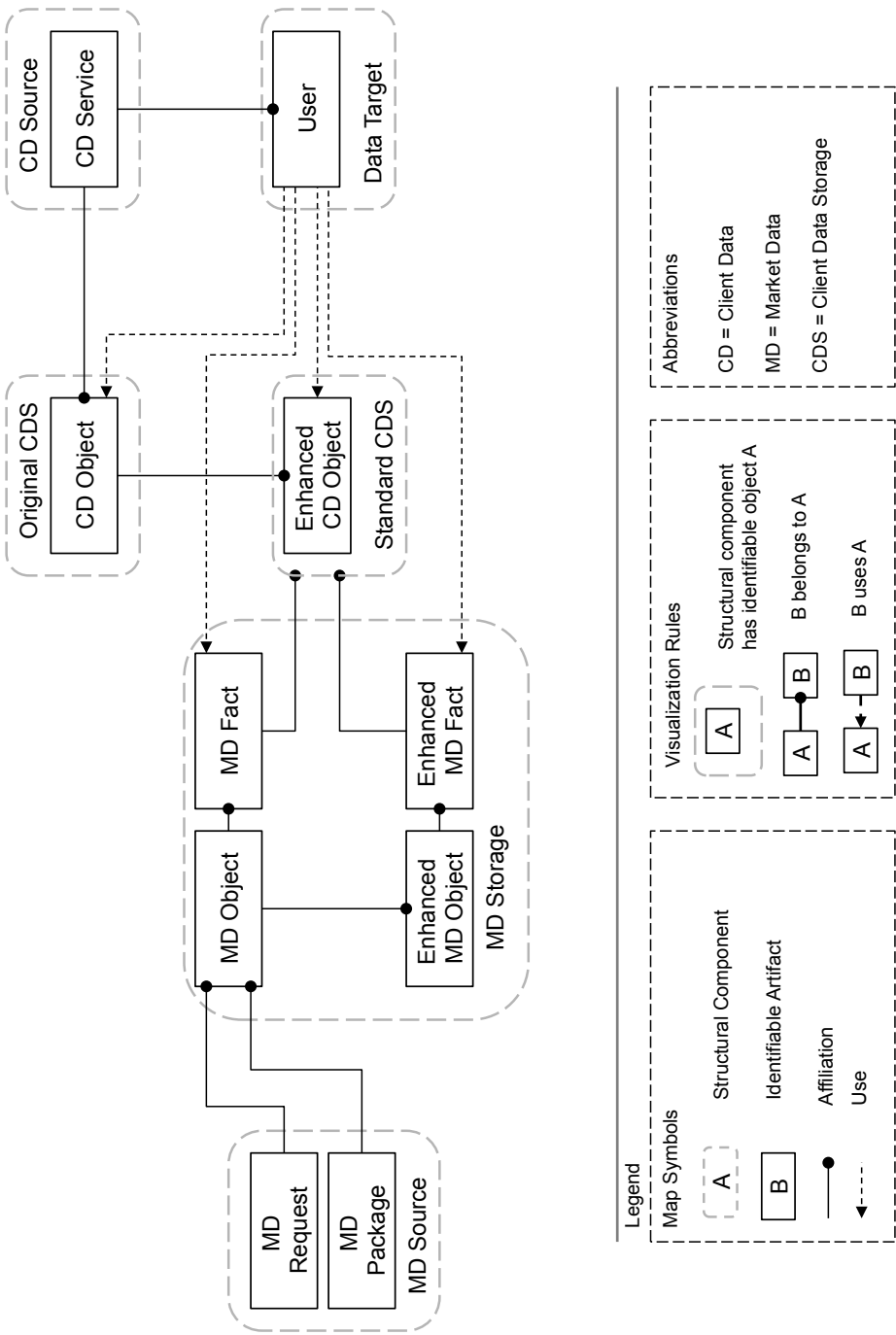


Figure 4.12: Authorization Viewpoint on the System

4.3.4 Stakeholder viewpoint

The communication viewpoint illustrates the communication channels between the stakeholders.

Client The Client is sending his client data to the DW, expecting to receive some kind of service from the application owner, resulting in some kind of an information gain. He has three communication channels: The primary communication channel is the Account Manager. There is also a communication channel regarding the sent client data for the DW and one with the Application Owner regarding the receipt of a service product.

Account Manager The Account Manager serves as interface between the client and the company's internal stakeholder, taking care of all formalities and legal affairs.

Vendor The Vendor is sending his market data to the MDP, receiving monetary compensation.

Procurement Manager The Procurement Manager serves as interface between the vendor and the company's internal stakeholder, taking care of all formalities and legal affairs.

Operational Manager The Operational Manager monitors the daily processes in the MDP or DW. If problems occur, the Application Owner and if necessary the Client need to be informed. The Operational Manager also serves as first-line of support in the respective system.

Database Engineer The Database Engineer mainly communicates with the Operational Manager about the daily processing of the data and whether changes need to be implemented.

Application Owner The Application Owner communicates, depending on the data, with the respective system stakeholders. He also communicates the service product to the client.

Management Board Member Serves as control instance.

Auditor Serves as control instance. Evaluates whether the company adheres to the auditor's respective requirements.

VP visualizes information of the Stakeholder IM.

Support

In the selected company no classic first-level-support exists, respectively a division, which specifically concerns themselves with that topic. Support inquiries about data quality, data completeness, etc., are rather solved by stakeholder within the systems, which are contacted, if discrepancies are identified. The following four Stakeholder from the stakeholder viewpoint are assigned to resolve the inquiries:

1. Application Owner: The instance that usually notices something is wrong with the market data. His first step is to look up, whether the MD is also wrong in the MD Client. If not, then the operation manager in the DW needs to be contacted. If the MD is wrong, then the operation manager in the MDP needs to be contacted.
2. Operation Manager (DW): Is the instance that has to evaluate first of all if the fact originated in the MDP or from the Client. Wrong client data is ignored and simply overwritten with MD. For wrong MD the operations manager in the MDP is contacted.
3. Operation Manager (MDP): The operations manager in the MDP evaluates first of all, if the MD was sent erroneous by the vendor or if the MD was processed with errors.
4. Vendor: Is solely contacted by the operations manager

A big problem for the support is that the sequence is not always adhered to and inquiries not sufficiently described.

