NOVA IMS Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

How data governance frameworks can leverage data-driven decision making: a sustainable approach for data governance in organizations

Patrícia Isabel Penteado Roque Bento

Dissertation report presented as partial requirement for obtaining the Master's degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

HOW DATA GOVERNANCE FRAMEWORKS CAN LEVERAGE DATA-DRIVEN DECISION MAKING: A SUSTAINABLE APPROACH FOR DATA GOVERNANCE IN ORGANIZATIONS

by

Patrícia Isabel Penteado Roque Bento

Dissertation report presented as partial requirement for obtaining the Master's degree in Information Management, with specialization in Knowledge Management and Business Intelligence

Co-advisors: *Miguel de Castro Neto, Nadine Côrte-Real*

October 2021

ACKNOWLEDGEMENTS

I would like to express my gratitude to my NOVA IMS co-advisors, Professors *Miguel de Castro Neto* and *Nadine Côrte-Real*, to all the collaboration and guidance throughout this project, contributing to my growth in this master.

To my family, to have encouraged and giving me strenght, understanding the importance of this work and making it easier to complete my master with all their support.

To my friends, who have giving me important feedback and support, never doubting of my worth, even when I have. Guiding me through the right path and pushing me to be my best self to deliver this work.

Forever grateful.

ABSTRACT

With the technological advances, organizations have experienced an increasing volume and variety of data, as well as the need to explore it to stay competitive. Data governance importance emerges to support the data flow, to record and manage knowledge derived from data, as well as establishing roles, accountabilities, and strategies, which further results in better decision-making. Through the definition of strategies to manage data in a consistent manner, data governance establishes the path to an enterprise-wide standardization, providing unchallenging access, management, and analysis of data to derive useful insights. Research on data governance frameworks is limited and lacks a key perspective: how can firms ensure sustainability on their programs. Data governance programs can only be continuously valuable if supported by a holistic framework focused on sustainability. To understand this gap, five frameworks are presented, analyzed and evaluated according to an assessment matrix based on eleven critical success factors for data governance. As a result of this study, where we offer a more comprehensive assessment tool, both researchers and practitioners can understand the maturity level of each critical success factor in the reviewed frameworks and identify which areas need further exploration and how to accomplish higher data governance maturity levels.

KEYWORDS

Data governance framework; sustainability; value creation; decision-making; assessment matrix; information management strategy.

INDEX

FIGURES INDEX

TABLES INDEX

LIST OF ABBREVIATIONS AND ACRONYMS

BI	Business Intelligence
CDO	Chief Data Office
CMM	Capability Maturity Model
CMMI	Capability Maturity Model Integration
CSF	Critical Success Factor
DG	Data Governance
DGI	Data Governance Institute
DGO	Data Governance Office
DMM	Data Management Maturity Model
DQM	Data Quality Management
GDPR	General Data Protection Regulation
IMS	Information Management School
IS	Information Systems
IT	Information Technology
MDM	Master Data Management
MVOT	Multiple Versions of the Truth
ROI	Return of Investment
SVOT	Single Version of the Truth

1. INTRODUCTION

1.1. Background and Problem Identification

With technologies' evolution, and in the speed that people live today, there is an increasing demand from consumers to practical, fast, accessible, and personalized goods. As competition arises, enterprises are pressured to make good decisions and deliver value for its customers, with no space for mistakes, or held back innovation and growth. Companies struggle to keep up with this pace, constantly seeking to innovate and improve its reach. In this context, data rises as a new mean for competitive advantage for organizations that can actually leverage its value, with strategies as fact-based decision making, risk control, costs reduction, and deeper knowledge of its own business (Attard & Brennan, 2018). Leveraging data as an asset to obtain competitive advantage and respond to the customers increasing demanding is one of the objectives of data governance.

In the 90's decade, data has joined other company assets as a strength for enterprises (Al-Ruithe, Benkhelifa, & Hameed, 2019), such as financial and human (Smallwood, 2014), representing a competitive advantage for each company. Due to its highly subjective component, data has the potential to generate powerful and unique insights that lead to faster, easier, and better decision making. Nowadays, it represents one of the main assets in a company, providing the path to derive unique business decisions based on facts that are available and continuously updated within the company (Abraham, et al., 2019).

While working with data seems crucial to sustain market position, it has risen several challenges for organizations, switching from an easily accessible asset to a demanding, high-risk one. The advance of technologies has also enabled the collection of data in various formats, previously unexplored (Haneem, Kama, Taskin, Pauleen, & Abu Bakar, 2019), which has led to disparate data across the organizations. This can have negative consequences on decision-making since the decisions may be made from incorrect data (Abraham, Schneider, & vom Brocke, 2019), and data problems can often represent fundamentally business problems (Lee, Madnick, Wang, Wang, & Zhang, 2014). Although data is accessible in every organization, it may not be leveraged in the right manner, and exploring and dealing with that data can have some bottlenecks, including costs, risks, and liability (Smallwood, 2014).

Nevertheless, it represents a leading differentiator to companies that use data as an asset (Attard & Brennan, 2018) and can actually retrieve value from it, aiding to achieve its goals, improve financial performance (Attard & Brennan, 2018), and increase the sustainability of competitive advantage (Smallwood, 2014). The value of data is not visible only on the enterprise level; in fact, the treatment of data enables organizations to improve its decisions, benefiting society (Brous, Janssen, & Vilminko-Heikkinen, 2016). On the other hand, the lack of understanding data's worth in an organization can bring difficulties and even failure (Alhassan, Sammon, & Daly, 2018).

Paramount to every organization in the current century, in many economic sectors, data represents the only form of economic growth and the "central nervous system of an organization" (Ladley, 2020). According to Bhansali (2014), the growth of an organization has relied not only on expanding its business world-wide but also by leveraging intrinsic knowledge. Moreover, it is stated that an appropriate data strategy is necessary for the success of those organizations (Bhansali, 2014).

In this context, data governance has created the path to facilitate companies to extract real value from data, without compromising business operations. In the simplest manner, it has been paramount in a company's structure to ensure the data's safety, quality, and consistency, establishing strategies that go from the extraction of the data to the stakeholders that make the decisions on that data.

To ensure reliable decisions, many companies are limiting their data governance, by exercising control over the quality of their data and compliance with regulations (Janssen, Brous, Estevez, Barbosa, & Janowski, 2020). Although data quality is fundamental in a data governance initiative (Alhassan et al., 2018), it relies heavily on the value of that data, which is context-driven and has yet no consensual definition (Attard & Brennan, 2018). For this reason, it is of great importance that the data governance program aligns with the goals (Al-Ruithe, Benkhelifa, & Hameed, 2016), needs and resources of the organization (Smallwood, 2014).

The first step to succeed from a data governance initiative is to be conscious that it implies a mindset shift and a transformation within the organization (Kiron & Ransbotham, 2017). According to Koltay (2016), data governance requires a multi-dimensional approach to be successful, from clearly defining its dimensions, that include processes, roles and accountabilities, stewardship, and change management. Implementing a data governance initiative and thriving from it is very challenging due to its ambiguity and abstraction, demanding a strong commitment and willing to face the difficulties and surpass the barriers that appear along this journey. However, organizations that do understand the power and potential of data have the opportunity to achieve excellence (Gupta & Cannon, 2020).

Although more companies are now becoming aware of the importance of data governance and investing in data governance solutions (Fu, Wojak, Neagu, Ridley, & Kim, 2011), the lack of interest in this subject leads to an erroneous approach in what regards data governance processes, focusing the investments in technology and legal obligations, rather than on defining the necessary strategies and application steps. In fact, introducing technology without proper organization and preparation can lead to risk increase (Kim & Cho, 2018). Furthermore, there is limited research on sustainable data governance frameworks (Abueed & Aga, 2019), which minimizes the importance of the subject in the companies.

Accordingly, the problem identified in this study is the absence of interest from the companies in exploring data governance as a beneficial program to boost its decision-making capabilities. More specifically, when further analyzing the origins of this disinterest, it was found that there is a lack of sustainable data governance frameworks for the organizations (Chang & Cheong, 2007).

1.2. Study Objectives

Along with reinforcing the importance of data governance as an asset (Abraham, et al., 2019), this study will contribute to the literacy of data governance through a literature review on this topic, as well as a review of the existing data governance frameworks with an assessment tool, based on the dimensions considered fundamental in a data governance initiative, to critically analyze the solutions available and suggest improvement opportunities.

The aim of this study is to provide an assessment tool, that can deliver a holistic view on the available data governance frameworks and how they contribute to the maturity of data governance in organizations, based on the dimensions found in the literature and identified as critical success factors for data governance. For this matter, five data governance frameworks will be analyzed and evaluated according to this assessment tool.

Accordingly, the question to answer is: “What are the critical success factors of data governance, and how do data governance frameworks contribute to its maturity level in organizations?”

More specifically, the objectives for this study include:

- Identifying the critical success factors of data governance through the literature;
- And reviewing the existing frameworks and assessing its contribute to the dimensions identified.

These goals rise as a strategy to answer the question established, going beyond and presenting solutions to the problems encountered.

2. METHODOLOGY

In line with the objectives identified in the previous section, the methodology chosen for this study followed an exploratory research design with the development of a conceptual model – an assessment matrix for data governance frameworks.

A research design can be defined as a logical plan connecting the research questions to the data collection and, further, to the conclusions made on that data. Accordingly, it initiates with the study questions, which can aid in identifying the methodology to be followed. In this line, as mentioned previously, this study's research question is related with data governance critical success factors and its maturity levels in data governance frameworks. Another component of research design is to properly identify the unit of analysis, which, in this study, corresponds to five data governance frameworks.

To achieve the goals mentioned before and answer the established questions, a science research framework was used to conduct the steps in this study. The first approach regards a literature research to identify the gaps and problems with the topic, as well as identifying and critically analyzing the existing solutions. Further, with the theoretical basis established, an assessment tool was used to evaluate data governance frameworks according to the critical success factors of data governance. Finally, this work provides insights on the state of the data governance frameworks and the critical success factors, as well as suggestions of improvement opportunities for the frameworks.

The literature review is organized as follows. Firstly, the terms data and data governance were explored in order to find different authors' definitions and considerations on this topic. Next, different solutions on the literature were presented; then, the relation between data-driven and data governance was explored. Moreover, the necessary dimensions for a data governance framework to succeed, according to each author were compared and analyzed. Lastly, the maturity model that served as a basis for the assessment tool was presented.

The consecutive steps were the definition and design of the assessment matrix, based on the literature review, and its implementation to review five data governance frameworks according to the critical success factors, and based on five maturity levels. Lastly, insights on overall scores of each framework and each critical success factor were provided, as well as improvement opportunities.

3. LITERATURE REVIEW

The aim of this literature review is to assess the presence of data governance (section 3.2) and data governance frameworks (section 3.4) in the literature, as well as to provide guidance on the best practices to implement data governance initiatives (section 3.5), which will contribute to the development of the assessment matrix (section 3.6) for the available frameworks. To understand the importance and role of data governance in a data-driven organization, and how it can represent a competitive advantage, this literature review includes the identification of data-driven culture and the role of data governance in this matter (section 3.3). Logically, all these areas are supported by data and its position as an enterprise asset (section 3.1.).

3.1. DATA

Data can be defined as the representation of unprocessed and unrelated facts, constituting the raw material for information (Zins, 2007). Accordingly, it is considered one of the building blocks of Information Science (IS) and information itself (Zins, 2007). From an historical point of view, data have originated from distinct legacy transactional structures, being further recognized as a business by-product, with its value being assigned to processing transactions and applications (Al-Ruithe et al., 2019). This tendency shifted, and data began to be identified as a valuable asset, potentiating data-driven decision-making, influencing both operational and strategic decisions (Alhassan et al., 2018).

