Universiteit Leiden

ICT in Business

Data Governance design for financial services organizations

Name: Foivos Anastasopoulos
Student-no: s1606042

Date: 26/08/2016

1st supervisor: Dr. E. (Enrique) Larios Vargas
2nd supervisor: Ph.D. Candidate B. (Bas) van Stein

MASTER'S THESIS

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Acknowledgements

I would like to express my deep appreciation for my academic supervisors, Enrique Larios Vargas and Bas van Stein, as they provided me with guidelines, suggestions and continuous support along the duration of this research thesis.

Furthermore, I want to highlight my gratitude towards my supervisors at *Company,* for the opportunity to work on this research project and their trust, as well as for their valuable assistance, insights and recommendations throughout the research process.

All of my supervisors' assistance to resolve encountered issues towards finalizing this project was critical.

Special thank you to all the people at *Company* that showed eagerness and enthusiasm to cooperate with me and provide assistance whenever needed. As far as I am concerned, the environment was ideal to perform business applied research, improve my knowledge in a variety of aspects and get acquainted with a mind-amusing intelligence of many people in this organization.

Last but not least, I want to thank my family and friends for their endless support throughout my life, in good but also hard times.

Scientific Abstract

Data is increasingly considered an *organizational asset* and modern organizations desire to retrieve business value and mitigate risks from their expanding information pools. *Data Governance (DG)* is a critical domain for *financial services organization*s considering the intensiveness of IT and data-related operations, as well as growing pressure from industry regulation. Surprisingly, financial organizations reportedly struggle to integrate DG into their structures and business model and create value from their programs. This research thesis studies the design and implementation of a Data Governance framework in the organizational and informational setting of *Company,* a small to medium sized asset management organization, using a case study methodology.

Data Governance is unique to the targeted environment, thus characteristics and priorities of the organization should guide the design of the framework (Criteria). No single framework from scientific research and the practitioner's domain was found to satisfy all requirements that stem out of *Company's* environment. Therefore an approach to synthesize a framework is discussed, mapping the derived criteria to specific DG-related capabilities that are included as components in the framework. An implementation plan for the synthesized DG framework is demonstrated, providing a structured iterative method to develop DG capabilities of an organization. The configuration of the DG organization as well as the focus areas of the program are also studied in respect of *Company's* organizational and data-related characteristics.

Generalization of the conclusions for DG design is possibly applicable for this case study as long as the target organization shares similarities with *Company* in terms of the business model, organizational characteristics, IT landscape and data environment. However, further research in similar organizations is vital to confirm generalization potential.

Contents

1. Introduction

1.1. Digitization Era

Digitization as a term is simply the conversion of analogue information into digital form. Nevertheless, following the example of (Loebbecke & Picot, 2015), the primary focus should not be on the technical level but studying the impact of digital transformation to the evolution of business, economic and societal models.

Managing to integrate innovative developments such as cloud computing, big data analytics, mobile platforms and social media networks into a company's business operation can be a source of competitive advantage, as many business leaders and executives describe it in conducted surveys (Harvard Business Review, 2014). Data analytics is defined as a way to analyze and interpret digital information, and can enhance the decision making process by providing new data sources to users and enabling them to easily navigate through them. Considering this huge growth of the available information at any given moment as well as the explosion of information systems processing capabilities, data analytics can be the primary game-changer factor in a wide variety of industries.

Furthermore, the decentralization of information access within contemporary organizations and corporations and the diffusion of the decision making activities in lower tiers of a company's structure enable managers and employees to handle a wider range of incidents and make better informed decisions without the need for constant intervention from their superiors (Margetts & Dunleavy, 2013). A solid parallelization can be made between the impact of automation capabilities for manufacturing workers and the levels of change that digitization combined with data analytics brings for knowledge workers, as data storage, processing and exploration capabilities are becoming available at low costs and great scale.

As firms and organizations try to reap the benefits of new digital technologies, a new form of strategic approach is generated that is based on the convergence points between IT and Business strategies and handles the integration of innovative technologies with an organization's business model and objectives in a more holistic manner, investigating changes and issues for products, services and business models as a whole (Matt, Hess, & Benlian, 2015). The process of digital transformation is a constant and complex venture for the majority of contemporary organizations, therefore it is vital to create a strategic plan that will guide organizations' response to the emerging concept of treating data and data-related organizational resources not as a strictly technical subject area, but rather as a business field with technical aspects.

1.2. Data as an Asset

Nowadays, an increasing number of organizations are adopting the belief that data can be a valuable asset to them, defined as data sources with a special business value to the organization (van der Meer, 2015). A simple example of data as an organizational asset is the electronic information that online retailers, for instance Amazon, store for their clients at the moment they create an account and complete a transaction (Lake & Crowther, 2013). Buying habits data can be combined with other information such as age, address, gender and cross-referenced with streaming data from social media platforms in order to create complete customer profiles. These information can also be sold or traded to advertising agencies for further profit.

Data is not an asset only for private organizations. State intelligence agencies are also seeing data as an important, intelligence asset by using data mining methods as well as monitoring, tracking and analyzing algorithms to detect undesired behavior and other anomalies.

It is a fact that companies, organizations and institutions create data at an extremely fast pace, as globally generated data is doubling every 18 months. The rise of enterprise information technology tools such as Enterprise Resource Planning (ERP), Customer Relationship Management (CRM) and other systems led to an increased number of data collection points and new collaboration and information exchange patterns for organizations and their customers, partners, suppliers and competitors. The combination of more and better data sources with flexible integration infrastructure and the progress of Business Intelligence (BI) tools and innovative data analytics capabilities contribute greatly to the fact that managerial decision making is increasingly relying on data-based analytics instead of the managers' "gut feeling".

This new trend of data-driven decision making has been an interesting area of studies and surveys for researchers (Brynjolfsson, Hitt, & Kim, 2011), reaching the conclusion that the organizations that have managed to incorporate data into their decision making processes have reached significantly improved overall productivity by 5-6%, as well as similar levels of higher profitability, improved financial performance and increased market share.

1.3. Financial Organizations

Companies and organizations of the financial services sector have always been some of the biggest producers and users of data and information. The urge of traditional banks and other financial organizations for digital transformation is now more profound than ever, recognizing that the development of digitized, omni-channel customer engagement capabilities are more than critical for future survival in this extremely competitive industry and with a customer base that is increasingly dependent on digital channels, instant communication patterns and always online service delivery (Oracle Financial Services, 2015).

The exponential growth of digital transactions performed worldwide at any given moment and the rise of digital currency platforms, for instance the Bitcoin and Ethereum platforms, that promise a decentralized, more democratic way of organizing the global economy with alternative, non-hierarchical control mechanisms are additional parameters that add up to create an innovative, tech-oriented financial services landscape. FinTech startups and alternative sources of financing such as crowdfunding platforms gain popularity as they are better synchronized with the digitization era and the emerging technologies behind this trend. Only in 2015, crowdfunding markets across Europe doubled in size compared to the previous year and totaled the amount of €6 billion, while worldwide they summed investments of $34.4 billion, increased by 112% from 2014 (Finish Ministry of Finance, 2016).

On top of that, especially in the domains of accounting, economics and finance, new regulations and arising demands from investors and other stakeholders require increased transparency in financial operations and reporting that fosters trust in the constantly changing global economy, creating this way a new epoch of digital financial information. Setting paper-based reports aside is not enough to reap the benefits of the digitization era, as the great numbers of spreadsheets, word documents and databases create a chaotic environment in many financial institutions and compose a time consuming, error-prone reporting process that is based on manual tasks and activities (Hoffman & Rodriguez, 2013).

Considering the sensitivity of financial data and the corresponding risks from possible wrong usage, most companies used an ad-hoc approach to deal with such issues where highly skilled and well-paid employees apply manual checking to all kinds of electronic documents to ensure compliance and reliability of information. Despite the fact that they could devote this time and resources to more productive activities, this approach becomes extremely difficult as the amount of documents and required tagging escalates. Thus, relevant research indicates a new direction for corporate financial reporting processes in order to reduce human intervention and automate data processing and analysis tasks, thus decreasing error-prone manual tasks that can lead to regulatory implications, increased risk and costly rectification activities (Ventana Research, 2008).

While it is widely recognized that all industries and business sectors already are or will be strongly influenced by the trends in digitization of business processes and data analytics, there is evidence that some specific industries, as for example the financial services industry, are more suitably positioned against this innovation stream (McKinsey Center for Business Technology, 2012). Figure 1 below illustrates the comparison of several industrial sectors of the U.S. economy using two different indexes:

1) *Potential for value creation using big data technologies*, taking into account the industry's competitive conditions (market turbulence, performance variability), structural form (number of potential business partners/customers, transaction intensity) and the quantity of available data.
2) *Ease of capturing the created value*, considering the capacity of the specific industry in data analytics skills (number of employees), amount of IT investments, accessibility of the data and the levels of data-driven decision making within this industry at the managerial level.

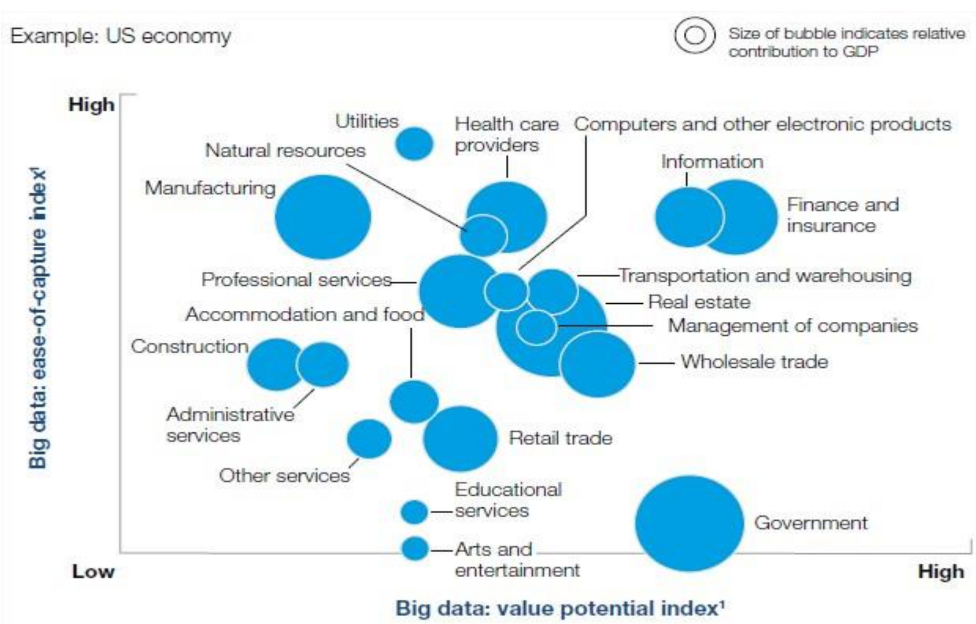


Figure 1 Financial Industries positioning in respect of reaping the benefits of big data technologies (McKinsey Center for Business Technology, 2012)

Financial institutions often demonstrate heavy investment in IT projects and strong focus to externally available or internally created information, leading to the existence of huge data pools that reside in their information systems and databases. Furthermore, most financial organizations are by nature data-intensive considering the real time data streams created and exchanged globally at any given timeslot of a day, not to mention the fact that, reasonably, the decision making style is focused on data and information that is double checked for accuracy, reliability and consistency and is not dependent to the "instinct" of any manager.

Another dimension that needs to be considered is the increased levels of analytical skills amongst employees found in companies of the financial services industry, as well as the high levels of competency in IT-related skills that can be noticed in these employees, even if they are not part of the IT function of the organization. The above can help the reader understand the reason that financial services industries, along with traditional information-oriented sectors, are positioned in a very favorable way towards taking advantage of the benefits of the big data analytics era and realizing the business value that can be retrieved from their data-related resources.

1.4. Self-Service Business Intelligence

In the process of facilitating analytical capabilities that enable end-users to perform well-informed decision making activities, the necessity to blend and create new relationships between both stationary and situational data arises (Abello et al, 2013). *Stationary* are data stored in an organization's Data Warehouse environments and other information systems, able to be directly integrated in the decision making process and owned by the decision maker himself. External data that can be related to the market, competition or to the client base and reside outside of the company's information systems and databases are called *situational* data.

These data can be structured (external databases) or dispersed across heterogeneous sources on the Web, usually have a narrow focus, are domain-oriented and often have a short lifespan for a subset of decision makers with specific needs. The concept of incorporating situational data into the decision making process gives birth to a novel range of Business Intelligence applications referred to as *situational BI*, *on-demand BI* or *collaborative BI*, however the term *Self-Service BI (SS-BI)* seems to gain ground in order to highlight the provided advanced analytical capabilities to end-users without any involvement by analysts, designers, programmers or any IT-staff in general.

Self-service Business Intelligence is defined by (Imhoff & White, 2011) as "The facilities within the BI environment that enable BI users to become more self-reliant and less dependent on the IT organization, focusing on four main axes:

- *Easier access to source data* for reporting and analysis
- *Easy to use BI tools and improved support for data analysis features*
- *Easy to consume and enhance BI results.* Simpler, customizable and collaborative end-user interfaces
- *Faster deployment options* such as appliances and cloud computing".

In the following Figure 2, an overview of the architectural design for modern Business Intelligence platforms is demonstrated (Parenteau, Sallam, & Howson, 2015). External databases as well as structured and unstructured data that reside in an organization's information systems are available for direct access via a Self-contained BI platform and by using flexible integration and data modeling techniques are stored into the in-memory engine. Subsequently, through data analytics and visualization methods and techniques performed by the business users, new dashboards, conclusions and insights are generated and shared with other members of the team or even cross-functionally with other teams and business units that possibly work with the same data. Optionally, business glossary terms and other data definitions are inserted into the BI platform through a common semantics layer that ensures consistency of generated data models, analytic dashboards and other data visualizations with the structure and organization of the existing data that can be found in the data warehouse infrastructure of the organization.

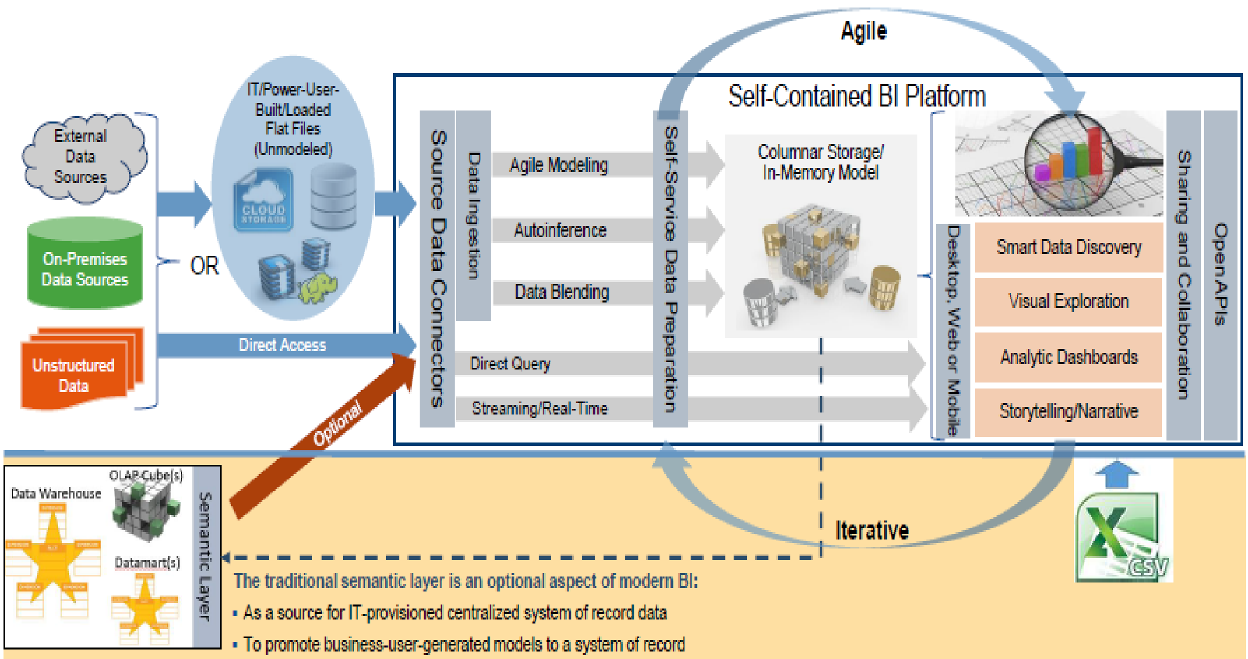


Figure 2 Modern BI Architecture (Parenteau, Sallam, & Howson, 2015)

Relevant research efforts and organizational suggestions summarize the IT/Business conflict in this respect as two opposing forces at work: the demand from IT to monitor and control the creation and distribution of data and created dashboards on the one side and the increasing need for information workers to have access to the required data sources and the freedom to change data viewpoints and operate ad-hoc analytical queries when the necessity arises.

This can lead to missed opportunities and loss of business momentum as BI users are forced to keep up with IT's slower pace while the business requirements are constantly changing and being able to follow this change can be a source of competitive advantage for many organizations. As several executives describe the situation, a shift of the ownership and governance of data assets from the IT department to the business user community needs to be made to increase flexibility and business responsiveness, but how is it possible to maintain proper IT control and establish data governance mechanisms in this novel environment?

1.5. Data Governance in Organizations

It is intriguing that while data is increasingly considered an important corporate asset and organizations reportedly struggle to provide an efficient way of organizing and administering their data, most companies lack a structured, enterprise-wide strategy for data and information governance along with the necessary standards, policies and processes.

Back in 2008, in a relevant survey conducted by the Economist Intelligence Unit amongst senior executives around the world on the benefits, challenges and risks relevant to developing an enterprise-wide information governance strategy and where 1 out of 4 respondents originated from the financial services industry (The Economist Intelligence Unit, 2008), only 38% of the participators confirmed the existence of a formal information governance structure and strategy for their company. The main obstacles identified include difficulty to specify benefits, costs and risks tradeoff from deploying a cross-organizational information governance initiative (40%), problems with enforcing corporate wide policies (39%) and support from functional directors and business line managers (35%).

Nowadays and even though Data Governance research has evolved and enriched, only about 40% of financial organizations report that their relevant programs are integrated into the business and maturity of DG initiatives has risen above the average (WBR & Informatica, 2015). Additionally, only 37% of organizations have created formal bodies or committees for their DG programs. Furthermore, only 21% of financial institutions managed to see real benefits from their DG programs in less than 12 months, whereas 30% have not yet delivered any results from their initiatives, even though this might have to do with their inability to measure and assess progress of Data Governance. Primary barriers reported are lack of awareness regarding DG value within the organization (75%), lack of information sharing and collaboration across different functions (64%) as well as change management and cultural issues (60%) (Information Governance Initiative, 2015).

Focusing on asset and insurance management companies (The Economist Intelligence Unit, 2015), an effective data strategy can benefit organizations in improved investments decisions, accurate risk assessment and improved risk management in general, faster and reliable internal and external reporting, as well as data-related cost rationalization. However, only 13% of the participants in the survey confirm that they can entirely capture and exploit the business value of their data.

In another survey conducted by PwC consultancy organization in a number of industries including the financial services sector (Messerschmidt & Stuben, 2011), one of the focus areas was the impact of data governance initiatives on Data Quality (DQ). As data is perceived as an organizational asset, the levels of data quality can consequently determine the potential business value that companies can exploit from their information pools. If data are incomplete, inconsistent or unreliable, every activity or tasks performed from information workers on these data can only result in poor decision making, mislead conclusions and increased risk especially in context of SS-BI environments.

In Figure 3, we can study the improvement of data quality levels per data domain in case a centralized governance model and platform are adopted and deployed. It is obvious that some categories of data such as financial data, customer data or product information can be dramatically benefited from a centrally coordinated data governance organizational structure and centrally controlled Master Data Management (MDM) systems.

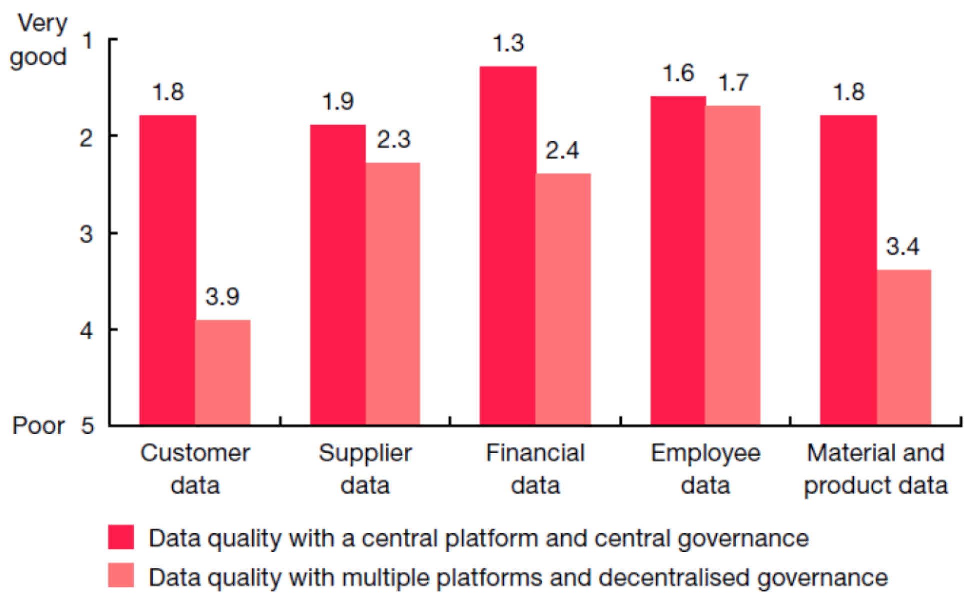


Figure 3 Impact of centralized data governance on average data quality levels per data domain (Messerschmidt & Stuben, 2011)

An effective DG framework and organizational model is considered vital for the success of investments in data analytic projects as demonstrated in (Aranow, 2014). Indeed, there are dangers that lurk below the deployment of intuitive and easy to use SS-BI tools due to the increased freedom provided to the end users, as they have direct access to enterprise data assets and infrastructure without sometimes the necessary technical expertise to deal with occurring implications. As a result, data governance practices, security mechanisms as well as IT-controlled monitoring and oversight are crucial for the successful deployment of SS-BI systems (Horwitt, 2011). Moreover, proper governance offers a chance to securely eliminate information barriers between different teams or departments of an organization, facilitating this way improved information sharing, increased transparency that also applies to the growing regulatory implications and more effective decision making processes that span throughout the organization's critical business processes.

Furthermore, several contribution areas for Data Governance in SS-BI environments are identified from TDWI (Imhoff & White, 2011):

- In order to provide *easier access to data sources*, data assets have to be identified and classified according to their origin, value and associated risk. Additionally, access to the data sources should be standardized for end-users to allow IT to supervise the process and monitor performance.

- When it comes to *easier to use BI tools and improved support for data analysis features*, governance could be applied in the form of prebuilt BI components such as reporting templates, customizable dashboards and predefined queries that could be shared and reused by end-users.
- In respect of *Easier to consume and enhance BI results*, rating and annotation mechanisms for the created reports and dashboards so that end-users can determine the value of these artifacts and integrate them into their decision making process depending on the occurring business requirements.
- Regarding *faster deployment options,* data governance has a crucial role through standardization and optimization of the ETL process that feed data to data warehouses, data marts and other repositories destined for SS-BI usage, ensuring trustworthiness, reliability and accuracy of datasets used by information workers.

1.6. Problem Definition

As discussed in the previous sections, modern organizations are presented to the challenges of digital transformation that impacts greatly various industries and business sectors. Financial organizations specifically face extremely increased demands nowadays in data collection, storage, integration, processing and analysis for their information systems and infrastructure. Considering the sensitivity of information in this industry as well as the emerging requirements from SS-BI environments, financial services organizations need to establish solid foundations in order to effectively govern organizational data assets.

1.6.1. Participating Organization

The company of this case study is a small to medium sized asset management services and investment solutions provider with multinational presence. However for the purposes of this research thesis the attention was focused on a specific location, which comprises around 85% of total working force (Headquarters).

The organizational structure consists of a number of business and operational functions, with the most important for the scope of this study being the IT department of the organization that operates under the Chief Information Officer (CIO) and also includes the Investments Back Office services. Other important functions for this case study are the Investments Front Office (core business function) and Sales & Marketing business units that have a standalone presence in the organizational chart of the company. Considering the importance of risk management as well as compliance to regulations and legislation for companies of the financial services industry, the department Compliance & Risk Management is also a crucial function of the organization with a high stake in data resources and relevant operations.

Company can be characterized as an IT-intensive organization, as currently more than 15% of employees are part of the IT department, that is responsible for delivering data, applications, interfaces and systems to the business, maintaining and enhancing the IT architecture of the organization and ensuring everything is up and running in a monitored and controlled, secure environment. An additional number of employees work in IT-related domains such as Data Management and the operational aspect of Company's information supply chain in general, for example data incidents in trades processing and execution, performance monitoring and so on. On top of that, technology and innovative solutions are an essential part of the company's day to day operation and long term vision in every line of business.

The informational landscape of the organization contains several different systems, applications and databases with some of them developed in house and others as package solutions purchased by software providers or cloud deployments. The data from source systems are delivered to the architecture's information bus by several IT teams at fixed intervals or after request from the business users, depending on the nature of the data and corresponding requirements. As the complexity of the IT architecture rises by the addition of new systems and technological solutions and the availability of data sources increases exponentially following the trends of digital transformation, the organization faces new challenges on managing its data and data-related resources. Requests from the business units for novel business intelligence platforms that will provide them with increased flexibility and improved business agility, such as Qlik Sense data visualization platform, introduce new demands of closer collaboration and understanding between IT and business teams and a new concept of conjoint responsibility when it comes to managing enterprise data assets.

Currently, central control and oversight regarding practices for governing and managing data that span through multiple systems and functions of the organization are performed by professionals in the IT department and its variants, such as the Enterprise IT Architecture and Data Management teams. This process is driven by established architecture principles, whereas formalized guidelines such as the spreadsheet policy provide principles to information workers for appropriate data usage and manipulation and avoidance of relevant risks. However, governance and management data practices are executed in an ad-hoc, case by case manner depending on emerging requirements and priorities, whereas no formal roles, responsibilities, standards and processes are established. Table 1 demonstrates the overview of the company's profile:

Table 1 Company Overview

Company	
Industry	Financial Services
Business	Asset Management
Size	Small to medium (<1000 employees)
IT Intensive	Yes: Around 25% of employees work for IT department and related teams
Data Intensive	Yes: Large availability of internal & external data sources, data driven decision-making, complex data landscape
Trigger	Self Service Business Intelligence platform (Qlik Sense)
Data Governance	Ad-hoc: No formal structures, standards and processes

1.6.2. Research Questions

As a data-intensive organization, *Company* desires to discover the hidden value in its existing and expanding data pools. As already mentioned, the envisioned IT architecture includes enhancement of existing data analysis and visualization capabilities via deployment of *Qlik Sense* data visualization platform that enables business and market analysts to turn data into valuable insights without the intervention of the IT department, providing them with enriched data analytics capabilities through a self-service environment.

As data is considered an enterprise asset with specific business value, it is essential to organize data-related resources in a way that ensures optimal usage. Central coordination and oversight has to be established to minimize risk in data-based investment decisions and reporting processes, foster information sharing and reuse of Business Intelligence artefacts such as report and dashboard templates. This way the organization can fight against information silos and management by departmental spread marts that increase risk and cost, whereas data-related processes can be formalized for shake of efficiency, as "shadow" data workflows often drain valuable resources from the IT department of the company.

Moreover, currently data ownership is tied to the system owners that create the data, therefore is unclear in several cases that multiple applications or business processes work with the same data. Furthermore, data owners should guarantee the completeness and reliability of the data itself in respect of the selected data quality metrics, but "fit for purpose" regarding the business context of data usage also needs to be safeguarded. Additionally, the desired Self-Service Business Intelligence environment is characterized by the ability provided to the users to add new data sources and combine them with the existing ones in order to create novel business insights, highlighting the necessity to create a set of rules that guides the business users through this complex data environment in case they lack the necessary technical expertise.

This thesis project includes research in scientific and the practitioner's domain in order to formulate proposals that find answers to the above challenges, which can be summarized in the following research question:

"*How to design a Data Governance framework for organizations of the financial services industry and how can it be effectively deployed and implemented?"*

The above research question can be further analyzed in the following sub-questions that shape a high level representation of the concepts for investigation and the expected focus points ought to be the major pillars of contribution for this thesis:

- *Question 1:* Which Data Governance framework is most suitable for a financial services organization such as *Company*?
- *Question 2:* How the selected Data Governance framework can be deployed within the organization to effectively manage *responsibility and accountability of* data sources and created reports and dashboards?
- *Question 3:* How can we define mechanisms that safeguard data quality and usage and provide a way to enable next-gen, self-service data analytics for end-users, while preserving suitable levels of control from the IT perspective?
- *Question 4:* How can we visualize the interaction of data items and entities with the organization's critical business processes, from the source systems to the end-user application layer?

1.7. Thesis Outline

Chapter 2 of this thesis describes the research design and process, Chapter 3 contains the performed literature review on Data Governance and related areas and in Chapter 4 the selected DG frameworks and models are evaluated taken into account parameters that are considered essential for successful DG in general as well as for the target organization. Also, an approach for organization-specific Data Governance design in demonstrated. Chapter 5 describes the implementation process for a formal DG program in the target organization of this case study based on the synthesized framework. Chapter 6 includes the conclusions of this research in respect of the defined research questions, describes constraints and limitations encountered and proposes fields of future research.

2. Research Design

This chapter describes the research design of this thesis, and more specifically the method and techniques used as well as the stages of the research process.

2.1. Research Method

This thesis follows a qualitative, case study research methodology, as it is the most suitable approach considering the nature of the research project. Case study research enables the researcher to perform the study in a natural environment, gain knowledge around the organizational reality of governing data assets and create theories from practice (Cheong & Chang, 2007; Benbasat, Goldstein , & Mead, 1987). Furthermore, with this approach the research can focus on answering the "which" and "how" questions in order to explore and comprehend the complexity of the issues under investigation, as well as the reason that these happen.