4.4 Information Models

”Contains an information model fragment including the definitions and descriptions of the used information objects” [S Bu 08].

4.4.1 Structural Information Model

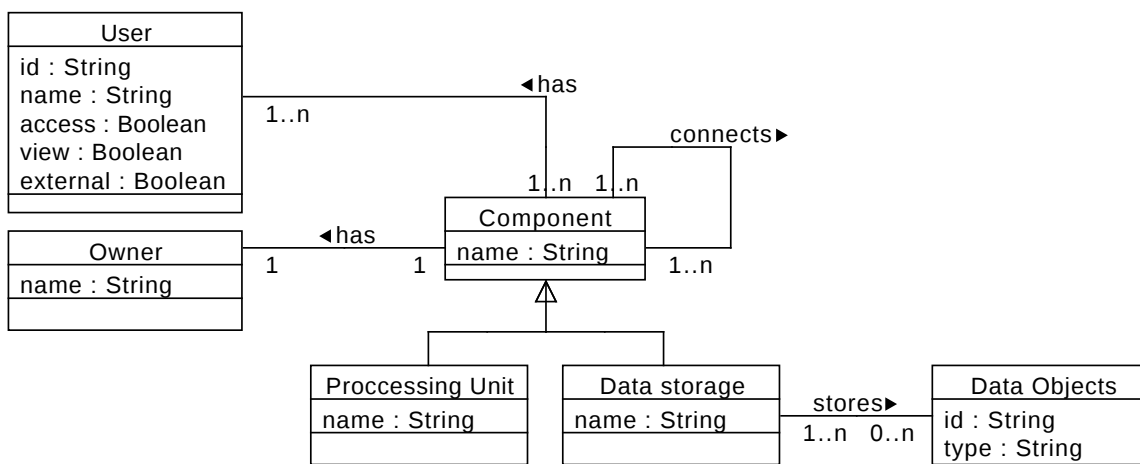


Figure 4.14: Structural Information Model

- **Component:** A component is an "individual unit of computation" [IEEE 00] and can be an application, a database or a directory. The component is part of a system.
- **Key Component:** Key components represent central parts of a system, mostly databases, where all the surrounding components depend upon.

The IM's associations:

- **Connection:** The connection represents the layout of the components within the system and serve as orientation. It also illustrates that components are interacting with other components.

4.4.2 Data Flow Information Model

- **Component:** A component is an "individual unit of computation" [IEEE 00] and can be an application, a database or a directory. The component is part of a system.

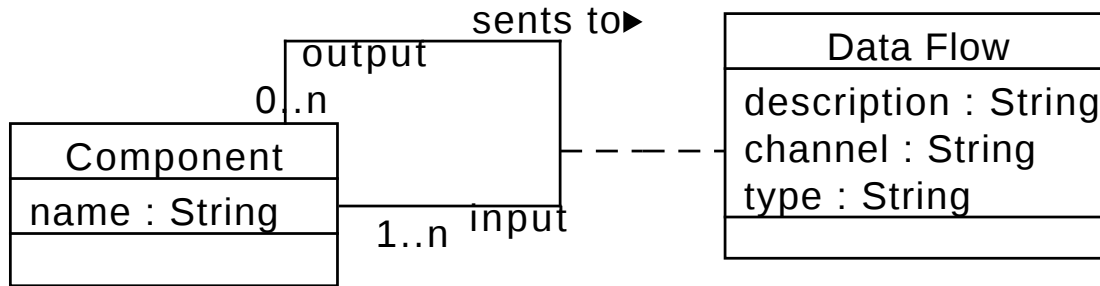


Figure 4.15: Data Flow Information Model

The IM's associations:

- Data Flow: Is the flow of data between an input component and an output component. The data flow elements in the core data flow are represented by jobs in the selected system, which are being logged and thus made analyzable.

4.4.3 Infrastructure Information Model

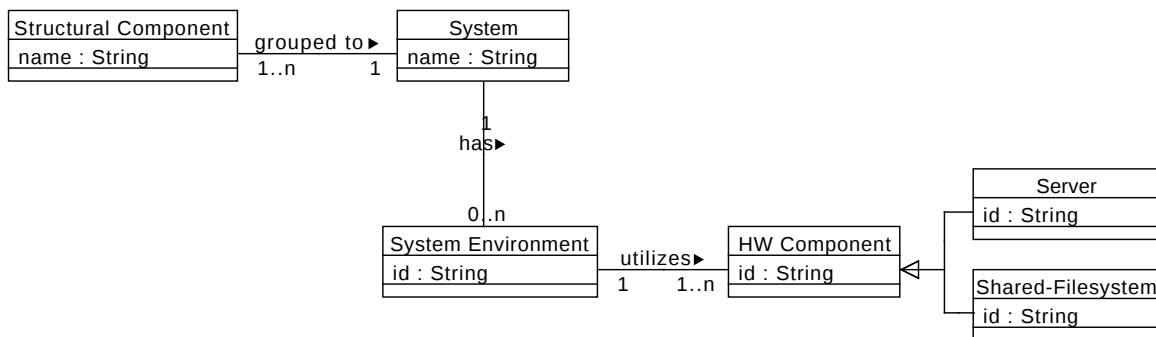


Figure 4.16: Infrastructure Information Model

- System (Software): A group of software components utilizing hardware components.
- System Environment: The physically separated, grouped hardware components the system utilizes.

- Hardware Component: The hardware components the software utilizes.