In order to make decisions based on data, it must be available for the organization to extract its value with processes that include storage, integration, interpretation, and analysis of this asset (Thompson, Ravindran, & Nicosia, 2015). These processes are enabled by the technological innovations that facilitate the creation of vast data repositories, where data can be stored and further taken advantage of. With these repositories and the increasing volume, variety and velocity that data is being generated and stored, however, concerns with security, protection, and quality of data are risen. In general, the more data an organization has, the more difficult it becomes to manage it, leading to data breaches, disparate data, and poor data quality. Moreover, the tendency of growing in data volume and disparate sources causes inconsistencies, paramount to be diagnosed and addressed to prevent decisions based on incorrect data (Abraham et al., 2019).

In the more recent years, there has been an increasing concern with data privacy and security, which has led to the creation of standards for handling personal data, forcing organizations with legal compliance to these standards, such as GDPR (General Data Protection Regulation). The aim of this specific regulation is to prevent the violation of fundamental rights through the misuse of personal data (European Commission, n.d.). Complying with these regulatory requirements force organizations to keep track of its data and analyze what, where, how, and why data is used (Abraham et al., 2019). Moreover, with the introduction of more self-service reporting and analytics, the need for an organization-wide understanding of data has been emphasized (Abraham et al., 2019).

On this topic, Yallop & Aliasghar (2020) have distinguished legal and ethical perspectives. Regarding the legal perspective, to comply with regulations on data collection, storage, usage and disclosure, organizations must adopt and follow appropriate processes. On the latter, this compliance is seen as insufficient to protect organizations from the discontent of customers in what regards the use of their personal information. This enlightens the need to go beyond legal compliance, by establishing standards and shields to protect the ethical risks that customers face, as well as the risks regarding security and reputation of the organization itself (Yallop & Aliasghar, 2020).

Data quality is also another challenge that organizations face when complying with legal and regulatory provisions. In fact, data quality is crucial in achieving many business requirements, such as unified and automated processes, effective reporting and integrated customer management, with the notion of a 360^{o} customer view (Otto, 2011b). However, organizations have difficulties in discovering, understanding and exploit the potential of data in creating value (Benfeldt, Persson, & Madsen, 2019).

3.2. DATA GOVERNANCE

Data Governance is an ongoing topic on the literature, both scientific and practical (Nielsen, 2017). However, its presence in the literature is not unified (Abraham et al., 2019); and Al-Ruithe et al. (2019) defend that it is still under researched and not widely practiced by organizations. Benfeldt, Persson, & Madsen (2020) state that the research on data governance has lacking attention on its introduction in complex organizational realities by practitioners, and that narrowing the focus on defining principles and assign accountabilities – without considering value, capabilities, collaboration, overview, practices, and politics –, can cause failure in even the more detailed data governance policies. In the same line, Nielsen (2017) expressed that theoretical methods and case studies dominate this field, in opposite of practice-oriented methods as design science, action research and experiments, and that the latter should be encouraged. In Nielsen's study, it is also identified the anchor between IT and data governance that is still very present, which, in conjunction to what was described before, leads to the lack of recognition of data governance as a general management discipline (Nielsen, 2017).

The practical component of data governance is, however, growing with the exploration of data governance benefits in new technologies and concepts, such as artificial intelligence (Janssen et al., 2020) and cloud-computing (Al-Ruithe et al., 2016, 2019). Although various articles focus on different industries, and data governance has been mainly explored and implemented in the corporate sector (Koltay, 2016), data as an asset is the common factor that states the necessity for a data governance initiative.

In the reviewed articles, it was found various definitions for data governance, as well as the description of this term as a working-term, as it has not yet been assigned a formal definition. Researchers define data governance in different manners, based on other definitions and adapting them to their own contexts (Abraham et al., 2019; Al-Ruithe et al., 2019; Mullon & Ngoepe, 2019). The Data Governance Institute (DGI), in a more practical approach, advises that organizations adapt their definition in accordance to its culture and environment, avoiding resistance from the stakeholders involved (Thomas, 2006).

Some studies inspire their definition of data governance in the definition of IT governance by Weill and Ross (2004) as "the decision rights and accountability framework for encouraging desirable behaviors in the use of IT" (e.g. Otto, 2011; Weber, Otto, & Österle, 2009). Khatri and Brown (2010), in a similar manner, define data governance as referring to "who holds the decision rights and is held accountable for an organization's decision-making about its data assets", focusing on the "*loccus of accountability of decision making*" (Khatri & Brown, 2010).

Data governance is often mentioned by researchers and practitioners as a means to improve data quality or related to data quality management (Nielsen, 2017; Otto, 2011c). Weber et. Al (2009) mentions data governance as designating roles and responsibilities for its decision areas, establishing organization-wide guidance and principles for data quality management, as well as guaranteeing compliance with both the organization's strategy and legal requirements (Weber et al., 2009). The connection between data governance and data quality management is a result of considering data as an asset – which is also identified as the main driver for data governance (Panian, 2010) –, whereby data governance refers to valuable data assets, being that this value is dependent on the quality of data, which is ensured by its management that is consequently governed by data governance (Otto, 2011c). Accordingly, it is stated that the lack of clarification of the roles and responsibilities among data stakeholders reduces the efficiency of data quality management and, hence, enlightens the need for a formal data governance model (Cheong & Chang, 2007).

Based on open coding to search for common characteristics of data governance's definitions found in the papers reviewed for their articles, in their study, Abraham et al. (2019) define data governance as a cross-functional framework for managing data as an asset, defining the decision rights and accountabilities for decision making, and formalizing data policies, standards and procedures, as well as monitoring compliance. According to the latter, it can be retrieved that data governance is considered cross-functional, highlighting the importance of a joint effort from all the organization; structured and formalized data management; identification of data as an asset; determination of the what, who, and how of decision-making; the need for consistent data policies, standards and procedures with the organization's strategy and promotion of desirable behavior; and, finally, the implementation of control to ensure that the latter are adopted (Abraham et al., 2019). This is considered a very complete definition and, thus, will be followed in this study.

According to Otto (2011a), the goals of data governance can be divided in formal goals, such as enable decision-making, increase operational efficiency, and support business and IS integration; and functional goals, as establish data stewardship and data life-cycle management, and implement data standards and metadata management (Otto, 2011a). In other papers is also considered "ensuring the quality and proper use of data, meeting compliance requirements, and helping utilize data to create public value" (Janssen et al., 2020), and maximize the value of data assets (Otto, 2011b).

As for its benefits, it enables company-wide collaboration, implements data management framework and formalization, determines the spectrum of who, what, and how of data-driven decision-making, establishing policies, standards, and procedures on data (Abraham et al., 2019). Moreover, by encouraging desirable behavior in treating data as an asset, it supports organizational goals (Benfeldt et al., 2020); and boosts data sharing, promoting effective data use (Ransbotham & Kiron, 2017). Another avail of data governance identified by practitioners is the capability to improve and maintain the quality and use of data (Alhassan et al., 2018), enabling the rationalization of information by organizations and achievement of a set of consistent and coherent views of the organization (Were & Moturi, 2017). Webber et al. (2009) identified benefits specifically for data quality management, but that can be extended to data governance, such as the implementation of organization-wide accountabilities, the definition of roles and responsibilities, the creation of guidelines and standards, and, finally, ensuring compliance with both legal requirements and organization's strategy (Weber et al., 2009).

Withal, data governance is very challenging. The most common mentioned challenges with data governance – specifically data quality management –, are data collection and sharing, where vast amounts of disparate data may be stored and integrated, with questionable quality, as well as the rise of data glitches due to environment changes. Data collection rises as well the misuse of sensitive data, which should be minimized, and, in the eventuality of its collection, data should be secured to prevent its violation (Janssen et al., 2020). Another challenge regards data stewardship and accountability, with the identification of who should play each role and bear which responsibilities. Moreover, in any organization, it is paramount that all personnel is fully committed and involved, and that the management and senior-level executive provide the necessary sponsorship (Al-Ruithe et al., 2019).

On the public sector, the identified challenges in conducting data governance include difficulty in perceiving value of data and data governance, enabling collaboration between different departments, fostering capabilities due to different data management maturity, and having track of data (Benfeldt et al., 2020). In addition, it can be challenging to acquire and possess the capabilities to build data governance principles (Benfeldt et al., 2020). Similarly, Attard and Brennan (2018) identified as challenging defining and measuring data value, modeling and optimizing data governance (Attard & Brennan, 2018).

In some papers, it is mention the lack of interest from organizations in conducting a proper treatment of data, rather concentrating their efforts on the results and use of technologies, neglecting the impact that data quality, management, and accountability, among others, can have on such results and, consequently, on the organizations' prosperity.

With data being considered an organization's asset, it is paramount that its worth is perceived and that it is treated in the correct manner, in order to be able to benefit from it. In this sense, a data governance initiative may be crucial in the running of an organization, aiding to determine data's worth through answering the questions: where, how, who, when, and why data is used/integrated (Alhassan et al., 2018). The importance of a data governance initiative is perceived and claimed by many authors, in both enterprises and governmental institutions (Abraham et al., 2019), as it can increase the effectiveness and success of data management through the definition of policies and procedures (Cheong & Chang, 2007), forming the groundwork for innovation processes with its data sharing and quality capabilities (Ransbotham & Kiron, 2017).

It is stated that it is important to identify the data paths within an organization, establishing a causal relationship between the different steps, in order to identify what caused a disruption in the final results and whose responsibility is it. Proper data governance can, and should, provide these paths that ultimately form the data's network within an organization (Janssen et al., 2020). In other perspective, a data governance initiative can also help overcome challenges that organizations face, such as compliance with regulations (Khatri & Brown, 2010; Otto, 2011b), unreliable and incomplete data (Kim & Cho, 2018), and disintegrated architectures and legacy systems (Benfeldt et al., 2019).

3.3. DATA-DRIVEN AND DATA GOVERNANCE

To be data-driven is to manage data as an asset (Ladley, 2020). This includes incorporating and identifying data as the raw material in the decision-making process, with emphasis in testing and experimentation (Berndtsson et al., 2018; Hannila et al., 2019). Authors as Berndtsson et al. (2018) and Watson (2016) focus their papers in the use of analytics and, particularly, its integration into business processes as one of the drivers for data-driven decision-making culture. In a data-driven organization, data motivate decision-making, problem solving and business activity (Watson, 2016). Since decisions are being made based on data, poor data quality can affect negatively these decisions, hence, it is paramount to monitor each stage of the data life cycle (Diván, 2017). Moreover, it can be necessary to implement changes in data management practices and organization structure, getting rid of data silos, and even govern data organization-widely, in order to transition for data-driven decision-making (Hannila et al., 2019). In this line, data-driven decision-making depend on data quality and data governance (Diván, 2017).

In the literature, the role of data governance within the data-driven domain is still under-researched and requires further analysis (Otto & Lis, 2020). However, with the emerging of digital strategies and new business models, data governance is recognized as imperative to manage data-driven business models sustainability, rising as essential to organizations in all sizes. Even with this recognition, organizations usually lack the proper tools for its implementation (Otto & Lis, 2020). Data governance, and specifically relational mechanisms as communication and mentoring, translates of high importance in a data-driven culture (Vial, 2020).

As a data-driven approach provides opportunities for organizations to increase organizational performance and competitive position (Berndtsson et al., 2018), and, since data governance enables and supports effective data-driven culture in organizations, it can be deduced that, when in presence of effective data governance and data-driven decision-making, the organizational performance and competitive market position increase.