Case study as a research method gained recognition and popularity in the research community due to the limitations of quantitative methods in explaining behavioral and organizational problems in a holistic and in-depth manner (Zainal, 2007). In addition to this, applied research using a case study can be a useful tool when the focus is on a specific situation and the description of the implementation of a program (Rose, Spinks, & Canhoto, 2015), such as a Data Governance program.

Finally, case study research approach is suitable for investigating topics that few previous research has been conducted, and indeed as indicated by the Gartner's hype cycle and its positioning in the *Peak of inflated expectations* stage (Logan, 2012), Data Governance is an emerging topic and relevant research efforts have only reached a low level of maturity. Additionally, not many research efforts that study Data Governance in a particular organizational environment exist, and only a small subset of them investigate Data Governance in the context of special requirements that come from financial services industry.

Mainly 3 reasons for single case study design are proposed (Rose, Spinks, & Canhoto, 2015; Benbasat, Goldstein , & Mead, 1987), in our case 2 out of 3 can be confirmed. The study is *critical* in some way (theory testing) as Data Governance theory is evaluated in the context of a particular organizational and informational setting and is *unique* to the targeted environment and its requirements, even though generalization of the findings in similar organizations is desired and pursued.

Several other research on Data Governance within organizations have also find the case study approach relevant and suitable in order to investigate the topic in the lenses of organizational reality (Weber, Cheong, Otto, & Chang, 2008; Cheong & Chang, 2007; van der Meer, 2015). To the best of our knowledge, only two research studies found are completely focused on financial services organizations (Traulsen & Tröbs, 2011; Faria, Macada, & Kumar, 2013), whereas one of the case studies in (van der Meer, 2015) studies DG maturity in a bank.

2.2. Research Design

The conducted research followed the process design demonstrated in Figure 4. Initially, the intention of the organization to deploy a SS-BI platform was used as a trigger to explore and collect requirements from multiple perspectives regarding data governance practices. Afterwards the topic was investigated using a literature study, which led to the process of evaluating the found DG frameworks and models based on company-specific criteria. Using the criteria and the evaluation outcome as an input, a novel DG framework was synthesized and the corresponding implementation plan was designed. Finally, the findings were assessed in respect of the research questions and the research conclusions were drafted.

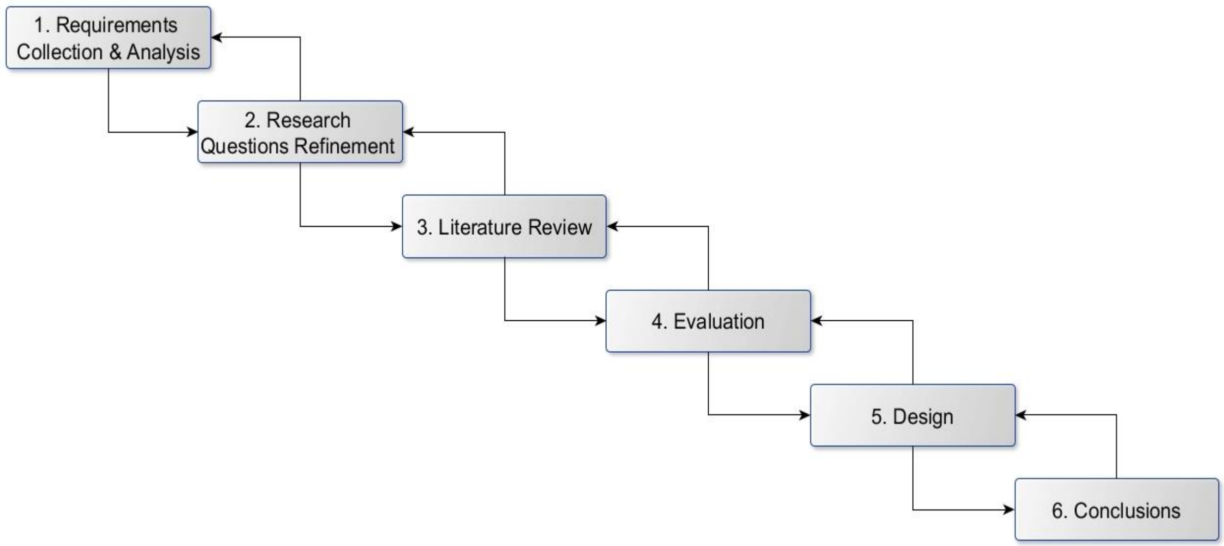


Figure 4 Stages of the research process

However, the research process did not follow a strictly sequential approach and at several points backward steps were performed, for example to receive feedback from the stakeholders, refine the research questions, review emerging literature and implement changes in the design. The following Table 2 shows the sections of the thesis that correspond to each stage of the research process.

Table 2 Sections that the stages of the research process are discussed

Research Stage	***Discussion***
1. *Requirements Collection & Analysis*	Sections 1.6, 2.2.1
2. *Research Questions Refinement*	Section 1.6.2
3. *Literature Review*	Chapter 3
4. *Evaluation*	Chapter 4
5. *Design*	Section 4.7, Chapter 5
6. *Conclusions*	Chapter 6

2.2.1. Stage 1: Requirements collection & analysis

In order to understand the organizational setting of the research and collect information about the data landscape, IT architecture, core business processes and so forth, multiple stakeholders from various functions and hierarchies within the organization were contacted, as illustrated in Table 3.

Table 3 Stakeholders matrix

	Function	***Position***
IT	Head of IT	CIO
	IT Architecture	Executive Director, Enterprise IT Architect
	Delivery Team A	Director, Team lead, Other IT professionals
	Delivery Team B	Director, Other IT professionals
	Delivery Team C	Director, Other IT professionals
	Quality Assurance	Director, Specialist
	IT Operations	Director
	Data Management	Director
	Data Quality	Manager
	End-User Computing	Team lead
	Information Security	Enterprise IT Architect
	Compliance & Risk Management	Senior Manager
Business	Investments Front Office	Portfolio Managers
	Investments Back Office	Implementation Managers
	Sales & Marketing	Specialist

Informal, unstructured interviews with these stakeholders were conducted in order to explore concerns regarding current data practices and requirements in respect of the desired future state. As mentioned, the intention to deploy a SS-BI platform was used as a trigger to facilitate relevant discussions about organizational data practices. One of the most interesting takeaways from this process was the confirmation of Business & IT as two opposing forces at work in the company's organizational reality, as IT demands increased control and security over data assets and Business calls for improved access and availability of data sources.

A necessary remark is that stakeholders from *Compliance* & *Risk Management* function are categorized in the *IT* domain due to the alignment of demands towards increased control, whereas stakeholders from the *Investments Back Office* department are considered *Business* representatives despite the technical nature of their job, as increased access and availability of data sources is vital from their perspective.

The qualitative methods used to collect data and information about the investigated organization were enhanced with limited quantitative data through questionnaires that required input from stakeholders In order to calculate the weights for the criteria used in section 4.4. The complete list of data collection methods used in this research thesis are demonstrated below:

- *Documentation*: IT architecture drawings, Process maps, Spreadsheet policies, Other formal and informal documents.
- *Archival records*: Organizational chart of Company, Data records.
- *Interviews*: Open ended, unstructured interviews with major stakeholders, Questionnaire (Appendix A).
- *Direct observation*: Decision making style, Business & IT interaction.

2.2.2. Stage 2: Research Questions

Based on information collected from stakeholders and other sources as described, the research questions of this thesis were gradually refined to their final form as already presented in section 1.6.2. Research question 1 is answered in Chapter 4, Research Question 2 is discussed throughout Chapter 5, Research Question 3 in sections 1.5, 5.6, 5.9 and 5.12 and finally, Research Question 4 is answered in sections 5.8 and 5.10.

2.2.3. Stage 3: Literature Review

An extensive literature review was performed to shape an overview of Data Governance concepts, principles, frameworks and models. Related areas such as Data Management and Data Quality were also explored in respect of their interrelation with DG. The major keywords used included: *(Information) Governance, Data (Information) Management, Data Stewardship, Data Quality, Information Security & Compliance, Self Service Business Intelligence Governance, Analytics Governance.*

The literature review included exploration of sources in both *scientific* and *grey* literature:

Scientific literature was explored mainly by using Google Scholar in combination with the selected keywords and possible combinations. Furthermore, scientific databases relevant to Information Science, Information Management, Computer Science, Management Science and so on were also scanned. Some examples amongst others are the IEEE Computer Society, Science Direct and ACM Digital Library. Exploration of scientific literature initially was performed without any filters, and subsequently was focused on recent publications of the last few years. Reference lists of the found material were further explored for additional papers.

Grey literature was searched including consultancy suggestions, organizational best practices, case studies, presentations and other material from DG practitioners. Grey literature is essential to understand the application of scientific concepts in realistic business environments and gain insights on how Data Governance is implemented and used in modern organizations.

A detailed list of all literature items reviewed for this research thesis is demonstrated in the Bibliography section of this document.

2.2.4. Stage 4: Evaluation

After Data Governance frameworks and models were collected in the literature review stage, another round of meetings with a subset of the initial stakeholders was performed in order to inform them about the literature review findings and evaluate their relevance to the targeted organization. In this context, the idea of creating a set of criteria to evaluate the available material came into action, that represent both widely accepted important aspects of DG but also company-specific requirements. Criteria were developed and weights were assigned to them based on input from the stakeholders (section 4.4). For this purpose, the questionnaire of Appendix A was distributed to 13 stakeholders to confirm the criteria list's validity and request their opinion on the prioritization (weights) of the criteria. In total, 10 out of 13 stakeholders responded (77%) and no new criteria occurred from this process. Subsequently, the selected DG frameworks were evaluated both qualitatively in text as well as quantitatively against the selected criteria in section 4.5 and the evaluation outcome is discussed in section 4.6.

2.2.5. Stage 5: Design

An approach to design an organization-specific framework is discussed along with an indication of the result of this synthesis process (section 4.7), whereas in Chapter 5 a suitable implementation plan for effective deployment of an official DG program within the targeted organization of this research is presented. Even though the described methods and implementation structure are designed with the intention to be easily generalized and applied to other organizations with similar characteristics and requirements, an effort has been made to identify examples that stem out of Company's environment in order to increase relevancy of the outcome.

2.2.6. Stage 6: Conclusions

The conclusions of this research in respect of the defined research questions are discussed in Chapter 6, and Section 6.5 contains an overview of constraints and limitations encountered during this research along with motivation for future research of related topics.

3. Data Governance

This chapter contains an extensive literature review on definitions and drivers for Data Governance, as well as the interrelation of these initiatives with relevant domains such as IT Governance, Data Quality and Data Security (DSec). Moreover, the concept of Data Stewardship (DS) is introduced and several Data Governance frameworks and models are presented and described in detail.

3.1. Data Governance Definitions

For reasons of clarity, a comparison between the terms "data" and "information" is essential. As (Otto, 2011) describes, data can be considered the "raw material" of information, whereas information can be defined as data in context or processed data, however he indicates that practitioners frequently use the terms interchangeably. The evolution of business models is conquered by technology and strongly influenced by the outcomes of digital transformation, leading frequently to the misinterpretation of information as product of technology in an objective manner (Faria, Macada, & Kumar, 2013).

However, information is the result of subjective interpretation of objective facts (Beijer, 2009), and this conclusion is supported by the shift in organizational practices regarding Data Governance from technology-oriented to people, policies and processes-oriented, as companies come to the realization that they can only retrieve the full business value of their information assets with proper governance of data access, analysis and protection. In order to avoid undesired misconceptions and enhance clarity, this thesis follows an approach that equates data and information and the subsequent definitions of data governance and information governance, as described in (Ladley, 2012).

Several Data Governance definitions can be found in scientific research:

1. (Cohen, 2006) defines data governance as "the process by which a company manages the quantity, consistency, usability, security and availability of data",
2. (Weber, Otto, & Osterle, 2009) extend IT Governance definition and conceive Data Governance as a framework for decision rights and accountabilities to promote desirable behavior in the use of data.
3. (Khatri & Brown, 2010) include the conceptualization of data as an asset in their definition: "Data governance refers to who holds the decision rights and is held accountable for an organization's decision making about its data assets."

When it comes to the practitioner's community and shaped definitions:

1. The IBM Data Governance Council (IBM Software Group, 2007) defines data governance as "the political process of changing organizational behavior to enhance and protect data as a strategic enterprise asset".
2. (Thomas G. , 2006) describes data governance as a strategy that "refers to the organizational bodies, rules, decision rights and accountabilities of people and information systems as they perform information-related processes. Data governance sets the rules of engagement that management will follow as the organization uses data".

3. (Newman & Logan, 2006) refer to data governance as “the collection of decision rights, processes, standards, policies and technologies required to manage, maintain and exploit information as an enterprise resource”.

Considering the above definitions and their convergence points respectively, Data Governance can be summarized to contain the following basic pillars:

- Data is a company asset, organizations must retrieve business value, minimize risk and seek ways of further exploitation.
- Definition of specific responsibilities and decision rights on data assets, role distribution to apply the decisions and specification of the exact tasks need to be performed for the decision making activities.
- Develop enterprise wide principles, standards, rules and policies for data management practices. Consistency, credibility and accuracy of the data are vital to ensure value-adding decision making.
- Establish the necessary processes that provide continuous monitoring and control of data practices and help to enforce data-related decisions across different organizational functions and business user categories.

3.2. Data Governance Motivation

Establishing a data governance framework with all the necessary responsibilities, standards and policies in order to safeguard the digital assets of an organization is a challenging task, since mindsets of people have to be possibly altered to adopt this initiative and resistance to change issues have to be overcome. For this reason, the impact of a data governance effort has to be evaluated and communicated across the organization, supported by relevant business drivers that justify the resources allocated in this project and provide the necessary motivation for employees to commit to the development and protection of a company’s data assets.

The most common business drivers for organizations to invest in data governance initiatives are the following (Panian, 2010; WBR & Informatica, 2015):

- ***Growth- Business Agility:*** Data are frequently dispersed across all kind of source systems and databases, making it harder for organizations to preserve a consistent view of the valuable information owned. By organizing data and information in a more consolidated manner, business users of a company’s systems and applications can arrive to market insights and recommendations to customers faster and with increased efficiency, providing this way improved services and higher levels of customer satisfaction that can in turn lead to increased market shares and continuous growth. Moreover, data quality as a core aspect of DG has been proven to have a positive impact on business growth of organizations, increased customer satisfaction and overall improved business efficiency.
- ***Lower Costs- Operational Efficiency:*** Considering the timeless effort of corporations to minimize costs and the importance of operational efficiency in this venture, we can identify new ways to achieve the targeted goals by streamlining and automating when possible all data-related business processes. The majority of these processes are flowing through multiple systems,

databases and applications of several business units of an organization, such as Finance, Human Resources, Sales & Marketing and others. Providing a consolidated and clearer view of the data and information owned can assist in the elimination of many manual tasks and the resolution of several redundancy issues that keep the costs and complexity at high levels.

- ***Risk Management- Compliance & Security:*** Especially for organizations and institutions operating in the sensitive and highly regulated financial services industry, risk management as well as compliance to external regulations and internal policies introduces additional requirements for transparency in data usage and reporting operations based on these data. These demands are impossible to be satisfied without a consistent and holistic overview of a corporation's data assets and as a result, data governance initiatives have to be recognized as crucial elements for ensuring compliance by defining all the necessary data standards, policies and processes and by shaping a framework with clear roles and responsibilities that control the application of these policies.

At the time being, the majority of organizations have adopted an application-centric approach in their data management practices and methods, due to the fact that control of data is usually assigned to the IT department of these organizations and are managed in the context of how they will serve best the targeted applications and corresponding user interfaces for these applications. (Informatica, 2006; Panian, 2010).

However, the belief that data, including customer, employee or product data, are an enterprise asset that has to be available to different functions and business units of an organization and has to be shared across several applications and business processes is gaining ground in adopted data governance and data management practices. This transition to a business process oriented data architecture is illustrated in Figure 5 and comes to remove limitations in creating value from enterprise data assets.

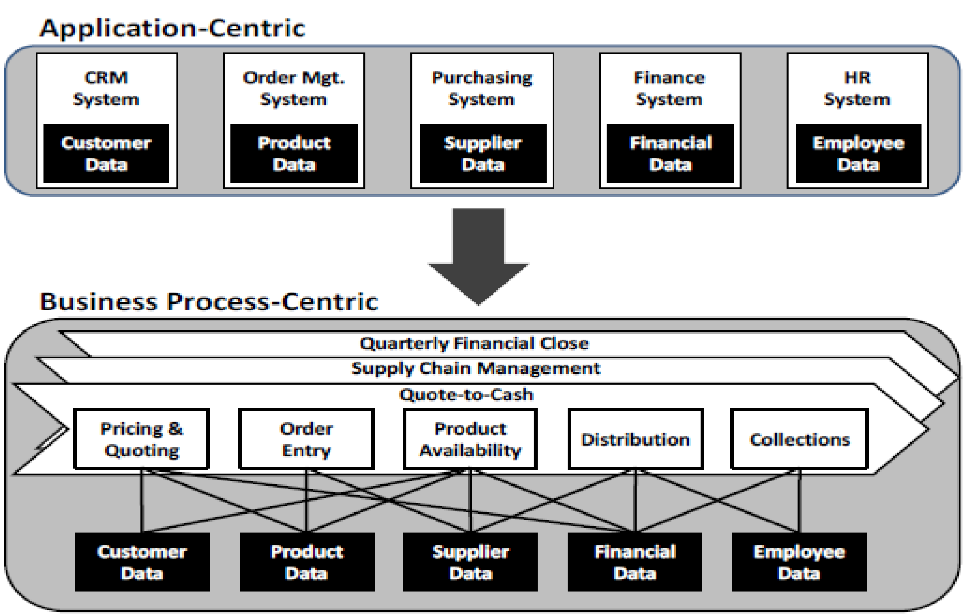


Figure 5 Application-oriented vs. Business process-oriented data architecture (Informatica, 2006)

According to executives from organizations (O'Neill, 2015), keeping track of data flows trails for auditing purposes is one the biggest hurdles for contemporary organizations, due to the increasing complexity and diversity of the origin, types and target systems of these data flows that introduce new risks and compliance issues for companies without a proper data governance plan established. Organizations must eliminate "shadow" data-related processes and visualize the data paths from the source repositories and databases to the end-user systems and applications. This way, they can subsequently apply the necessary standards, rules and policies on the visualized data workflows and establish effective data practices that assists them in managing their data assets and taking advantage of their business value.

3.3. Interrelation of Data Governance with relevant domains

(Russom, 2006) from The Data Warehouse Initiative recommends to establish an executive-level governance committee responsible for supervising application of data-related standards, rules and policies across different functions of the organization. Formulating an enterprise wide viewpoint on an organization's data assets is important to coordinate efforts and combine knowledge and insights from both the technical details and business purpose of the data.

(Wende, 2007) places both Data Governance and IT Governance as functions that have to be aligned with Corporate Governance guidelines and principles as shown in Figure 6, but distinguishes the two terms by emphasizing that DG refers to closer collaboration and improved understanding between IT and Business professionals. Data Quality Management is defined as the day-to-day decision making regarding collection, organization, storage, process and exploitation of high quality data and is a core part of Data Governance. DQM is the actual decision making on data practices according to the guidelines and principles provided by the framework for DG.

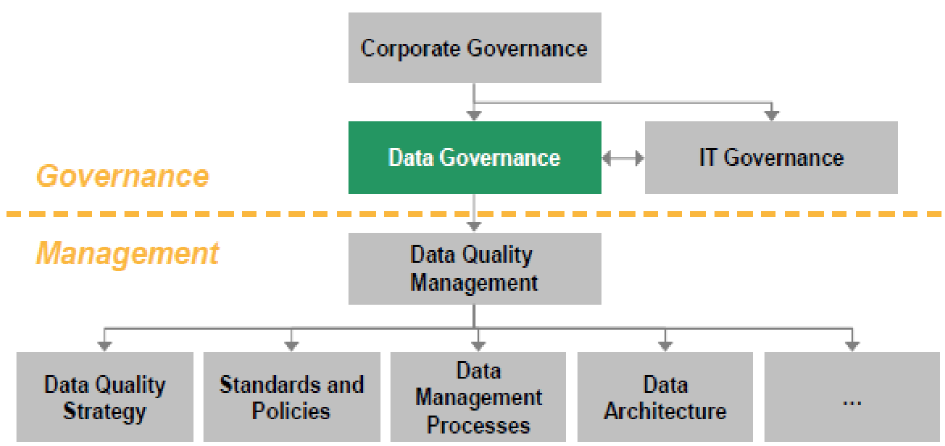


Figure 6 Data Governance and related domains (Wende, 2007)

In Figure 7, (Hoying, 2011) from Stanford University provides a schematic representation of the relationships between entities and concepts that belong in the data governance environment. More specifically, Data Stewardship (DS) is a large part of the Data Governance (DG) and is the primary enabler of Data Quality (DQ) efforts such as Data Quality Management (DQM). However, DQ projects can be deployed as an initiative outside of the scope of the DG domain. Finally, Data Security (DSec) is the major component that defines requirements and demands for specification of DG policies and standards and in a lower level, DS tasks and responsibilities.

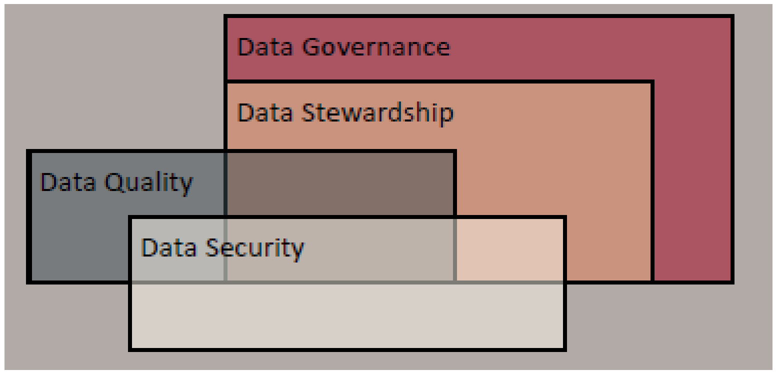


Figure 7 Relationship between DG, DS, DQ & DSec (Hoying, 2011)

As (Otto, 2011) puts it, the findings around the strong interrelation between Data Governance and Data Quality occur from the perception of data as an organization asset with a specific value and, similarly with physical assets, they need to be managed in a way that this value is maintained and enhanced.

3.4. Data Stewardship

As organizations try to make the transition to a mentality of embracing data as a corporate asset, the realization that data quality and data governance initiatives should be a business issue, guided and driven from business requirements and objectives comes to the surface (Friedman, 2007).

Data stewardship is a quality-control discipline, representing the formalized responsibility regarding management and use of the data in order to optimally address business requirements. Data stewards are responsible for improving the quality of the data as an enterprise asset to maximize its business value and control associated risks (Soares, 2010). Essentially, data stewardship is the concept of assigning responsibility for management of the organizational data resources to the people that work with the data on a day-to-day basis, as they are the ones that know their business value, usage context and inherent risk.

Gartner indicates that Data Stewardship is most effective when it is positioned close to the sources and systems that data are captured and maintained (Friedman, 2007). However, a frequent misconception found in companies that try to apply Data Governance is the equation of Data Stewardship with Data Ownership, leading to unsuccessful or with minimum impact governance programs. Ownership refers to restriction and control of access, whereas stewards are responsible not for owning the data but for performing data-related activities in order to support business processes with data of acceptable quality in a timely and reliable manner.

(Marco, 2014) analyzes data stewardship from a more realistic organizational perspective. Data is recognized as an important organizational asset and consequently, the role of data steward has grown significantly over the years to ensure enterprise data and metadata are suitably defined and understood across the organization. The data steward acts as the connectivity point between IT and the business, aligning decision supporting and operational systems with the business requirements and end-user's demands. Data stewards should be visible, respected and influential so that the importance of the Data Governance program is adequately communicated to the rest of the organization.

However, not all data stewards in a data governance organization should be the same. A modern company owns data in a wide variety of fields, financial, operational, sales & marketing, employee information that fall into data privacy limitations as well, and therefore, it is important to be handled and administered by a data steward that has knowledge of the context and unique characteristics. He proposes to organize the data governance organization in several areas of interest within the enterprise that are called *subject areas*, representing logical grouping of items of interest to the organization. Each subject area includes *Business Data Stewards* responsible for the business interpretation of the available data (e.g. attribute definitions) as well as *Technical Data Stewards* that oversee technical data-related issues such as data transformations or other activities of the ETL process.

In order to further elaborate on how the various data stewardship teams are organized, valuable insights can be found in organizational best practices as discussed in (Dyche & Polsky, 2014). The authors perform an analysis of Data Stewardship concepts as a central connectivity point within organizational environments of increased data complexity and volume, enforcing accountability via assignment of specific decision rights, defining clear data ownership and providing increased visibility in data-related workflows. The main argument here is that parameters of the application environment will determine the way Data Stewardship will be applied in this environment, describing five different approaches:

- ***Data Stewardship by Subject Area***, where data stewards are domain and subject area experts.
- ***Data Stewardship by Function***, where data stewards focus on data tied to an organization´s specific function or department.
- ***Data Stewardship by Business Process***, an approach that is more suitable for corporations with a clear view of their data-related critical business processes.
- ***Data Stewardship by Systems***, a highly IT-centric approach with data stewards assigned to the systems that generate or capture the data he is responsible for.
- ***Data Stewardship by Project***, a faster and more practical approach for organizing Data Stewardship teams within a company.

Naturally, the above are not "one size fits all" models that can guarantee success in all cases, but rather an indication of how organizations could start deploying their Data Stewardship programs depending on their characteristics, requirements and priorities. Hybrid versions that combine elements from two or more DS models are also possible to be developed, for example appointment of *Data Stewards per Function* with a lower layer of additional Data Stewards responsible for one or more *Subject Area* of this function. Another interesting variation would be the *process-by-function combination* in which each organizational department that participates in a business process assigns a Data Steward to this process. This approach includes multiple stewards for each business process and functional stewards might operate in multiple processes at the same time, establishing this way a more collaborative culture in data management activities.

3.5. Building Blocks for Data Governance

This subsection demonstrates several Data Governance frameworks found in scientific research and the practitioners' domain. Effective Data Governance requires suitable data policies, guidelines and standards have to be designed and developed in synch with the organization's business model, strategic direction, values and culture.

A DG framework is a high level representation of elements and components of DG that demonstrate data-related principles and standards and provide the guidelines to enforce them with suitable configuration of formal roles and responsibilities for the overall data strategy of the organization. The usefulness of a DG framework is praised by several researchers and practitioners as it provides a way to introduce the necessary abstraction layers that make it easier to communicate DG concepts and domains to audience of diverse background and expertise, from both the business and the technology world (Aranow, 2014; Blosser & Haines, 2013). The DG framework can be used by organizations as a reference point for the progress and evolution of their information-related programs and projects.

(Panian, 2010) suggests six *Key Data Attributes* that are usually associated with digital challenges that contemporary organizations face:

- ***Accessibility***, all enterprise data have to be accessible no matter their structure or source systems and applications.
- ***Availability***, all data are available to end users and applications, when, where and how needed.
- ***Quality***, data satisfy requirements of the business in terms of completeness, reliability, accuracy and integrity.
- ***Consistency***, ensure consistency of the meaning of data and eliminate different views on the same data by establishing common business semantics across all systems, processes and business functions of the organization.
- ***Auditability***, enforce traceability of the data and suitable control of dataflow so that auditing demands are satisfied and relevant implications are avoided.
- ***Security***, ensure secure access to the data and define proper access and authorization layers.

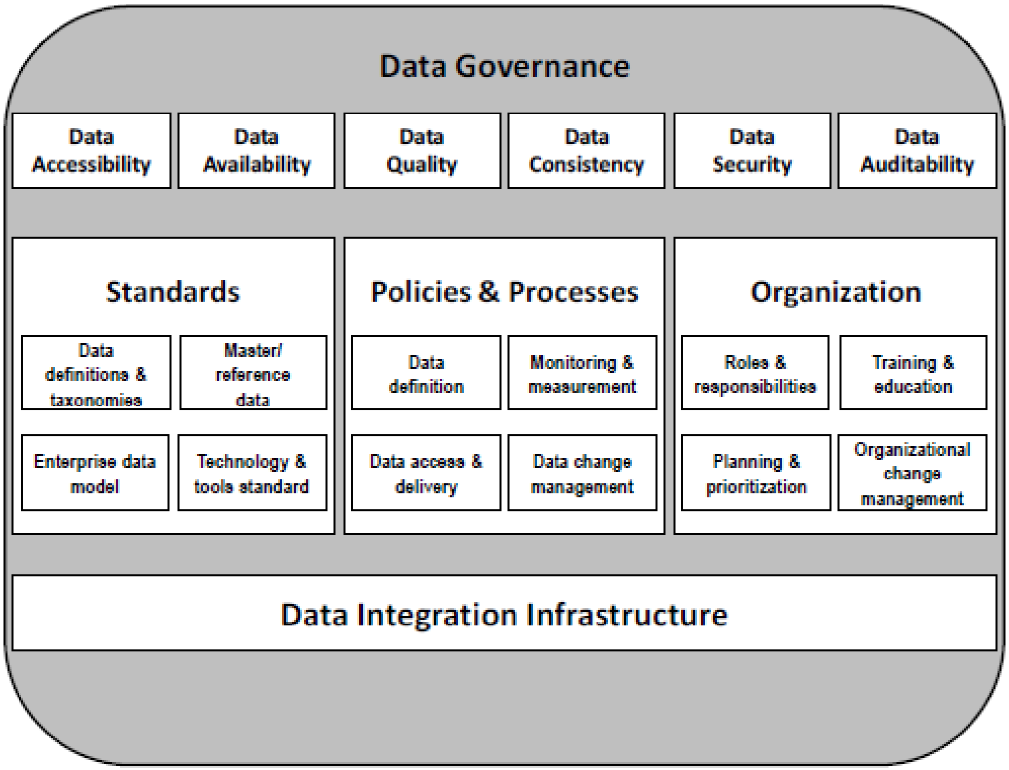


Figure 8 Basic Components for a Data Governance Framework (Panian, 2010)

Furthermore, as Figure 8 above shows, he describes four major domains for data governance:

- ***Standards***, referring to the definition of common data definitions and taxonomy for different business units, development of enterprise wide data models, specification of master data and all other technical details and standards related to the data.
- ***Policies & Processes***, as business rules for data usage, access and delivery have to be defined and the necessary monitoring and measurement mechanisms have to be provided in order to control performance and handle incidents during the data's lifecycle in various systems and applications.
- ***Organization & Roles***, meaning the assignment of roles and responsibilities that connect the enterprise data with a virtual, data governance organizational structure. Blending demands from both business and IT perspective is also a crucial aspect of the data governance organization as control of data flows has to be ensured concurrently with providing the data in a way that will promote the business-oriented data usage.
- ***Technology***, the technological infrastructure responsible for data integration tasks. Many organizations deploy data governance programs with an ad hoc, manual approach using spreadsheets, diagrams and other documents, however the real benefits of a data governance initiative can only be visible by using the automation capabilities technological solutions can provide.