4.4.4 Authorization Information Model

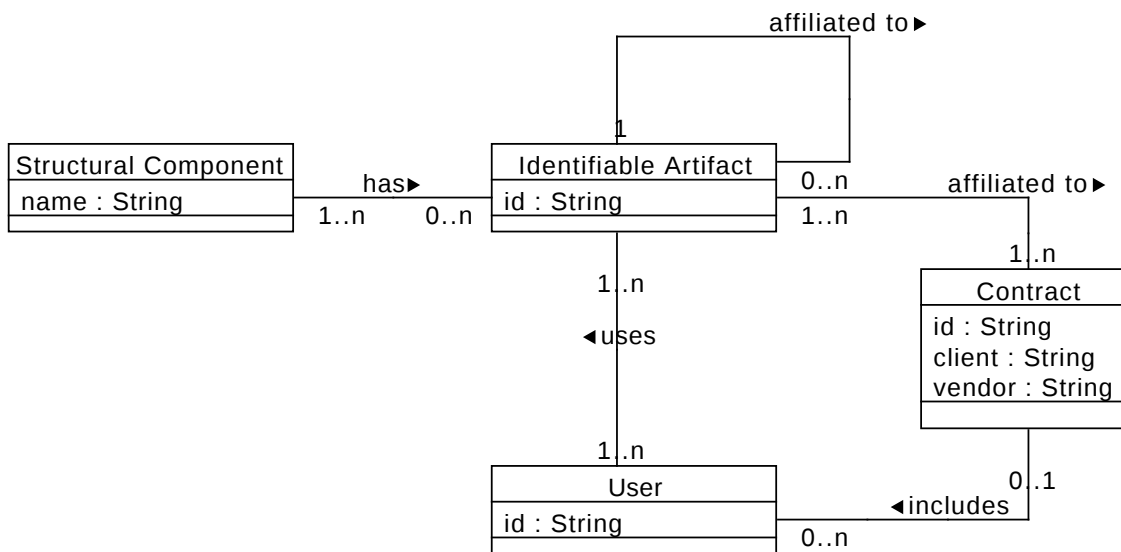


Figure 4.17: Authorization Information Model

- Structural Component: A component is an "individual unit of computation" [IEEE 00] and can be an application, a database or a directory. The component is part of a system.
- Identifiable Artifact: Some identifiable artifact that enables to identify its source and its destination.

The IM's associations:

- Affiliation: An identifiable artifact is affiliated with another identifiable artifact. The "bigger" artifact being the one affiliated with.
- Use: The user (also an identifiable artifact) can use only certain identifiable artifacts.

4.4.5 Stakeholder Information Model

The IM's classes:

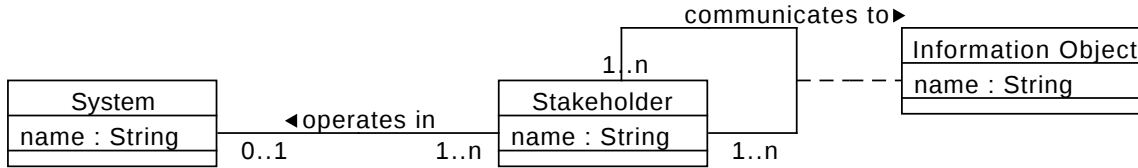


Figure 4.18: Communication Information Model

- Stakeholder: A stakeholder is "an individual, team, or organisation [...] with interests in, or concerns relative to, a system." [M La 05]
- System: A group of components.

The IM's associations:

- Communication: This association depicting the communication between two stakeholders.
- Control: This association depicts the stakeholder's task to control the processes of the system.

4.5 Data Governance

Due to data governance significance for market data it will be included into the IA as separate point.

4.5.1 Theory of Data Governance

Information Definition and Information Governance "are the most difficult challenges companies face" [M Go 10] and are included in core principles that guide an IA [M Go 10]. The goal is to answer the following questions [M Go 10]:

- What information do you have, where is it stored, and what value does it have?
- How does your business use it and for what purpose?
- How good (or bad) is the information quality?
- What information do you keep and what do you archive?

These questions are answered by the following objectives of information definition & governance [M Go 10]:

- Defining governance processes, infrastructure and technology
- Establishing common data and information domain definitions
- Monitoring and ongoing improvement of data quality
- Establishing necessary executive sponsorship, organizational policies, and cross-organizational oversight

How can trusted data be achieved? [V Kh 10] propose a framework for data governance decision domains and divide data governance into five decision domains:

Data principles Data needs to be classified as an asset and therefore certain decision need to be made and put into principles for consistency. Decisions include the business uses of the data, possibilities for sharing and reusing the data and many more.[V Kh 10]

Data quality Data quality is achieved when the data can be trusted. To make data quality possible, requirements need to be gathered and definitions about the data and its use need to be made. Decisions include the standards for data quality or describe the process to establish data quality.[V Kh 10]

Metadata Data on its own is often not interpretable. Data about the data needs to be added in order to give it a meaning. An example for metadata would be a timestamp. A price value is worthless without a timestamp, because only then statements about its validity can be made. Decisions include the process to continuously update the types of metadata and how it will be defined consistently.[V Kh 10]

Data access Data is distributed on a need to know basis and therefore the access requirements need to be defined. Decisions include the business value of data, periodic monitoring and audit compliance.[V Kh 10]

Data lifecycle The data lifecycle determines "the definition, production, retention and retirement of data." Decisions include how and where the data is stored as well as the duration.[V Kh 10]

4.5.2 Governance controls

Governance controls can be interpreted as an enterprises implemented functions that assure data governance. The selected company has already some governance controls in place. Data principles are not yet in place and the selected company does not yet see the benefits of such. Metadata is already heavily used by the selected company. To every data object is metadata added, which is required for the processing of the data object. The metadata accumulation includes an exhaustive logging of every job performed at a structural component and is saved in the data catalog. This logging enables the first step of data quality, namely the successful run of a job and the respective existence of the data object at a certain structural component. Data access requirements are all handled by the user authorization and is mostly prevented by giving users only access to views and not the actual tables, where the data object is stored. The data lifecycle in the selected company can be identified with the structural viewpoint, the data flow viewpoint and the data storage viewpoint. Market data is also never deleted in the selected company. Not even the data that was falsely sent by the vendor and is later corrected. It would only be interesting, when the imported data from the vendor would fluctuate, but since the amount of data imports is only growing and rarely shrinking, there is no point in focusing on this aspect.