3.4. Data Governance Frameworks

The increased interest in data governance raised both practitioners and researchers to invest their efforts in designing data governance frameworks that would facilitate organizations to implement a data governance initiative. Fu et. al (2011) stated that an effective data governance framework can bring benefits for organizations, such as defining a clear mission, attain clarity, enhance reliability on data, maintain scope and focus, and quantify success criteria.

In the literature, the presence of data governance frameworks is self-evident. From practitioners to researchers, the explicit need for a data governance program has risen the interest in designing a data governance framework that brings value to the organizations. These frameworks range from conceptual to industry-specific, such as health, big data or cloud-computing. Others also focus in specific disciplines or activities of data governance, such as data quality management and data sharing.

In their study, Alhassan et al. (2018) identified that the majority of publications that were reviewed are included under the “define” action, followed by “implement” and, lastly, by “monitor”. This enlightens the need for more research on data governance and data governance frameworks that are focused on the implementation and monitoring of these initiatives. In this line, Cheong & Chang (2007) identified business as the data governance initiative driver as it uses data to make decisions. Khatri and Brown (2010), however, advocate that designing data governance involves identifying fundamental decisions that are important for the organization and the human force that should be making them, deterring the quotidian decisions.

Some authors defend that data governance can be context-driven and propose that the framework should be adjustable according to the needs and goals of the organization. For instance, Weber et. al (2009) suggest that data governance encompasses parts of IT governance, and that the contingencies identified by a few authors in the literature regarding IT governance also affect data governance. In this sense, similarly to what is proposed in IT governance frameworks, the authors defend that there is no universal approach for data governance, proposing a flexible data governance model, with a contingency approach and a main focus on accountabilities, composed of roles, decision areas and responsibilities that form an assignment matrix. The identified contingency factors include performance and competitive strategy, market regulation, decision-making approach, organizational structure, and diversification breadth (Weber et al., 2009). Although this model establishes the path for organizations to design their own data governance program, it does not encompass the assessment, implementation and monitoring of a data governance initiative in order to constitute a long-term solution for the organization.

Similarly, Cheong and Chang (2007) present a scalable data governance structure based on different roles and responsibilities, such as Data Governance Council, Data Custodian, Data Steward, and User Groups, according to different organizational levels. This structured framework mitigates the risks of data management, including the engagement with business and IT at strategic, tactical and operational levels, ensuring that both these areas are informed and IT initiatives are aligned with the objectives set for data governance (Cheong & Chang, 2007). This model is useful in identifying the roles in each organizational level and their associated tasks but does not contribute in implementing data governance in the organization.

In a broader perspective, Khatri and Brown (2010) inspired their data governance framework in the IT governance framework proposed by Weill and Ross (2004), enriching it with a set of five decision domains and its "*loccus of accountability*". These decision domains are Data Principles, Data Quality, Metadata, Data Access, and Data Lifecycle, which, according to the authors, are interrelated in the following manner: data principles establish the prerequisites for the intended uses of data, setting the standards for data quality that introduce the way data is interpreted (metadata) and accessed (data access). Finally, data principles are operationalized based on decisions that define the manufacture, retention, and withdrawal of data (data lifecycle) (Khatri & Brown, 2010). The importance in defining decision domains lies in the need to assign the right responsibilities and roles (Alhassan et al., 2018). This framework provides means to assess and identify the decision domains when designing a data governance model but lacks the implementation component of a data governance initiative.

The above-mentioned frameworks, despite lacking some aspects to implement and monitor data governance initiatives, constitute the basis for more recent models in data governance. Several authors inspire their frameworks in these authors' researches. As an example, Alhassan et al. (2018) use the decision domains established by Khatri and Brown (2010) as the decision domains to be considered for data governance. In an attempt to fulfill these gaps, practitioners developed frameworks that were business-driven and with a stronger implementation aspect, where industry associations and software vendors played their part.

The Data Governance Institute presented a framework divided in three core levels: rules and rules of engagement, people and organizational bodies, and processes. The first level embeds the structural components, with the definition of the main aspects in the data governance program, such as its mission and vision – "why" –; focus areas – "what" – and its goals, success measures and funding plan; data rules; decision rights; accountabilities; and, finally, control mechanisms. On the people and organizational bodies level – "who" –, there are three main roles to be established: data stakeholders, Data Governance Office (DGO) and data stewards. Lastly, the processual level – "when" and "how" – includes all the processes in data governance, split in reactive, proactive and ongoing. It also includes the connection between data governance and management, with business and IT processes that touch data (Thomas, 2006). This framework encompasses the assessment, implementing and monitoring phases of a data governance initiative, enhancing the importance of clearly define and identify what are the problems to be solved, how to address them, and how success can be measured before taking the next step of designing and implementing the program. Filling the gaps identified above and establishing a path to assist organizations in this journey, the main challenge lies on the communication required to involve and motivate stakeholders in achieving their goals (Thomas, 2006). As complete as this framework presented by DGI may be, it can lead to discouragement of organizations as it does not align the powerful tools that technology can provide in dealing with data, leaving a gap in the implementation and monitoring phase.

SAS Data Governance Framework is an example of a contribution by software vendors. Designed to overcome data governance challenges – with the adequate depth, scope and adaptability –, this framework can be divided in three levels of focus: on the strategy level, Corporate Drivers include decision making, operational efficiencies and compliance, identifying the need for data governance; on the organizing framework level, Data Governance – includes program and guiding objectives, decision rights and decision-making bodies – and Methods – includes people, process and technology – refer to the development and monitoring of policies that drive Data Management outcomes – data quality, definition, architecture and security; the latter, along with Solutions – data integration, preparation, monitoring, visualization – and Data Stewardship – roles and tasks – refer to the tactical level, including the quotidian processes and technology to support those processes (SAS Institute Inc, 2018).

On the top of SAS Data Governance Framework, is represented the section Corporate Drivers that, by aligning the data governance activities with the organizations' drivers and strategies, contributes to increase the interest of the organization and measure the initiative's success, linking it with key business goals. As for the challenges identified with the organization's culture and recruitment limitations, Data Governance section includes planning tools to build a roadmap that aligns data governance with the culture of the organization and its resources. Data Stewardship section refers to the assignment of this role, as the keepers of data throughout the organization, with the authority to oversee data that represents the bridge between business and IT. To implement the policies created by data governance, the section Data Management represents the set of functions of both business and IT which, thus, should be based on a holistic design. The Methods section includes the three main areas of a data governance program: people, process and technology. Finally, the Solutions section describes the tools used to assist the implementation and automation of the data governance initiative (SAS Institute Inc, 2018).

The SAS Data Governance Framework, although providing a very complete solution, it can seem overwhelming when introduced to organizations. In this sense, it is advised that this implementation is done incrementally, respecting the timelines and priorities defined by the organization, demonstrating its value through "quick-wins". With a higher focus on implementing and monitoring phases, this framework can more easily capture the interest of the organization, constituting a sustainable solution when thinking long-term.

Some of the more recent frameworks for data governance presented in the literature are industry or context-specific, as mentioned, where some are inspired by former frameworks, as Panian (2010) (e.g., (Kim & Cho, 2018)), and Khatri & Brown (2010). In this sense, the gaps identified in the available frameworks are not being filled, instead, those are being narrowed to fit the needs for specific topics. Moreover, in the cases where those gaps disappear, the framework is so narrow that it is not extensible to other organizations and, consequently, not representing sustainable options. Moreover, frameworks from data governance specific areas, such as data quality or master data management, do not allow for a sustainable option in a data governance initiative as they lack fundamental aspects for it to be succeeded. In detail, highly centralized approaches, as master data management, can restrain data's flexibility and, hence, its strategical applicability (DalleMule & Davenport, 2017).

3.5. DATA GOVERNANCE SUCCESS FACTORS

In the literature it is found divergent considerations of what are the best practices in a data governance framework, in part due to its distinct nature. Frameworks focused on roles and accountability will, naturally, identify the best practices for designing a framework within this spectrum, and so forth.

In their study, Alhassan, Sammon, & Daly (2019) presented a set of seven Critical Success Factors (CSF) for data governance, as the areas of activity in data governance that require the most attention to have a successful data governance program. The first CSF regards 'employees data competencies', where the highlight is the capability and awareness of all people in the organization – from senior executives to entry-level – to handle data manners in which they are involved. In this line, it is suggested that certain actions are conducted, such as continuous training oriented for their involvement with data governance activities and increase employee's awareness to critical and sensible data. The second CSF is 'clear data processes and procedures', which, if not guaranteed, leads to lack of trusting in data. The suggested actions are to embed those processes and procedures into the system, and to reevaluate and update the current ones.

The third CSF regards 'flexible data tools and technologies' – software and hardware that affect data – , which has a "significant impact" on other CSFs, such as 'clear data processes and procedures'. Accordingly, it is recommended to have appropriate IT infrastructure and integrated data, where advanced technologies automate the validation of data, and testing procedures that still enable flexibility to future adjustments. The fourth CSF is 'standardized easy-to-follow data policies', which relates with both the implementation of data policies to increase trust from employees, and the clarity that these data policies are presented in. The latter is mentioned specifically to unify, simplify and renovate these policies to make sure it is both understandable and appealing to all employees. In this case, the tools and technologies should also be implemented to monitor and update data policies.

The fifth CSF regards the 'established data roles and responsibilities' – highly mentioned in the literature –, emphasizing the unclarity on what are the roles and what responsibilities each role has that the employees face. To surpass this problem, establishing a committee and identify data owners can be crucial. The penultimate CSF is 'clear inclusive data requirements' – such as data regulations –, which includes its understanding by the business owner and its clear explanation to IT, avoiding inconsistency between what is presented – such as reports – and what is required for the business owner – for its analysis. Lastly, 'focused and tangible data strategies' involves the importance of considering data as an asset and the related activities, and the planning for data governance, with an orientation for achieving short-term and long-term objectives. The advised action here is to establish an executive team to encourage organization-wide structure for data governance.

Although the mentioned study is based on a single specific case study in the banking sector, the authors compare their findings with the existing literature and relate them with the five decision domains reported by Khatri & Brown (2010), establishing comparisons with other papers that allow for the extent of these findings to other industries and contexts. Notwithstanding, it is recognized by the authors that these specific topics and, in particular, actions to be taken, should be further investigated and valued in the literature. It is notable that tools and technologies related to data are highly mention in both as a critical success factor for data governance and as a means to automate and standardize the other factors, giving it an important position in a data governance program.

In other perspective, Cave (2017) used the results of a case study in the health industry and identified five strategies for implementing data governance practices. In accordance to the study mentioned above, a 'committee to provide structured oversight' is suggested – as in the fifth CSF –, as well as 'effective and strategic communications' and 'compliance with regulations' – included in the last CSF –; stakeholders' training – first CSF –; and 'benchmarking and standardization' – as the fourth CSF.

Notwithstanding, Cave (2017) contributed with novel practices, such as 'obtaining stakeholder buy-in', which consists in involving and reaching agreements from all the stakeholders involved through debates and feedback. This allows for stakeholders to have a voice in the changes that are happening and learn from one another, contributing to increase its interest and commitment in data governance activities. Another addition to the practices found before is 'effective and strategic communications', stating that the communication should play an important role and be aligned with the organization goals. This communication is necessary organization-wide, as well as from external organizations (Cave, 2017).

The common ground between these authors, although from very different industries, illustrates the importance that Data Governance has across all industries and in different environments. Both used single case studies but, overall, the results are similar in terms of what is important for a data governance initiative to be successful, whether it is on banking or health sectors.

In a more general manner, in their systematic literature review, Al-Ruithe et al. (2019) identified 20 CSF in the literature, mentioned by the authors as the most important CSF for data governance. To the ones identified above, the follow can be added: 'assess data governance situation', and 'define the sustaining requirements'. The first CSF is assessing the current state of data governance in the organization before implementing a data governance initiative. The second CSF aims to ensure the continuity of data governance and constant improvement, and can also be classified as the sustainability needed within the data governance initiative implementation to provide for a long-term solution for the organization.