The Data Governance Institute proposition (Thomas G. , 2006) demonstrated in Figure 9, provide a similar categorization for the basic components of a data governance framework across three main fields:

- ***People & Organizational Bodies (WHO)***, referring to various data stakeholders and owners within the organization, the Data Governance Office responsible for definition of data standards, guidelines and issue resolution and the Data Stewards that are assigned with the operational supervision of data governance activities and report the occurring issues to the DGO for discussion and resolution.
- ***Rules, Policies & Standards (WHY & WHAT)***, that have to be specified and communicated across the organization and will be the point of reference for safeguarding data usage according to the company's business model, mission and objectives, combined with the necessary monitoring and control mechanisms that will measure the effectiveness and efficiency of the data governance initiative.
- ***Processes (WHEN & HOW)***, that have to be established for the alignment of Business strategy with IT control to provide an enterprise wide viewpoint on the organization's data flows and to define a way to attach the defined metrics and business goals as well as the assigned roles and responsibilities to the company's business processes that span multiple functions and teams.

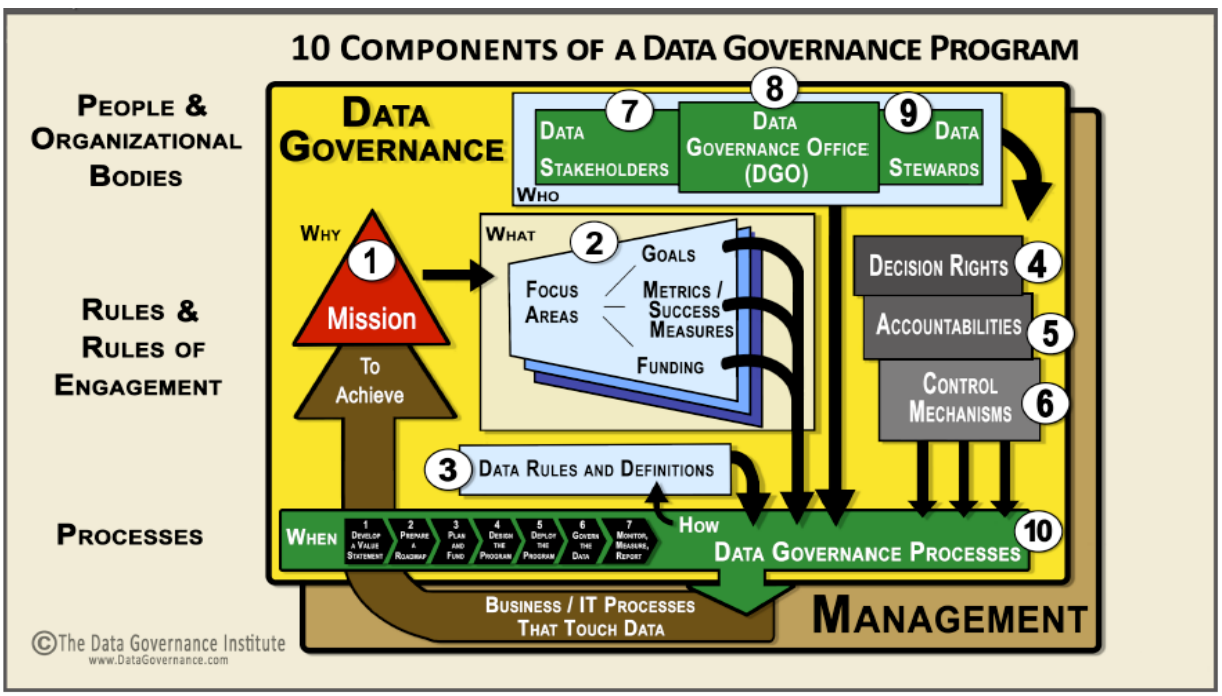


Figure 9 Data Governance Framework, Data Governance Institute (Thomas G., 2006)

(Cheong & Chang, 2007) propose a framework organized in 3 dimensions: *Organizational Bodies and Policies, Standards and Processes* and *Data Governance Technology,* distributing the internal components that correspond to DG capabilities and activities along these dimensions as presented in Table 4.

Table 4 Data Governance components (Cheong & Chang, 2007)

Organizational Bodies & Policies	***Standards & Processes***	***Data Governance Technology***
Governance Structure	Data Definition & Standard (Metadata Management)	Metadata Repository
Data Custodianship	Third Party Data Extract	Data Profiling tool
User Group Charter	Metrics Development and Monitoring	Data cleansing tool
Decision Rights	Data Profiling	
Issue Escalation Process	Data Cleansing	

IBM (IBM Software Group, 2007) provides a slightly different classification for the major components of an effective Data Governance framework as illustrated in Figure 10, focusing on the interconnection and shaped relationships between four major entities:

- ***Outcomes***, describing the techniques and mitigation measures for risk identification, prioritization and avoidance (***Data Risk Management & Compliance)*** as well as the process and methods by which data assets are organized in a corporation in order to enable maximized ***Value Creation*** from the business.
- ***Enablers***, referring to balancing responsibility between business units and the IT department regarding data governance (***Organizational Structures & Awareness)***, ***Stewardship*** as for the quality control domain for data asset enhancement, risk mitigation and organizational control and ***Policy***, the formal representation of the desired organizational behavior around data assets that is enforced by stewards and the rest of the data governance organization.
- ***Core Disciplines***, specifying the methodology used to assess and improve the quality, reliability and integrity of data (***Data Quality Management)*** along with a collection of policies and rules for ***Information Lifecycle Management*** referring to data collection, usage, retention and deletion. Finally, the ***Information Security & Privacy*** component describes the practices and controls used by an organization to protect data assets and mitigate risk from a security and data privacy perspective.
- ***Supporting Disciplines,*** in terms of how data sources, systems and applications are interrelated in the organization's ***Data Architecture***, enabling availability and delivery of data to the suitable end-users, the tools and techniques used to create a consistent, enterprise wide set of definitions for business semantics, data types and terminology as well as the data models attached to the company's repositories ***(Classification & Metadata)***, and the processes and metrics in place for measuring data value and the overall efficiency of the deployed data governance framework ***(Audit Information, Logging & Reporting).***

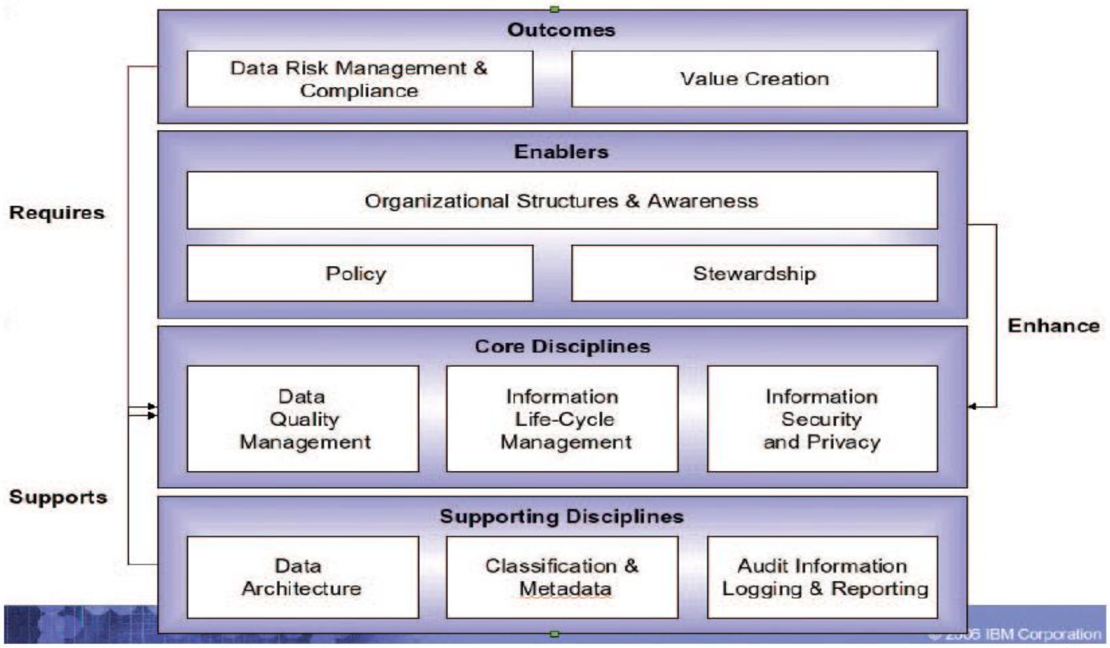


Figure 10 Elements of Effective Data Governance (IBM Software Group, 2007)

The Data Management Association (DAMA-DMBOK2 Framework, 2014), defines 11 knowledge areas related to data management activities and practices, which are depicted in Figure 11. ***Data Governance*** is considered the central component of the Data Management domain and according to this framework can be described as the planning, oversight and monitoring tasks over management of data and usage of data or other data-related resources. ***Data Architecture*** knowledge area has to do with the integration of the overall structure of the data with the organization's enterprise architecture, whereas ***Data Modeling & Design*** field includes analysis, design, building, testing and maintenance activities of data models. ***Data Storage & Operations*** deals with management and deployment of a company's physical data assets that also have a structured form and ***Data Security*** with privacy, confidentiality and access aspects of data sources.

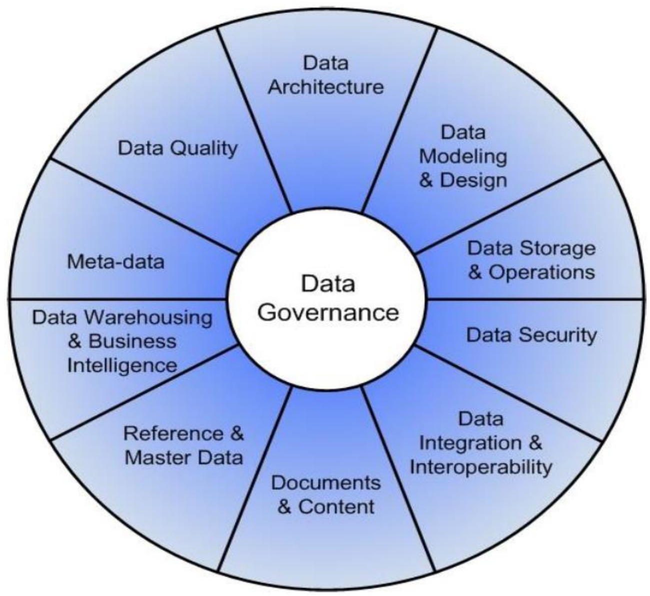


Figure 11 Data Management Knowledge Area Wheel (DAMA-DMBOK2 Framework, 2014)

ETL tasks for data extraction, transformation and loading as well as other activities regarding data replication, federation, virtualization and delivery are part of the ***Data Integration & Interoperability*** component of the framework. ***Documents & Content*** knowledge area is about storage, protection and indexing tasks that enable proper access to unstructured data such as digital documents and physical records. Standardized definitions about data and their potential usage are responsibility of the ***Reference & Master Data*** field in order to reduce redundancy and improve data quality, while enabling access to decision support data for reporting and management of data processing & analytics are part of the ***Data Warehousing & Business Intelligence*** knowledge area. ***Metadata*** subdomain has to do with gathering, categorization, integration, delivery, monitoring and maintenance of the metadata, that is the activity records of the data itself, and finally ***Data Quality*** includes the definition of related metrics that help to monitor data usage and evaluate impact of data management activities to the quality of the organization´s data assets.

Additionally, in order to provide a logical and consistent definition of the various knowledge areas, seven Environmental Elements are selected that can be grouped into a larger category based on another type of descriptor: ***People, Process*** or ***Technology*** as demonstrated in Figure 12.

Figure 12 Environmental Elements of Data Governance (DAMA-DMBOK2 Framework, 2014)

In an effort to elaborate further on the actual content of data-related decision domains, (Khatri & Brown, 2010) present an approach that supports the concept that management and governance of data assets should be aligned and correlated to the management and governance of IT assets, showing this way the direct relationship between Data Governance and IT Governance models previously developed from other authors (Weill & Ross, 2004). This framework focuses on the fundamental decisions that need to be made regarding organizational data assets and by who.

Initially, the authors discuss the concept of key organizational assets that need to be governed, extending and adapting existing definitions from other research efforts. More specifically, six categories of organizational assets are defined as shown in the following Figure 13, with the difference that they differentiate between IT and information assets, with the former referring to the technological infrastructure that support business processes (e.g. computer systems, databases etc.) and the latter to facts and information with a perceived business value that reside in the infrastructure.

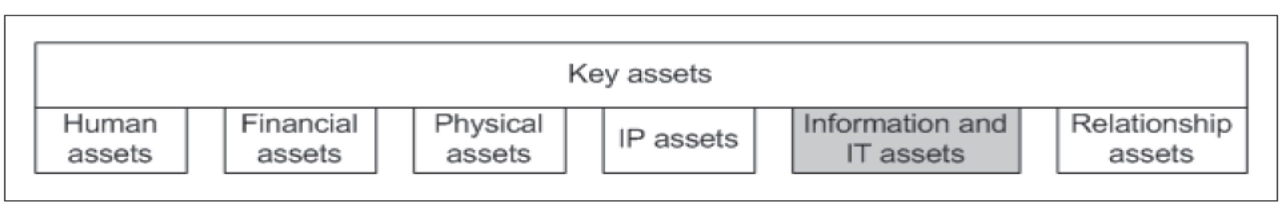


Figure 13 Key Organizational Assets (Khatri & Brown, 2010)

Subsequently, they analyze the Data Governance domain from a decision making perspective around an organization's data assets and identify five interconnected decision domains. ***Data Principles*** are the foundation of the DG design, setting the direction for all other data-related decisions and defining the requirements for the desired use of data. These requirements establish the standards for ***Data Quality*** within the organization, which form the basis for how data interpretation (***Metadata***) and ***Data Access*** is performed by the users. Finally, decisions around data production, retention and deletion are assigned to the ***Data Lifecycle*** data domain.

Furthermore, there are detailed descriptions of which decisions fall under each Data Governance domain and who should be responsible or accountable for these decisions across a number of different data or information-related roles that can be found in modern organizations as shown in Table 5.

Table 5 Data Governance Domains, Decisions & Roles (Khatri & Brown, 2010)

Data Governance Domains	***Domain Decisions***	***Potential Roles or Locus of Accountability***
Data Principles: *Clarifying the role of data as an asset*	• What are the uses of data for the business? • What are the mechanisms for communicating business uses of data on an ongoing basis? • What are the desirable behaviors for employing data as assets? • How are opportunities for sharing and reuse of data identified? • How does the regulatory environment influence the business uses of data?	• Data owner • Data steward • Data producer/supplier • Data consumer • Enterprise Data Committee /Council
Data Quality: *Requirements for intended use of data*	• What are the standards for data quality with respect to accuracy, timeliness, completeness and credibility? • What is the program for establishing and communicating data quality? • How will data quality as well as the associated program be evaluated?	• Data owner • Subject matter expert • Data quality manager
Metadata: *Semantics or "content" of data for interpretability*	• What is the program for documenting the semantics of data? • How will data be consistently defined and modeled so that it is interpretable? • What is the plan to keep different types of metadata up-to-date?	• Enterprise data architect • Data modeling engineer • Enterprise Architecture Committee
Data Access: *Specifying access requirements of data*	• What is the business value of data? • How will risk assessment be conducted on an ongoing basis? • How will assessment results be integrated with the overall compliance monitoring efforts? • What are data access standards and procedures? • What is the program for periodic monitoring and audit for compliance? • How is security awareness and education disseminated? • What is the program for backup and recovery?	• Data owner • Data consumer • Chief information security officer • Information Security Architect • Enterprise Architecture Committee
Data Lifecycle: *Definition, production, retention and*	• How is data inventoried? • What is the program for data definition, production, retention, and retirement for different types of data?	• Enterprise data architect • Information supply chain manager

retirement policies of data	• How do the compliance issues related to legislation affect data retention and archiving?

3.6. Organizational Aspects of Data Governance

According to (Thomas G. , 2006), three categories of data governance approaches can be identified depending on the level of formal control as well as the positioning of the decision-making focus:

- ***Governance via Management***, where no separate organization for data governance exists and responsibilities are not formalized. Decision making and definition of data-related terms and rules is assigned primarily to managers across the organization according to the already existing structure of the company and are performed in a more ad hoc manner.
- ***Governance via Stewardship***, a more formal approach with specific roles and responsibilities assigned to a new data governance organizational structure. This governance organization is constituted by a hierarchy of data owners and data stewards with the responsibility for the definition of data standards, taxonomy, rules for data usage and change of the data assigned to the stewards, whereas the data owners are accountable for the reliability and trustworthiness of the content of the data itself.
- ***Governance via Governance***, an approach with the clearest distinction between governors -that define data related rules and resolve the occurring issues- and data stewards, which are mainly responsible for working with the data, follow the defined rules for data usage specified by the governors and raise data-related issues to them for resolution. The created data governance organization is integrated and acts complementary with the existing organizational structure of the company and cooperates closely with IT and Business functions management in order to shape a holistic overview of data-related issues and challenges.

Scientific research suggests that it is vital to shape a “DG virtual organization” (Ladley, 2012) that is responsible for implementing the building blocks of data governance framework as described in the previous section and make essential data-related decisions in an unbiased manner that will help the whole enterprise maximize the value retrieved by their data assets (Weber, Cheong, Otto, & Chang, 2008).

3.6.1. Data Governance Virtual Organization

Several frameworks and models provide valuable insights on organizational aspects of DG. In an effort to clarify specifically the DG and DQ domains, (Weber, Otto, & Osterle, 2009)provides a detailed overview of all roles and responsibilities identified in relevant scientific literature. Table 6 demonstrates a set of four different roles and one committee, along with a description of the role, the organizational layer it can be assigned to and alternative names found in the literature for these roles.

Table 6 Data Governance Roles Descriptions (Wende, 2007)

Role	***Description***	***Organizational Assignment***	***Identification in Other Sources***
Executive Sponsor	Provides sponsorship, strategic direction, funding, advocacy and oversight for DQM	Executive or senior manager (e.g., CEO, CFO, CIO)	Strategic information steward, executive sponsor, executive council
Data Quality Board	Decides for corporate wide standards and controls its implementation	Committee, chaired by chief steward, members are business unit and IT leaders as well	Business information stewardship Team, Data governance Council, data governance committee, GRCS board
Chief Steward	Puts the board's decisions into practice, enforces the adoption of standards, helps establish DQ metrics and targets	Senior manager with data management background	Master data coordinator, director of data management, chief steward, corporate steward, lead stewards
Business Data Steward	Details the corporate wide DQ standards and policies for his or her area of responsibility from a business perspective	Professional from business unit or functional department	Information professionals, business information steward, business data steward, subject area steward, master data lead, domain steward, business Steward, subject matter expert
Technical Data Steward	Provides standardised data element definitions and formats, profiles source system details and data flows between systems	Professional from IT department	Database steward & information architecture steward, technical steward, source system data steward

(Cheong & Chang, 2007) focus on the organizational aspects of DG and recommend 3 different levels of categorization regarding levels of engagement between IT and Business functions and staff, *Strategic, Tactical and Operational,* as illustrated in Figure 14. The major differences with the previously analyzed propositions is that there is no distinction between *Business* and *Technical Data Stewards* and separate lines of communication between the Data Governance virtual organization and the IT functions of the company are demonstrated in the context of aligning business demands and objectives for higher flexibility and responsiveness with IT requirements for increased control and standardization.

Additionally, the new role of *Data Custodian* is introduced, sharing similarities with the *Chief Steward* role mentioned before but at a lower, tactical level and oriented around divisional business operation and data management. The data custodian is responsible and accountable for enterprise data assets and the resolution of issues that occur in group meetings, unless there are cases of conflict between different stakeholders from various business functions, where the issue resolution is transferred to the *Data Governance Council*. They are also responsible for endorsing data integration and data management plans, transforming strategic initiatives to tactical plans as well as stakeholder and change management.

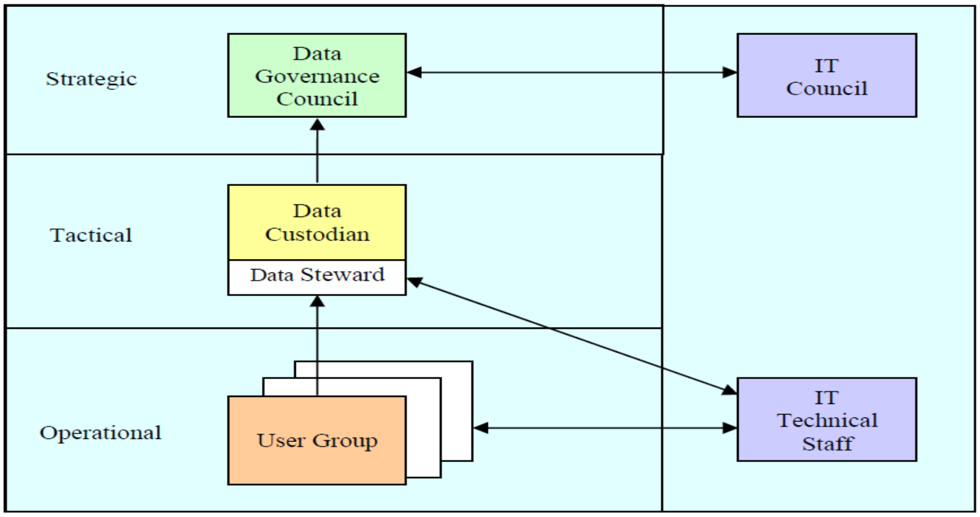


Figure 14 Data Governance Structure (Cheong & Chang, 2007)

A very interesting form of organizational data governance model occurs from this proposition, as the above data governance structure can be initially applied to specific business teams, projects or units and can be later scaled and adjusted to include other business divisions, introducing this way a *Federated Data Governance Model* as depicted in Figure 15, with the Data Governance Council acting as the "organizational glue" of the various local data governance structures.

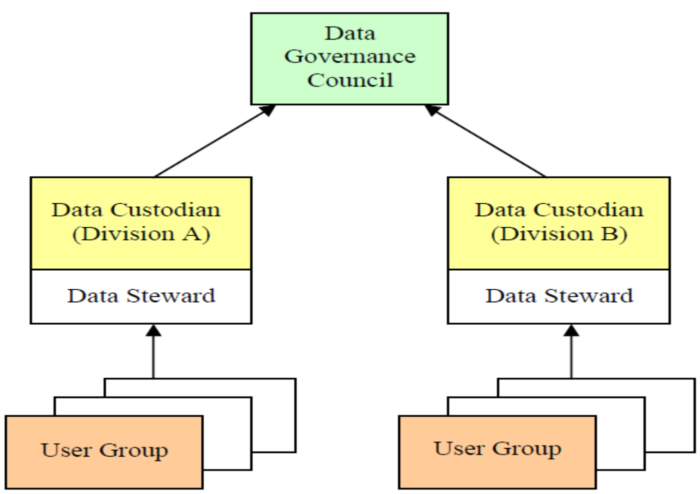


Figure 15 Federated Data Governance Model (Cheong & Chang, 2007)

A detailed DG organization blueprint is described by (Marco, 2014). As illustrated in Figure 16, the Data Governance Council has the general oversight of data governance activities by shaping policies, procedures and standards that safeguard data quality and usage in cooperation with the Data Stewardship Group and the Subject Area Groups. This council is led by the executive sponsor selected to promote data governance initiative within the organization and is constituted by the Subject Area Groups Chief Stewards and other essential business or IT staff, always in relation to organizational parameters such as culture, functional structure and business priorities.

Furthermore, the composition of the DGC can be rotational, covering this way extended areas of the corporation and assisting in developing a data governance culture to more people across the enterprise. Larger companies can also include an Enterprise Oversight component in their data governance organization that looks over the DGC, in the form of a strategic steering committee or a project management office, however in medium-sized or small organizations this layer can be replaced by a smaller subset of executive sponsors. The Data Stewardship Coordination Group connects Steward Teams with different focus areas, ensuring that the technical specification and business description of data coming from Subject Area Groups are properly distributed to the correct focus area and corresponding Steward Team.

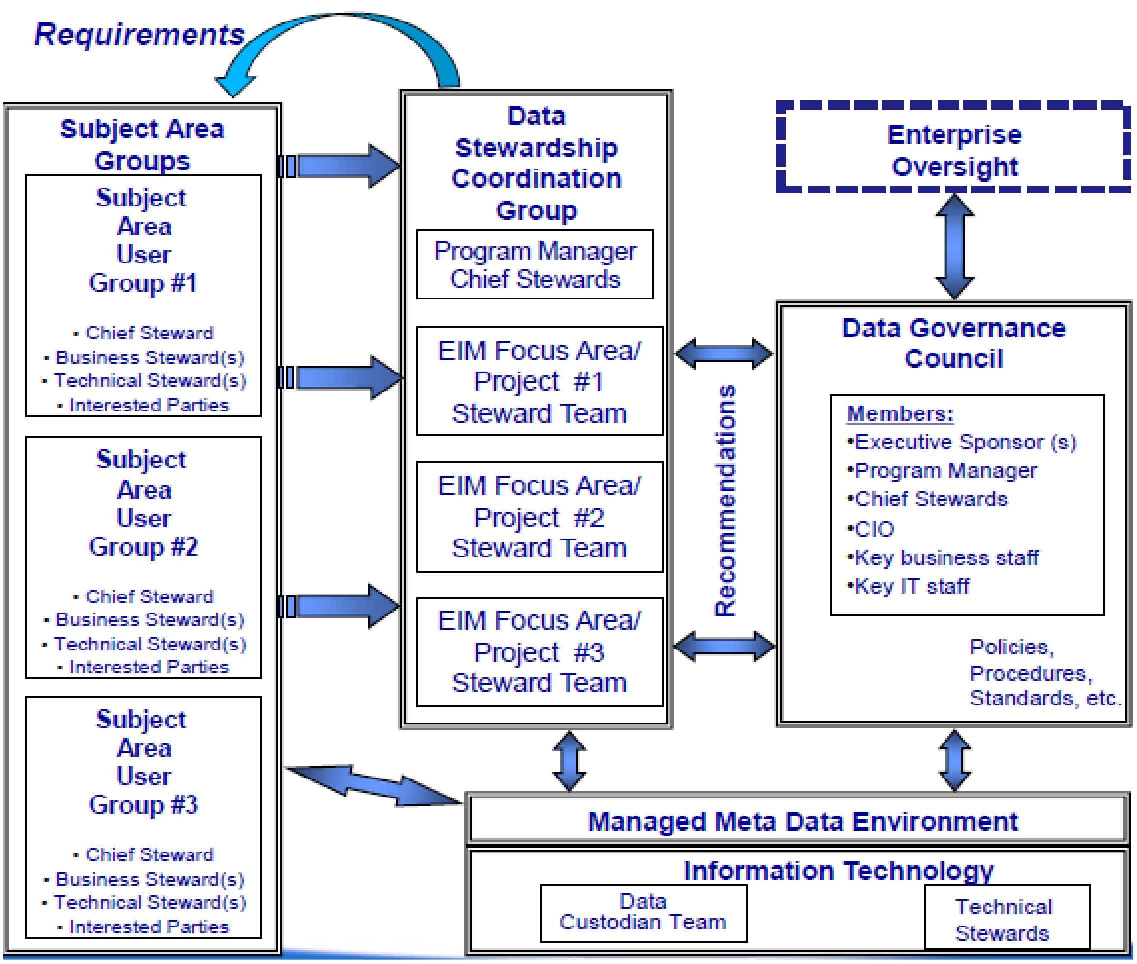


Figure 16 Data Governance Organization: Subject Area Design (Marco, 2014)

3.6.2. Roles & Responsibilities

(Wende, 2007) comes up with Data Governance model that follows the logic of a *Responsibility Assignment Matrix (RAM)*, such as the RACI chart that adopts the following notation:

- ***Responsible ("R"):*** Role responsible for executing a specific DG activity.
- ***Accountable ("A"):*** Role accountable for authorizing a decision focusing on a specific DG task.
- ***Consulted ("C"):*** Role that has to be advised and provide input and support regarding a particular DG activity.
- ***Informed ("I"):*** Role that has to be informed about a specific DG activity or decision.

The following RACI matrix in Table 7 is an indication of how a Data Governance Model would be formulated in an organization taking into account the different DG roles identified in literature and a set of decision areas and related activities in the data governance domain, with a focus on Data Quality Management aspects that lies in the core of successful DG.