The data quality, however, is the central concern of almost all system stakeholder and is also required for compliance with Solvency II. The implementation of such data quality assurance measures is however hard to implement. The only simple quality check on the MD that is performed and monitored in the selected company is, whether the data is sent by the vendor at all and whether the data successfully passes through the core data flow chain. If the market data is erroneous, it is in principle discovered by clients or by the application users in the application landscape. Basically is the vendor trusted to send the correct data and if the data is wrong the vendor is trusted to send the respective correction data.

The actual data quality evaluation processes are out of scope of this research and is a point for future work, but the roadmap to reach data quality will be presented.

The sequence of activities of a data quality methodology can be broken down to three phases [C Ba 09]:

1. State reconstruction: collects "contextual information on organizational processes and services, data collections [...] and quality issues."
2. Assessment and measurement: "measures quality of data collections along relevant quality dimensions. The measurement refers to the actual measured value along

quality dimensions. Assessment refers to the process of comparing this the previously measured value to a reference value. [C Ba 09]

3. Improvement: "concerns the selection of the steps, strategies, and techniques for reaching new data quality targets" [C Ba 09].

The steps of the assessment phase include[C Ba 09]:

- Data analysis: "to reach a complete understanding of data and related architectural rules"[C Ba 09].
- Data quality requirement analysis: interview data viewers to "identify quality issues and set new quality targets"[C Ba 09].
- Identification of critical areas: the most important databases and data flows are selected and quantitatively assessed [C Ba 09].
- Process modeling: "provides a model of the processes producing or updating data"[C Ba 09].
- Measurement of quality: "selects quality dimensions [...] and defines corresponding metrics"[C Ba 09].

The quality dimensions include: completeness, consistency, currency, volatility and timeliness of data.[C Ba 09] After having quality controls identified, they need to be added to the processes within the system. This can be done through process controls or the redesign of already existing processes[C Ba 09]. The selected company's preferred technique is to control the existing processes. The feasibility is presented in the governance viewpoint and the governance information model.

4.5.3 Governance viewpoint

The governance viewpoint visualizes the possibility to implement governance controls and which structural components need to be considered for such an application. The governance controls are enabled through the data catalog in the MDP and the DW. The data catalog can serve as information source for the log information of the data processing or the data itself, thus enabling controls. Existing governance controls mainly only include whether the information is complete and that all the processes were performed. The governance controls are detached from the core data flow to enable a parallel processing to enable a shorter core data flow chain.

VP visualizes information of the Governance IM.