Critical success factors	*Description*	*Suggested actions*	*Reference*
Employees data competencies	Capability and awareness to handle data manners.	Continuous training oriented for the involvement in data governance activities. Increase employee's awareness to critical and sensible data.	Alhassan, Sammon, & Daly (2019)
Clear data processes and procedures	If not guaranteed, leads to lack of trusting in data.	Embed processes and procedures into the system, and to reevaluate and update the current ones.	Alhassan, Sammon, & Daly (2019)
Flexible data tools and technologies	Software and hardware that affect data.	Have appropriate IT infrastructure and integrated data, with automatation and testing.	Alhassan, Sammon, & Daly (2019)
Standardized easy-to-follow data policies	Implementation and clarity of data policies, unifying, simplifying and renovating these policies.	Tools and technologies should also be implemented to monitor and update data policies.	Alhassan, Sammon, & Daly (2019)
Established data roles and responsibilities	Emphasizes the unclarity on what are the roles and what responsibilities each role has that the employees face.	Establishing a committee and identify data owners.	Alhassan, Sammon, & Daly (2019)
Clear inclusive data requirements	Understanding of data, avoiding inconsistency between what is presented and what is required for the business owner.	Communication and clarity between business owner and delevopers.	Alhassan, Sammon, & Daly (2019)
Focused and tangible data strategies	Importance of considering data as an asset and the related activities; and the planning for data governance, with an orientation for achieving short-term and long-term objectives.	Establishing an executive team to encourage organization-wide structure for data governance.	Alhassan, Sammon, & Daly (2019)

Obtaining stakeholder buy-in'	Involving and reaching agreements, contributing to increase interest and commitment in data governance activities.	Debates and feedback.	(Cave, 2017)
Effective and strategic communications	Importance of communication and alignment with the organization goals.	Disseminating communication aligned with organizational goals.	(Cave, 2017)
Assess data governance situation	Assessing the current state of data governance in the organization before implementing a data governance initiative.	Data governance maturity level assessment.	Al-Ruithe et al. (2019)
Define the sustaining requirements	Ensure the continuity of data governance and constant improvement, providing for a long-term solution for the organization.	Definition of actions needed to ensure the initiative's sustainability.	Al-Ruithe et al. (2019)

Table 1 - Critical Sucess Factors (CSF)

3.6. CMMI Maturity Models

A maturity model is a conceptual framework, representing "phases of increasing quantitative and qualitative capability changes of a maturing element in order to assess its advances with respect to defined focus areas" (Kohlegger, Maier, & Thalmann, 2009). In the literature, there is a considerable variety and widespread of maturity models within the Information Systems spectrum, as for example, the Capability Maturity Model, which is basis for other maturity models (Kohlegger et al., 2009).

The Capability Maturity Model (CMM) is a framework that portrays the key elements of an effective software process, describing an "evolutionary improvement path from an ad hoc, immature process to a mature, discipline process" (Paulk, Weber, Garcia, Chrissis, & Bush, 1993). This maturity model was developed to fill the gap in managing software processes that organizations were experiencing in that time, being since then adapted by many industries as a framework to manage different types of projects. This maturity model created by the Software Engineering Institute was the first CMM for software organizations; however, there is a variety of CMMs with the common premise to improve processes in an organization. In this line, in the 90's decade, CMMs have been matured for multiple disciplines, such as Software Engineering and Integrated Product and Process Development (CMMI Product Team, 2006).

Most organizations required improvement in different groups within the organization, which implied the use of multiple models. The lack of integration of these multiple models raised costs problematics in training, assessment, and improvement actions, which prompt the development of the Capability

Maturity Model Integration (CMMI) that combines multiple source models to solve the problematic around using multiple CMMs.

The CMMI has two different approaches: continuous and staged. The primary relates to the improvement of processes related to a single or a group of process areas, using capability levels to portray improvement in an individual process area. The staged approach defines an improvement path for the organization through maturity levels, where each level provides a set of process areas that portray different organizational behaviors. Comparatively, the continuous approach provides flexibility in the order of improvement, allowing for specification within each process area with the possibility to have different rates for different processes. This approach is recommended by its creators to apply when the processes to be improved and its dependences are clearly defined. On the other hand, the staged approach relates to an organization-wide predefined and proven improvement path, providing a specific capability to the organization in each maturity level. It is recommended when there is uncertainty on where to begin and the processes that need improvement (CMMI Product Team, 2006).

The two approaches mentioned above both use levels to describe an evolutionary path in the organization, characterizing its improvement "from an ill-defined state to a state that uses quantitative information to determine and manage improvements that are needed to meet an organization's business objectives" (CMMI Product Team, 2006). Nonetheless, these approaches differ on the types of levels used, as mentioned before. The continuous approach uses Capability Levels – incomplete, performed, managed, defined, quantitatively managed, and optimizing – to incrementally improve the processes of a process area. The staged approach uses Maturity Levels – Initial, Managed, Defined, Quantitatively Managed, and Optimizing – to achieve process improvement across multiple process areas (CMMI Product Team, 2006).

"A maturity level consists of related specific and generic practices for a predefined set of process areas that improve the organization's overall performance" (CMMI Product Team, 2006), enabling the prediction of such performance on a set of disciplines. These are measured by the achievement of specific and generic goals. The CMMI has five maturity levels: Initial, Managed, Defined, Quantitatively Managed, Optimizing.

According to CMMI Product Team (2006), the maturity levels are defined as follow:

- Level 1: Initial

Organizations in this level have poor or no maturity in their processes. These are ad hoc and chaotic, consonant with the unstable environment that is frequently supporting such processes. There is a recurrence of exceeded budgets and missed schedules, and a tendency to overcommit, withdraw, and the inability to repeat success.

- Level 2: Managed

In this level, processes planning and execution is aligned with the established policy, and include skilled people with adequate resources to produce controlled results and relevant stakeholders. These processes are monitored, controlled, reviewed, and evaluated according to their descriptions. The existing practices are maintained during times of crisis, without abandonment. Here, the overcommitment does not happen since it is established with stakeholders and adjusted as needed.

- Level 3: Defined

This level is a reinforcement of what is already established in the previous, with more rigorous process descriptions and proactivity in its management. In this level, processes are properly defined and understood, described in standards, procedures, tools, and methods. The set of standard processes is established in this level and improved over time, contributing to organization-wide consistency. The processes in each project are adjusted in conformation to specific guidelines.

- Level 4: Quantitatively Managed

In this level, quantitative objectives for quality and process performance are settled and used as criteria for process management. Quality and process performance are measured in statistical terms and managed throughout the process life, and these measures are incorporated to assist fact-based decision making. This maturity level adds the processes quantitative predictability to the already available qualitative predictability found in the previous level, by identifying and correcting the special causes of variations in processes – cause of defect that is not inherent to the process.

- Level 5: Optimizing

The focus of this level is continuous improvement, with a quantitative look on the processes' common causes of variation – based on normal and expected interactions in the process –, through original technology and process improvement. There is a quantitative perspective on processes, its results, and improvement objectives, where these objectives are established and adjusted according to the business goals, and used as criteria in process improvement management. The impacts of such improvements are measured and evaluated according to the defined improvement objectives.

The CMMI model's recurrent improvement and expansion over the years has led to the development of a Data Management Maturity model (DMM) in 2014. In the DMM, there is a data perspective for each Maturity Level, as follows: in level 1, data is managed as a requirement for projects' implementation; in level 2, data management as a critical infrastructure asset is seen as important; in level 3, at the organizational level data is handled as paramount for successful mission performance; in level 4, data is a source of competitive advantage; in level 5, data is perceived as critical for survival in competitive markets. In Data Governance literature, the presence of maturity models has been explored as a means to assess the organizations' maturity in data governance and the CMM has been used as a reference, particularly by using its Maturity Levels.

3.7. SYNOPSIS

Data has shifted from a business by-product to a valuable asset that potentiates data-driven decision-making. To make decisions based on data, it must be at its full value and in a trusting position. In this process, data quality, security and privacy rise as concerns for this new organizational strategy.

To tackle these concerns, data governance emerges a cross-functional framework to manage data as an asset, maximizing its value, guaranteeing its proper use and quality, and ensuring compliance with legal requirements and organizational strategy. Despite its clear benefits, various authors consider data governance very challenging, as it encompasses all data's lifecycle, as well as the organization's processes, people, and organizational culture. Moreover, its presence in the literature is still not sufficient and is discordant, lacking the practitioner's intention in introducing it to complex organizational realities. Practice-oriented methods as design and action research are encouraged. The abstraction present in data governance leads to the lack of a formal definition, with different approaches tackled by various authors. Yet, common characteristics as cross-functionality, framework, definition of decision rights and accountabilities, formalization of policies and procedures, and monitoring of compliance are identified by several authors. The characteristics behind data-driven culture are identified and connected with data governance, contributing to understand both the role of data governance in a data-driven organization and in its competitive market position. Moreover, the gap in research on this topic was mentioned by some authors, justifying its inclusion in this study.

Along with data governance, frameworks to facilitate its implementation have emerged from many authors. Common to the various studies found is the verbalized need for a framework when dealing with data governance. From the first contact with data in an enterprise, to the last action made based on that data – as well as the value that was added through the process –, it is required ground-basic strategies that provide an enterprise-wide uniformity and consistency of this asset, allowing for its easy, fast, and managed access. Without this process, there is an amplified risk, both with business and compliance, with data not being leveraged at its full value and not managed properly following the legal obligations.

Nevertheless, the frameworks proposed encountered obstacles when implemented, namely the resistance from the organization to change, the costs associated, and the time and effort required. Here, as well, there is an explicit need for frameworks that are more focused on the initiative's implementation and monitoring. In this line, five frameworks were introduced, with different characteristics. The first encompasses a flexible contingency approach, the second a scalable structure to identify roles and responsibilities, the third a broader framework based on decision domains of data governance and its "*loccus of accountability*", the forth a practitioner's contribution with three core levels – rules and rules of engagement, people and organizational bodies, and processes, and the last a software vendor's framework divided in strategic, organizational framework, and tactical levels with the aim to provide the adequate depth, scope and adaptability. Industry and context-specific, and single data governance area focused frameworks were not included in this literature review due to its narrow spectrum and, hence, incompatibility with sustainable options – which constitutes one of this work's focus.

At last, the success factors for a data governance initiative were identified based on three complementary articles ((Cave, 2017), (Al-Ruithe et al., 2019),(Alhassan et al., 2019)), as well as the description of maturity models and its maturity levels. These established the basis for the assessment matrix, that aims to evaluate the mentioned frameworks.

In conclusion, this literature review contributed substantially for the established objectives of this work, by providing essential foundations for the next chapters.

4. ASSESSMENT MATRIX

An assessment matrix is a tool used to evaluate the level of a phenomenon in different cases. It is composed by rows and columns, each representing different dimensions, and different maturity levels. The assessment matrix proposed in this work differs from other maturity models in the literature as it does not evaluate the implementation of the data governance initiative in an organization; but rather the definition and design of such initiatives, focusing on how they can contribute to the success of data governance in any organization. In other words, the objective is to assess the proposed models and its potential success, and not the organization's data governance maturity.

In this sense, by combining the maturity level of an organization with the maturity level of a data governance framework, it can provide an optimal pair to successfully fulfill the needs of that organization in a personalized manner. This match can increase the probability of success of the data governance initiative in the organization and contribute to the sustainability of such program within the organization. As mentioned before, the sustainability of a data governance initiative is a concern and it is lacking a vision within this spectrum. Accordingly, this assessment matrix rises as an opportunity to fill that gap and provide guidance in a more practical approach, as well as creating a new paradigm of maturity models within data governance. The purpose of this assessment matrix is to provide a clearer understanding of the available frameworks, with its assets and liabilities, in order to facilitate the choice by the organization.