Table 7 Data Governance Responsibilities (Wende, 2007)

	Roles					
Decision Areas	*Executive Sponsor*	*Data Governance Council*	*Chief Steward*	*Business Data Steward*	*Technical Data Steward*	...
Plan DQ initiatives	A	R	C	I	I	
Establish DQ review process	I	A	R	C	C	
Define data producing processes	-	A	R	C	C	
Define Roles & Responsibilities	A	R	C	I	I	
Establish DQ policies and standards	A	R	R	C	C	
Create business data dictionary	-	A	C	C	R	
Define IS systems support	-	I	A	C	R	
...	-					

Naturally, in the process of applying these data governance models and the accompanied responsibilities distribution in practice, a number of parameters have to be taken into account that refer to the organization's size, culture, business model and strategy, as well as the company's organizational structure and decision making style. This contingency approach for a company-specific configuration of a data governance framework is demonstrated in the following Figure 17, as presented by (Wende & Otto , 2007).

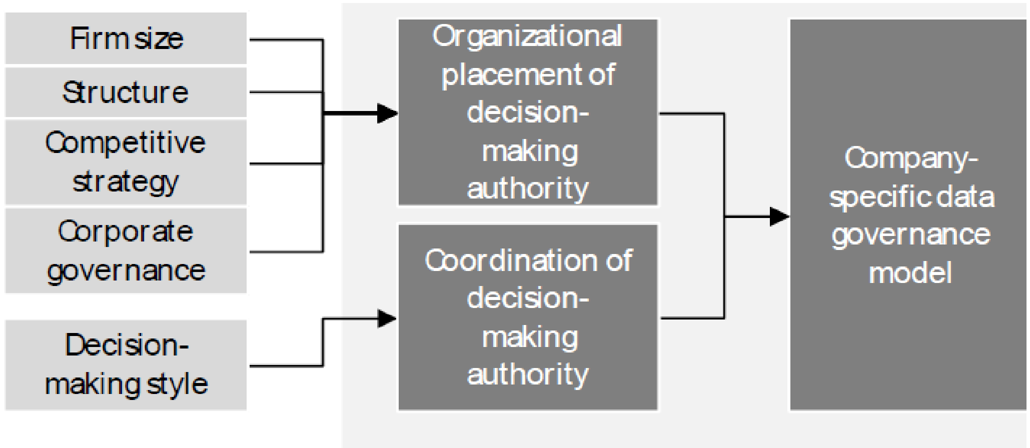


Figure 17 Contingency approach for data governance (Wende & Otto , 2007)

The various factors that influence the application of data governance frameworks in realistic business environments can be concentrated into 2 design parameters that have an impact on the form of the organization-specific data governance model under development: ***Organizational placement*** and ***Coordination*** of the ***Decision making authority.*** As described in Table 8, in terms of organizational placement we can make a distinction between *centralized* and *decentralized* design, referring to if the decision making authority for data governance issues will be positioned to the IT department or will be dispersed to the various business functions and teams.

We can see that in centralized schemas, responsibility and accountability is assigned also to the higher levels of hierarchy, such as the Data Governance Council or even the Executive Sponsor for some major decisions, whereas a decentralized approach would place the focus of decision making to the Business & Technical Stewards of IT and Business units, introducing a more consulting role for the Executive Sponsor, Data Governance Council and so on. The centralized form would fit smaller firms and a more conservative approach, with greater IT control and higher levels of standardization, while the decentralized form is associated with larger firms, aggressive strategies and increased levels of flexibility and responsiveness to the market with customized solutions for each business unit.

Table 8 Organizational structure for decision making determines level of (de)centralization for Data Governance (Wende & Otto, 2007)

Role	***Centralized Data Governance Design***	***Decentralized Data Governance Design***
Executive sponsor	"A" in some decisions of major relevance	"C" (recommending, not commanding)
Data Governance Council	Many "A"	Many "C", "I", no "A" alone
Chief steward	Many "A"	"C", (might even be obsolete)
Business & technical data steward	"R", few "A"	Many "A", "R"

Regarding coordination of decision making in modern organizations, we could distinguish between the *hierarchical* and *cooperative* data governance framework design as shown in Table 9, with the former enforcing a vertical, top down decision making approach with pyramid-like structure and coordination via communication of the superiors with their direct subordinates in the hierarchy. The cooperative approach establishes formal and informal coordination principles horizontally in the organization and across various business units, promoting collaboration and a consensus-oriented style in decision making processes. On the contrary with the hierarchical data governance design, the cooperative form never places the responsibility for decisions to a single role, introducing higher levels of transparency and a more democratic culture.

Table 9 Coordination of decision making determines the levels of cooperation for Data Governance (Wende & Otto, 2007)

Role	***Hierarchical Data Governance Design***	***Cooperative Data Governance Design***
Executive sponsor	"A" in some decisions of major relevance	"A" (conjointly)
Data Governance Council	"A" (separately)	Many "C" and "A" (conjointly)
Chief steward	"A" (separately)	"C", "I", few "A"
Business & technical data steward	"R", "I", few "C"	Many "A" (conjointly) and "C"

4. Data Governance Frameworks Evaluation

In this chapter the discussed Data Governance frameworks and models are evaluated regarding their suitability for Company's organizational environment and data landscape. Furthermore, an approach to design a company-specific DG framework is described.

4.1. Data Governance Organization & Contingency Theory

As described in section 3.6.2 of this thesis, in each organization a number of special characteristics and parameters can be identified that shape a unique environment with its own critical hotspots and pain points that impact the potential success of deploying Data Governance.

The contingency theory claims that the relationship between a characteristic and the organization's effectiveness is determined by contingencies. In other words, characteristics such as the organizational structure (e.g. centralized or decentralized) of a company can have a positive impact on this company's efficiency and productivity if the configuration of the DG design takes into account parameters that make this company unique regarding this characteristic . The contingency approach in our case refers to the fact that every organization demands a DG configuration that matches with a collection of context factors, a set of unique characteristics that describe this specific organizational environment and demonstrate the relationship between the design of the DG model and the potential success of this model in respect of the envisaged results (Wende & Otto , 2007).

An interesting research project on data governance with the participation of six international companies from various industries (Weber, Otto, & Osterle, 2009) applies this contigency approach to data governance and shows the relationship between a number of contigency factors and the company-specific configuration of a data governance model.

The authors propose seven contingency factors that influence the success of a DG design and perform an evaluation of an example *Company A* on the basis of these seven contingency factors. In respect of this thesis' context, this example can help the reader understand what could be the impact of the presented contingency parameters to the configuration of the DG organizational design. Table 10 illustrates the outcome of this assessment along with the expected consequence on the company-specific DG configuration.

Table 10 Assessment of example company A in respect of contingency factors (Weber, Otto, & Osterle, 2009)

Contigency	***Assessment Company A***	***Consequence for the DG design***
Performance Strategy	Profit	Trend towards centralization → concentration of competencies and knowledge in one Master Data Management (MDM) department in the corporate headquarters
Diversification Breadth	Related (+)	
Organizational Structure	Centralized (+)	
Competitive Strategy	Propsector (-)	Conflicting interests result in

		difficulties to enforce MDM mandates at regional and national levels → in the data quality board regional representatives discuss and solveconflicting issues
Degree of Process Harmonization	Globally harmonized (-)	Trend towards centralization and harmonization influence on the scope of MDM → harmonized product, customer and vendor MD are key to globally harmonized processes (matching of identification numbers is not sufficient)
Degree of Market Regulation	Highly regulated (+)	Expertise regarding country regulation regulation is provided by local data stewards
Decision Making Culture	Consensus- building	Many “C” in the data governance model point to a cooperative approach to data governance

In another scientific article that focuses on the organizational dimensions of Data Governance and the various design alternatives that companies can choose from when deploying their DG programs, two different case studies on DG initiatives of large telecommunication firms are investigated (Otto, 2011). The findings of this research effort suggest that there is no “off-the-shelf” approach for organizing DG, but the configuration depends on a number of external and internal organizational parameters

The contingency approach presented above from the corresponding research studies frame the process of adjusting the Data Governance design to fit a specific organizational setting in a strictly decision making perspective, for example if the decision making authority will be placed centrally in case of a centralized organizational structure and relatively narrow product and services diversity or it will support a decentralized style to satisfy unique functional requirements.

Another example would be a multinational corporation with globally harmonized business processes where the design and enforcement of Data Governance policies and processes could be done centrally and then diffused to the various local business units and departments. On the contrary, in case of heterogeneous business processes this would not be possible and the specification of Data Governance policies and processes would have to take into consideration special characteristics and demands of each local unit and corresponding processes and then possibly be approved from a more centralized decision making body. The same adjustments can be made for the structure and form of Data Governance committees and bodies in relation to an organization’s size, with the demand for additional Data Governance bodies in case of a multilayered vertical management structure that is often noticed in large organizations.

4.2. Data Governance Domains & Contingency Theory

However, besides organizational aspects and how the DG virtual organization will be configured into the targeted application environment for optimal fit, it should be also important to study the impact of organizational contingencies to elements, components and domains of DG that are not connected to the configuration of the DG virtual organization.

As research case studies and several organizational examples indicate (Aranow, 2014; Blosser & Haines, 2013), before a company deploys a Data Governance initiative and starts designing tailor-made Data Governance roles and committees suitable for its organizational particularities, it is very common that a Data Governance framework such as the ones reviewed in the previous chapter of this thesis is used as a foundation of the DG design and provides a baseline on what would be core domains and components of the DG program.

A framework is nothing else than the specification of Data Governance as a concept with a number of lower level components that compose a strategy for managing data as an organizational asset. In other words, a DG framework is a way to introduce a necessary abstraction layer that helps organizations and DG researchers and practitioners to organize, present and communicate DG concepts in an easier for non-expert audience to understand manner.

This is especially important if we consider the wide range of data-related subdomains that fall under DG "umbrella", such as data definitions, data models, metadata, master data, information security and lifecycle and so on as well as the strong interrelation of DG with other domains of modern business operation such as the overall business strategy and individual objectives, business intelligence projects, process management, compliance and audit mechanisms and the like.

For example, financial organizations might face increased pressure from industry regulations that need to be translated into requirements for their DG design and finally design and develop specific policies and processes that differ at a large extent with the ones established in a company of a different industry, for instance retail sales. In this case, data traceability for audit purposes from the source systems to the end-users applications is a higher prioritized objective for the financial organization, and this has to be depicted in the principles and elements of the DG design.

In this respect, another example could be identified in the case that an organization intents to implement DG in the context of a specific project or initiative, for instance Master Data Management or a Business Intelligence project, with the latter sharing similarities with this case study and the trigger for DG occurring from the intention to deploy Qlik Sense data visualization platform. Depending on the organization's special requirements regarding DG, the lower level elements of the implementation details that relate to specific DG domains such as Data Quality Management, Metadata & Master Data Management, Data Analytics and so on should take a different form and represent demands and priorities that stem out of the target organizational and data landscape.

Since it is generally accepted that DG is dependent on contingency factors and no "magic recipe" or "one size fits all" approach exists, then also the selected for implementation DG framework should also represent organizational particularities and requirements of the informational setting. This way, communication of DG concepts to the people of the organization can be more direct and effective as they can relate DG concepts and lower level components in the framework with the data landscape of their organization and the information working culture currently in place and possibly even instantly recognize the improvement potential and what it takes to get there.

Consequently, before any formal and official announcements for a DG program are made and designs of the so called DG "virtual organization" begin, it is essential to select a framework that demonstrates an optimal fit with the organization destined to be implemented.

4.3. Description of the evaluation process

Figure 18 illustrates a schematic representation of the evaluation process that is described in this section. As a logical sequence, the first step in the process of evaluating existing DG frameworks and selecting the most suitable one would be to **(1)** Create a set of criteria that can be derived from widely accepted DG-related requirements, but also from organization-specific demands that frame unique characteristics that have to be taken into account in the process of designing DG for a particular organizational and informational setting. After the criteria list is finalized, the next phase would be to **(2)** Evaluate the DG frameworks presented in the literature review against this criteria list. This way, we can arrive to a relatively safe conclusion about **(3)** If an already existing framework would adequately satisfy the defined requirements and priorities for successful deployment and implementation of DG in the selected organization.

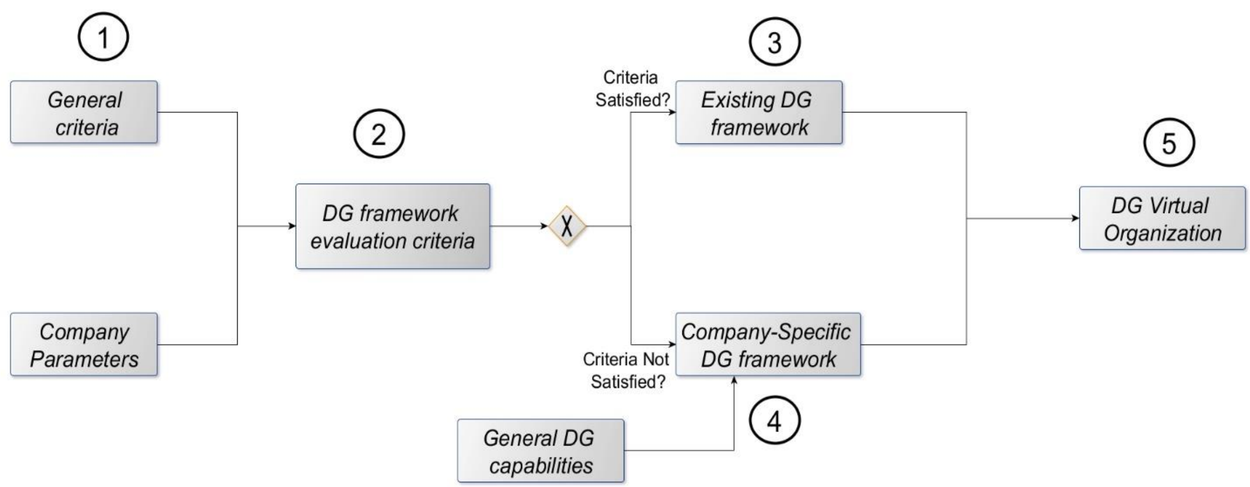


Figure 18 DG frameworks evaluation process

In case such an optimal DG framework does not exist in literature **(4)**, it has to be synthesized by putting together important for our purposes components of existing frameworks so that we have a composition of DG components that satisfy the defined requirements and priorities. In this effort it is important to include not only components that are connected to the criteria list, but also related to general DG capabilities that were not included in the evaluation process in the first place as they have to do with data-related standards, policies and processes that more or less are considered important in all DG frameworks, research studies and organizational best practices. Examples of such DG capabilities would be Metadata

& Master Data Management, Data Quality Management and Information Lifecycle Management that are considered essential pillars of DG for the majority of researchers and practitioners, therefore including them as criteria in the evaluation process would be redundant.

Finally **(5)**, the last step is the configuration of the DG organization design, the so called DG virtual organization. This step is disconnected from the evaluation of the frameworks itself and will be executed anyway, regardless if the selected DG framework lies in existing literature or occurs as a result of the evaluation process.

4.4. Criteria Development

All of the general DG criteria and Company parameters were gradually refined by several meeting iterations with major data stakeholders in the company until we arrive at a final criteria list. A subset of the stakeholders described in Section 2.2.1 were contacted for this purpose, that were identified as key business and IT human resources in the organization that participate in activities and processes that "touch" data, as a result their input and active involvement are crucial for the success of a Data Governance initiative.

Lower level criteria that were perceived as important by the stakeholders were aggregated to a more general level to keep the criteria list concrete and meaningful for the evaluation process. For instance the criterion Business Intelligence/Data analytics focus contains the lower level criteria data accessibility and availability, as these attributes are critical for the successful deployment of BI projects and tools.

The most left layer of the evaluation process flow diagram consists 2 different categories of criteria that will guide the evaluation process: ***General criteria*** are factors that are considered vital for any Data Governance program regardless of the industry and the organization:

- ***Alignment with Business Objectives:*** Specification of DG program's values, vision and mission. To what extent the frameworks under evaluation help in alignment of DG with the business strategy and objectives of an organization so that is easier to show the potential business value and ensure the necessary levels of active participation from business? Essential for justification of the DG program to ensure appropriate resources allocation (people, time, effort) and funding when necessary.

- ***Clear Roles & Responsibilities Formalization:*** Are the various roles & responsibilities for the DG virtual organization and data-related decisions formalized and described adequately in the frameworks? Do the frameworks contain suggestions on the structure of the DG organization, who would be more suitable to participate in each committee or role (e.g. DG Council, Data Steward etc.)?

- ***Organizational Structure & Decision-Making Culture:*** Do the DG frameworks take into account the organizational structure and decision-making style of a company in their suggestions for designing DG? If data-related decisions are made by top management or lower tier employees make a difference on the configuration of the DG organization and the way the DG roles and responsibilities will be distributed. Also decision-making culture of the organization (consensus-oriented or hierarchical) will impact the distribution of decision rights for data-related decision domains. Moreover, size of the company and alignment of the DG organization with current corporate structures have to be taken into account.

Company Priorities refer to a collection of attributes that establish a connection between the evaluation process of DG frameworks and the principles, data working culture and requirements for DG that can be met in the organization itself. As a result, *Simplicity* for example is a very important principle for selecting an optimal framework as DG concepts and guidelines have to be communicated across multiple teams, functions and hierarchies throughout the organization, including employees with limited technical knowledge and/or awareness regarding Data Governance domain.

- ***Simplicity:*** How simple is the framework in terms of number of components and interrelations between them, how easy is for non-expert audience from different backgrounds, functions and teams in Company to comprehend and communicate the framework's concepts. This is important as the DG framework will be used as an artifact of communicating what DG means for Company throughout the organization and as a reference point as evolution of DG progresses.

- ***Practicality & Implementability:*** Practical value of the framework in Company's organizational and informational landscape. Can data stakeholders and employees of various teams, business units and organizational hierarchies effectively relate concepts and components of the framework with the data environment and working culture of the organization? Also are implementation details covered adequately with relevant guidelines and recommendations?

- ***Flexibility:*** How extendable is the framework to include DG-related requirements and priorities that emerge in the future? How easy it is to add or remove a component from the framework and how much existing components and their interdependencies will be impacted? Also, does the framework take into account possible extension of the scope for DG? (e.g start with a focus on data sources & requirements of a limited scope, then expand to include additional user categories and data items).

- ***Focus on Business Intelligence/Data Analytics:*** How important are BI/Data Analytics aspects for Data Governance according to the frameworks under evaluation? To what extent the frameworks contain information about principles, guidelines and techniques that will foster information sharing and support accurate and fast reporting and decision-making operations that are based on effective organization and utilization of the available data sources? At a lower level, this criterion will be used to evaluate DG frameworks against data-related attributes such as data access & availability that are vital for effective data analysis and decision making, as well as data consistency & integrity (accuracy) that are critical for reliable reporting considering the various source systems involved.

- **Focus on Regulatory Compliance & Security:** How important are regulatory compliance & security aspects of data for the DG frameworks under evaluation? As the available DG frameworks have a generic nature, they don't always contain appropriate levels of information on compliance & security aspects that are crucial for financial organizations that desire to deploy DG programs. At a lower level, this criterion will be used to evaluate DG frameworks against data-related aspects such as data security as well as data auditability & traceability across the whole lifecycle from the creation to the consumption layer.

The table below demonstrates general criteria and company-specific parameters related with deploying a DG initiative in the organization under investigation for this thesis research. Each criterion has a detailed description that is provided in the text above and is matched with a specific *weight*, depending on how essential is this specific criterion for the company and the impact that this particular criterion will have on the evolution of the DG design within the organization.

The weights were collected from the stakeholders by distributing the questionnaire of Appendix A and asking them to fill in weights and submit any relevant remarks or new criteria. Finally, the average scores from all stakeholders were calculated for each criterion and the resulted values are presented in Table 11 below. Furthermore, the weights collection phase did not result to any new criteria as input from the stakeholders, providing validation for the soundness of the criteria set.

Table 11 Criteria for DG Framework Evaluation

	Criterion	***Weight***
General Criteria	*Alignment with Business Objectives*	11%
	Clear Roles & Responsibilities Formalization	12%
	Organizational Structure & Decision-Making Culture	9%
Company Parameters	*Simplicity*	14%
	Practicality & Implementability	18%
	Flexibility	10%
	Focus on Business Intelligence/ Data Analytics	9%

Focus on Regulatory Compliance & Security	18%
Total	100%

4.5. Evaluation of Data Governance frameworks

In this section, the criteria list shaped in the previous step will be used to evaluate selected DG frameworks of the literature review phase. Every framework is assessed in relation to the defined criteria giving us an indication of how effectively this framework addresses the requirements we have selected as most important. For the shake of readability, the following abbreviations are used for each DG framework:

- Panian : (Panian, 2010)
- DGI : The Data Governance Institute (Thomas G. , 2006)
- IBM : (IBM Software Group, 2007)
- DAMA-DMBOK2 : Data Management Association (Cupoli, Earley , & Henderson, 2014)
- Khatri & Brown : (Khatri & Brown, 2010)
- Wende : (Wende, 2007)
- Cheong & Chang : (Cheong & Chang, 2007)

Since we are interested in evaluating the available frameworks against a list of selected criteria, it makes more sense to perform the comparison of all frameworks in respect to each criterion, so that we can conclude about their potential suitability in a more direct manner that is connected to the extent they address the selected priorities. This also makes the comparison of the frameworks easier and more intuitive.

Below an initial qualitative comparison of the frameworks is described in text, followed by a quantitative assessment in () that is based on the qualitative characteristics of the frameworks. This way we have a more practical evaluation outcome by providing specific scores (grades) for the available DG frameworks, while at the same time the qualitative comparison provides detailed justification of how each score came up. The grading scale is ranged from 0 (worst score) to 5 (best score), however all frameworks were found to satisfy the criteria at a basic level and none of them was assigned a zero (0) score.

Alignment with Business Objectives:

DGI (5), IBM (5), Panian (4) and DAMA-DMBOK2 (4) describe data as an organization's most valuable data asset and greatest source of risk at the same time. The frameworks identify business value and envisaged benefits of DG in alignment with our case's objectives in areas of business growth, operational efficiency and risk management through the discussion of specific examples and use cases. However, Panian and DAMA-DMBOK2 demonstrate a more technical orientation and related with business value components are not adequately represented in the framework design.

Khatri & Brown (3) also provide a decent coverage regarding business value of DG with the corresponding use cases for data quality in support of improved compliance and security, business intelligence systems and other areas of data management, however in a somewhat limited range compared to other frameworks.

Finally, Wende (2) and Cheong & Chang (2) emphasize the twofold nature of DG in both business and IT aspects and consider DQ as the core discipline of the "data as an enterprise asset" concept. However, their analysis lacks depth and there is no specific and elaborate connection of DG with the business value an organization can realize.

Clear Roles & Responsibilities Formalization:

Wende (5) and Cheong & Chang (4) provide specific and detailed descriptions and proposals on roles, responsibilities and organizational bodies of the DG virtual organization in alignment with the existing IT management structures of an organization. However, Wende's proposal is preferred as it contains a wider range of propositions for DG-related role distribution within an organization.

Khatri & Brown (3) identify several examples in existing structures and positions of an organization that roles & responsibilities for DG could be assigned, however data stewardship is not described adequately and their suggestions in some areas lack structure and content compared to higher evaluated frameworks.

DGI (2) analyzes data stewardship and ownership and some example structures for the DG organization are discussed, but roles & responsibilities are not as elaborately covered as in other frameworks, and no specific propositions are made at a lower level. IBM (2) and DAMA-DMBOK2 (2) describe the concept of data stewardship and the need for clear roles and responsibilities at all organizational layers, however their propositions lack the depth and detail of other frameworks.

Finally, even though Panian (1) recognizes clear roles and responsibilities as an important success factor for DG, no specific roles and responsibilities are proposed in his framework.

Organizational Structure & Decision-Making Culture:

Wende (5) develops an elaborated contingency theory for optimal configuration of the DG organization and decision making responsibilities in respect of 2 contingency parameters: Organizational placement and coordination of decision-making authority. Khatri & Brown (4) and Cheong & Chang (4) provide some recommendations on the configuration of the DG organization and relevant decision-making between centralization and decentralization, but in a not so elaborated manner.

DAMA-DMBOK2 (3) and DGI (3) consider organizational aspects of DG as very important in their propositions, emphasizing the need for alignment with the environment's culture, size and decision making style. However, their analysis stays at a general level. IBM (2) and Panian (2) focus on aspects of business and IT collaboration and the need for cross-functional support for effective DG whereas culture and other organizational factors are also considered, however no specific recommendations are described.

Simplicity:

Panian (4) and Cheong & Chang (4) keep their framework simple and comprehensive, following a categorization of DG components and domains in similar axis, however there is an overlap in the component distribution that compromises clarity. Furthermore, the orientation of their suggestions emphasizes technical data-related aspects and it is possible business people will find it hard to identify relevancies.

DGI's (3) suggestion is understandable and organized around axis consistent with relevant DG research, but multiple interconnections between the components hinder simplicity of the design, whereas lower level components are not adequately specified. IBM (3) organize DG components in: *Supporting disciplines, core disciplines, enablers* and *outcomes,* a unique way compared to other frameworks that promotes understandability, but is inconsistent with DG research and scientific definitions.

Khatri & Brown (3) provide a simple yet one dimensional view on DG, as their framework's design focuses on data decision domains and lacks depth in the interconnection with organization and process related aspects. Wende (3) focuses on organizational aspects of DG and DQM and this enhances simplicity of the discussed proposals, but this comes at expense of content as little insight is provided in respect of other related domains.

DAMA-DMBOK2 (2) provides a detailed coverage of DG domains, but simplicity is compromised due to the technical orientation in combination with the specified in the framework but complicated interconnection with organizational and other aspects.

Practicality & Implementability:

Cheong & Chang (4) develop their framework with a more practical view on DG and how a company can successfully initiate a formal DG program with the corresponding organizational bodies. However, their view on a specific organization might make their approach a bit "distant" from another company's environment and its particularities and characteristics.

IBM (4), DGI (3) and DAMA-DMBOK2 (3) frameworks demonstrate a high level nature with abstract connections between the DG components. Also an issue is the exact content of each component as possibly makes it harder for data stakeholders in an organization to identify their objectives due to different interpretations. However, the above are compensated by the description of implementation plans and methodologies that address the majority of low level aspects of DG, increasing the practical value of their suggestions. IBM's framework is accompanied by the most complete and structured implementation proposal.

Panian (3) and Khatri & Brown (3) present DG components that address data-related requirements in a relatively practical manner due to the lower level nature of their design, however little insight is provided on how the DG framework should be deployed within an organization.

Wende's (3) proposal displays high practical value on the organizational aspects of DG and relevant decision making on DQ, but the same does not apply for other core domains of DG whereas deployment proposals lack the analysis that can be found in other suggestions.

Flexibility:

IBM (4), DGI (3) and DAMA-DMBOK2 (3) demonstrate a high level nature that promotes flexibility and contain wide coverage on the vast majority of requirements that stem out of DG-related fields in the present but also in the future. Moreover, scalable implementation of DG is emphasized. However, IBM's suggestion provides a more clear view of the potential impact of changes in the framework structure and component composition.

Panian (3) provides a multidimensional perspective on DG that allows integration of emerging demands, while a pilot DG initiative with extendable scope is identified as a good practice to show the value of DG quickly and gain support at multiple levels. However the lower level focus on technical aspects limits flexibility of the design. Cheong & Chang (3) proposal can be enhanced with additional components in 3 axis and is generally flexible, but their focus on a specific organization and relevant requirements can be a strong limitation.

Wende (2) focuses mainly on organizational aspects and emphasizes DQM as a core component of DG, thus flexibility of the proposal in terms of emerging requirements and future innovative initiatives is limited. Similarly, Khatri & Brown (2) provide a framework organized in data decision domains and this one dimensional perspective constraints flexibility of their suggestions.

Focus on Business Intelligence/ Data Analytics:

Khatri & Brown (5) demonstrate a strong interconnection between DG and BI/data analytics and this is also reflected in the strong focus of their framework on data related attributes and relevant use cases (e.g. data quality, data access in respect of the business use of data). DAMA-DMBOK2 (5) and DGI (4) also show strong interconnection of their suggestions on DG with BI/Data analytics projects, however DGI stays at a more general level of detail.

Even though connection of DG with BI is not too detailed in IBM's (3) framework as in other proposals, components that directly relate to successful Business Intelligence and Data analytics projects are included (e.g DQM, Data architecture, business semantics and metadata).

Panian's (2) framework does not elaborate much on the connection of DG with BI and data analytics, but at a lower level many data attributes (e.g. data accessibility, availability, consistency and quality) included as components are key aspects of any BI project.

Wende (1) and Cheong & Chang (1) refer to business intelligence in respect of one the major business operation domains that DG can have a great impact by providing quality data to the users, but apart from that no further information or specific guidelines are provided on how to better support business intelligence or data analytics projects and systems.

Focus on Regulatory Compliance & Security:

IBM (5) considers Compliance and Security aspects as core components and focus areas of effective DG for modern organizations. DGI (4), Panian (4) and Khatri & Brown (4) also demonstrate strong interrelation of DG with compliance and security domains, however these areas are not so emphasized.

DAMA-DMBOK2 (3) and Cheong & Chang (3) provide decent coverage of the topic, but the representation in the design and contained analysis is not so strong as in other frameworks, whereas Wende (2) mentions support to regulatory compliance and security as an area that DG can contribute via increased quality of the provided data, but the framework suggested lacks a closer and more specific view on these aspects.

Table 12 shows a quantified evaluation of the frameworks that is based upon the qualitative comparison performed in text. The grading scale used is from 0 to 5, however there was not a case in which a framework scored 0 in a specific criterion as there was at least a basic coverage of all criteria in all the studied material. In order to calculate the total score of each framework, the weight of each criterion was taken into account besides the individual scores of each framework to depict the difference in the criteria prioritization that was taken as an input from the stakeholders.