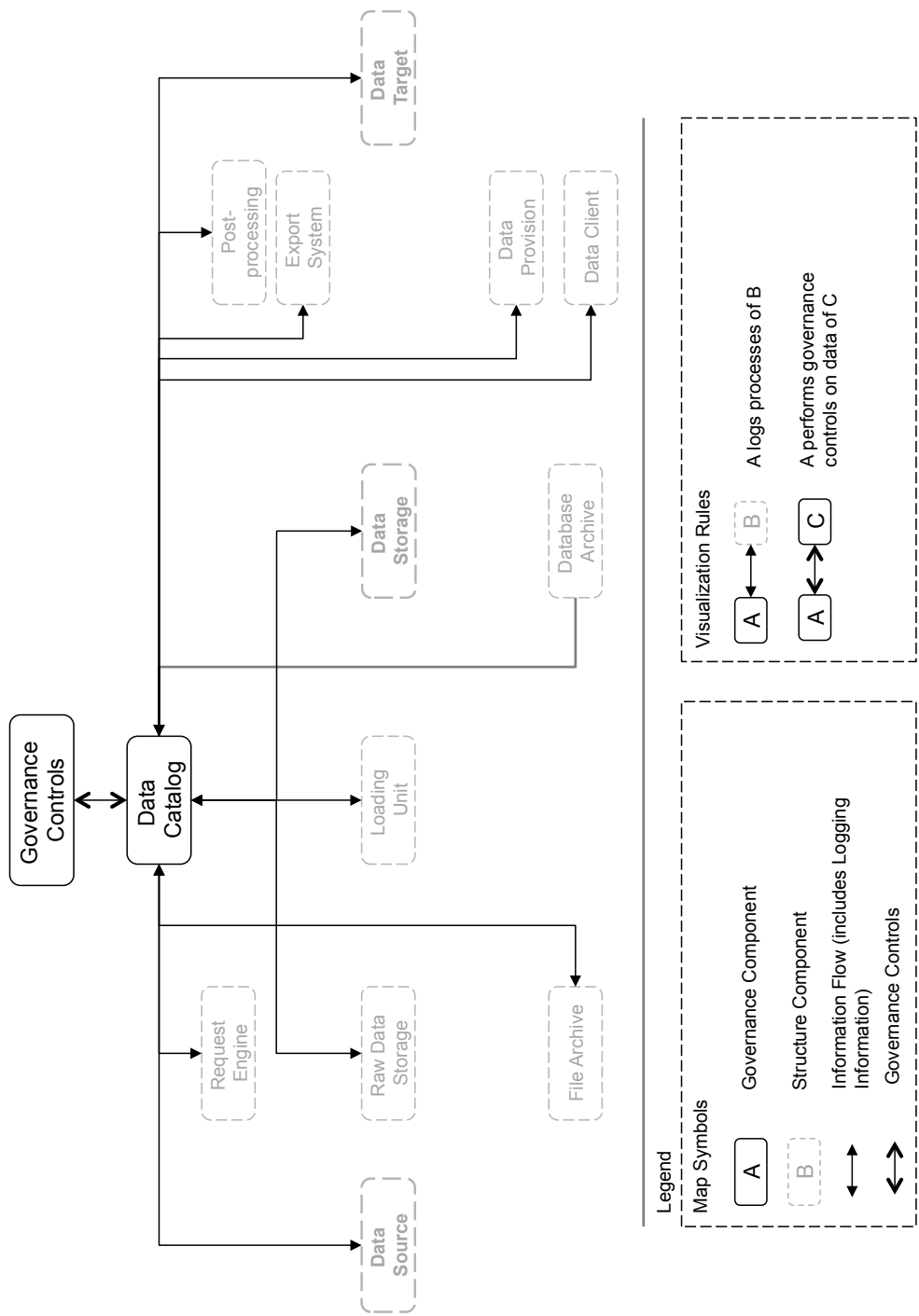


Figure 4.19: Governance Viewpoint on the Market Data Pool

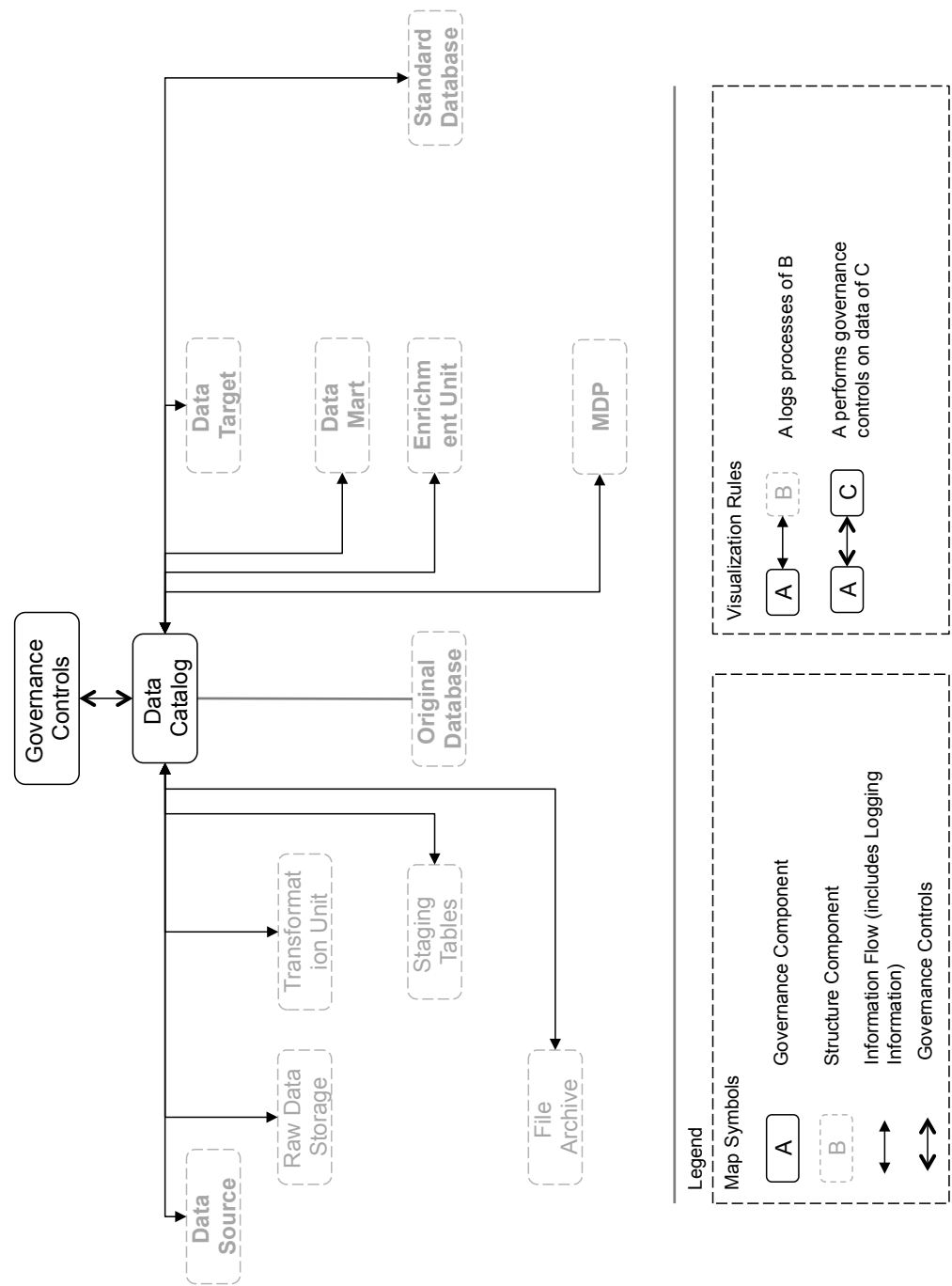


Figure 4.20: Governance Viewpoint on the Data Warehouse

4.5.4 Governance Information Model

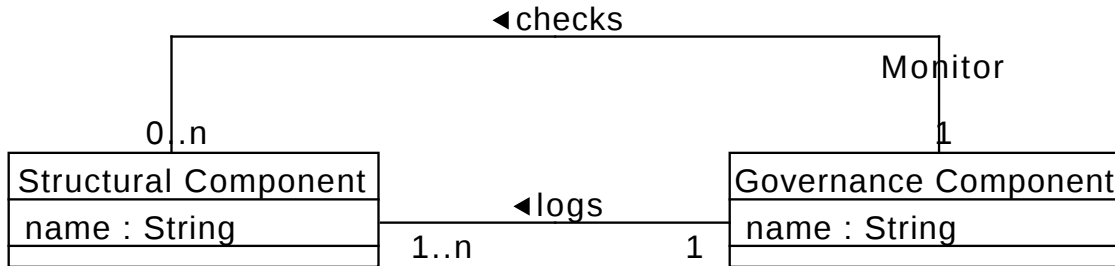


Figure 4.21: Governance Information Model

The IM's classes:

- **Governance Component:** A governance component is a component needing to implement governance controls. Governance controls include the monitoring of the successful run of the processes.
- **Structure Component:** A component is a "individual unit of computation" [IEEE 00] and can be a application, a database or a directory. The component is part of a system.

The IM's associations:

- **Information Flow:** The Information flow includes the logging information of the structural components and should also enable the access of data.
- **Governance Control:** Interaction between governance components. One providing the information, another providing the logic.

4.6 Interim Conclusion

The IA is composed of the system overview, the viewpoints, the information models and also a data governance component. Having defined structural components makes it possible to map the previously collected requirements. This IA then enables system stakeholder to know where to gather the information to answer their concerns. However, it is not a roadmap with every step describing what exactly needs to be done, but rather

a reference point for orientation within the system.