In their work, Rivera et al. (2017) also shifted the paradigm of assessment matrix from organizations to the assessment of maturity models themselves. This provides an overview of some of the existing maturity models and a basis to decide on which to apply to the organization. The same rationale is proposed here, replacing maturity models with data governance frameworks. However, as other maturity models found in the literature, the maturity levels are based on the Capability Maturity Model Integration, with a focus on data governance principles, processes and/or domains.

4.1. ROWS – CRITICAL SUCCESS FACTORS

The combination of the three approaches on the critical success factors for data governance led to a final list of 11 critical success factors in a data governance initiative. Due to its nature, it is possible to group them into four main areas: people, data, tools and technologies, and assessment and sustainability.

People

By definition, data governance is a cross-functional framework for managing data as an asset, defining the decision rights and accountabilities for decision making, and formalizing data policies, standards and procedures, as well as monitoring compliance. Retrieving keywords from this sentence, such as 'cross-functional', 'decision rights and accountabilities', and 'formalizing', the need for people in this area rises naturally.

As perceived in the literature review, a data governance initiative has a strong foundation in the people that are involved – from the funding and support by top management, to the identification of roles and responsibilities, and all the stakeholders involved that impact its performance. People play a crucial role in a data governance program, in both technical and relational aspects, being the main focus of various data governance frameworks found in the literature. In fact, it is mentioned that organizational issues play a more important role to the success of data governance than technical issues (Korhonen, Melleri, Hiekkanen, & Helenius, 2013). Accordingly, the identified success factors for a data governance initiative are intimately connected with this first area of success – people.

In a data governance initiative, the most important aspects in this area are: training, roles and responsibilities, buy-in, and communication. It is paramount that these aspects are translated into clear objects so that its assessment is straightforward.

4.1.1. CSF 1 – Training

This first aspect enhances the impact that people's capabilities and awareness can have in a data governance initiative success. It is described in the CSF 'employees data competencies', which includes as suggested actions: continuous training, and increasing awareness to critical and sensible data.

In a perspective of sustainability, leveraging the available resources can be more advantageous than investing in new ones. In this sense, it is important to, not only evaluate and measure the resources at hand, but also to create a path to avail and take advantage of those resources. Employees' training on data and technological competences can increase their technical skills and avoid the need for external entities to deal with problems that may arise, reducing the organizations' costs at the same time as it motivates and contributes to the continuous learning and evolution of the people in the organization. Emotional intelligence capabilities can also be leveraged in a data governance initiative as the awareness and buy-in of stakeholders are considered to be critical in its success.

Organizations may argue that training is time and resources consuming, as well as representing additional costs, but, at the long-term, this will pay-off as it will contribute to a sustainable solution of data governance, avoiding its failure and the loss of investment that was put to it. It can be seen as a redirection of investment and resources rather than additional costs. As for time, if the data governance initiative fails, which is mention in the literature as recurrent, it would constitute a greater failure.

4.1.2. CSF 2 – Roles and Responsibilities

As one of the main focus in the majority of data governance articles and frameworks, the definition of roles and accountabilities in data governance initiatives are a fundamental aspect for its success. Here, the clear definition of roles and responsibilities can increase the trust on data and its processes. It is described in the CSF 'established data roles and responsibilities', where it is suggested that a committee is established and data owners are identified.

In other perspective, data governance is a novel paradigm within organizations, which can contribute to uncertainty regarding its features, namely what roles are needed, what profile should fit that role, and its attached responsibilities. Most roles in a complex organizational environment are already well-established and clearly defined, contrary to data governance, which is still perceived as a project to fulfill the data needs and not as part of the organization strategy. For that matter, organizations tend to resort to external sources as consultant enterprises to provide this type of services. While this approach may cut down time and costs on bureaucracy, it also wastes the opportunity to take advantage on the organizations' resources and exploit the knowledge of its employees, which are already aligned with its goals and vision.

4.1.3. CSF 3 – Buy-in

As mentioned, people play a very important role in a data governance initiative, which can impact its success. From technical aspects in data processes to ownership of data and decision makers based on that data, stakeholders are involved in every aspect of the project. In this line, their position regarding how the project is being conducted can influence their performance in each of these responsibilities. The CSF 'obtaining stakeholder buy-in' reinforces the need to involve stakeholders in the project and reaching agreements to guarantee that it runs smoothly and everyone is on board. To aid in this process, debates and feedback are suggested.

4.1.4. CSF 4 – Communication

In the same line, communication represents an important feature of these initiatives, as it can avoid errors due to misunderstandings and gives confidence to stakeholders. Not only communication has to be effective to guarantee that the tasks and procedures are perceived, but it also has to be strategically aligned with the organizations' goals and disseminated organization wide. It contributes to the transparency and trust of the program.

Data

Data is the core of any data governance project, where everything is built around these unprocessed facts, with the aim to forge business value from it. Each process, role, responsibilities and decisions are made based on the data that is available, in an effort to transform it into valuable and reliable information that can be used to inspire business decisions and elevate the organization.

On this note, it represents a key feature for a data governance project to be successful. Here, the main aspects are: strategy, processes and procedures, policies, and requirements.

4.1.5. CSF 5 – Strategy

In any project, it is fundamental to clearly define the strategy that is going to be undertaken. The strategy is what defines the approach and conduct to be adopted in order to achieve the project's objectives – the plan to be followed that will lead to success. A data governance project is no different, clearly defined strategies are helpful to guarantee that every person involved is informed and aware of what is being implemented, and why.

The CSF here represented is the 7th – 'focused and tangible data strategies'. One of the main aspects in this CSF lies on the fact that it is important to ensure that the strategy works towards data being considered as an asset, with a profit paradigm in mind. In other perspective, it also has a strong component on planning and objective achievement orientation. This is also sustained by the CSFs 'organizational', 'accountability', 'environmental', 'organizational culture change', 'develop change management plan', 'develop business case for data governance', and 'aligning data governance with the overall organization context' found in the Al-Ruithe et al. (2019) article.

4.1.6. CSF 6 – Processes and procedures

With a clear strategy and the appropriate conditions for stakeholders to embrace this project, the practical component of a data governance project initiates in defining and implementing the right processes and procedures to ensure that everything flows and that the goals are achieved. In this line, these processes and procedures should be clearly defined and aligned with the objectives and strategy, as well as embedded on the system, since the success of the project depends on its outcomes. These aspects are particularly important to ensure that the outcomes of these processes and procedures are relevant, updated and aligned with the necessities of the organization, and that the stakeholders involved understand and are on board with them. In a general manner, this will increase the trust on data's quality, security, and relevancy.

Accordingly, the second CSF – 'clear data processes and procedures' – reflects this important component of a data governance project. As well as the CSF in Al-Ruithe et al. (2019) article, 'Develop processes, procedure guidelines, principles, policies, and stander to support the data governance'.

4.1.7. CSF 7 – Policies

Policies can be defined as "a set of guidelines or rules that determine a course of action". These guidelines establish what actions should be taken and how, guaranteeing that these are properly implemented and aligned with the goals and strategy of the organization. The lack of guidelines and rules in a project can lead to imprecise actions, with different outcomes that what is expected. Specifically in data governance, this can have severe impacts, compromising the trust, security and privacy of data.

In this line, the CSF 'standardized easy-to-follow data policies' enhances the importance to have guidelines and rules related with data, that are unified and clear within the organization. This will ensure that everyone involved is conscious of what actions to take and that there is rigor and consistency in these actions, increasing the trust in the outcomes.

4.1.8. CSF 8 – Requirements

Despite all the benefits and relevance that data governance has in any organization, the focus of its implementation by organizations has been to meet requirements and comply with legal obligations, such as data regulations. Notwithstanding, this is, in fact, an important aspect of data governance, as these regulations guarantee that there is awareness of the importance and sensibility when dealing with data.

In this success factor – 'clear inclusive data requirements´ –, it is also included the clarity and inclusivity of the requests made by business owners and end-users. These users, that make decisions based on data, require that the information is accurate and relevant for such decisions. Accordingly, it is important that the people who are involved in the process that derives these outcomes from data have a clear understanding of those requirements, in order to extract and deliver the proper value from the available data. In this line, it is crucial that decision-makers conduct these decisions with inclusivity and diversity, extracting the most value from data and deliver different and broader perspectives to the business.

4.1.9. CSF 9 – Tools & Technologies

Technologies were, in fact, some of what sparked the statement that data are now having in the day-to-day life. With its evolution, the volume, variety and velocity that data are integrated continues to scale up every year, allowing the organizations to possess inconceivable amounts of data. In this line, technologies have a crucial role when dealing with such abundance – enabling its integration, treatment and monitoring –, empowering the organizations to extract actual value from idle data.

The CSF ‘flexible data tools and technologies’ resonates with other CSF, having an important role on data policies, processes and procedures, and requirements. This category is identified as a critical success factor due to the importance of data integration, embedding policies, processes, and procedures into the system, and the automation and testing of data processes. Not only does it accelerate data lifecycle’s, enabling the extraction of value from more data and in a faster way, but it also increases the trust of such data and all related activities through techniques as monitoring, automation, and testing of systems. The flexibility can have different advantages, prospering the investment made by having other capabilities that can be used in different areas within the organization, as well as allowing to incorporate future changes in such processes. It is, however, important to have appropriate IT infrastructure and integrated data. The Al-Ruithe et al. (2019) CSF ‘monitor tool and measure metrics’ regards the monitoring and measuring of data governance performance in each stage of the data governance initiative, and can be included within this CSF.

Assessment & Sustainability

4.1.10. CSF 10 – Assessment

Prior to the implementation of a data governance initiative, its presence should be assessed within the organization. This guarantees that the program is adapted and adjusted according to the resources, needs, and objectives regarding data governance of the organization, allowing for a better use of the available assets and a more precise definition of the strategy, implementation plan, and data governance program.

4.1.11. CSF 11 – Sustainability

In the simplest manner, sustainability can be defined as “the quality of being able to continue over a period of time”. As mentioned before, the sustainability in a data governance initiative has been one of the main bottlenecks found in its implementation and long-term success. From the literature, Al-Ruithe et al. (2019) identified as a CSF the definition of sustaining requirements, which must ensure the continuity and improvement of data governance in the organization over time.

The sustaining requirements can be specific to each organization, depending on its structure, resources, needs, and objectives. The important thing is that these requirements should be identified, defined, and documented so that they are included in the data governance initiative.

4.2. Levels – Maturity Levels

In the line with the Literature Review, the maturity levels used for this assessment matrix are based on the CMMI maturity levels. Originally, that framework was designed to assess the maturity of an organization in one or more process areas, which has since then been recreated several times in the literature through adjustments to process areas or narrowed to a single discipline. For this study, these maturity levels were adapted to a new reality: the assessment of frameworks it selves.

The main objective of a Data Governance framework is to successfully implement data governance in an organization. Accordingly, the maturity levels will be defined in consonance with the capability of such frameworks to contribute to the organizations' maturity in data governance. Moreover, as one of the objectives of this study is to identify sustainable frameworks and actions within Data Governance, the staged approach from the CMMI will be used as it defines an improvement path for the organization, without the need to clearly define the processes to be improved first hand. The latter affirmation is even more relevant since Data Governance relates with different process areas.