Table 12 Quantified Evaluation of Data Governance Frameworks

Criterion	*Weight*	*Panian*	*DGI*	*IBM*	*DAMA*	*Khatri & Brown*	*Wende*	*Cheong & Chang*
Alignment with Business Objectives	11 %	4	5	5	4	3	2	2
Clear Roles & Responsibilities formalization	12 %	1	2	2	2	3	5	4
Organizational Structure & Decision-Making Culture	9%	2	3	3	3	4	5	4
Simplicity	14 %	4	3	3	2	3	3	4
Practicality & Implementability	18 %	3	3	4	3	3	3	4
Flexibility	10 %	3	3	4	3	2	2	3
Focus on Business Intelligence/Data Analytics	9 %	2	4	3	5	5	1	1

Focus on Regulatory Compliance & Security	18 %	4	4	5	3	4	1	3
Score	100 %	3.00	3.36	3.72	3.03	3.35	2.70	3.24

4.6. Discussion on the evaluation outcome

The outcome of the evaluation process confirmed our approach and initial assumptions that there is no single Data Governance framework from scientific literature and organizational best practices that can satisfy the full set of criteria that occur from a specific organizational environment and data landscape.

Indeed, none of the frameworks scores close to maximum (grades 4, 5) for all the selected criteria. Furthermore, even though some frameworks score well in the majority of the criteria, they demonstrate below average score (≤2) for at least one of them. IBM (3,72), DGI (3,36) and Khatri & Brown (3,35) appear to be the stronger performers for the specific evaluation process. Nevertheless, despite their strengths in specific aspects, all 3 of them demonstrate vulnerabilities in others.

For example, IBM's framework that collected the greatest score amongst the evaluated frameworks has a strong focus on regulatory compliance and security aspects (weight 18%) and is valuable regarding its practical usefulness mainly due to the detailed description of implementation options (weight 18%). Also, it provides an optimal fit with potential business value of DG (weight 11%) for the areas that can be identified in the target organization as compliance & risk management are considered top priorities regarding the desired outcomes. On the downside, the framework scores average regarding its simplicity (weight 14%). Furthermore, clear roles & responsibilities are not defined and described in detail (weight 12%) and organizational aspects are not covered in depth (weight 9%).

Furthermore, a general remark has to be made regarding the relevancy of the framework with the company's setting, as some stakeholders could not identify their objectives easily and intuitively in the components. For example, *Data Architecture* component in IBM's framework refers to the architecture of data resources, systems and applications that enable data availability and distribution to the business users, however a more meaningful and representative naming convention could be *Data Architecture & Analytics,* to emphasize the importance of available, accessible and properly organized data for analytical projects and Business Intelligence projects such as the deployment of a SS-BI platform within *Company*. Another example is the *Audit Information, Logging & Reporting* component that refers to measuring value and risks of data assets, as well as the progress and success of DG. In this case, a clearer name that provides a better match with the described content could be *Monitoring & Measurement.*

4.7. Synthesis of suitable Data Governance framework

The evaluation of the available DG frameworks concluded that there is no single framework that can satisfy our full list of criteria that stem out of Company's organizational environment and data landscape. As already mentioned in section 4.3, creating a new framework that is not reinventing the wheel but rather is a composition of important components, best practices and DG capabilities that can be found in the DG frameworks of the literature review phase can provide us with a solution that addresses the majority of our requirements.

In order to reflect criteria and priorities that come out of Company's environment into the newly created framework, a bottom-up approach should be taken into consideration, organizing a series of workshops with stakeholders from multiple functions and hierarchies that will leverage different opinions and data-related requirements and will result in agreed-upon definitions for Data Governance, identify the most important relevant capabilities, as well as the structure that the synthesized framework should have to increase its relevancy with the targeted organizational and informational environment and as a result, its practical value for the company.

The participants of this workshop could form a Data Governance working group that bears the responsibility to introduce DG throughout the organization, design the most suitable framework for implementation as well as build a solid business case that will assist in providing justification for the program and its demands in resources, time and effort from employees, managers and executives. More information about the configuration and goals of the DG working group can be found in the next chapter 5, as it is has an important role in the initial phases of designing and deploying an effective DG program within an organization.

An approach that can be used by the Data Governance working group to facilitate discussion towards creating a Company-specific DG framework is presented in this section, describing a process of matching each criterion from the list we used to evaluate the DG frameworks to a specific DG capability, which in turn is included as a component to the framework under development.

Table 13 demonstrates the matching process, where criteria are mapped to DG capabilities based on how each criterion would be satisfied within an organization by the corresponding capability. For example, *Simplicity* of the DG framework would be achieved if the framework enables and promotes *Awareness* on how DG is conceptualized in a particular organization, what it really means and how is planned to be implemented.

Similarly, *Practicality & Implementability* would be satisfied by 2 DG capabilities with the most practical value: *Monitoring & Measurement* in terms of metrics development to measure efficiency of information workflows, data-related operational performance and progress of DG in general and *Information Workflows* referring to the aspect of managing organizational data assets effectively by having a clear view of how data flow through various systems and interact with core business processes, in other words how data that drive the business model can be organized in the best possible way.

However, as it is analyzed in the description of the evaluation process, a number of important capabilities that would be vital for success of any DG program are not included in the criteria list to make the comparison of the frameworks more meaningful, as these aspects are adequately covered in all of them. Therefore, these important DG capabilities are added complementary in the capabilities that come up from the matching process, creating a complete set of components that can be included in the design under development.

The occurring capabilities were taken from several DG frameworks that were reviewed in Sections 3.5 and 3.6, with the necessary modifications to increase relevancy with the targeted environment as much as possible. IBM's framework (IBM Software Group, 2007) that was evaluated higher in Section 4.5 was found to contain several capabilities that with the necessary adjustments could increase the value of the synthesized framework for *Company's* organizational environment and data landscape.

Table 13 Matching criteria to DG capabilities, Capabilities adapted from DG frameworks reviewed in Sections 3.5 & 3.6

Criterion	***Capability***	***Description***
Alignment with Business Objectives	*Value Creation*	Connection of DG with business strategy, methodologies to retrieve and quantify value from organizational data assets in respect of risk management, business agility and operational efficiency.
Clear Roles & Responsibilities formalization	*Roles & Responsibilities*	Clear, specific & formalized Data Governance roles & responsibilities that consist the DG virtual organization and are responsible for creating, approving and enforcing standards, policies and processes.
Organizational Structure & Decision Making Culture	*Decision, Access & Authorization Rights*	Appropriate distribution of formal DG roles & responsibilities throughout the organization so that decision, access & authorization rights regarding enterprise data assets reflect the organizational structure and decision making culture.
Simplicity	*Awareness*	Organizational awareness for Data Governance concepts , principles and the new culture of collaboration and mutual responsibility between business & IT in the direction of governing data to mitigate risks and maximize their business value.
Flexibility	*Readiness*	Proper foundation in organizational mentality around data practices that will make it easier to integrate emerging requirements and organizational changes, and also absorb future innovative initiatives more effectively.
Practicality & Implementability	*Monitoring & Measurement*	Oversight and control performed by the DG organization using process & operational efficiency success metrics for data-related performance monitoring & incident handling. Measurement of progress in data-related capabilities.
	Information Workflows	Levels of formalization, transparency & standardization of data-related workflows and their interaction with core business processes of the organization.
Focus on Regulatory Compliance & Security	*Compliance & Security Management*	Policies, practices & control mechanisms in the process of mitigating risks from wrong data usage or unauthorized access, protect organizational data assets and ensure compliance with regulations.

Focus on BI/Data analytics	*Data Architecture & Analytics*	Architectural design of enterprise data assets and corresponding repositories, systems and applications. Methods and tools to establish common data definitions and enterprise wide data models. Formal, written specification (standards & policies) of desired organizational behavior around data usage.
General DG capabilities	*Master Data & Metadata Management*	Master data & metadata repository to provide "single version of truth" for critical data items as well as context & activity records of the data from business and technical perspective.
	Data Quality Management	Techniques and policies to measure & improve quality of data in production and non-production environments.
	Information Lifecycle Management	Methods and policies for collection, archiving and deletion of data and information.

To enhance simplicity of the framework and provide a more concrete and intuitive design for DG, the decision to consolidate capabilities and reduce the number of components was made. Indeed, a number of overlaps or conceptual connections were easily identified, as *Awareness* and *Readiness* as concepts and DG capabilities that both refer to organizational aspects are strongly interconnected, in other words the one leads to another. As a result, only the former was included in the final design. Furthermore, *Decision, Access & Authorization Rights* refer to how *Roles & Responsibilities* for DG are distributed optimally within the organizational structure and decision making culture, as formalization alone is not enough. Therefore it is possible to merge the capabilities.

Also, we notice that some general DG capabilities have strong correlation with *Data Architecture & Analytics* component. Business and technical *Metadata* that enhance consistency between different systems and repositories as well as *Master* data items that provide a single version of the truth are crucial aspects of an organization's data architecture and vital for success of any data analytics project or platform. *Data quality* is very important and the choice in the design is to emphasize this aspect by including DQM as a separate component, as benefits from increased DQ have a positive impact in several business areas, as for example more effective marketing & sales initiatives, accurate investment decisions or fast and reliable reporting. Finally, *Information lifecycle management* policies for data retention and deletion are also included as a stand-alone component, as it impacts and interrelates with many aspects of DG such as DQM, Compliance & Security and others.

The synthesized framework contains 9 components in total, that as already mentioned were taken and adapted from several DG frameworks of the literature review (Sections 3.5 and 3.6). Based on Data Governance definitions that were discussed in Section 3.1 and in order to increase consistency of the design with scientific theory, the decision to organize the DG framework around 3 axis was made: *1) People & Organization, 2) Standards & Policies* and 3) *Processes.* Figure 19 demonstrates an indication for the outcome of the synthesis process described above:

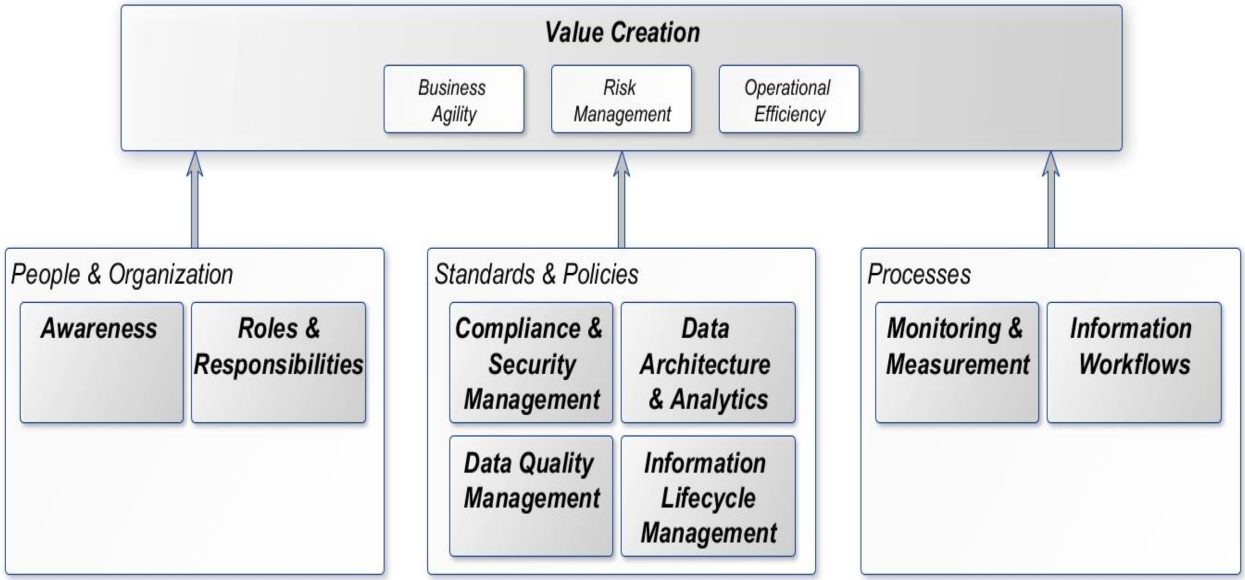


Figure 19 Synthesized DG framework, Capabilities adapted from DG frameworks reviewed in Sections 3.5 & 3.6

The prioritization in the criteria weights provided as input from the stakeholders can be depicted in the framework by focusing initial efforts and iterations of the DG program on developing the corresponding capabilities. For example, the maturity assessment as well as the action roadmap described in sections 5.3 and 5.4 can have a narrowed scope for the most important capabilities, in our case: *Compliance & Security Mechanisms* (Regulatory Compliance & Security Focus), *Information Workflows* and *Monitoring & Measurement (*Practicality & Implementability) and *Awareness* (Simplicity).

Nevertheless and despite this is not confirmed by the stakeholders' input, scientific research and best practices suggest that an official DG program must be accompanied by formalized roles and responsibilities, and it is presented as a critical success factor (Panian, 2010; Weber, Otto, & Osterle, 2009; van der Meer, 2015; Cheong & Chang, 2007). People are a central aspect of DG, without clear accountabilities and central coordination of efforts and resources the motivation and dedication in changing established data practices might not be sufficient. Therefore, it is recommended that the initial planned actions will target on creating and gradually fine-tuning the DG organizational design.

5. Design of a Data Governance Implementation Plan

This chapter describes the actions and activities that should be performed in order to effectively deploy a program that addresses core DG competencies. Several implementation suggestions exist in DG literature (Soares, 2010; Thomas G. , 2006; Ladley, 2012), with the first 2 of them correlated with the DG frameworks that were evaluated in the previous chapter of the thesis.

For the purposes of this thesis IBM's proposition was selected (Soares, 2010), as it provides clear, concrete and detailed description of all the intermediate stages required to deploy a DG initiative. This was depicted in the evaluation outcome as well, as IBM's framework was amongst the top performers for criterion *Practicality & Implementability,* partly because of the detailed implementation plan. The complete plan adjusted to the needs of this thesis is illustrated in Figure 20 below:

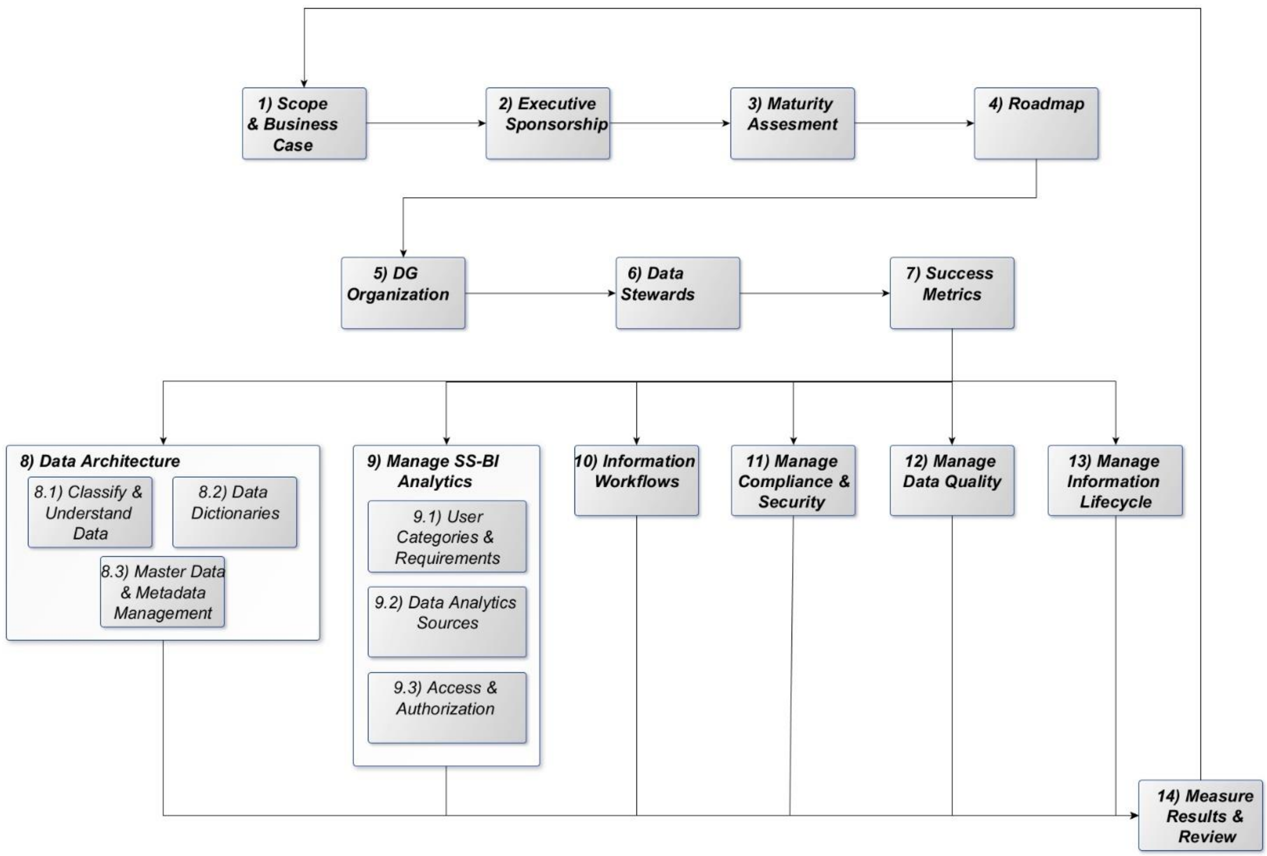


Figure 20 Overview of DG Implementation, adapted from (Soares, 2010)

Nevertheless, valuable information, suggestions and examples were used from all 3 implementation plans that are referenced at the start of the chapter, so that a complete and consistent with scientific research and organizational best practices solution can be presented. In order to provide a better match of the plan with the targeted environment taking into account company-specific demands and priorities, a number of components were rearranged or renamed, while others were created additionally.

The first 7 stages of the deployment plan are considered sequential regarding their execution as they help to explicitly define business value and objectives of DG, ensure support from the executive board based on the envisaged benefits, assess the current maturity of the organization in respect of important DG capabilities and build a roadmap of actions and initiatives that will assist in reaching a higher maturity level. All the above actions and relevant discussions can contribute to increase organizational awareness on the new culture of governing data assets and the underlying business value (***Awareness)***. As already discussed (Section 1.5), lack of awareness regarding value of DG within organizations is reported as the number one barrier for progress of Data Governance. Furthermore, the design of the DG organization and establishment of Data Stewards are vital to guarantee proper distribution of responsibilities and accountabilities regarding data-related decision making (***Roles & Responsibilities***). Finally, metrics for data-related business agility, risk management and operational efficiency need to be established that are based on the envisaged value and provide a baseline to measure progress and success of the DG initiative (***Monitoring & Measurement)***.

After these steps are designed and implemented, an organization can start governing its data assets in a structured and formalized manner. In order to rationalize the information landscape, data should be classified depending on their criticality and impact for the business, and data terminology and underlying relationships need to be mapped accordingly. Moreover, metadata and master data management is important to increase consistency of data across sources and systems, improving effectiveness of data analytics as well ***(Data Architecture & Analytics).*** Metadata specification is also crucial for visualizing data flows through systems and business processes in order to identify delays and inefficiencies ***(Information Workflows).***

Policies for data manipulation and reporting need to be extended and enhanced to ensure compliance with regulations, while data assets should be also protected from unauthorized access and usage ***(Compliance & Security Management).*** Metrics for Data Quality are another important area for effective DG incorporated in the corresponding policy ***(Data Quality Management)***, while data and information need to be archived or deleted based on specific standards and rules ***(Information Lifecycle Management)***. Finally, the results regarding progress of the DG program have to be measured against the initially defined objectives of the business case as well as the specified success metrics (***Value Creation)***, and the necessary adjustments that will help DG to evolve and mature in the organization in subsequent iterations of the program have to be made.

Practitioners of DG have to keep in mind that we have to do with an ongoing program of continuous improvement on managing, maintaining and exploiting business value of organizational data assets and not with a project that usually have a finite duration with a certain end point. Therefore, iterative implementation cycles have to be established that will help the DG design to evolve as the organization gradually learns and improves its capabilities to govern enterprise data assets (Ladley, 2012). Each of the intermediate stages is described in the corresponding section of this chapter.

5.1. Scope & Business Case

In many organizational examples, the root cause of failure or delay for DG programs is recognized as a weak or non-existent linkage to business value (Soares, 2010). Data Governance is after all a business program, with IT serving a supporting role. Therefore, it is crucial to identify the areas that effective DG will bring real value to the organization and build a solid business case that will persuade executives, managers and employees from several business units and organizational hierarchies about the usefulness of the program and will engage them to devote time and effort as well as provide continuous support.

The business case should support a clear vision and mission to support the drive for cultural change needed, identifying the potential benefits in relation to specific business projects and initiatives as well as costs and possible risks that come with the DG program (Ladley, 2012; Thomas G. , 2006).

Example Vision statement:

Company desires to increase its agility in incorporating innovative technologies and advancements in order to reap the benefits of digital transformation and be in sync with emerging trends.

Example Mission statement:

Company desires to develop a new culture of increased awareness, improved collaboration and conjoint responsibility between business and IT teams in the effort of organizing, safeguarding and enhancing organizational data assets to increase their business value and reduce enterprise risk.

After a long term vision and a relatively short term mission for governing organizational data assets are formulated, it is crucial to identify areas of value than a properly designed and deployed DG program can bring to the whole enterprise. In general, value for information projects can be found in all areas of business operation that are based on data and information (Ladley, 2012), while several categories of value contribution can be shaped: tangible and intangible direct benefits regarding areas that data support business directly, indirect benefits that refer to other initiatives and projects that have an increased possibility of success with effective DG in place.

A very useful approach for initiating a DG program in an organization is to follow an iterative, project-based method that will show the value of governing enterprise data assets in a narrowed scope that is easier to manage and control (Soares, 2010). This way, quick wins of the program can be demonstrated that will also help in the direction of gaining support for the program justification and the corresponding resources allocation and funding, as the organization realizes the potential value of the tasks at hand. After that, scope can be extended and include other projects, business units and data domains. In this case study, DG was initially scoped to include the *Investments Front Office* domain, as the core function of the Company's business model, the *Sales & Marketing* department to improve established customer channels and business opportunities and the *Investments Back Office* function due to the cross-functional and data-intensive nature of their operations.

As it is already concluded from the literature review (Panian, 2010; WBR & Informatica, 2015; Ladley, 2012; Soares, 2010; Thomas G. , 2006) and with a more specific focus on financial services organizations, governing enterprise data assets in a more effective manner demonstrates business value in 3 areas: *1) Risk Management, 2) Business agility* and *3) Operational efficiency*. Regarding Company's business model and informational landscape, a large variety of envisaged benefits from the above literature sources that can be tied to a DG program can be identified, presented in the following list:

1. Risk Management:

- Increased *trust & confidence* in data & data-driven decision making by improving data quality and enriching metrics used.
- *Accurate & timely* regulatory reporting.
- *Increased transparency & visibility* in data operations, ability to demonstrate audit trail.
- Reduced risk from *data misuse* and *incorrect decisions* based on wrong data.
- Reduced risk from *regulatory compliance issues*, *decreased fines* from possible compliance audits that went wrong.
- Gartner predicts that, by 2017, one third of Fortune 100 organizations will experience an information crisis, as a result of ineffective governance and decreased trust in their enterprise information (Information Builders, 2014). As volume of structured and unstructured data grows year by year, reducing risk for *data loss & leakage* becomes a critical aspect of modern organizations as research indicates that 4 out of 5 companies have experienced data loss incidents in the past.
- Reduced risk from a *possible security breach*: With average costs for every exposed record with sensitive information increased by 23% since 2013 (SecureWorks, 2016), and average total cost estimated around $7 million, it is obvious that organizations should protect their valuable data assets in the best possible way. Furthermore, it is reported that the financial consequences of a data security breach today amount to an average 29% damage of brand reputation, 19% loss of revenue and 21% lost productivity (IBM Global Technology Services, 2014). Especially for the financial services industry where trust plays an extremely important role for customers and partners, it is not hard to imagine the impact of such an event in an organization's customer base, revenues and even long term presence in the market.

2. Business Agility:

- Improved *business agility & investment decision making* by increasing timeliness & reliability of data.
- Increased *effectiveness of Business Intelligence* projects via the development of common data models, data definitions, business semantics and metadata specification across systems and applications. According to relevant research studies (The Economist Intelligence Unit, 2015), actual data analysis now corresponds to around 30% of time in data analytics and business intelligence efforts, with the majority of time devoted to preparation, modelling and integration tasks.
- Accurate, efficient and faster internal and external *reporting capabilities*.
- Improved *information sharing* between different teams and functions of the organization.
- *Enhanced collaboration* between IT and business functions.

- Positive impact on organization's *revenues* due to improvements in data accuracy and data quality that helps in reducing cost of missed business opportunities, retaining old customers due to increased customer satisfaction and pursuing new ones through novel business insights e.g. integrated customer management (360 degree view of the customer).
- Support future *innovative initiatives* regarding new business ideas, systems and technologies with increased success possibility than before.
- Easier *deployment of new systems and applications* by standardizing data models, definitions & business terminology
- Effective *configuration of application integration points*.
- Fight against perception that data & information initiatives always "fail" and that allocating resources in such programs is wasteful.

3. <u>Operational Efficiency:</u>

- *Improved productivity & efficiency* regarding IT & business processes that "touch data", e.g. less manual tasks & workarounds, standardized processes.
- *Lower costs* by gradually eliminating ROT (Redundant, Obsolete, Trivial) data e.g. duplicate data, documents, backups, SharePoint documents, and emails. This can also provide a small boost for DW & BI projects and their ROI. Studies indicate that at least 40% of corporate data can be defined as ROT (Recall, 2015). Considering that the amount of organizational data increases at a higher rate than storage costs decreases, the impact in cost savings can be great.
- Potential *cost savings* estimated around 10-20% for securities databases depending on the initial situation. An example can be found in the Luxembourg market (Deloitte Advisory & Consulting, 2014): With an average cost per security entered estimated at €80 to €120, the total cost for a DB with around 80,000 security entries would be estimated between €6.5 million and €8.8 million per year, including market data purchase as well as infrastructure and human resources costs required to manage the acquired data. This means potential cost savings between €650,000 to €1,76 million per year for each available database.
- *Cost savings* from decreasing departmental DBs and spread marts.
- *Cost savings* from better cost allocation of data deliveries currently assigned to IT even though requested from business teams.
- *Consolidation of business systems and applications*: Simplification of the organization's IT architecture by reducing the interdependencies and number of point-to-point interfaces between systems & applications of the architecture.
- Increase *efficiency & speed* of information delivery to end-users (DDAs).
- Improved data quality: *avoid costs* related to poor data quality e.g. time and manual effort needed to fix data errors.
- Improved *application performance* as more effective information lifecycle management can reduce the amount of data in the production environment.
- *Smaller backlog* for IT teams, *shorter delivery time* for business users requests.
- Fight against *growth of "shadow IT"* as a result of poor perception of IT from the business.

- Improved *understanding of* data dependencies & relationships via visualization and simplification of complex data flows.
- *More effective integration* of external/3rd party DBs through the development of consistent data models, business terminology, codes, file formats etc.

However, despite the plurality of potential benefits, costs and risks that stem out of a deployment of a DG program also have to be identified (Ladley, 2012).

- ***Costs*** refer to time and effort needed from current staff of the organization, including executives, business and IT directors and managers, other data stakeholders and data consumers (end-users) and the necessary education and possible training that they need to follow to increase awareness of DG concepts and what it means for the whole enterprise. Moreover, the organization needs to investigate the emerging demands for new positions that require a specific budget, either full-time (e.g. Chief Data Officer) or part-time (e.g. data stewards with a form of bonuses attached to relevant tasks apart from existing job description and corresponding salary). Additionally, even though at initial stages of deploying a DG program investments in new technology and tools is not considered necessary, as the organization matures in respect of DG capabilities the demand for acquiring software solutions that will help increase effectiveness of executed actions might be brought in the surface.
- ***Risks*** that can be a result of the DG program itself are categorized in *Business Risks*, referring to failure to address business requirements via the DG initiative, *Compliance & Security Risks* that have to do with security breaches and violations of regulations due to an ineffective DG program and *Cultural Risks*, in case the organization fails to align effort and resources to support Data Governance that will lead to continuing the wrong data management practices that resulted to the need for a DG program in the first place.

5.2. Executive Sponsorship

In order to increase the potential for successful deployment of a DG program within an organization, commitment and support to the new culture of managing data and information as an enterprise asset is vital at a very senior management level (Ladley, 2012). The executive support provides the strategic direction for the Data Governance initiative to ensure alignment with the corporate and business strategy and helps to resolve political conflicts that are a frequent phenomenon in large information management initiatives with participants from various part of the organization that have different goals and demands.

Furthermore, executive sponsors are responsible to facilitate the necessary cross-functional coordination and provide the motivation that will support the organizational change needed (Thomas G. , 2006). Ensuring funding for the DG program is an absolutely vital aspect, where an engaged executive sponsor's contribution can be valuable for continuous and effective support of the program with the necessary resources. On top of that, executive sponsors are responsible for providing oversight to the program, monitoring its progress and promoting the necessary adjustments to ensure alignment with the business objectives and strategy.

In a survey amongst asset and insurance management companies (The Economist Intelligence Unit, 2015), 69% of organizations with C-level support of their overall data strategy report that current challenges and requirements can be at least adequately managed, whereas the same number for companies without C-level support dropped to 57%.

A variety of suggestions on the business areas that the executive support should come from depending on the direction and demands of the organization for the DG program can be found in literature (Soares, 2010):

- ***Information Technology***: The trend of assigning ownership of the DG program to the IT organization and the *Chief Information Officer (CIO)* is gaining ground in the past few years, considering the increased importance of IT and technology for multiple industries and business areas. However, as already discussed, DG is a business program and participation of the business is crucial for effective design of the appropriate rules and policies that will allow maximum utilization of the business value of organizational data assets. Involvement of the IT is vital in a supporting role of the business and therefore, accountability for the DG program should be shared to facilitate effective communication, coordination and collaboration and avoid possible political obstacles.
- ***Business:*** Consequently, depending on the industry, several business domains demonstrate active involvement and engagement in DG initiatives. For an organization with the business model of Company, the *Investments* domain can provide valuable input for the design and development of a DG program that focuses on critical data that lay in the core business model of the company.
- ***Risk, Compliance and Security:*** Especially for organizations of the financial services industry where privacy, security, compliance and risk management have an essential role, many enterprises have assigned ownership of the initiative to the *Chief Information Security Officer (CISO), Chief Risk Officer (CRO)* or similar. The recent global, systemic financial crisis have brought in the surface new requirements for increased transparency in data operations and improved trust and confidence to the data for effective risk management.
- ***Sales & Marketing:*** Internal and external data and their effective governance to increase understanding of the market, customers and partners are critical for any business and industry, thus the *Chief Marketing Officer (CMO)*, *Sales Executive* or similar can play an important role in advocating and supporting the DG program.
- ***Corporate function:*** Several organizations have deployed DG as a corporate function under the *Chief Executive Officer (CEO)* umbrella in order to emphasize the importance of the program as well as provide consistency across the enterprise regarding the strategy, objectives and activities of the initiative.