The information that can be derived from the descriptions in chapter 2, the system overview in section 4.1 and from the viewpoints in section 4.3 seems to be partially redundant and maybe unnecessary. In any case was this done intentionally. It depicts the process of understanding and getting to know the system in addition to enabling the traceability of including and excluding elements. The architectural description evolves from bigger components into smaller, more specific components, which represents the actual knowledge transfer about the system for stakeholder.

Chapter 5

Evaluation of IA

In this part the previously constructed research artifact, the IA is being evaluated. For the evaluation it is important to recall the purpose of this IA. This IA's main purpose is the creation of a structure and defined components in the system environment of the selected company. Having done this, enables system stakeholder to gather information and to fully exploit this information. It also furthers the understanding of the system and serves as reference point, whether all the structural components were considered. For the evaluation of the IA, the previously identified and selected stakeholder were interviewed again and they needed to evaluate, whether their concerns could be answered or not, what could have been done better and what information demands arose in turn after having seen the IA. The evaluation of the IA is based on architectural analysis methods.

5.1 Theory of Architectural Analysis

"Organisational effectiveness cannot be achieved through local optimisations, but is realised by well-orchestrated interaction of organisational components." For the creation of "such an integrated perspective," we need not only a description technique for architectural models, but also model-based analysis techniques [M La 05]. Architectural analysis supports the decision-making process and creates well-informed stakeholders [M La 05].

[M La 05] firstly differentiates between the "types of analysis input and results: functional (e.g., structural and dynamic properties) and quantitative (e.g., performance and costs)." Simulation "can be seen as the execution of a model" [M La 05].

Quantitative analysis answers questions such as "how quick" or "how cheap".

Architectural models "do not provide sufficient information to perform detailed quantitative studies" [M La 05].

Functional analysis "is performed to gain insight into the functional aspects of an architecture. Among others, it is used to understand how a system that conforms to an architecture works, to find the impact of a change on an architecture, or to validate the correctness of an architecture" [M La 05].

Quantitative simulation "is used to make statistical statements about the quantitative measures of a system based on multiple runs." Simulation "can be seen as the execution of a model" [M La 05].

Functional simulations "are useful to illustrate the dynamic behaviour of an architecture." The goal of functional simulation is to "gain insight into the properties and behaviour of an architecture." Furthermore enables a functional simulation "a better common understanding of the architecture and therefore plays "an important role in the communication between the stakeholders" [M La 05].

5.2 Architectural Analysis

For the architectural analysis the functional simulation was used as a basis. The goals of functional simulation conform to the goals of the IA. So, if the functional simulation results in positive feedback from the stakeholder, it can be said to satisfy its original goal. What was actually done during the evaluation (functional simulation) of the IA? The interviewed employee's of the selected company have a double role, while being a stakeholder they are also the experts knowing and working with the system on a daily basis. They were asked a set of questions to confirm the correctness of the IA and whether it actually serves its purpose. It is to be noted that the stakeholders asked for concerns and the interviewed employees for the evaluation are actually the same people.

Each interviewed expert (an employee of the selected company and at the same time system stakeholder) was asked the following set of questions:

1. What is your first impression of the IA in its entirety, after having seen it for the first time? Is it practicable and understandable?
2. Are the elements of the models correct and the level of detail sufficient enough?
3. Are your concerns sufficiently answered?
4. What would you have done different?

The upcoming presented answers, are not the exact representations of what the interviewed employee answered, but were written down in note form during the interviews and expanded on directly afterwards. The IA was previously sent to each interviewed employee for a first look and reading if desired, but when asked whether they have read it, all answered that they have taken a look, but not more. Therefore, did I present the IA to each stakeholder during the IA. This does however influence the evaluation negatively, since an explanation is always easier to understand.

External Auditor

The interviewed employee representing the external auditor was actually not external and not an auditor itself, but rather an employee in charge of external auditors, if the company is being audited. An external auditor was not available. The external auditor is actually a special case, due his high importance. For the interview, not only were the IA artifacts prepared, but also the audit requirements in table 3.1 were exemplary instantiated. I was able to fully answer all the requirements for the audit requirements with the help of the IA as base map and for orientation. The actual information instantiations could not be included and presented in this thesis, due to being company specific information not to be disclosed to the public.

Background of the interviewed employee: Has an academic background in business management, some financial knowledge and little IT knowledge. On the basis of his background he should have problems gathering all the information himself, because of the extensive SQL knowledge needs. However, with predefined SQL queries the data collection would be possible.

1. The interviewed person was surprised by how many components the systems consists of, but likes the simplicity of the viewpoints. Furthermore, was the interviewed person surprised by how big the MDP actually is and how similar the DW and MDP are structured. The single viewpoints are understandable, but the IA in its entirety would not be understandable after a first reading, also are some things not clear enough by itself.
2. Whether the elements of the models are correct was not directly answered, due to little knowledge of the actual system.
3. The market data flow is clearly represented and especially the instances of the data in the MDP and DW are very good pictured. The compliance to vendor requirements were sufficiently answered from his point of view and would be very happy to be

able to gather the information by a single person. Asking several different people for information always takes longer. The information source, as to whom to contact for questions could have been a little better, since the database engineer and the operations manager do not know everything about every component in their system, but is sufficient in most occurring cases.

4. Has never done something similar and knows too little about the systems to be able to create such models and can therefore not judge.

Overall, could his concerns be mostly answered, what is sufficient for a model and did give positive feedback.

Operation Manager MDP

For the Operations Manager were exemplary SQL queries prepared to demonstrate the feasibility of governance processes with the information from the Data Catalog.

Background of the interviewed employee: Has an academic background in information management, good financial knowledge and good IT knowledge apart from programming.