Maturity level	*Description*	*Level*
Initial	There is little to no reference to the CSF and/or it is not sufficiently structured.	1
Managed	It is described and added in the form of a documented plan. The components of this plan are established according to the business or data strategy. A review is part of this plan as a means to evaluate and improve its performance.	2
Defined	The plan is well characterized and described in standards, procedures, tools, and methods. These standards are established and improved over time. It is seen as important to the success of the program.	3
Quantitatively Managed	Use of quantitative measures to evaluate the program organization-wide. Quantitative objectives for this plan's quality and performance used as criteria for its management. Predictability on the programs' results.	4
Optimizing	The plan is constantly reviewed and monitored for its appliance and is embedded into the system. This CSF is one of the main concerns when implementing the program.	5

Table 2 - Description of maturity levels for the assessment matrix

4.3. COLUMNS – DATA GOVERNANCE FRAMEWORKS

The Assessment Matrix columns contain the frameworks that will be assessed for each CSF on the different maturity levels described above. As mentioned in the Literature Review section, with the expanding interest and necessity of data governance – both in organizational and research settings –, the need for a data governance framework is advocated by many authors. Although there is a considerable number of data governance frameworks in the literature, most are focused on either industry or matter specific.

The election of the data governance frameworks to include in this assessment matrix was based on criteria as: could not be industry specific (such as health industry), technology specific (such as big data or cloud), neither matter specific (such as data quality). Additionally, several articles – including literature reviews – were reviewed to determine the frameworks that were more prominent in the literature ((Nielsen, 2017);(Al-Ruithe et al., 2019);(Alhassan et al., 2018)). Finally, a combination of different characteristics – as accountability vs business-focused, academic vs practice sources, and others – was considered to guarantee diversity in the frameworks. Along those lines, five data governance frameworks were selected to be assessed against the CSFs identified previously, based on the different maturity levels. The frameworks are divided in Academic – developed by researchers in an academic perspective of data governance, and Practice-oriented – developed by industry association or software vendors in an organizational perspective.

Academic

4.3.1. Framework 1 – One Size Does Not Fit All

In 2009, Hubert Oesterle, Boris Otto and Kristin Weber published the article “One Size Does Not Fit All—A Contingency Approach to Data Governance”. In this article, mainly focused on Data Quality Management (DQM), the authors state that data governance can aid organizations to implement organization-wide accountabilities for DQM that include both IT and business professionals (Weber et al., 2009). By transferring concepts from IT governance, the article presents a responsibility assignment matrix based on three components: data quality roles, decision areas, and responsibilities.

The responsibility assignment matrix is organized as follows: the rows identify the key decision areas, the columns indicate the roles in DQM, and the cells contain the types of interaction between each decision area and each role. These types of interaction can be ‘Responsible’ – responsible for the execution of a particular activity; ‘Accountable’ – accountable for authorizing a decision regarding a particular activity; ‘Consulted’ – may be consulted to provide input and support to an activity or decision, and ‘Informed’ – may be informed of the output of an activity or decision. In this article, it is included the description and correspondent organizational assignment to a set of roles that are considered important within a data governance initiative implementation.

	DQM Role 1	DQM Role 2	DQM Role 3	...	DQM Role n
DQM Task 1	[R|A|C|I]	[R|A|C|I]	[R|A|C|I]		[R|A|C|I]
DQM Task 2	[R|A|C|I]	[R|A|C|I]	[R|A|C|I]		[R|A|C|I]
DQM Task 3	[R|A|C|I]	[R|A|C|I]	[R|A|C|I]		[R|A|C|I]
...		DQM Responsibilities (assignment of roles to tasks)			[R|A|C|I]
DQM Task n	[R|A|C|I]	[R|A|C|I]	[R|A|C|I]	[R|A|C|I]	[R|A|C|I]

Figure 1 - Data Governance model. Weber et. al (2009)

In addition, as referred by the authors, this model addresses DQM on three horizontal layers: Strategy, Organization, and Information Systems. On the ‘Strategy’ layer, the data quality strategy associates data management with business drivers, establishing strategic objectives; which is combined with the business case for data management and the installation of a maturity assessment. According to the authors, main tasks include: develop a data quality strategy and objectives, define a portfolio of data quality initiatives, formulate the business case, and conduct an assessment and establish a review process. The 'Organization' layer regards to organizational and process-related features, establishing measurement and control systems with the aim to monitor and improve the performance, as well as defining and assigning roles and processes related to data management. Main tasks include: measuring data quality through the definition of the appropriate metrics, define roles and responsibilities, establish data owners, define data management processes, and assign ownership for those processes. The third layer – 'Information Systems' – incorporates the development of the data model, the design of data architecture, and the definition of system support. The main tasks are: identify data objects, define a data glossary, assign quality dimensions to those data objects, establish data storage and infrastructure, define standards for data distribution across systems, and establish cleansing methods and tools.

Finally, the authors suggest two parameters to design the data governance model – organizational structuring of DQM activities and coordination of decision-making for DQM activities –, which define the extremes between centralized and decentralized, and hierarchical and collaborative, respectively. The organizations must find its balance between each of these extremes in order to define and assign the responsibilities to each role. The main idea behind this framework, as the name suggests, is that the data governance model is individual and personalized for each organization. Through the guidelines and framework design, the authors expect that each organization customizes this model to its own context.

The choice of this framework for this assessment matrix lies on the fact that sustainability in a data governance initiative is the main objective of this study. An individualized and personalized model can provide some space to the organizations to adapt this implementation to its own context, allowing for a greater involvement and, therefore, interest. People are keener to do something that they understand and where they are involved, which would be the case in this specific framework. Furthermore, it enables the use of the available resources within the organization – such as employees for the specific roles –, which contributes to the sustainability of this implementation. Notwithstanding, this approach may require some level of maturity and understanding of what the resources are, needs, and objectives of the organization regarding data governance.

4.3.2. Framework 2 – Data Governance Structure

Vanessa Chang and Lai Kuan Cheong published the article "The need for data governance: A case study" in 2007, that explores the relationship between IT Governance and Data Governance, and the latter's relevance to effective data management. The authors propose a Data Governance Structure and Framework to promote effective data management and emphasize collaboration between IT and business (Cheong & Chang, 2007).

The Data Governance structure proposed in this article emphasizes the engagement from business and IT in strategic, tactical, and operational levels. With this engagement, it is ensured that both IT and business are informed and that the IT initiatives are aligned with the data governance objectives. This framework is based on specific roles and responsibilities, that comprise the mentioned levels. At the Strategic level, there is the Data Governance Council role, whose members are executives with interest in data management and are responsible for "endorsing policies, resolving cross divisional issues, engaging the IT council at the strategic level, strategically aligning business and IT initiatives, and reviewing budget submission for IT and non IT related projects" (Cheong & Chang, 2007); the Tactical level includes the Data Custodian, who is responsible for data management and accountable for its quality, for resolving controversy in user group meetings, for endorsing data management and data cleansing plans, for ensuring data is fit for purpose, as well as adapting strategic plans to tactical, change and stakeholders management. Still in the Tactical level, the Data Steward role comprises members knowledge of business processes and data requirements, as well as IT knowledge that helps translating business into technical requirements. Data Custodians lead these stakeholders, who are responsible for implementing the tactical plans, and assisting Data Custodians in change management, stakeholders management, asset related information systems management, and project management. These stakeholders manage user group meetings and train and educate the data users; finally, on the Operational level, the User Group members are stakeholders from different departments who are involved in data's collection, process and report. With the help of technical IT staff expertise, these meetings are where urgent operational data issues are discussed. These members, also referred to as data users, are responsible for informing data related problems, making functionalities requests for efficient data collection, and designate reporting requirements.

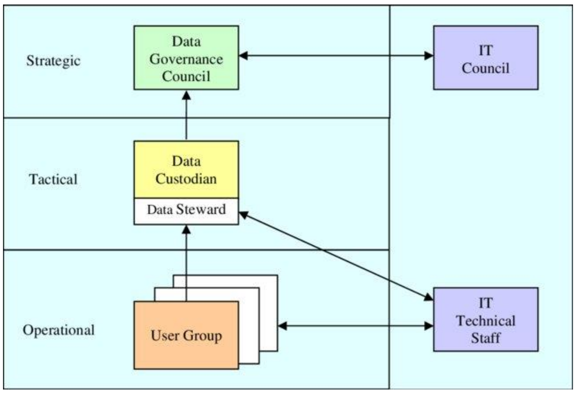

Figure 2 - Data Governance Structure. Cheong and Chang (2007)

This framework was chosen to this assessment matrix due to its encompass with IT, demystifying the notion that data governance is conducted by the IT department, as well as its detailed description of the various roles and respective responsibilities across the different levels.

4.3.3. Framework 3 – Data Governance matrix

In 2010, Carol V. Brown and Vijay Khatri published the article "Designing Data Governance", with the aim to develop an overall framework that can be used by both researchers to focus on important data governance problematics, and practitioners to develop an effective data governance approach, strategy, and design (Khatri & Brown, 2010).

The framework proposed by these authors is inspired in the one that Weill and Ross designed for IT Governance and is supported by a data governance matrix that practitioners can use to design their own data governance. This matrix comprises five decision domains – Data Principles, Data Quality, Metadata, Data Access, and Data Lifecycle – and the *locus of accountability*, between centralized and decentralized. The overarching framework comprises five decision domains along with their importance and guidelines to follow, as well as the *locus of accountability of decision making*, which represents the roles for each decision domain.

Decision Domain / Locus of accountability	Data Principles	Data Quality	Metadata	Data Access	Data Lifecycle
Centralized	✓				
				✓	✓
			✓		
		✓			
Decentralized					

Figure 3 - Example of a Data Governance matrix. Khatri and 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/trustee • Data custodian • Data steward • Data producer/supplier • Data consumer • Enterprise Data Committee/ Council
Data Quality • Establishing the requirements of 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 • Data quality analyst
Metadata • Establishing the semantics or "content" of data so that it is interpretable by the users	• 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 • Enterprise data modeler • Data modeling engineer • Data architect • 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 beneficiary • Chief information security officer • Data security officer • Technical security analyst • Enterprise Architecture Development Committee
Data Lifecycle • Determining the definition, production, retention and retirement of data	• How is data inventoried? • What is the program for data definition, production, retention, and retirement for different types of data? • How do the compliance issues related to legislation affect data retention and archiving?	• Enterprise data architect • Information chain manager

Figure 4 - Framework for data decision domains. Khatri and Brown (2010)

In Figure 4, it is depicted the overarching framework proposed by the authors. This framework was very important in the literature as, since then, many authors use these data decision domains to design their own data governance framework. This study contributed with the data decision domains for data governance, which allows for the right roles and responsibilities to be defined and assigned for a data governance initiative.

This framework was chosen due to its important contribute to the literature, as well as the design of other data governance frameworks. Having both components of a data governance matrix and a framework for data decisions, it is very complete and brings a whole new perspective and paradigm, focusing on data decision domains.

Practice

In this work, Practice frameworks are defined as frameworks business or organizational-focused, developed by software vendors and industry associations. In their study, Al-Ruithe et al. (2019) performed a systematic literature review and refer that, on non-cloud data governance, there are few frameworks, mainly being developed by industry associations such as DGI and IBM.

4.3.4. Framework 4 – DGI Data Governance Framework

Gwen Thomas, from the Data Governance Institute, prepared an article "The DGI Data Governance Framework" that explains in detail the framework designed by the DGI and its main concepts. This framework comprises 10 universal components of a data governance program, distributed among 3 areas: People & Organizational Bodies, Rules & Rules of Engagement, and Processes.

'Rules & Rules of Engagement' include the following components:

Mission and Vision – The idea of triangulation is here presented by the author as the Data Governance mission can be divided in three parts: proactively define and align rules; provide ongoing protection and services to data stakeholders; and reactively resolve issues. For the program vision, the authors recommend designing a clear, compelling vision of what would the organization look like with a mature data governance program, in order to inspire and encourage data stakeholders.