However, the motivation for a DG program rarely starts top-down. Usually, a group of people in various functions throughout the organization are the initiators, as they are concerned by the huge growth of available data and are focused on improved management of them to maximize their potential business value and mitigate the underlying risks (Soares, 2010). After these individuals are identified as major stakeholders of the DG initiative, they can form a preliminary DG working group that will work on the details of the program (business case, DG framework, implementation plan etc.) and will push towards ensuring executive support and engagement. This DG working group, possibly with some critical additions and/ or subtractions of stakeholders can be transformed to the Data Governance Council as a tactical level organizational board that will refine the DG framework for implementation and control its deployment within the enterprise, as it is further described in section 5.5 of this chapter. The composition of the DG working group can also be similar with that of the DG Council.

Examples for participants regarding the technical perspective of data operations are of course leaders from various teams in the IT department that are responsible for delivering data, applications and systems to the business users, IT architects, leaders from the data management and business intelligence teams, as well as compliance, security and privacy officers. In respect of the business viewpoint, suitable participants for the workshop can be identified in team leaders from the Investments Front Office domain, leaders from the Investments Back Office function as well as the Sales & Marketing business units, considering the initial scope for DG.

5.3. Maturity Assessment

The best way to initiate a Data Governance program and create a list of most important actions, tasks and activities towards governing an organization's data assets is to evaluate current status of relevant capabilities (Soares, 2010). For this purpose, several maturity models can be found in literature, with the majority of them being based on the Capability Maturity Model Integration (CMMI) that was initially developed by the Software Engineering Institute of Carnegie Mellon University as a methodology to design, measure progress and fine-tune an organization's software development process (CMMI Architecture Team, 2007).

Now the model is transferred to the university's subsidiary CMMI Institute with an extended scope on how to measure and build value-adding capabilities around people, process and technology related aspects of modern companies and institutions and their projects. The model provides a way to identify weaknesses and strengths, select starting points, prioritize relevant actions to improve important capabilities and measure the progress of this process. 5 different maturity stages are included in the model (IBM Software Group, 2007) described as following, while the schematic illustration of the CMMI is demonstrated in Figure 21:

- *Maturity Level 1 (Initial/ Ad-hoc):* Ad hoc, reactive issue resolution, non-standardized processes and environment.
- *Maturity Level 2 (Managed):* Successes are repeatable and partially based on documented and standardized processes and practices. However, the standardization might not apply for all operational areas of the organization and as a result, there is still a risk for project derailment in terms of cost and time predictions.
- *Maturity Level 3 (Defined):* Defined and standardized processes that provide consistency for project execution throughout the organization. However, depending on the business unit and relevant requirements and implications differences in standardization levels might exist.

- *Maturity Level 4 (Quantitatively Managed)*: Design and application of statistical or other quantitative methods to measure performance in processes of various operational areas of the organization.
- *Maturity Level 5 (Optimizing)*: Optimization of standardized processes. Pursue of continuous improvement by establishing enterprise-wide quantitative objectives that are constantly reviewed and adjusted according to changing business objectives and goals.

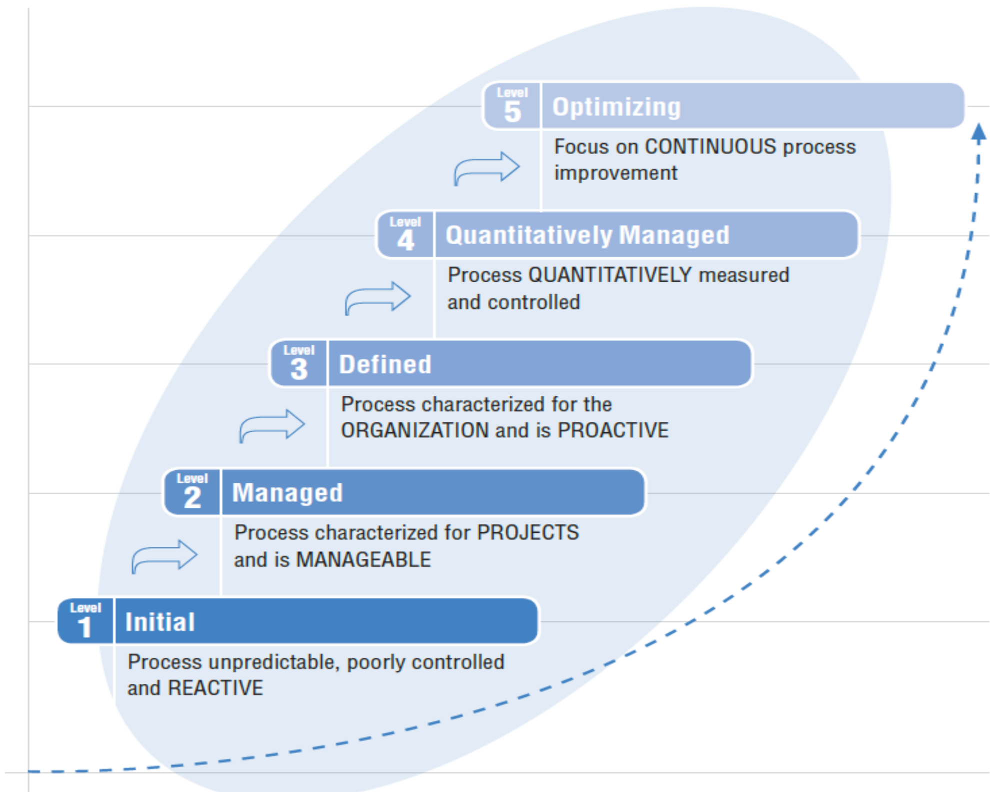


Figure 21 Capability Maturity Model (IBM Software Group, 2007)

Other data governance maturity models can also be found in literature, however they are all based on the CMMI approach and the differences have to do mainly with alterations in the number and names of the different levels of maturity included in the model. For instance, Microsoft suggests 4 maturity levels: *Basic, Standardized, Rationalized, Dynamic* (Salido & Voon, 2010), Kalido the same but with different naming conventions: *Application-Centric, Enterprise Repository-Centric, Policy-Centric, Fully Governed* (Chen, 2010). Therefore, an organization that desires to deploy a DG program and assess its current maturity has different options for selecting the most suitable DG maturity model, however in their essence they all share many similarities are they are following the same logic and approach. A very elaborate discussion on existing DG maturity models as well as the development of a novel structured, quantified approach to measuring DG maturity via a number of case studies in specific organizations can be found in (van der Meer, 2015).

The above CMM model can be used to evaluate current maturity of the organization in respect of the Data Governance capabilities included in the selected for implementation DG framework. Initially, a scope adjustment might be necessary to evaluate current status against a limited set of capabilities that are considered the most important or urgent (Soares, 2010). Moreover, the assessment can be focused to specific functions, teams or geographical units of the organization as a full, enterprise-wide attempt might at first seem intimidating in terms of time and resources needed. Furthermore, it might be the case that a single rating is impossible to be derived for the whole organization. Most companies select a specific timeframe regarding the transition from the current state to the desired, future state of maturity that is between 12 and 18 months and can be considered as the first phase of the DG program, depending on the scope for the capabilities as well as the organizational functions under evaluation.

An effective way of conducting a maturity assessment is to organize a workshop with the participation of the appropriate mix of business and IT professionals, resulting in the same group composition with the already mentioned DG working group. This way, it is easier to ensure commitment and active involvement from all critical for DG parts of the organization. The participants for the workshop in respect of Company's organization can be identified in functions that motivation for DG exists as improvement of the methods and practices in data operations can have crucial impact on the business value retrieved from the available data assets.

In Table 14 a number of example questions for the participants of the workshop are demonstrated, as a method to identify current maturity of the organization in a quantified manner. The questions are distributed across the various DG capabilities included in the selected framework to assess maturity of each selected capability and for the shake of this example questions to evaluate maturity in respect of 4 key capabilities are presented. The capabilities names and content have been derived from the synthesized framework presented in section 4.7 of this thesis, and for the shake of the example maturity assessment demonstrated here, a number of most critical DG capabilities have been taken into account that also reflect priorities of the organization.

For each capability only 3 questions are defined in the table, the list of questions should be enriched after extensive consultation with all major data stakeholders to cover all data domains and issues that are considered in scope. A more detailed description of this example can be found in (Arizona, Department of Education, 2014):

Table 14 Data Governance Maturity Assessment Example, adapted from (Arizona, Department of Education, 2014)

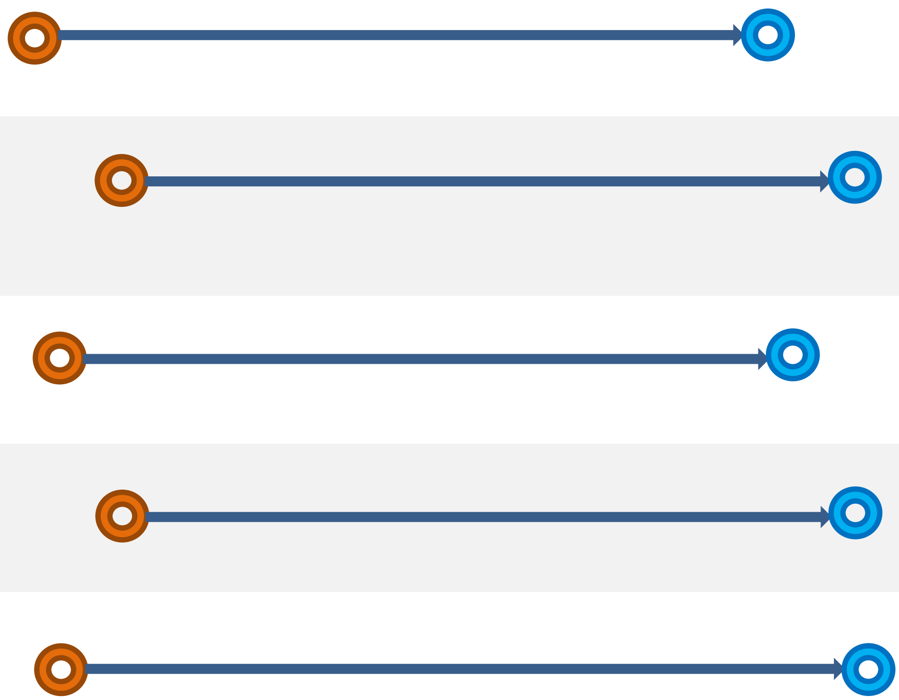
Awareness	**Yes**	**No**
1. Is there an easily accessible online repository (e.g. SharePoint) with information for the DG framework and relevant communication tools to increase awareness of DG program throughout the organization?	3	0
2. Have you developed a charter for the DG program with specific definitions and clear objectives in sync with executive demands to ensure sufficient support?	2	0
3. Have you organized workshops and other change procedures in the direction of developing a culture of collaboration and conjoint responsibility to effectively govern organizational data assets?	2	0

	Yes	No
4. …	…	…
Score	…	…
Compliance & Security Mechanisms	**Yes**	**No**
1. Have you established a standardized process regarding data access, processing and analysis so that unauthorized access and changes in source systems and databases are prevented and compliance with regulations ensured with the appropriate internal controls?	2	0
2. Have you identified sensitive data items with high associated risk and tagged accordingly the databases and systems these are located?	1	0
3. Are there data access restriction rules to authorized business and IT users (data stewards), with all login activity traced?	2	0
4. …	…	…
Score	…	…
Information Workflows	**Yes**	**No**
1. Have you identified critical data items and come to an agreement regarding their official accepted location in systems and databases of the enterprise IT architecture?	2	0
2. Have you documented information workflows for critical business processes?	2	0
3. Are you able to demonstrate an audit trail for business critical data through various sources, systems and applications with all intermediate transformations?	2	0
	…	…
Score	…	…
Monitoring & Measurement	**Yes**	**No**
1. Are there clearly specified metrics connected to business & technical KPIs regarding data-related organizational performance?	3	0
2. Have you established reporting processes that make it easier for executives and the DG council to track status and perform control?	2	0
3. Are there fixed intervals of metrics review with the necessary adjustments?	2	0
4. …	…	…
Score	…	…
Roles & Responsibilities	**Yes**	**No**
1. Is there a Data Governance program supported by executive level sponsor(s) with clear strategy, business objectives and action roadmap that has been designed with the participation of all departments and functions with a stake in governance of enterprise data assets?	3	0
2. Is there an established Data Governance council that will define and control the implementation of the DG program?	2	0
3. Has the organization appointed data stewards in alignment with IT department as well as major business units and the corresponding core business processes and systems?	2	0
4. …	…	…
Score	…	…

After the responses in the questions of the assessment are collected and the total scores for each capability are calculated, the current maturity as well as the actions needed to reach the desired state in the future can be mapped according to different score scales that have to be agreed upon. For example, assuming that the maximum score of each capability would be 20 in case all questions are answered as *Yes*, the convention shown in Table 15 can be used to evaluate the organization's maturity:

Table 15 Maturity Level Calculation, adapted from (Soares, 2010)

	Maturity Level				
Data Governance Capability	**Initial 0< Score ≤3**	**Managed 4≤ Score <8**	**Defined 8 ≤Score <12**	**Quantitatively Managed 12≤ Score <16**	**Optimizing 16≤ Score <20**
Awareness					
Compliance & Security Mechanisms					
Information Workflows					
Monitoring & Measurement					
Roles & Responsibilitie s					
…					



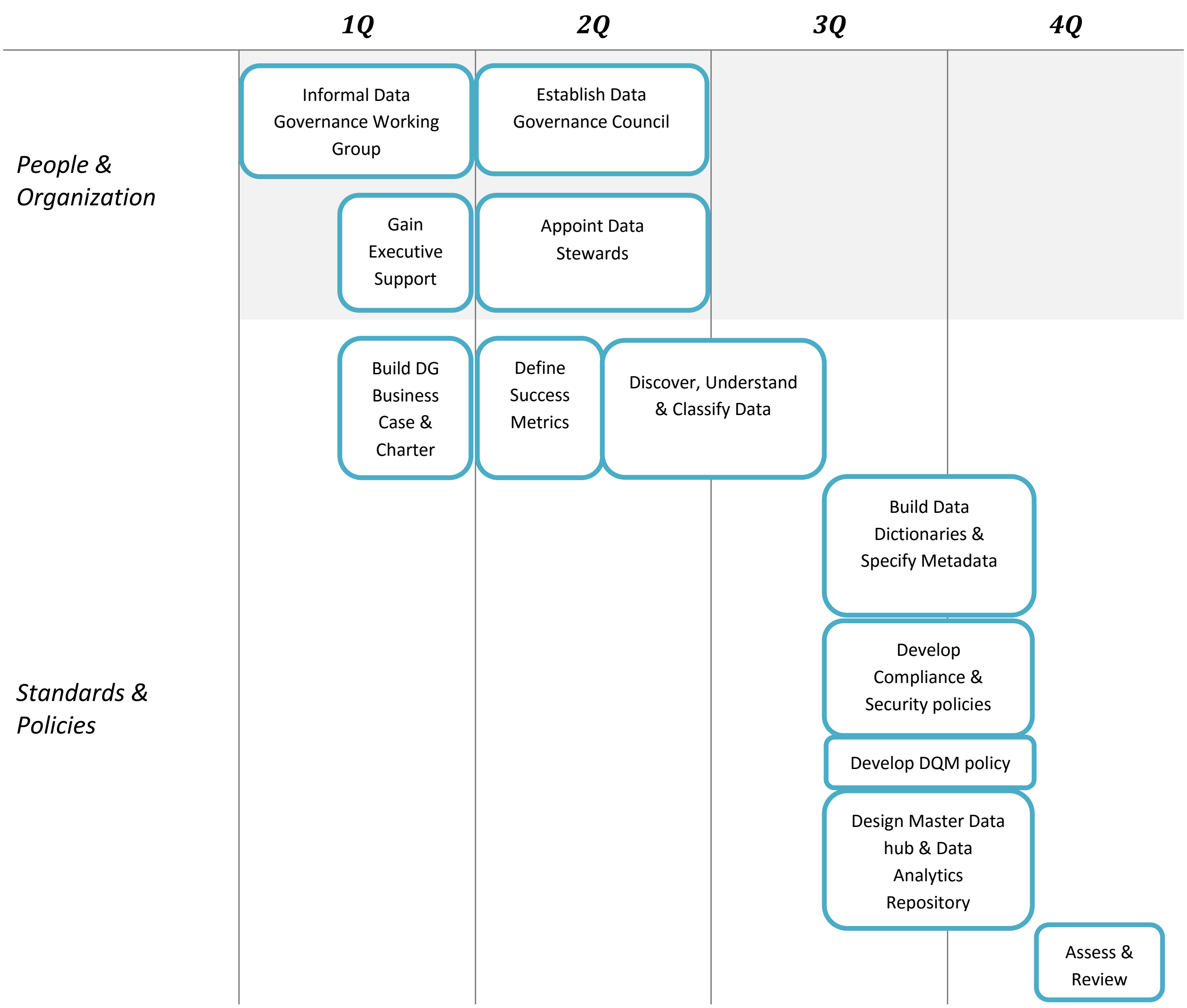
Where is the current state, and the desired state in the specific timeframe selected.

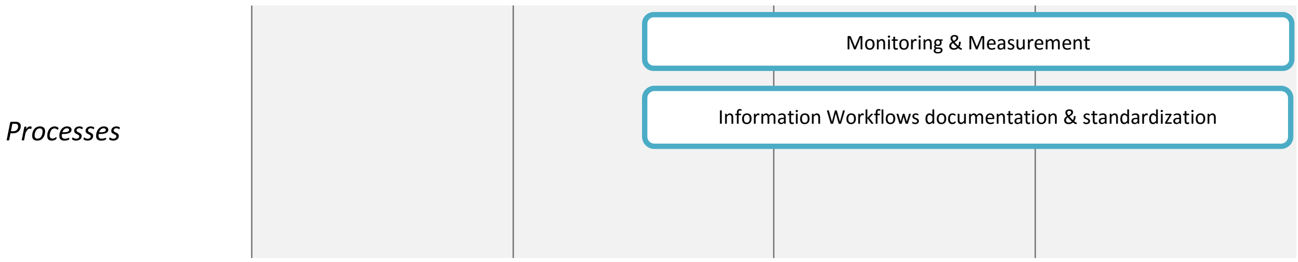
5.4. Roadmap

The desired outcome of this step is to create a collection of activities the organization needs to perform in order to support the transition from current to the desired state of maturity in respect of the selected DG capabilities (Soares, 2010). Since no formal DG organizational bodies and roles have been established yet, the DG working group that also conducted the maturity assessment is the most suitable place to design and plan the deployment roadmap as well. After the results of the assessment are summarized and the current maturity is evaluated, the next step is to prioritize the capabilities that improvement is most urgent, or the business value identified has the greatest impact potential for the organization.

This would be useful to provide the "quick wins" that were also discussed in the business case section. However, the required input and resources for the actions necessary to increase the framework's capabilities has to be taken into consideration, as initiatives with larger scope and increased demands in resources also carry the proportionate risk. In this context and since there is already an agreed-upon score for each action that can be taken as an input from the maturity assessment, the DG working group can plot the various activities along a number of milestones that are distributed across the duration of the deployment. The actions can be categorized in 3 dimensions: 1*) People & Organization, 2) Standards & Policies and 3) Processes*, so that the roadmap is consistent with DG definitions and the corresponding categorization of DG capabilities in these 3 aspects.

Table 16 Data Governance Roadmap, adapted from (Soares, 2010)





A timeframe of 12 months is selected, so that results can be assesed as early as possible and review of the deployment process with the necessary corrections and adjustments can be performed. The DG working group is responsible for focusing on the value of data as an organizational asset, quantify relevant benefits, risks and costs as much as possible and build a solid business case that will assist in gaining support, active engagement and funding from senior management and executives. The first designs of the DG organization should be then formed, including the DG council as well as the process of appoining Data Stewards, while success metrics should be developed based on the business case, in order to evaluate progress of DG within the organization. After DG organizational structures have been formalized, the organization can concentrate resources on DG projects in a centrally coordinated, structured manner.

5.5. Data Governance Organization

People and organizational bodies is one of the most important aspects of Data Governance, because after all the responsibility of protecting, maintaining and enhancing the value of enterprise data will be assigned to them. As a result, the design of the Data Governance organizational blueprint is an essential process for effective deployment of the program and its best fit with the application environment (Soares, 2010).

A number of contingency factors that impact the configuration of DG can be identified in the organizational and data landscape of Company as demonstrated in Table 17. These factors occur from similar research efforts (Weber, Otto, & Osterle, 2009; Otto, 2011) but also from the environment of this case study. Moreover, each contingency factor is correlated with a relevant *criterion* from section 4.4, along with a description of how the criteria are satisfied in the configuration of the Company DG Organization and the focus of the DG program.

Table 17 Company Contingency Assessment, adapted from (Weber, Otto, & Osterle, 2009; Otto, 2011)

Contingency Factor	***Assessment***	***Satisfied Criterion***	***Impact on DG design***
Mandate for DG	Cross-functional coordination between IT and Business	Alignment with Business Objectives	CIO & Business executives should lead the initiative
Current DG practices	Informal, reactive approach	Clear Roles & Responsibilities Formalization	RACI matrix to formalize responsibilities for data-related decision making
Organizational Structure	Centralized	Organizational Structure	Centralized responsibility in the DG Council

Decision making style	Consensus based	Decision making style	Multiple "C" in the RACI matrix , Representation of all key domain stakeholders in DG Council
Company Size	Small to medium	Simplicity, Practicality & Implementability	3-layer DG Organization, Small DG organization bodies, Emphasis on direct communication of DG concepts, decreased "bureaucracy" and fast execution
Scope for DG	Core business functions with greatest impact from data, several other functions in future scope	Flexibility	Core business functions representation in the DG Council, Chief Domain Stewards, Easy to extend for other business functions & domains
Culture	Practical	Practicality & Implementability	IT Operations key stakeholder in the DG Council. Emphasis on operationalized DG, quantified methods to measure progress and success. Focus on business impact of information workflows
Industry & Business model	Strong emphasis on security, regulations & risk management	Focus on Compliance & Security	Information Security and Compliance & Risk Mgt functions key stakeholders in DG Council, Priority in relevant policies
Emerging requirements	Distributed and Complex IT & Data environment, Self Service-BI, relatively diverse business user demands depending on the function	Focus on BI/Data Analytics, Flexibility	IT Architecture & Data Mgt functions key stakeholder in DG Council, Functional break-down of Data Stewardship teams, Business Stewards that request increased access & availability in cooperation with Technical Stewards bear responsibilities for data management and BI artifacts ownership

Based on the assessment of Company displayed above and the corresponding conclusions regarding the impact on the DG design, a formal DG organizational structure was created as illustrated in Figure 22.

However the exact composition of these bodies can be adjusted accordingly to specific characteristics. For instance, a really engaged executive sponsor can as well lead and participate in the DG council, whereas a company that wants to establish a decentralized model can include divisional or subject area data stewards in its DG council composition instead of their team leaders to provide a better alignment with the business requirements and increase flexibility and agility of the model.

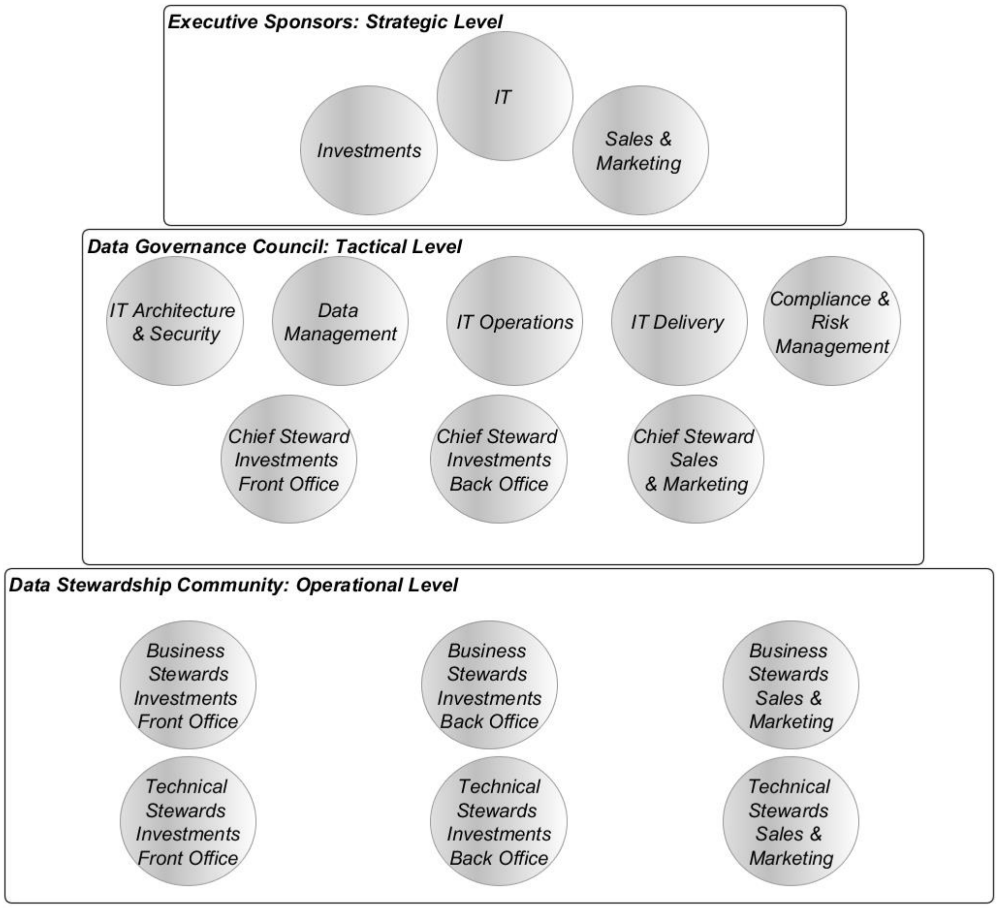


Figure 22 Company DG Organization

After the organizational design for DG has been created, responsibilities for relevant decisions can be distributed using the RACI matrix in combination with the factors *Organizational placement* and *Coordination* of the decision making authority that are proposed by (Wende, 2007), finding the balance between centralization and decentralization as well as between hierarchical and cooperative decision making style regarding DG-related decisions.

A RACI matrix example for Company in respect of several DG-related decision domains is presented in Table 18. Domain Chief Stewards (e.g. Investments Front Office, Investments Back Office, Sales & Marketing) is embedded in the DG Council and bear the responsibility to communicate data-related decisions to his/ her stewards and escalate issues raised by the business stewards to the Council. The equivalent of Chief Domain Steward from the IT perspective is represented in the council by functions such as IT Architecture & Information Security, Data Management, Compliance & Risk Management.

Table 18 RACI Matrix for Company, adapted from (Wende, 2007)

Decision domains	*Executive Sponsor(s)*	*DG Council*	*Business Steward*	*Technical Steward*
Budget, Strategic business objectives, Alignment with IT & Corporate Governance	R	C	I	I
DG Framework implementation, Project authorization & prioritization,	A	R	C	C
Metrics development, Monitoring & Measurement	I	R (IT Operations)	C	C
Official data sources, Data architecture principles	I	R (IT Architecture)	C/I	C/I
Data Quality policy & standards, Master Data, Information Lifecycle policy	I	R (Data Mgt)	C/I	C/I
Data Usage, Regulatory compliance & privacy policy, Audit trail	I	R (Compliance & Risk Mgt)	I	I
Access & Authorization policy	I	R (Information Security)	I	I
Business data models, Data definitions, Calculation formulas, Business metadata, Data profiling, Dashboard ownership	-	A (Chief Domain Steward)	R	C
Technical data models, ETL process, Data flows, Technical metadata, Data profiling	-	A (IT Architecture, Data Mgt)	C	R

A necessary remark is that responsibility for policy approval and data-related decision making lies in the DG Council, but the actual implementation should be delegated to IT & Business professionals of the function in (), after consultation sessions that will result in specific guidelines.

5.6. Data Stewards

Data stewards can be identified as people spread throughout the organization that work with the data on a day-to-day basis and are given the responsibility to manage them as organizational assets since they have the maximum of knowledge about their usage, characteristics, risk, business meaning and value (Soares, 2010).