1. Had a very good first impression and was surprised by how simple the MDP could be illustrated. All the viewpoints are easily understandable, the information models would be harder without explanation. The best feature of the viewpoints was that they illustrate really good the interaction of client data and market data.
2. For the MDP all structural elements are illustrated and correctly. The knowledge of the DW and its processes is not as good as of the MDP, but there are no elements that seem to be in conflict with her knowledge about the elements. The level of detail is perfect and could not have been done better. Abstract enough to understand it, but detailed enough to actually gain insight from it.
3. By enabling the monitoring of the structural data deliveries through a central governance and the prepared SQL queries would be very helpful. At the moment is the monitoring very tedious, since several different applications with different knowledge requirements exist. In addition is there no process in place, which can be adhered to and therefore every employee has his own technique to perform the monitoring. Also the data quality checks would be great, but hard to implement and actually getting used. An event based viewpoint for the market data interface maintenance would have been a nice addition. Transparency for the client could not be better illustrated, than with the viewpoints of the IA, but the meaning of it cannot be derived from the IA.

4. As already said, is the level of detail perfect. It would have been nice to also gain some information about the application landscape, but it is understandable to have been excluded. It remains to be seen whether it would actually be used or be the cause for any changes, but it would be desirable, especially regarding the monitoring of the data processing.

Overall liked the Operation Manager the IA and would really like to see a central governance application. Another event viewpoint for changes in the market data interface would have been nice though.

Database Engineer (DW)

For the interview with the database engineer was a very tedious, exemplary instantiation of the authorization viewpoint prepared. This instantiation was very time consuming for only a small portion of market data and client data. This is no work to be done manually. Either it is being implemented and adhered to by controls in the applications or there is no practical use in this viewpoint. It does however improve the understanding of the client data and market data and their interaction.

Background of the interviewed employee: Has an academic background in computer science, and sufficient financial knowledge through years of work experience.

1. The instantiation of the authorization view really impressed the interviewee and hopes that this does not need to be instantiated on a regular basis or on a big portion of data. The interviewee really liked the look and simplicity of the viewpoints.
2. All the structural elements in the DW are correct and included. The level of detail is also well picked.
3. The market data interface change could have been addressed more directly, but the viewpoints combined do help to create at least an understanding. An event based viewpoint would have been preferred as well. The same applies to the new requirement market data interface. The approach for the user authorization is interesting, but due to the sheer amount of data the selected company receives and processes, this is very tedious and would require both a highly efficient, optimized implementation, as well as more powerful hardware. Most likely, neither will be available.
4. The interviewee would have liked for the interfaces and channels between the components to be better and the addition of an event-based viewpoint for market data changes and requirements

The Database Engineer expected less from the IA and really liked the level of detail of the viewpoints. The interviewee mentioned several viewpoints, he would like to be added, since he really liked those. They would also be perfect to give new employees a good understanding of the system.

Database Engineer (MDP)

For the Database Engineer (MDP) was an exemplary data instantiation for single processes generated and the duration of a whole run of the core data flow chain. These information is easily generated with SQL queries, if you know the data flow with its structural elements. The data flow represents jobs that are being run in the system and logged. Hence the easy information collection. The interviewee was also explicitly presented with the requirements.

Background of the interviewed employee: Has an academic background in computer science and has gained a lot of knowledge in finance over the years.

1. The simplicity was the first thing being registered by the interviewee. Models are very well concentrated on the core aspects of the system. The information models could be a little bit better concerted with the viewpoint models, having found minor errors in the viewpoints use of the information models.
2. All the required and essential structural elements are included and other viewpoints are correct as well. The level of detail is very good, except for the stakeholder viewpoint. Here, the wrong impression is conveyed to the reader, that a systems only consists of the represented stakeholder, but actually encompasses many more.
3. The process performance is very good answerable, as demonstrated with the exemplary instantiation. Only the on time is not really answerable, but this information needs to be delivered by the client. Regulation adherence is not directly addressed by a viewpoint, but rather the previously gathered requirements for the IA. Those are very interesting for the interviewee and are much more specific, than he was aware of. The system understanding concern is also answered most satisfactory and could not have been done better.
4. Except for some minor inconsistencies the IA serves its purpose and represents everything he expected it to be.

The Database Engineer did not gain any insight from the architecture, but still liked the IA, since he considers such an architecture to be important. It helps people not as familiar with the system and the data to learn and give a point of reference. Such a point

of reference is also very nice for developers, because it can promote more interaction between them and lessen the knowledge monopoly on certain parts of the system.

Procurement Manager

The procurement manager was also presented with the same exemplary instantiation of the authorization viewpoint as the Database Engineer (DW). The interviewee was surprised by how many dependencies existed and the complexity.

Background of the interviewed employee: Academic background in finance, with SQL knowledge.

1. Very positively surprised by the simplicity the system is broken down and regards the viewpoints as very helpful. It gives a very good understanding of the system and the interaction of market data with the clients data, without requiring extensive IT knowledge. Also very positively registered was that the viewpoints are not scattered among several pages, but every viewpoint seems to be enclosed by itself. The interviewee also liked the gathering of the requirements. Only very few people know of requirements and the concerns of stakeholder are also interesting to get some insight in other peoples information demands.
2. As far as this is judgeable by the interviewee the IA seems to be correct and the level of detail is also good.
3. All of the stakeholders concerns can be answered by the authorization viewpoint, which is very complex and laborious to instantiate. It is feasible as the instantiation proofs, but not practicable.
4. The interviewee said that there needed to be another way to map the market data, client data, data source and data target.

The procurement manager regards the IA as good, especially since it communicates a good understanding of the system, but would like another authorization viewpoint. Usually the

stakeholder Account Manager should have been included as well, but sadly no one was available in time for an interview.

5.3 Interim Conclusion

All of the interviewed stakeholder found the IA useful and sufficient enough for its purpose. Be it for understanding of the system, as reference point for information gathering or requirement catalog. Very well received was the list of requirements. Few stakeholder knew of the extent of the requirements regarding market data and find it helpful to have a knowledge about requirements. The viewpoint models were by far the most popular artifact of the IA, whereas the information models were far less considered and time spent on. Very well received was also the scoping and level of detail of the particular viewpoints. However, only very few considered it in its existing form as aid for their day to day work. It would need to be built upon. Only the operations manager (MDP) was very fond of the idea to have a central governance application, mainly for monitoring the processes.

It can be said that the IA's accuracy is very good and also the break down of the system was well chosen. It includes a high-level map of the information requirements from the system stakeholders & the vendors and also divides the system into major information categories with the means of viewpoints.