Goals, Governance Metrics/Success Measures, Funding Strategies – The program's objectives should follow the SMART technique: Specific, Measurable, Actionable, Relevant, and Timely. To choose which goals to set, it is recommended that the effects of data governance efforts be predicted in four areas: Programs, Projects, Professional Disciplines, and People. Asking the right questions, as "how can our efforts help organization's programs increase revenue and value", can lead to the definition of the proper goals for the data governance program. Measures follow the same structure as goals and, as Thomas (2006) said, "Everyone involved in Data Governance should know what success looks like, and how it's being measured". The author also proposes to use the following formula to create value statements: "If we do A, then we should expect B, with a result of C; otherwise, we should expect D, with a result of E." (Thomas, 2006). Regarding the funding strategies, along with key stakeholders, it is important to define ways of funding that include investments in a Data Governance Office, a Data Analyst/Architecture – to help define rules and data, and research problems that must be solved –, Stewardship activities, as well as protocols for Business and IT staff who participate in data related thematic.

Data Rules and Definitions – Data policies, standards and definitions are included in this component, as well as compliance requirements, and business rules. The focus here can be to create new/gather existing rules and definitions, aligning and prioritizing existing conflicts, along with address gaps and overlaps, and institute and formalize rules specific to certain definitions.

Decision Rights – The decision rights attribute the responsibility to the stakeholders who will make each decision, and its assignment must be facilitated, and even documented and stored, by the Data Governance program. For compliance-based programs, the author defends that these decision rights should be assigned to the Board of Directors with input from Legal department. As for the other programs, these must be more carefully assigned, through negotiations if needed.

Accountabilities – Posterior to the definition of decision rights, and the decisions it selves, it is necessary to assign the parts that each stakeholder will have in each step of the action on that decision, and when. The Data Governance program is expected to help define accountabilities and embed them into everyday processes. In compliance-related cases, organizations tend to have a centralized group who develops the necessary requirements and then disseminates them to the stakeholders. Nevertheless, in other environments there is also the use of cross-functional teams, with governance coordinators understanding and following the organization's protocols to engage staff, assign tasks, and provide status to management.

Controls – Data often constitutes a risk for the organization, and this risk must be managed. The organization's strategies to manage the risk are made operational through controls, which can be preventive, detective or corrective. These controls can be recommended in the Data Governance program at different levels of the controls stack, such as operating system, database, and others, to support organization's objectives. In other perspective, existing general controls, such as change management, policies, and training, can be adjusted to support governance goals.

'People & Organizational Bodies' include the following components:

Data Stakeholders – Data Stakeholders are members of different areas within the organization, who affect and are affected by data-related decisions, whether it is by creating and use data, or setting rules and requirements for data. These expectations must be met by the Data Governance program, defining who should be included in data-related decisions, who should be consulted before the formalization of such decisions, and who should be informed of those decisions after they are implemented. As a best practice, it is recommended that a group of executive stakeholders constitute a Data Governance Board.

Data Governance Office – The DGO facilitates governance activities as making data-related decisions, defining data, monitoring compliance, and resolving issues. This authority is also responsible for collecting metrics and success measures and report them to the stakeholders, providing continuous "stakeholder care" with communication, access to information, record-keeping, education, and support.

Data Stewards – A Data Stewardship Council is a group of stakeholders who make data-related decisions, set policies, specify standards, or even make recommendations to the Data Governance Board. It is recommended that a hierarchy of stewards is implemented, dividing the Data Stewardship Council into working groups to address specific data issues or decisions.

‘Processes’ include the following component:

Proactive, Reactive, and Ongoing Data Governance Processes – Data Governance Processes should be standardized, documented, and repeatable, supporting regulatory and compliance requirements for Data Management, Privacy, Security, and Access Management (Thomas, 2006). The DGI recommends formal, documented and repeatable processes for: aligning policies, requirements and controls; decision rights; accountability; performing stewardship; change management; data definitions; resolving issues; data quality requirements; building governance into technology; stakeholder care; communications; and measuring and reporting value.

All these components are arranged in the DGI Data Governance Framework, divided in the three areas that include those components, and in the “who”, “why”, “when”, and “how” elements of a Data Governance program. The following narrative can be depicted from Figure 5: The mission (“why”) defines the focus areas and the goals, metrics/success measures, and funding strategies; the latter’s influence the Data Governance processes, along with data rules and definitions (“how”), and the people (“who”) involved; in turn, people involved define decision rights, accountabilities and control mechanisms that will influence the processes; these processes provide input to Business/IT processes that touch data, ensuring that the mission defined is achieved; the program’s life cycle (“when”) consists of 7 steps to implement each effort; finally, management is “behind” all these components, whether providing support, funding, input, or even receiving feedback or reporting status of the program.

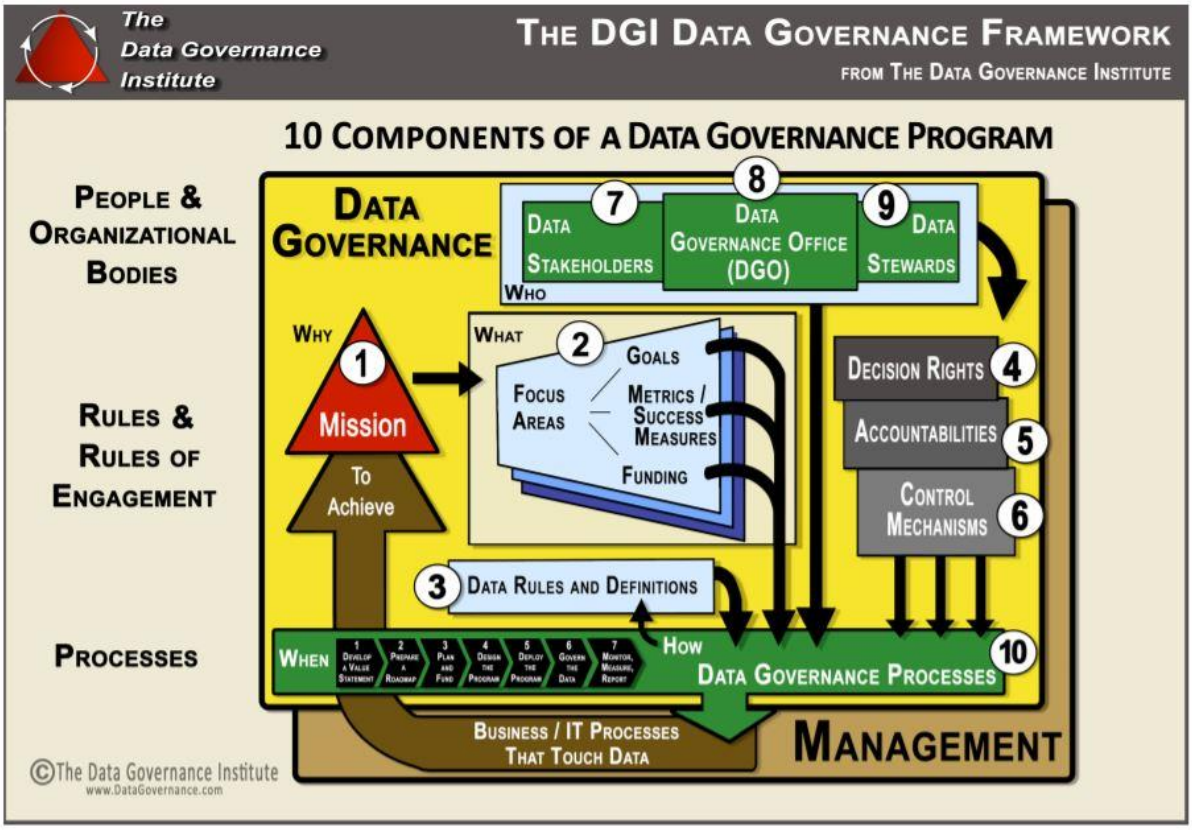

Figure 5 - DGI Data Governance Framework. Thomas (2006)

4.3.5. Framework 5 – SAS Data Governance Framework

SAS Data Governance Framework emerged from the efforts that the software vendor SAS designed to provide the depth, breadth and flexibility to overcome data governance failure points, such as the poor definition of data governance, or being depicted as a finite project or academic exercise. This framework is designed having in consideration both holistic – all data aspects are taken into account – and pragmatic – phased tactical deployment with 'quick-wins' – approaches. (SAS Institute Inc, 2018)

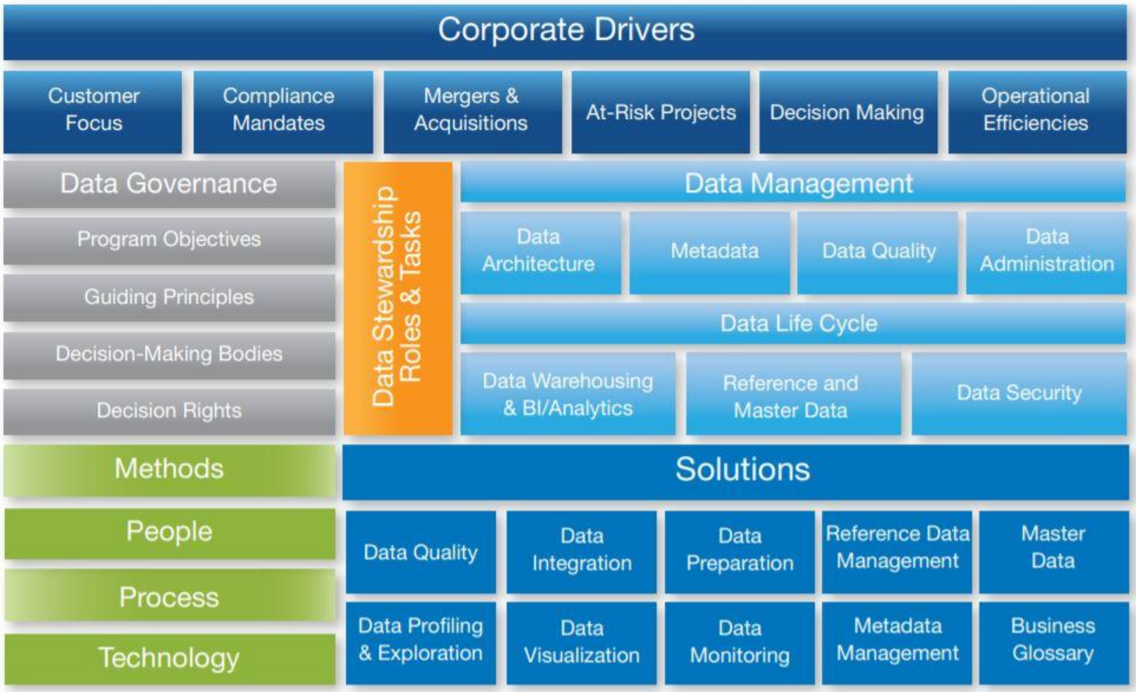

Figure 6 - SAS Data Governance Framework (SAS Institute Inc, 2018)

This framework is divided in three main aspects of Data Governance: strategy, organizing framework, and tactic, with six sections – Corporate Drivers, Data Governance, Data Management, Data Stewardship, Methods, and Solutions. The strategy level includes the Corporate Drivers section, which tackles the business drivers and strategies that highlight the need for data governance. On the organizing framework level are included Data Governance and Methods sections, with a focus on development and monitoring of policies that drive data management events. Lastly, Data Management, Solutions, and Data Stewardship sections are included in the tactical level, related to data management ongoing procedures. As specifically mentioned in the SAS Data Governance Framework White Paper, the organization's data governance strategy must tackle its short-term necessities while contributing for a long-term governance proficiency. In order to achieve such abiding proficiency, it is important to assess the different levels of governance maturity existent in the organization and having them in consideration when developing a long-term plan, contributing for the sustainability of data governance.