A *Data Stewardship Community* can be seen as a forum of coordination, collaboration and knowledge exchange between data stewardship teams of different subject areas, functions or systems. The leaders of the various stewardship teams can coordinate activities for maximum resources and effort utilization, whereas a *Chief Data Steward* is possible to be appointed as head of all stewardship teams as their size and scope increases and the maturity of DG in the organization advances. In section 3.4 of this thesis and according to (Dyche & Polsky, 2014), 5 different ways to appoint data stewards are described and an analysis of advantages and disadvantages of each approach in respect to Company's organizational setting and data environment is demonstrated below:

- ***By Subject Area***: Possible subject areas could be portfolio data, prices, benchmarks, customer, product and employee domains. Clear scope for data-related responsibilities is provided, whereas it is easier to find suitable stewards within the organization with experience in the usage context of the data and their value. Furthermore, the domain knowledge of the steward will evolve over time as he gains experience in data management practices of his subject area, while information sharing is enhanced as the stewards act as connectivity point between multiple teams that work with the data of his subject area. However, the size and complexity of the subject area can lead to difficulty in finding qualified data stewards and possibly a further categorization in subdomains will be necessary. Additionally, this approach fits best in organizations that have above intermediate maturity in DG capabilities and good knowledge of the available data sources and ways to organize them effectively in subject areas.
- ***By Function:*** Similarly, functional data stewardship offers clear scope of responsibilities around the data a specific department uses (e.g. Investments Front Office, Investments Back Office, Sales & Marketing, HR etc.) and even more stewards are aligned with the existing organizational structure. Also, they have advanced knowledge of the data content and context and it is easier to attach their activities to specific business initiatives and objectives of their business unit. On the downside, this model can lead to duplication of efforts and information silos as stewards from different functions might work on the same data (e.g. portfolio data) without being motivated for cross-functional collaboration, thus a stable and relatively mature DG environment would be more suitable for this approach to leverage the disadvantages.
- ***By Business Process:*** Assigning stewards with this model requires a clear view of the company's business processes and data flows that span through multiple systems and functions, an effort that is currently under progress and has not reached the required maturity levels within Company

for the effective deployment of this approach. Furthermore, boundaries for data responsibilities might be hard to define as multiple business processes need and work with the same data. Nevertheless, even in the current situation it can help the organization in the direction of mapping its business processes explicitly and define stewardship as an extension of process definition, also making it easier to measure progress and success in respect of process-related metrics.

- ***By Systems:*** This approach shares similarities with current definition of data ownership in Company that is assigned per system and a corresponding business manager that is responsible for quality of the data. However, this way of deploying data stewardship does not resolve existing problems around ownership as data are produced or provided to the users by a number of different systems, thus making it hard to clearly define responsibility for the data. Additionally, the system-oriented approach might lead to the perception that responsibility for the data belongs to IT, resulting in disconnection of data stewardship with the business. It might be though a practical and effective way to introduce stewardship in the organization, assuming that proper balance between IT and business responsibility can be established.
- ***By Projects:*** This might be a fast approach to assign data stewards in organizations with low DG-related maturity as project teams are already in place, however it cannot provide a consistent approach to data stewardship as specific data-related knowledge might be hard to find in members of all project teams. Moreover, as projects have a specific lifecycle, data stewardship has to be redefined every time a project ends which is incompatible with the ongoing nature of DG and DS.

Considering the above, a hybrid function-subject area organization of the data stewardship community based (Dyche & Polsky, 2014) might be the most suitable proposal, as it is easier and more intuitive to initially organize DS per function and fine-tune the design as knowledge about the enterprise data and experience in governing them effectively increases.

Current data ownership is tied to owners of systems of applications that produce or capture the data and usually are managers of the relevant domain. An option is to identify current data owners as Chief Stewards of the domain that oversee lower level Data Stewards and are responsible for day-to-day data practices. Subject areas for each domain can then be identified and cross-functional collaboration forums can be established to avoid creation of information silos and redundancy in data and resources spent. The suggested model is depicted in Figure 23. More elaborate discussions within Company and involved stakeholders can offer us more insights regarding the most suitable manner to organize and deploy data stewardship teams in the organization.

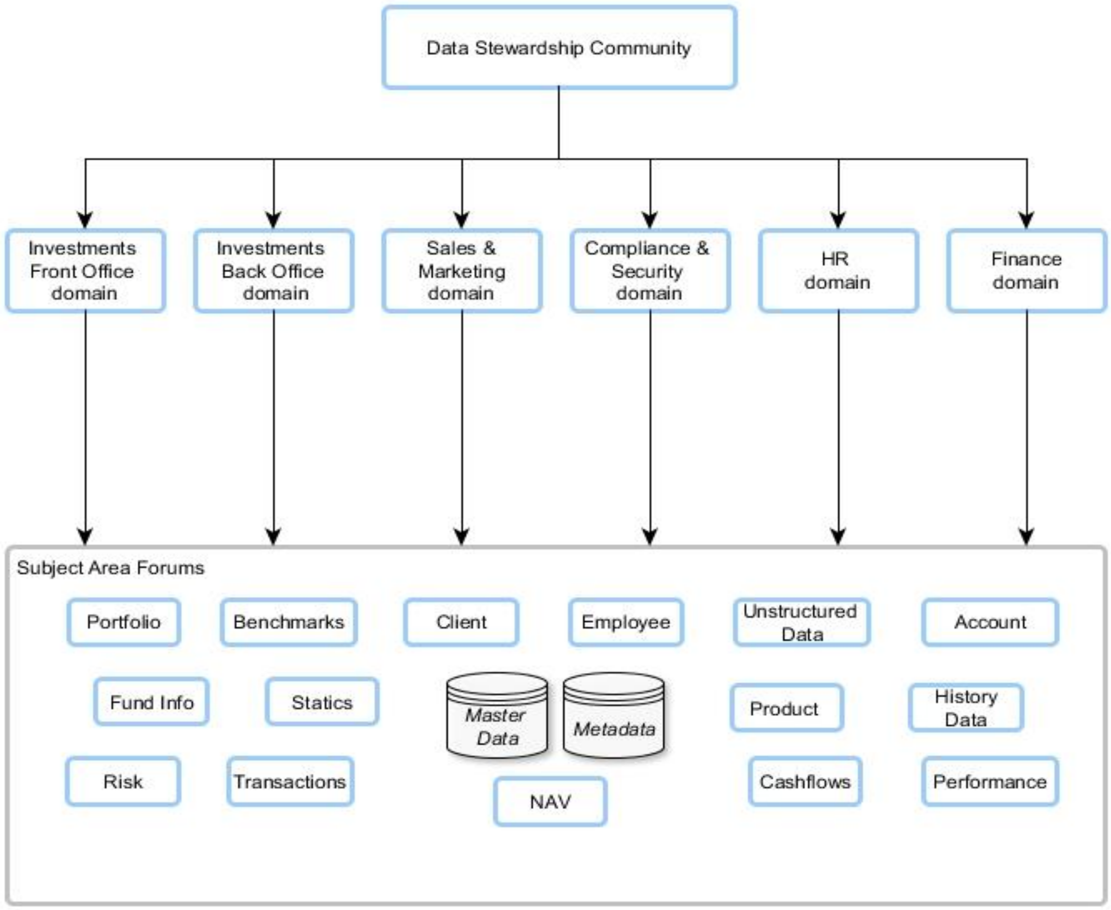


Figure 23 Data Stewardship teams deployment, Hybrid DS model based on (Dyche & Polsky, 2014)

5.7. Success Metrics

In order to measure, monitor and control the progress and success of the program, a collection of metrics need to be shaped that is focused on data-related organizational performance (Thomas G. , 2006). DG progress needs to be evaluated in respect of aspects that have to do with the organization's business model, people, processes and every other area that improving efficiency of data operations can have a great impact.

Data Governance Key Performance Indicators (KPIs) have to be established in relation to business requirements and objectives, but also technical aspects (Soares, 2010). *Business-driven DG KPIs* are data-related metrics that directly or indirectly support business KPIs. For instance, if portfolio performance against the benchmark is a business KPI and a specific column in the data sources that is critical for calculation of portfolio performance is found to contain several nulls or out of range values, the improvement of this specific data element in respect of data completeness and accuracy can be a business-driven DG KPI. Other aspects that can be connected to business performance for the formulation of a complete list of DG KPIs are the following (Thomas G. , 2006; Ladley, 2012; Soares, 2010; Ajilitee, 2012):

- % percentage of total time spent on data analysis as a fracture of time spent on data collection and preparation.
- % percentage of reporting processes/ generated reports that provide proof of compliance with regulatory requirements.
- Increase in customer satisfaction and retention rates as a result of the Data Governance program and improvements in customer data and relevant processes.
- Improvement in cross and up selling results due to more accurate and consistent customer data.
- Overall cost savings due to improvement in data-related inefficiencies and more effective utilization of relevant resources.
- Total data-related cost per trade performed (€), by calculating all associated data costs regarding infrastructure, 3rd party data purchase, human resources and so on.

Technical aspects of DG also have to be measured by developing *technical DG KPIs*. Some examples of the areas these KPIs could be focuses on are provided below:

- % percentage of documented data flows.
- % percentage of monitored in respect of performance data flows.
- % percentage of incidents that come from data-related problems
- % Percentage of time reduction for data issues resolution.
- # Number of obsolete data models that were removed.
- # Number of data sources that were consolidated
- # Number of data attribute definitions per data domain that were revised/agreed upon and the corresponding data dictionary was created.
- # Number of reports generated per month by business area.
- AIRT (Average Incident Response Time) for data-related incidents.
- Speed in integration of new data sources (#Number of days/hours).
- Total storage in gigabytes and total cost of storage (€).

The accepted baseline for all these metrics has to be established so that progress of DG can be measured and evaluated, also milestones and intermediate goals have to be defined to ensure the program is not derailed. The ultimate vision for operationalized and quantified Data Governance can take the form of a DG dashboard or scorecard with all available metrics that enable real time monitoring & control from the Chief Data Officer and the DG council. The detailed specification of these metrics can help to further develop the business case for DG, for instance by calculating advanced DG KPIs such as total data-related cost per trade in € and potential savings regarding this index by improvement in data-related inefficiencies.

5.8. Data Architecture

After the DG organization is established, data stewards are appointed in an appropriate manner and responsibilities and accountabilities are distributed amongst them, the organization is ready to start working with the available data resources at a lower level in a formal and standardized way. Considering the complex, distributed IT architecture and data environment of Company that is often met in financial services organizations, rationalization of data architecture and focus of initial efforts on business critical data domains and objects can have a great impact in demonstrating real value. Standardization and establishment of enterprise-wide common data definitions is one of the main objectives of Data Governance (Thomas G. , 2006), that has a significant impact in areas such as DQ, compliance & risk management and effective support of data analytics projects.

5.8.1. Classify & Understand Data

As a baseline for any improvement in data operations, it is vital to explore organizational data assets, identify the most critical for the business data domains and items and classify them according to their risk and potential value (Soares, 2010; Thomas G. , 2006). A very useful method for these activities is provided by Gartner (White & Friedman, 2016). Initially, the available data items should be categorized depending on their importance in respect of the applications, systems and business processes that use these data by using the information rings that are illustrated in Figure 24 below:

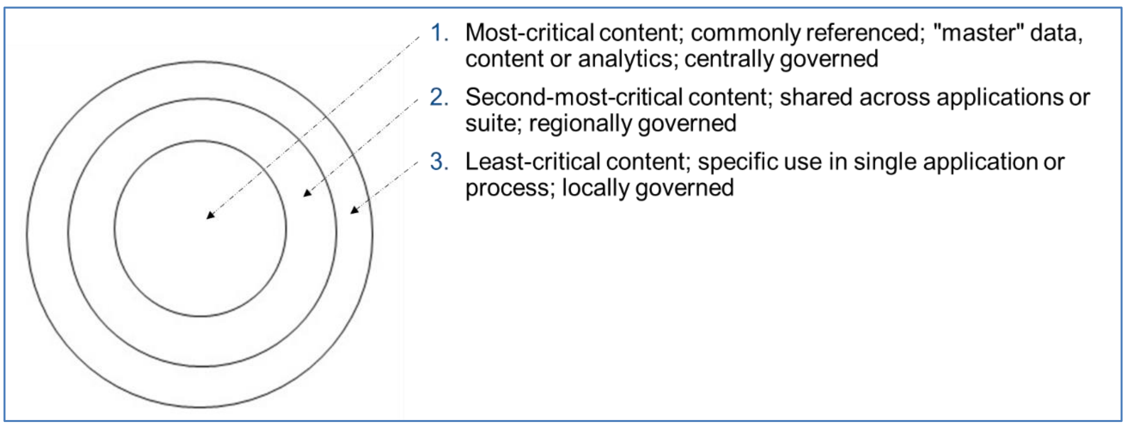


Figure 24 Gartner's information rings, (White & Friedman, 2016)

This way, the company can have a more clear view of the available data resources and their impact on business operation, enabling this way the prioritization process to begin regarding which data domains and items will be the focus point during the initial phases of the DG program. Narrowing down the scope of first efforts to those domains and items also provides a way to prove the value of the program without spending too much effort and resources that could be the case if the organization attempted to establish new governance practices for the whole scope of its data assets. For the selected domains and elements, overlaps across various data sources need to be spotted and possible duplicates and inconsistencies that are result of complex transformations as the data flow through different platforms and systems have to be dealt with.

5.8.2. Data Dictionaries

As the process of resolving these inconsistencies in the most critical data elements begins, the agreed definitions and other changes in data structure need to be registered and organized so that are available to the business at the required time and location. Consequently, consistent, enterprise wide business terms organized in repositories for each data domain that are called *data dictionaries* or *business* *glossaries* have to been defined (Soares, 2010; Thomas G. , 2006). Moreover, it is important to establish clear linkages between business terms and technical specification such as database column names, as well as integrate the data dictionaries with the corresponding business systems and applications. This way business employees will be able to locate data sources that correspond to the business information they need and also IT professionals can understand the business context of the data while they perform technical tasks on them. After all, enhanced communication and a collaborative culture of conjoint responsibility in respect of organizational data assets between business and IT is the essence of Data Governance.

5.8.3. Master Data & Metadata Management

The most critical for the business data domains and items that are shared across multiple systems, applications and processes should be organized in the corresponding repositories for direct access by the users to ensure enterprise-wide consistency. The same applies for effective management of metadata that provide knowledge for the full range of activities performed on the data assets throughout the organization, thus creating the demand for a metadata hub that enables central control and management.

5.8.3.1.Implement Master Data Management

Based on the classification of the data and the corresponding analysis regarding the impact of available data assets, the most critical and commonly shared throughout the organization data items and domains that a MDM project should focus on should be organized at a central repository that is used as a point of reference and breaks through departmental information silos. For the financial sector, the largest data management costs come from market data that are purchased from vendors and data that are shared and used from front office portfolio management systems to back office transaction settlement and reporting processes (e.g. instrument, client, positions etc.) (Deloitte Advisory & Consulting, 2014). Cost savings and general improvement potential in operational efficiency for asset management companies can be great by centralizing and consolidating market data flows, optimizing data providers' costs and rationalizing number of reference data fields in systems and databases.

Subsequently, systems and business processes that "touch" data items included in the master data domains need to be identified as the designed MDM hub should be connected to these systems and processes for effective information sharing and seamless data distribution. Regarding the architecture design of the MDM solution, the following approaches can be deployed (Soares, 2010):

- *Transactional architecture*: Based on Service Oriented Architecture (SOA) principles, this approach is highly integrated with systems and processes and changes are updated in the central data files before transmitted to the peripheral sources.
- *Registry architecture*: In case a transactional hub is not allowed due to regulations (e.g. healthcare, law enforcement) or other restrictions, this design contains only the pointers to the source systems in the master hub and not the actual data.
- *Analytical architecture*: Updates in the data files are implemented only in the central master hub that is used for analytical purposes and are not propagated to the source systems.
- *Hybrid architecture*: Combination of principles and design styles of all the above approaches.

5.8.3.2. Create Metadata Repository

Metadata is "data about the data", in other words activity records regarding when, in which system and from whom the data and other artifacts such as data models or the corresponding ETL processes are created or changed, what is their intended use as well as what is the impact from possible changes to these artifacts (Soares, 2010). Metadata help IT professionals to understand the business context of data usage, but also business users to get a better view on the complexity of the underlying transformations and other technical tasks performed on the data. Therefore, it is important to merge business and technical metadata in a common, central repository that is accessible to the whole organization. Moreover, accurate metadata specification is crucial to enable mapping of data flows across various systems and processes of the organization and demonstration of data linage and audit trail for compliance purposes.

5.9. Manage SS-BI Analytics

Considering the initial trigger for the Data Governance program within Company, relevant discussions started due to the intention to deploy Qlik Sense data business intelligence platform and the corresponding concerns regarding the freedom provided to the business users and how data and generated BI artifacts (reporting templates, dashboards etc.) can be effectively governed to increase the retrieved business value and minimize associated risks. BI tools and data analytics platforms are positioned at the end-point of information workflows, as a result the quality of the data that reach business users in the consumption layer is crucial for the success of any analytical project (Ladley, 2012). This way trust in the accuracy and consistency of the generated BI artifacts that are created from different parts of the business can be increased.

A knowledge exchange hub to improve sharing of information and best practices in data analytics projects across different teams and business units of the organization is a qualified solution to organize data for analytical purposes effectively. Intel created an integrated analytics hub for its sales and marketing teams (Intel IT, 2015) to overcome limitations of traditional data warehousing approaches. This way they can manage the increasing size and complexity of enterprise structured and unstructured data with data lake model implemented in a Hadoop distribution, removing the need for rigid data models that compromise flexibility to address emerging requirements.

The target was to improve data accessibility and availability and support consistency with multiple BI front-end tools to keep the design compatible with future initiatives and avoid lockdown of BI capabilities to a specific tool and architecture. The results demonstrated estimated quarterly cost savings of USD 170,000 and final sales of USD 576,000 as a result of marketing analytics optimization and automation across digital media channels. Furthermore, speed of data delivery dramatically improved and business analysts could access data at the time and format needed.

However, no matter the choices of an organization in respect of technology selection and implementation details, the steps described in the following sections consist a methodology to understand business demands and the corresponding data requirements. This way, the data, reporting templates and dashboards needed for BI can be organized in an integration layer that focuses on consistent and on-time distribution to the end-user layer based on predefined standards, policies and processes that contribute in resources and effort rationalization.

5.9.1. User Categories & Requirements

The first phase of governing analytics would be to identify the various user categories within the organization so that corresponding data requirements, reporting demands and business targets that need to be achieved via advanced analytics and visualization capabilities are recorded. Within Company, several user groups can be initially identified as data-intensive in respect of their business operation, creating the need for improved data access and availability and novel analysis and visualization capabilities provided by a self-service environment:

1. Investments Front Office
2. Investments Back Office
3. Sales & Marketing
4. Compliance & Security Office
5. HR
6. Finance

Since some of these domains include a number of teams with differences in data and reporting requirements, further categorization in subdomains might be necessary to shape a clear landscape of needed data items for each team and locate overlaps. This process shares similarities and can be tied to the gradual development of the data stewardship community, constantly refining the subdomains as DG maturity increases and arriving at an optimal organization of the subject area forums, the corresponding data sources and elements and the stewardship teams that bear the responsibility for protecting, maintaining and enhancing value of enterprise data assets. The Data Stewardship Community can evolve as a competency center for data stewardship best practices, assisting in transferring acquired knowledge and experience throughout all functions and domains of the enterprise.

The following Table 19 focuses on the user categories that demonstrate the greatest desire and urgency for advanced data analytics and visualization capabilities via self-service platforms so far and are also in the initial scope of the DG program, and shows a possible further categorization in subdomains that have different data and reporting requirements:

Table 19 Data analytics domains and subdomains

Domain	***Investments Front Office***	***Investments Back Office***	***Sales & Marketing***
Subdomain	1. Equity 2. Fixed Income	1. Process Management 2. Activity Based Costing	1. Sales 2. Marketing 3. Market Intelligence

The DS community can be broken down to different hierarchies within each domain so that the scope of responsibilities for stewards is rationalized and flexility to approve changes and resove isssues is indreased. An illustration of the evolved DS community is displayed in Figure 25.

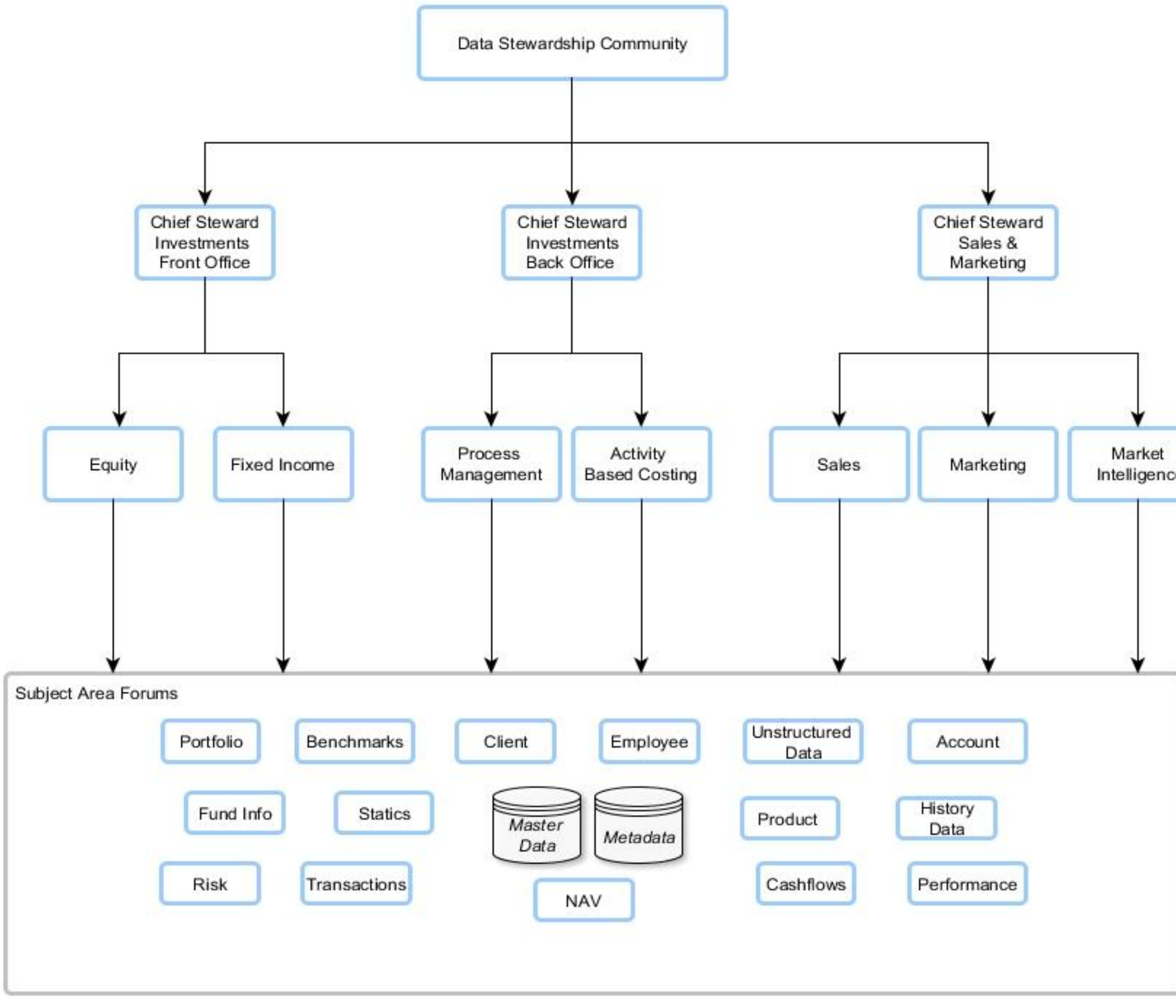


Figure 25 Data Stewardship Community, Hybrid DS model based on (Dyche & Polsky, 2014)

5.9.2. Data Analytics Sources

A series of case studies in organizations that desired to increase business agility and flexibility through Self-Service BI indicate that Data Governance is crucial for the success of the initiatives (TechTarget, 2016; Intel IT, 2013; Eckerson, 2009). After all, accurate, complete, consistent and on-time data organized in accessible data dictionaries with all relevant metadata specification are the lifeblood of any data analytics platform. As a result, it is important to organize the necessary data sources and elements and use them as a foundation for building BI capabilities that address business requirements while providing a clear view of the data landscape for effective monitoring and control. Data requirements for Qlik Sense project can be the starting point for this process that can be described as following:

- Identify data requirements for each domain and subdomain. Data requirements refer to which data items are needed as a first step to investigate overlaps and potential for information sharing, but also special requirements that may correspond to these data items as a next step (e.g. frequency the data are updated/delivered).

- Match the data items needed for each user category with the sources for these data, or else the systems that create or provide the data. This will help to identify cases in which the same data items are provided by different systems, leading to potential inconsistencies.
- The final step is to connect the source systems and the necessary data to the front-end layer of BI tools so that we have a clear view of the number and nature of connections that are necessary. This way data could be organized in an integration layer independent from Qlik Sense or any other BI platform itself with predefined data connections, increasing efficiency of the process, removing redundancies and duplication of work in data preparation, modeling and so on. Also all connections to the data would be known, so no need to manually add them exist for the business users. New connections or in general changes for the process can be requested by the data stewards of each user category, with the creation of a workflow that includes the necessary approvals for effective governance.

Table 20 demonstrates a first effort to record data requirements for the Investments Front Office domain, including different needs from 2 internal teams, *Equity* and *Fixed Income.* After the required data items are identified, the acceptable source systems have to be agreed upon according to data architecture principles and guidelines from the DG Council. This way data can be organized and integrated in a standardized way to avoid duplicates, invalid data and conflicting definitions. As we can see, multiple sources exist for some data, increasing the risk for potential inconsistencies and redundancies. Furthermore, there are cases where the source is not a system or database that is centrally governed, maintained and controlled, but a departmental database or even excel spreadsheets that are stored locally, are unknown to the rest of the organization and have increased risk for the quality of the data they hold.

Table 20 Data analytics requirements, Investments Front Office domain

Investments Front Office Domain			***Subdomains***	
Data Item	*Source System*	*Governed Source?*	*Equity*	*Fixed Income*
1. Instruments	Enterprise Data Management (EDM)/ Portfolio Mgt System (PMS)	Yes	X	X
2. Instrument statics and prices	EDM/ PMS	Yes	X	X
3. Indices/benchmarks	Benchmark Administration/ PMS	Yes	X	X
4. Positions	Investment Book of Records (IBR)/ PMS	Yes	X	X
5. Transactions	PMS	Yes	X	X
6. Rankings (instrument)	Quantitative Ranking (QR) database	No	X	X
7. EQ characteristics (instrument)	QR database	No	X	
8. FI characteristics (instrument)	Front-Office system (Fixed Income)	Yes		X

9.	Returns (instrument)	Front-Office system (Equity) / Front-Office system (Fixed Income) / Performance Measurement System	Yes	X	X
10.	Risk	Front-Office system (Equity) / Risk System	Yes	X	
11.	Cashflows	Unknown/Various	No	X	X
12.	NAV	Accounting Book of Records (ABR)/ PMS	Yes	X	X
13.	Other data (instrument)	Manual	No	X	X

Another remark is that the majority of data for this domain are needed for both internal teams, increasing the potential for information sharing and rationalization of data integration and modeling efforts to support the corresponding business requirements. Data analytics requirements for the other 2 domains, *Investments Back Office* and *Sales & Marketing* are demonstrated in Appendix B (Confidential).

5.9.3. Access & Authorization

Developing and adjusting policies, monitoring data usage and implementing changes in the data sources that will improve business interpretation and usage of data should be the focus of DG in the context of supporting SS-BI environments. Unnecessary controls and limitations can lead to workarounds, information silos and governance by spreadsheet that increases risk and in general a trend towards "shadow" IT processes that drain valuable resources and compromise transparency in data operations (TechTarget, 2016).

Integration of the data analytics repository with the DG program and connection of their objectives is essential. This can be achieved by appointing the suitable data stewards not only from the business but also from IT that can help in exploring and consolidating the data landscape of the organization in a more effective way. Two-fold data stewardship can foster collaboration, as IT professionals can understand better the business demands and users can comprehend the technical implications and impact of their choices in a self-service environment.

Decentralizing the responsibility for data-related decisions to the people that work with the data and know their content, context and characteristics accelerates the analysis process and removes obstacles and bottlenecks for the business, while at the same time appropriate levels of control are in place with formalized accountabilities and continuous monitoring from the IT. Finding the optimal balance between data accessibility and security is the new challenge for modern organizations.

In this respect, a hierarchy with multiple layers of stewards can be created for each domain and its subdomains, allowing the organization to assign specific decision, access and authorization rights to the users based on their domain and position in the data stewardship hierarchy.

This approach is similar with the reporting framework for SS-BI platforms suggested in (Eckerson, 2009), where role-based permissions configure which users can access and analyze data and information at different levels based on their needs. For instance, data visualizations without advanced drill down capabilities for managers and executives, dimensional data views and filters for analysts and detailed low level operational queries and reporting capabilities for employees that not only analyze and consume, but also produce data and have advanced business and technical knowledge of the source systems. To ensure drill downs and other analytical tasks on the data as well as report generation and sharing are done in a controlled environment, lower level data stewards need to request approval from the Chief Steward of their domain for tasks with big potential impact on the data and other teams that share these data and relevant reports.

5.10. Information Workflows

In many organizations, the data architecture contains several systems, applications and databases of different formats and technologies. Without proper documentation of the official data sources and the data relationships between different systems, many business users have difficulty to find the data they need and are unsure about the quality and currency of the sources. Furthermore, due to the distributed IT environment of Company that is based on scalability and performance of the numerous business applications and systems, architectural principles are set aside in some cases and data are stored in multiple locations.

Table 21 demonstrates an initial scan of Company's data architecture, including a number of business critical data items, their location in source systems as well as the applications and systems that these data are used.