Chapter 6

Conclusion

This research originated in the selected company's need to disclose information about its system and processing of market data. They were also interested what kind of requirements existed and would need to be adhered to. The research was started with a evaluation of different research processes and fitting the reference research approach to the topic. Subsequently was the foundation laid for the research. First of all an understanding of what market data actually is, was created. Then different approaches for creating an IA were evaluated and taken into consideration. The problem regarding what elements to include in the IA, was that there is not the one correct solution listing all the elements needed for an IA. The appropriate elements of an IA are totally depending on the topic, the stakeholder and the system and the IA is fitted to its purpose. The foundation was concluded with an introduction to Data Governance and elaboration on the reference architecture of a DW.

The requirements gathering for the IA was initially started by a literature research studying existing laws, affecting market data and financial reporting. For the selected company Solvency II turned out to be the most important, but it was hard to derive requirements for the market data from these laws. Consequently a different approach was taken, namely by studying the selected company's audit documentation. The focus here shifted further towards the vendor. The requirements derived from the performed audits were then enhanced by studying the license agreements. Due to the information demands still being too small in quantity, company internal stakeholder were identified and their concerns were added as requirements.

After all the requirements had been gathered the second artifact, the IA was constructed. To fit the purpose of the IA, a system overview was included. Then viewpoints enabling the creation of specific views on the system, answering the previously gathered concerns

and enabling the gathering of information for the requirements were modeled. Those viewpoints present the data stored in the information models. Due to the importance of data governance for market data this aspect was separately added to the IA, but included as a viewpoint and information model as well to align the suggested governing component with the existing system. Data quality as one of the five domains of data governance and as an very important topic among the stakeholders was focused on.

The resulting IA was then evaluated by performing functional simulations and showing the results to the stakeholder, receiving feedback and proofing the IA's utility. Overall were the stakeholder satisfied with the IA, but it would need to be added on to be of greater use, than just enabling insight into the company's system and being a list of requirements. However, that was exactly to goal of this research. It is to serve as a concept to check for feasibility, evaluating different approaches and be a point of reference.

Bibliography

- [A Ke 06] A. Kemper and A. Eickler. *Datenbanksysteme*. Oldenbourg, 2006.
- [B Wa 14] B. Waltl, A. Schneider and F. Matthes. “Deriving and Modeling Compliance Requirements from Legal Audits”. 2014.
- [Brau 10] M. C. Braun. “Internal Control over Financial Reporting”. November 2010.
- [Bund 14] Bundesanstalt fuer Finanzdienstleistungsaufsicht (BaFin). “Solvency II: Aufbau und Gesetzgebung”. Mai 2014.
- [C Ba 09] C. Batini, C. Cappiello, C. Francalanci and A. Maurino. “Methodologies for Data Quality Assessment and Improvement”. *ACM Computing Surveys*, Vol. 411, No. 3, p. 52, July 2009.
- [C Ge 13] C. Gerner-Beuerle, P. Paech and E. P. Schuster. “Annex to Study on Directors’ Duties and Liability”. Tech. Rep., LSE Enterprise, 2013.
- [G Mu 08] G. Mueller, S. Sackmann and O. Prokein. *Handbook on Information Technology in Finance*, Chap. 29, pp. 711–729. Springer, 2008.
- [IEEE 00] IEEE Computer Society. “IEEE Std 1471-2000: IEEE Recommended Practice for Architectural Description of Software-Intensive Systems”. Tech. Rep., The Institute of Electrical and Electronics Engineers, Inc., 2000.
- [J C 89] J. C. Brancheau, L. Schuster and S. March. “Building And Implementing An Information Architecture”. *DATA BASE*, pp. 9–17, 1989.
- [Kech 15] C. Kecher and A. Salvanos. *UML 2.5*. Rheinwerk Verlag, 2015.
- [Krcm 10] H. Krcmar. *Informationsmanagement*. Schäffer-Poeschel, 2010.
- [M Go 10] M. Godinez, E. Hechler, K. Koenig, S. Lockwood, M. Oberhofer and M. Schroeck. *The Art of Enterprise Information Architecture*. IBM Press, 2010.
- [M La 05] e. a. M. Lankhorst. *Enterprise Architecture at Work*. Springer, 2005.
- [Mitr 10] T. Mitra. “Documenting software architecture, Part 2: Develop the system context”. May 2010.

- [P Of 09] P. Offerman, O. Levina, M. Schoenherr and U. Bub. “Outline of a Design Science Research Process”. *DESRIST*, 2009.
- [Pier 05] R. Pierce. “Modeling an information architecture”. *Communication Design Quarterly Review*, Vol. 6, No. 1, p. 11, 2005.
- [S Bu 07] S. Buckl, A. M. Ernst, J. Lankes, F. Matthes, C. M. Schweda and A. Wittenburg. “Generating Visualizations of Enterprise Architectures using Model Transformations”. *Enterprise Modelling and Information Systems Architectures - An International Journal*, Vol. 2, No. 2, p. , 2007.
- [S Bu 08] S. Buckl, Alexander M. Ernst, Josef Lankes and Prof. Dr. Matthes. “Enterprise Architecture Management Pattern Catalog”. Tech. Rep., Software Engineering for Business Information Systems (sebis), 2008.
- [S St 10] S. Strecker, D. Heise and U. Frank. “Toward modeling constructs for audit risk assessment: Reflections on internal controls modeling”. *Modellierung betrieblicher Informationssysteme (MobIS 2010): Modellgestütztes Management*, 2010.
- [S T 95] S. T. March and G. F. Smith. “Design and natural science research on information technology”. *Decision Support Systems*, Vol. 15, pp. 251–266, 1995.
- [Simo 96] H. A. Simon. *The Science of the Artificial*. MIT Press, 1996.
- [V Kh 10] V. Khatri and C. V. Brown. “Designing Data Governance”. *Communications Of The ACM*, Vol. 53, No. 1, pp. 148–152, 2010.
- [V Ko 12] V. Koeppen, G. Saake and K-U. Sattler. *Data Warehouse Technologien*. mitp, 2012.