Table 21 High Level Data Architecture

Data Items	***Source Systems AS-IS***	***Usage AS-IS***
Portfolio	1. Custodian X 2. Custodian Y 3. Portfolio Administration System 4. Investment Book of Records 5. Accounting Book of Records	1. Historical Portfolio Database 2. Front-Office system (Equity) 3. Front-Office system (Fixed Income) 4. Securities Lending Solution 5. Portfolio Overlay Solution 6. Portfolio Management System 7. Reporting Solution 8. Risk System
Integrated Portfolio	Multi Manager Solution	
Value Portfolio	Portfolio Administration System	
FundInfo	Reporting Solution	1. Company.com 2. Historical Portfolio Database

Prices	Enterprise Data Management (EDM)	1. Portfolio Management System 2. Risk System 3. Reporting Solution 4. Multi Manager Solution 5. Front-Office system (Equity) 6. Front-Office system (Fixed Income)
Statics	Enterprise Data Management (EDM)	1. Portfolio Management System 2. Risk System 3. Reporting Solution 4. Multi Manager Solution 5. Front-Office system (Equity) 6. Front-Office system (Fixed Income)
Analytics	Enterprise Data Management (EDM)	1. Portfolio Management System 2. Risk System 3. Reporting Solution 4. Multi Manager Solution 5. Front-Office system (Equity) 6. Front-Office system (Fixed Income)
Benchmarks	Benchmark Administration	1. Reporting Solution 2. Front-Office system (Fixed Income)
Indices	Benchmark Administration	Reporting Solution
NAV	Portfolio Administration System	Enterprise Data Management (EDM)
Performance	Performance Measurement System	Reporting Solution

The complete architectural diagram that shows data sources and systems in respect of core business processes of the organization is available in Appendix C (Confidential). As we can see in the above table, there is limited improvement potential for portfolio data objects that are currently stored in 5 different source systems and databases, whereas the other data items have been already mapped to an official source according to architecture guidelines. However, data items are used in several systems and applications and in some cases conflicts with architectural principles can be identified. Therefore, the impact on performance in production environments and other possible implications in these cases have to be studied in order to assess the potential in simplifying the data landscape, consolidating the number of system interfaces and decreasing data replication.

As soon as all data items of the domains in scope are located only in the accepted by the architectural principles systems and databases, we can zoom in a specific business process and track the path of data between sources and destinations, and create detailed workflow models of the information paths of the organization. Focus on interaction of data with core business processes and how these can be supported in the more efficient way is the goal of visualizing the information workflows. An example of this that already existed in the company is demonstrated in Appendix D (Confidential). The information workflows can be further enhanced with the specified metadata that capture all intermediate transformations and paths, providing a way to track flow of data and information across from source to destination and demonstrate audit trail for compliance.

5.11. Manage Compliance & Security

The frequency and limitations of new regulations for financial institutions increase at a fast pace, as new demands from regulators and investors arise regarding transparency of investment operations (CapGemini, 2012). International Financial Reporting Standards 8 (IFRS 8) for publicly traded mutual and private equity funds, Basel III and Solvency regarding capital requirements and new regulations for trading of derivatives increase pressure to financial organization to incorporate the necessary controls in their data flows and data manipulation.

Regulatory compliance and information security are a crucial aspect and potential value area for financial services organizations that desire to deploy a DG program (Thomas G. , 2006; Soares, 2010), and this was also depicted in the input provided by the stakeholders within Company regarding their priorities. As a result, it is important to align objectives of the program with key stakeholders from this area, such the Information Security and Compliance & Risk management functions.

Based on their input and active engagement in the program, the existing spreadsheet policy can be expanded to include all data and information related assets (defined as artifacts), an activity that is already in progress within the organization. Furthermore, formal, standardized policies need to be developed and enforced in 3 dimensions (Soares, 2010):

- *Data privacy*: only authorized users and applications can have access to data identified as sensitive or private.
- *Security*: Ensure rights for access and changes in organizational DBs are only assigned to privileged users and are performed in sync with the established monitor and control procedures. Implement protection layers for external attacks and track all changes in values, database structures, security and access control objects and database configuration files.
- C*ompliance*: Establish automated workflows for internal and external reporting and audit. Ensure data and all information in risk reporting, investment decisions, trades, accounting procedures and others comply with all regulatory standards

5.12. Manage Data Quality

Data quality is defined as the domain that ensures data are fit for purpose for the intended use in operations, decision-making and planning (Lucas, 2010). Moreover, research indicates that poor data quality is to blame in around 40% of business projects and initiatives that cannot achieve the expected value (Gartner Research, 2007). Consequently, one of the first priorities for an organization that desired to increase trust in the available data is to develop a DQM policy based on data items and attributes with the highest business value (Soares, 2010). For example, a bank has as a business KPI regarding total risk exposure by industry. Standard Industry Classification (SIC) codes are used to calculate this KPI, so SIC codes are identified as a high business value attribute and a prioritized focus point for the DQM policy.

Existing data quality metrics have to be enriched with new dimensions so that the full spectrum of potential data issues is covered and controlled. The list below is an indication of data quality metrics that could offer valuable insights for the assets of a data-intensive organization as the one of this case study. Motivation for DQ metrics definition and corresponding adaptation to our case from (Friedman, 2007):

- *Data completeness*: % of critical data items with existent values in the specified mandatory fields.
- *Data uniqueness*: % of critical data items with unique identifiers established and duplicates removed.
- *Data validity*: % of critical data items that values are in an acceptable format (e.g. Employee_ID: 3 letters & 4 digits).
- *Data accuracy*: % of critical data items that reflect properties of the attribute they model (e.g. no negative values allowed, range_low ≤ X ≤ range_high).
- *Data consistency*: % of critical data items with consistent and agreed upon definitions, values and associated calculation formulas in all locations, systems and applications.
- *Data Integrity*: % of critical data items with intact expected relationships between multiple table, files and repositories.
- *Data availability*: % of necessary data items for a specific business process that are available to the user.
- *Data timeliness*: % of necessary data items for a specific business process that become available at the predefined delivery intervals.
- *Data security*: % of critical data items that are protected from unauthorized access via the corresponding standardized policy and process.

5.13. Manage Information Lifecycle

Information Lifecycle Management refers to a structured, policy-based approach in respect of data and information classification, collection, usage, archival and deletion (Soares, 2010). Proper management of information across its lifecycle can have a positive impact on storage and infrastructure costs that increase as records volume explodes and several DBs are duplicated for testing and backup purposes, improve the performance in production environments and limit risks from non-compliance with regulations. The DG council in cooperation with professionals from Content Management should develop specific rules and guidelines in the following direction:

- Data and information classification in structured and unstructured content.
- Archive, retention and deletion policy with different levels depending on the characteristics of the content managed.

5.14. Measure Results & Review

At the end of the selected timeframe that bounds the first iteration of the program but also at specific intermediate milestones, results need to be measured against the defined metrics for success (Soares, 2010; Thomas G. , 2006; Ladley, 2012). This way, progress can be evaluated based on the impact for value creation in the areas of risk management, operational efficiency and business growth, and benefits can be effectively communicated to stakeholders all over the organization and of course the executive sponsors of the program. The above is vital for renewed and ongoing support, funding and commitment to the program, considering its iterative and continuous nature.

6. Conclusions

This chapter concludes the research by answering the formulated research questions, describes limitations and constraints that occurred during the research process and proposes areas of future research.

6.1. Research Question 1

"Which Data Governance framework is most suitable for a financial services organization such as Company?"

Based on feedback from several data stakeholders and for an organization that shares similarities with Company in respect of the industry and business model, size, organizational structure, decision making culture and information environment, a DG framework must be *Practical & Implementable, Focused on Compliance & Security* aspects and *Simple* to comprehend. Moreover, it should define *Clear Roles & Responsibilities,* be *Aligned with Business Objectives,* should be *Flexible,* integrated in the *Organizational Structure & Decision Making Culture* and *Focused on BI & Data Analytics* projects support (*Section 4.4)*. Evaluation of literature frameworks concluded that no single Data Governance framework can satisfy the full set of criteria that occur from a specific organizational environment and data landscape. Even though a number of DG frameworks perform strong in specific criteria, all of them demonstrate vulnerabilities in others *(Section 4.5).*

Therefore, a novel DG framework is synthesized *(Section 4.7)*, that is *Simple* and *Flexible* by consolidating overlaps in capabilities to include only 9 components and providing a comprehensive structure of DG in 3 dimensions: *People & Organization, Standards & Policies* and *Processes* that allows enrichment with new components to address emerging requirements. Also, it is *Practical & Implementable* as it provides specific success metrics that can be further enhanced in order to create quantified methods of measuring progress of DG and it is focused on how core business processes are optimally supported by organizational data assets. Furthermore, a specific implementation process is suggested *(Chapter 5)*.

Focus on Compliance & Security is ensured by focusing initial resources on development of relevant policies and authorization mechanisms, while the framework is *Aligned with Business Objectives* that in Company's business model can be categorized in improved risk management, business agility and operational efficiency. *Clear Roles & Responsibilities* are formalized and adjusted to the *Organizational Structure & Decision Making Culture* for optimal fit, whereas *Focus on BI & Data Analytics* is supported by concentrating efforts on rationalizing the data architecture and improving data quality to support relevant projects. However, scientific validation of the above conclusions in Company's environment has not yet been performed.

6.2. Research Question 2

"How the selected Data Governance framework can be deployed within the organization to effectively manage responsibility and accountability of data sources and created reports and dashboards?"

An iterative implementation process should be established in order to develop DG capabilities contained in the synthesized framework *(Chapter 5)*. Initially, *Awareness* regarding DG concepts and value should be increased with the actions described, related to data resources *Roles & Responsibilities* have to be assigned within the shaped DG organization and *Monitoring & Measurement* processes for progress of DG need to be established based on the defined success metrics. *Data Architecture* of the organization should be rationalized to effectively support *Analytics* projects and provide a clear view of the data landscape in order to visualize *Information Workflows.*

The above can be the foundation for development of policies and standards regarding *Compliance & Security Management* as a core focus area of the program, as well as other significant components such as *Data Quality Management* and *Information Lifecycle Management.* In every iteration of the program, progress of the DG program have to be measured against defined business objectives and success metrics (*Value Creation)* and reviewed so that the necessary adjustments that will help DG to evolve and mature in the organization can be made.

In order to effectively establish responsibilities and accountabilities in respect of organizational data resources, the DG organization has to be configured based on characteristics of the application environment *(Sections 3.6.2, 4.1, 5.5).* These characteristics can also be correlated with the DG-related criteria provided as input from the stakeholders (Section 4.4), providing valuable insights in terms of which key data stakeholders should compose the DG organization as well as the exact way responsibilities should be assigned with the corresponding DG RACI matrix.

6.3. Research Question 3

"*How can we define mechanisms that safeguard data quality and usage and provide a way to enable next-gen, self-service data analytics for end-users, while preserving suitable levels of control from the IT perspective?"*

Data Governance is vital to effectively measure and improve data quality in an organization, by appropriate distribution of data-related decision making responsibilities within both business and IT functions and development of relevant standards, policies and processes (*Chapter 3, Section 5.12)*. Furthermore, several contribution areas of DG can be identified in SS-BI environments, such as easier to access data sources, cross-functional sharing of information and BI artifacts (e.g. queries, reporting templates etc.), standardized data loading and manipulation practices and increased transparency in data operations that also have a positive impact on compliance and security aspects *(Section 1.5)*.

Data Stewardship is a quality control discipline and a core domain of Data Governance. DS represents formalization of responsibilities regarding low level data-related decision making such as common data definitions and models, metadata specification and so on that promote business value exploitation of organizational data assets and mitigation of associated risks *(Section 3.4)*. DS can be best facilitated by close business & IT collaboration and cross-functional sharing of best practices in day-to-day data management, and an organization needs to find an optimal way to align data stewards with its structure, culture and business model. Two-fold DS (Business and Technical) can ensure appropriate data usage in terms of both business requirements and technical considerations *(Section 5.6)*. User categories demands need to be identified and data should be positioned in governed sources, while role-based access and authorization mechanisms have to be established *(Section 5.9)*.

6.4. Research Question 4

"How can we visualize the interaction of data items and entities with the organization's critical business processes, from the source systems to the end-user application layer?"

Initially, the most critical for the business data domains and items should be identified and organized in data dictionaries with enterprise wide definitions, calculation formulas and so on to ensure data validity, integrity and consistency. Considering the complex, distributed IT and data architecture of a financial services organization that needs to support scalability and performance across a variety of systems and applications, rationalization of the data landscape can have a positive impact on how core business processes are supported by information workflows. Metadata specification is also crucial for this purpose, as they can provide activity records regarding when, in which system and from whom the data and other artifacts such as data models or the corresponding ETL processes are created or changed, what is their intended use as well as what is the impact from possible changes to these artifacts *(Section 5.8)*.

As soon as all data items of the domains in scope are located only in the accepted by the architectural principles systems and databases, we can zoom in a specific business process and track the path of data between sources and destinations in order to visualize information paths of the organization in detail. The information workflows shown can be further enhanced with the specified metadata that capture all intermediate transformations and paths, providing a way to track flow of data and information from source to destination layer and demonstrate audit trail for compliance *(Section 5.10)*.

6.5. Future Research

Several limitations and constraints encountered in this research have to be enlisted and explored in respect of future research. Moreover, areas that motivation for future research occurred naturally during this project are also discussed in this section.

6.5.1. Constraints & Limitations in current research

Generalizability of case study

The case study was performed in one organization (Section 1.6.1), therefore generalization of the conclusions should not be attempted unless the environment shares similarities with the one of this research. Moreover, criteria development and their weights were based upon input from a limited number of stakeholders, thus the sample is not statistically significant (Sections 2.2.1, 4.4.). It would be interesting to see how the criteria list and corresponding weights would be shaped from a larger stakeholder group, or even from stakeholders in multiple organizations of the financial services sector.

The research could be also enhanced by creating a structured methodology to identify priorities for DG design within the targeted organizational and informational setting through the usage of questionnaires, creating a specific framework for the nature and range of answers from the stakeholders. This way, the criteria list could be more accurate and representative of the environment's characteristics and corresponding requirements, as currently it is largely based on intuitive responses and ideas of the stakeholders and interpretation of the researcher.

Subjective Bias

Bias of the researcher due to the exposure in the environment of the case study has to be taken into account in the research conclusions reached based on observations, as well as subjective bias in the evaluation of the frameworks as it was performed by only one person (Section 4.5). Also, bias can be identified in the conceptual connection of criteria to the DG capabilities as well, as it is a process largely dependent on subjective interpretation (Section 4.7).

Other Limitations

Time limitations also need to be considered as the research lasted 5 months in total, an extended timeframe might contributed in improvement of the conclusions validity.

6.5.2. Formalized Roles & Responsibilities and Data Governance success

Even though formalized roles and responsibilities for Data Governance are considered a critical success factor from many researchers and practitioners, further research on the topic is strongly suggested. Based on observations during this case study, indications that effectiveness of data management activities and sense of responsibility does not lie strictly in formalization, but in the motivation of people for change and their subjective interpretation of the situation. In other words, data stewards and other Data Governance roles might be equally effective and efficient without formal DG structures and bodies, as improved governance of enterprise assets will have a positive impact on the amount and nature of data-related work they have to do in current situation.

6.5.3. Quantified Data Governance and Measuring Progress

Organizations in general and financial institutions specifically struggle with the implementation of Data Governance and according to surveys, the most important barriers reported refer to difficulty to specify benefits, costs and risks from information-related projects as well as limited awareness for the value of Data Governance within the organization. Scientific research and best practices lack a systematic, quantified methodology to measure Data Governance capabilities that would provide a way to monitor and control progress and success of relevant projects and initiatives in a comprehensive, benchmark-style manner. This way, connection to the business value would be easier to be made and Data Governance concepts could be communicated in a clearer, more direct manner to audience from different backgrounds and organizational hierarchies.

6.5.4. Agile and Lean Data Governance

Considering the iterative and incremental approach for Data Governance, as the "boiling the ocean" method is strongly opposed by researchers and practitioners, it would be interesting to study the interconnection of Agile and Lean Six Sigma management methodologies and practices with Data Governance domains. Furthermore, operational efficiency is a core area of Data Governance and parallelization with Lean principles can be found in all data-related areas that *Waste* can be identified such as duplication of data, efforts and resources, error-prone manual tasks, inefficient data vendor management and others.

6.6. Final Considerations

Data is increasingly considered a valuable asset for modern organizations and needs to be governed appropriately. This research thesis studied Data Governance design in a financial services organization using a case study methodology. Several DG frameworks were evaluated against criteria that stem out of Company's organizational and informational setting, reaching the conclusion that no single framework can address the full set of requirements. Therefore, an approach to design a novel DG framework was proposed by mapping the derived criteria to DG capabilities that can be found in literature frameworks but were adapted to optimally fit the targeted environment.

An iterative implementation plan to deploy the synthesized framework within the organization was also described, focusing on the configuration of the DG organization as well as the priorities of the program in respect of Company's organizational and data-related characteristics. Effective Data Governance and two-fold Data Stewardship is vital to support emerging business requirements in SS-BI environments, while ensuring the necessary control mechanisms from the IT perspective. Moreover, it can help an organization rationalize its data landscape and have a clear view of the way data and information flow through several systems and repositories to support core business processes in the best possible manner.

Bibliography

Abello, A., Darmont, J., Etcheverry, L., Golfarelli, M., Mazon, J.-N., Naumman, F., . . . Vossen, G. (2013). Fusion Cubes: Towards Self-Service Business Intelligence. *International Journal of Data Warehousing and Mining, 9(2), April-June*, 66-88.

Ajilitee. (2012). *The Ultimate Guide to Data Governance Metrics: Healthcare Edition.*

Aranow, M. (2014). *Data Governance is Key to UPMC's $100 Million Investment.* The Advisory Board Company.

Arizona, Department of Education. (2014). *Data Governance Maturity Assessment.*

Beijer, P. (2009). Meaningfulness in Information Governance. *Primavera Working Paper*.

Benbasat, I., Goldstein , D., & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *Management of Information Systems*.

Bitterer, A., & Newman, D. (2007). *Organizing for Data Quality.* Gartner, Inc.

Blosser, D., & Haines, P. (2013). Data Governance at Chevron GOM: A case study. *PNEC 17th International Conference on Petroleum Data Integration, Data and Information Management.*

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? *Social Science Research Network*.

CapGemini. (2012). A Case for Enterprise Data Management in Capital Markets.

Chen, W. (2010). *Kalido Data Governance Maturity Model.*

Cheong, L. K., & Chang, V. (2007). The Need for Data Governance: A Case Study. *18th Australasian Conference on Information System.* Toowoomba.

CMMI Architecture Team. (2007). *Introduction to the Architecture of the CMMI Framework.*

Cohen, R. (2006). BI Strategy: What's in a Name? Data Governance Roles, Responsibilities and Results Factors. *DM Review*.

Cupoli, P., Earley , S., & Henderson, D. (2014). *DAMA-DMBOK2 Framework.* The Data Management Association.

Deloitte Advisory & Consulting. (2014). *Banking and asset management players are increasingly considering electronic data management to be a strategic activity requiring operational efficiency.*

Dyche, J., & Polsky, A. (2014). 5 Models for Data Stewardship. SAS Best Practices.

Eckerson, W. (2009). *TDWI Checklist Report: Self Service BI.* TDWI Reseach.

Faria, F. d., Macada, A. C., & Kumar, K. (2013). Information Governance in the Banking Industry. *46th Hawaii Conference on System Sciences.*

Finish Ministry of Finance. (2016, April 12). Why Crowdfunding Act is Needed? Retrieved from http://vm.fi/en/article/-/asset_publisher/miksi-tarvitaan-joukkorahoituslaki-

Friedman, T. (2007). *Best Practices for Data Stewardship.* Gartner Inc.

Gartner Research. (2007). *"Dirty Data"is a Business Problem, not an IT Problem.*

Harvard Business Review. (2014). *The Digital Transformation of Business.* Microsoft Corporation.

Hoffman, C., & Rodriguez, M. M. (2013). Digitizing Financial Reports – Issues and Insights: A Viewpoint. *The International Journal of Digital Accounting Research, Vol.13*, 73-98.

Horwitt, E. (2011). *Self-service BI catches on.* ComputerWorld.

Hoying, M. (2011). *Data Governance at Stanford: A Data Governance Overview.* Stanford University.

IBM Global Technology Services. (2014). *Building the business case for continuity and resiliency.* IBM Corporation.

IBM Software Group. (2007). *The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance.*

Imhoff, C., & White, C. (2011). *Self-Service Business Intelligence: Empowering users to generate insights.* TDWI Research.

Informatica. (2006). *Data Governance: Why it matters to the business.* Informatica White Paper.

Information Builders. (2014). *Driving Better Business Performance with a Practical Data Strategy.*

Information Governance Initiative. (2015). *Information Governance Initiative: Annual Report 2015-2016.*

Intel IT. (2013). *Delivering Self-Service BI, Data Visualization, and Big Data Analytics.* Intel Corporation.

Intel IT. (2015). *How Intel's IT Integrated Analytics Platform Helps Sales and Marketing.*

Karel, R. (2007). *Data Governance: What Works and What Doesn't.* Forrester Research.

Khatri, V., & Brown, C. (2010). Designing Data Governance. *Communications of the ACM*.

Ladley, J. (2012). *Data Governance: How to design, deploy and sustain an effective Data Governance program.* Morgan Kaufmann.

Lake, P., & Crowther, P. (2013). Data, an Organisational Asset. In *Concise Guide to Databases: A Practical Introduction.* Springer.

Loebbecke, C., & Picot, A. (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *Journal of Strategic Information Systems*, 149-157.

Logan, D. (2012). *Hype Cycle for Enterprise Information Management.* Gartner Inc.

Lucas, A. (2010). Corporate Data Quality Management: From theory to practice. *5e Confrence ibérique des systemes de gestion de l'information.*

Marco, D. (2014). Data Governance & Stewardship. EWSolutions.

Margetts, H., & Dunleavy, P. (2013). The second wave of digital-era governance: a quasi-paradigm for government on theWeb. *Philosophical Transactions of the Royal Society A 371: 20120382*.

Matt, C., Hess, T., & Benlian, A. (2015). Digital Transformation Strategies. *Bus Inf Syst Eng 57(5)*, 339-343.

McKinsey Center for Business Technology. (2012). Are you ready for the era of 'big data'? *Perspectives on Digital Business*. McKinsey & Company.

Messerschmidt, M., & Stuben, J. (2011). *Hidden Treasure: A global study on master data management.* PricewaterhouseCoopers(PwC).

Newman, D., & Logan, D. (2006). Governance Is an Essential Building Block for Enterprise Information System. *Gartner Research*.

O'Neill, M. (2015). The Impact of Mobile, Social and Big Data on Data Governance. (TechTarget, Interviewer)

Oracle Financial Services. (2015). Banking is Changing.. with or without the banks: Response to the millennials digital expectations. Oracle Corporation.

Otto, B. (2011). A morphology of the organisation of data governance. *ECIS.*

Otto, B. (2011). Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers. *Communications of the Association for Information Systems: Vol. 29, Article 3*.

Panian, Z. (2010). Some Practical Experiences in Data Governance. *World Academy of Science, Engineering and Technology*.

Parenteau, J., Sallam, R., & Howson, C. (2015). Modern BI and its Impact on the 2016 BI & Analytics Platform Magic Quadrant. Gartner, Inc.

Recall. (2015). *What is the ROI of an Information Governance Initiative.*

Rose, S., Spinks, N., & Canhoto, A. I. (2015). *Management Research: Applying the Principles.* Routledge.

Russom, P. (2006). *Taking Data Quality to the Enterprise through Data Governance.* The Data Warehousing Initiative (TDWI).

Salido, J., & Voon, P. (2010). *A Guide to Data Governance for Privacy, Confidentiality and Compliance.* Microsoft Corporation.

SecureWorks. (2016). *Determining the True Costs of a Data Breach.*

Seiner, R. (2014). *Non-Invasive Data Governance.* Technics Publications.

Sentance, D., Caporuscio, V., & Faraone, D. (2014). *Data Governance.* Pittsburgh: PwC.

Soares, S. (2010). *The IBM Data Governance Unified Process.* MC Press.

TechTarget. (2016). *Pair BI and Analytics With Data Governance.* SearchDataManagement.com.

The Economist Intelligence Unit. (2008). *The Future of Enterprise Information Governance.* The Economist.

The Economist Intelligence Unit. (2015). *Helping Asset Managers Navigate the Data Sea.* Northern Trust.

Thomas, G. (2006). *Alpha Males and Data Disaster: The case for Data Governance.* Brass Cannon Press.

Thomas, G. (2006). *The DGI Data Governance Framework.* Data Governance Institute.

Traulsen, S., & Tröbs, M. (2011). Implementing Data Governance within a Financial Institution. *Jahrestagung der Gesellschaft für Informatik.* Berlin.

van der Meer, L. (2015). *Measuring Data Governance: A structured method to assess data governance maturity.* Leiden: Leiden Institute of Advanced Computer Science.

Ventana Research. (2008). *Selecting the right XBRL Solution.* Clarity Systems.

WBR & Informatica. (2015). Transforming Financial Institutions through Data Governance. *Financial Information Management.*

Weber, K., Cheong, L. K., Otto, B., & Chang, V. (2008). Organising Accountabilities for Data Quality Management– A Data Governance Case Study. *Conference DW2008: Synergies through Integration and Information Logistics*.

Weber, K., Otto, B., & Osterle, H. (2009). One Size Does Not Fit All- A Contigency Approach To Data Governance. *ACM J. Data Inform. Quality 1, 1, Article 4*.

Weill, P., & Ross, J. W. (2004). IT Governance: How top performers manage IT decision rights for superior results. *Harvard Business Press*.

Wende, K. (2007). A Model for Data Governance – Organising Accountabilities for Data Quality Management. *ACIS*.

Wende, K., & Otto , B. (2007). A Contigency Approach to Data Governance. *12th International Conference on Information Quality (ICIQ-07).* Cambridge, MA.

White, A., & Friedman, T. (2016). Toolkit: How to Classify Information Assets to Be Governed in Applications. Gartner Inc.

Zainal, Z. (2007). Case study as a research method. *Jurnal Kemanusiaan*.

Acronyms

ABR: Accounting Book of Records

BI: Business Intelligence

BICC: Business Intelligence Competency Center

CRM: Customer Relationship Management

DB: Database

DG: Data Governance

DM: Data Management

DS: Data Stewardship

DSec: Data Security

DQ: Data Quality

DQM: Data Quality Management

ERP: Enterprise Resource Planning

IBR: Investment Book of Records

IS: Information Systems

IT: Information Technology

MD: Master Data

MDM: Master Data Management

PMS: Portfolio Management System

SS-BI: Self-Service Business Intelligence

QR: Quantitative Ranking

Appendices

A. Criteria Weights Questionnaire

This document contains analytical explanation and description for each one of the selected criteria that are going to be used for the evaluation of Data Governance frameworks found in relevant scientific literature and organizational best practices. The intention is to "survey" the opinion of major data stakeholders within Company regarding the importance of the criteria in the context of implementing DG in the organization. After reading the provided descriptions so that you understand what each criterion means for this specific research project and the evaluation process, you are kindly requested to fill in the empty column on the right end of the table with a percentage (%), depending on how important (or not) in your opinion this specific criterion is for deploying a DG program in the company. In case you believe a particular criterion is missing from the evaluation table, you can add a new row to the table with the criterion's name and weight of importance as well as a brief description in text of why this criterion is important to you so that we can investigate if it will be included in the final list.

General DG criteria*:* Factors that are considered vital for any Data Governance program regardless of the industry, other external parameters as well as internal characteristics of the organization.

Alignment with Business Objectives: Specification of DG program's values, vision and mission. To what extent the frameworks under evaluation help in alignment of DG with the business strategy and objectives of an organization so that is easier to show the potential business value and ensure the necessary levels of active participation from business? Essential for justification of the DG program to ensure appropriate resources allocation (people, time, effort) and funding when necessary.

Clear Roles & Responsibilities Formalization: Are the various roles & responsibilities for the DG virtual organization and data-related decisions formalized and described adequately in the frameworks? Do the frameworks contain suggestions on the structure of the DG organization, who would be more suitable to participate in each committee or role (e.g. DG Council, Data Steward etc.) and how responsibilities can be distributed to the created roles & organizational bodies?

Organizational Structure & Decision-Making Culture: Do the DG frameworks take into account the organizational structure and decision-making style of a company in their suggestions for designing DG? If data-related decisions are made by top management or lower tier employees make a difference on the configuration of the DG organization and the way the DG roles and responsibilities will be distributed. Also decision-making culture of the organization (consensus-oriented or hierarchical) will impact the distribution of decision rights for data-related decision domains.

Company Parameters: Factors that help us evaluate the different DG frameworks in respect of Company's principles, culture, working style and objectives.

Simplicity: How simple is the framework in terms of number of components and interrelations between them, how easy is for non-expert audience from different backgrounds, functions and teams in Company to comprehend and communicate the framework's concepts. This is important as the DG framework will be used as an artifact of communicating what DG means for Company throughout the organization.

Practicality & Implementability: Practical value of the framework in Company's organizational and informational landscape. Can data stakeholders and employees of various teams, business units and organizational hierarchies effectively relate concepts and components of the framework with the data environment and working culture of the organization?

Flexibility: How extendable is the framework to include DG-related requirements and priorities that emerge in the future? How easy it is to add or remove a component from the framework and how much existing components and their interdependencies will be impacted? Also, does the framework take into account possible extension of the scope for DG? (e.g. start with a focus on QlikSense user categories and related data sources & requirements, then expand to include additional user categories and data items).

Business Intelligence/ Data Analytics Focus: How important are BI/Data Analytics aspects for Data Governance according to the frameworks under evaluation? To what extent the frameworks contain information about principles, guidelines and techniques that will foster information sharing and support accurate and fast reporting and decision-making operations that are based on effective organization and utilization of the available data sources? At a lower level, this criterion will be used to evaluate DG frameworks against data-related attributes such as data access & availability that are vital for effective data analysis and decision making, as well as data consistency & integrity that are critical for reliable reporting considering the various source systems involved.

Regulatory Compliance & Security Focus: How important are regulatory compliance & security aspects of data for the DG frameworks under evaluation? As the available DG frameworks have a generic nature, they don't always contain appropriate levels of information on compliance & security aspects that are crucial for financial organizations that desire to deploy DG programs. At a lower level, this criterion will be used to evaluate DG frameworks against data-related aspects such as data security as well as data auditability & traceability across the whole lifecycle from the creation to the consumption layer.

Table 22 Criteria for evaluation of DG frameworks and corresponding weights

	Criterion	***Weight (%)***
General DG criteria	*Alignment with Business Objectives*	
	Clear Roles & Responsibilities Formalization	
	Organizational Structure & Decision-Making Culture	
Company Parameters	*Simplicity*	
	Practicality & Implementability	
	Flexibility	
	Focus on Business Intelligence/ Data Analytics	
	Focus On Regulatory Compliance & Security	
	Total	100