Corporate Drivers

Corporate or business drivers are the key factors, activities, and resources that have significant impact in the performance of a business. These can vary from organization to organization, but are commonly related to profitability and productivity, as Customer Focus, Compliance Mandates, Mergers and Acquisitions, At-risk Projects, Decision Making, and Operational Efficiencies. Contrarily to data governance activities, corporate drivers are commonly recognized as vital to the organization's performance as its results are measurable and easily comprehended. In this sense, attaching data governance activities and investments to corporate drivers will promptly associate data governance achievements with key business objectives. With this strategy, data governance activities are tied to organization's objectives, and its value is easily depicted by even the most reluctant stakeholders.

This section in the framework aims to tackle the data governance failure point of unclear return of investment (ROI) and difficulty in associating data governance activities with business value, as well as the perception that it is an academic exercise or finite project.

Data Governance

Other data governance failure points identified are related to organizational culture, decision-making culture, and staff constraints. In this section are highlighted the planning tools – Program Objectives, Guiding Principles, Decision-making Bodies, and Decision Rights – that tackle these issues, with the implementation of data governance accommodating the culture of the organization and its resources. In the same rationale, issues as overload of resources and unclear definitions of data governance and its activities can be tackled with these planning tools and a future vision as a mature organization in data governance.

Program Objectives – Consistent with other programs in organizations, data governance objectives must be clearly defined – and aligned with business objectives – with the aim to measure its performance. Data governance initiatives frequently have impacts organization-wide and, due to the risks involved when working with data, the organization policies must include data governance. This can be ensured by: include data stewards in planning and work teams, identify data risks and mitigation steps, and set standards for metadata capture.

Guiding Principles – Guiding principles enhance data governance support to the organization culture, structure and business goals, aiding to highlight to top management executives data governance value and funding necessities. These principles include: manage data as an asset, maximizing business value and reduce risk; clear and transparent communication on data governance policies and decisions; and adaptability of the data governance program according to the business unit.

Decision-making Bodies – Decision-making Bodies are composed by stakeholders that make decisions about data and its management. Having the right stakeholders and guaranteeing that these decisions reflect both the needs of business units and the organization is a driver for data governance success. Stakeholders can be part of different bodies, such as: data governance council, data steward team, enterprise data governance office. In the white paper is described that effective data governance requires that stakeholders are involved in data decisions, and that they are more likely to make better decisions when the data governance role is part of their jobs.

Decision Rights – Stakeholders in decision-making bodies can have different roles in data governance activities. In this framework, the RACI approach is recommended, as in the "One Size Does Not Fit All" framework. It can be applied in different activities, such as: approve or develop policies and procedures, and monitor compliance.

Data Stewardship

In the white paper, stewardship is defined as “an ethic that embodies the responsible planning and management of resources”. A data steward establishes and oversees the connection between other stakeholders and their related data governance activities, being responsible for data definitions, usability, questions and accesses, as well as enforcing data policies, monitoring data quality, and reporting metrics and issues. These stakeholders have capacity to understand both business and IT, being paramount in an organization with the granted authorization to oversee data.

Data Management

Data Management comprehends all the phases of Data Lifecycle, being defined as the set of disciplines to manage data as a valuable asset, through its collection, storage, security, and maintenance. In this framework, it is recommended that the inclusion of these disciplines is made incrementally, depending on the organization data strategy.

Data Architecture – Data Architecture is a structure that comprises the conceptual, logical and physical models involved in a data lifecycle, encompassing standards, rules, and policies. It defines the people, processes, and technology for data management, along with preventing duplication and reducing complexity. Examples of these models are data flows and policy documents.

Metadata – Metadata is data that describes and provides information on other data, such as its length, format, and source. Its management includes the description, usage, relationships and ownership of enterprise data. In data governance, automated technology to metadata management – as capturing, managing and publishing metadata – is paramount for the effectiveness of the program, providing a business glossary, lineage traceability, and reusable information for business analytics.

Data Quality – Data quality management comprises standards and procedures on data quality and its monitoring, cleansing and enrichment, being defined and established at the business level and in consonant with the business objectives, while its execution and implementation is managed by IT. Flexible automated tools and technologies, with standalone integration in transactional systems, are recommended to cleanse and enrich data in both batch and real-time modes.

Data Administration – Data administration is primarily a responsibility of IT, by setting and managing standards, policies and procedures on quotidian activities within the data architecture, such as monitoring procedures and batch schedules. These policies and procedures should have the input of business as well, being able to encompass service-level agreements used as targets to the defined metrics.

Data Warehousing, Business Intelligence (BI) and Analytics – A data warehouse is a repository of data that can store a great amount of history and current data. Feeding such data to BI and analytics activities allows for forecasting and get insights from the past, ultimately driving better decision making. In a data governance program, the right data to perform accurate analysis are embedded to the systems, through tools and techniques as data reporting and visualization, data movement, and sophisticated security.

Master Data – Master data is a set of data elements critical to business operation and analysis, such as customer or product. Master Data Management (MDM) is the discipline that manages these vital data elements through its consolidation, matching and standardization across transactional systems, providing higher data quality and coordinated data across all systems. MDM requires standard definitions, traceability between the different systems and the identification of common master data. All these requirements are enabled by data governance, which has a paramount impact on the success of MDM initiatives.

Reference Data – Reference data can be specified as a subset of master data, being unambiguous and non-negotiable, it is used to classify or categorize other data. This data is used by all transaction systems for storage efficiency and, accordingly, it should be managed to ensure consistency. Data governance helps defining, cataloging and capturing terms that support business activities, and assign ownership.

Data Security – Data security management ensures that data is protected and secured, through the definition of policies and procedures to establish access levels to data. It is recommended that these policies and procedures are applied through role-based model, granting access rights to different roles or groups. IT is responsible to implement these requirements in the system architecture, and business establishes the access levels and its connection to the different roles and groups.

Data Lifecycle – Data lifecycle comprises every data related activity since its collection to its archiving or disposal. Data management must encompass all these activities, guaranteeing that its maximum value is being retrieved and that it is consistently managed through all phases. Business defines the requirements for all the data across its lifecycle, such as its sources, movement, enrichment, uses and ultimate state, while IT is responsible for the implementation of such requirements.

Methods

Methods represent the elements necessary for data governance program implementation and execution.

People – The effectiveness of data governance in an organization is related with the people and their capabilities in performing data related activities, as it is paramount to have the right people with the right expertise to define policies and procedures and being accountable for them. The organization should have a proper documented structure that includes the different roles and accountabilities of each group and define them as an integral part of their job descriptions, such as data governance council, stakeholders, stewards and producers or consumers.

Process – According to the white paper, measurement and communication processes are common to all successful programs. Measurement processes illustrate the program effectiveness and prove its value and allow organizations to control data governance processes. Communication processes should be focused on reusability, broad acceptance and effectiveness, since complete, clear and concise communication is critical to any successful program. It is recommended to implement effective training on all features of the communication processes.

Technology – Technology and automated tools provide the means to enhance data volume and velocity within the systems in the organization, whilst sustaining the same policies and processes. Although data governance can be performed without such tools, these enable its expansion to larger volumes of data and reduce its time consumption, providing more consistency as well between the different systems.

Solutions

This section includes the tools that enable the implementation and automation of data governance activities, including data quality, data integration, data preparation, reference data management, master data, data profiling and exploration, data visualization, data monitoring, metadata management, and business glossary.

4.3.6. Frameworks comparision

Frameworks comparision	*1 - One Size Does Not Fit All*	*2 - Data Governance Structure*	*3 - Data Governance Matrix*	*4 - DGI Data Governance Framework*	*5 - SAS Data Governance Framework*
Context	Academic	Academic	Academic	Practical	Practical
Year	2009	2007	2010	2006	2018 (?)
Key strenghts	Flexible & customable Reusability of resources	Collaboration between IT & business Detailed roles and responsibilities	Important contribute to the literature Basis to other frameworks New perspective: data decision domains	Very complete Oriented for results Detailed explanations of the different levels of DG	Very complete Focus on DG problems & sustainability Holistic & pragmatic approaches with "quick-wins"
Improvement opportunities	May require an understanding of resources and needs Requires maturity in data governance to create the framework	Lack of practical guidelines Focused almost exclusively on roles and responsibilities	Lack of practical guidelines	Lacks the use of tools and technologies	Requires a maturity assessment Requires maturity in data governance to apply the framework

Table 3 - Frameworks comparision

4.4. CELLS – DATA GOVERNANCE FRAMEWORKS MATURITY LEVEL

Identified and described in detail the frameworks and the different maturity levels within each CSF, these must be combined to generate the values in the matrix cells.

4.4.1. Framework 1 – One Size Does Not Fit All

Training – Although people are a component that is very present in this framework, through the establishment of roles and responsibilities within data realm, the authors do not mention the importance of those people capabilities neither its development as part of the framework. Accordingly, its maturity level is Initial.

Roles and Responsibilities – Roles and Responsibilities area is one of the most present in this framework, being defined and described in detail. In the 'Organization' layer, one of the main activities described is to define roles and responsibilities, which places the framework in the 'Initial' maturity level. In the article, based on the literature, a table with different roles and their descriptions is provided with four roles and one committee; in addition, the responsibilities follow the RACI chart, being defined and guided accordingly. The definition of these roles and responsibilities within the organization is individual and personalized, according to their needs and requirements; moreover, the responsibilities assignment is guided through two design parameters, contributing for a plan well characterized and described in standards, procedures, tools, and methods. This last statement would place this framework within the 'Defined' maturity level, however, as described above, the maturity levels are cumulative, and it is necessary to comply with all statements in each level in order to move to the next. As it is missing a review of the plan on roles and responsibilities, in the 'Managed' maturity level.

Buy-in – In a decentralized model, "business and technical stewards decide autonomously for their area of responsibility. In addition to the top management executive sponsor, every business unit might have its own sponsor from the business-unit management" (Weber et al., 2009), which can contribute to stakeholders buy-in, as they are deeply involved in data related matters. The same rationale goes for the cooperative model, as can be perceived in this statement "enterprises relying on consensus-building should adopt a cooperative IT governance design" (Weber et al., 2009). Nevertheless, it is possible to have cooperative moments with centralized and hierarchical model, as depicted in the Company C present in the study, where "Although they have a very centralized organization structure, (…) their informal rules of decision-making are very cooperative, in that many employees agree jointly on day-to-day activities and decisions" (Weber et al., 2009). In conclusion, there is the assumption that stakeholders must be involved and agreements must be reached, however, the lack of structure and creation of documented plan leads to the 'Initial' maturity level.

Communication – Similarly to what was described in the 'Buy-in' area, the focus on communication in this framework is not easily depicted, as it is more focused on accountabilities. As such, it is placed in the 'Initial' maturity level.

Data Strategy – Strategy is one of the three layers described in the framework, establishing the link between data management and major business drivers, outlining strategic goals, constructing the business case for data management, and implementing a maturity assessment. Through the main tasks in this layer, it is depictable the maturity level on this area: "develop a corporate data quality strategy, including strategic objectives; define a portfolio of strategic corporate data quality initiatives; formulate the business case for corporate DQM; and carry out a status quo assessment and establish a review process." (Weber et al., 2009). In the previous statements, it is identifiable the importance of data strategy, its documentation through a portfolio, its relation to the organization culture through the business case, and its improvement through a review process, consisting of a standardized and well characterized plan. All these components lead to the 'Defined' maturity level.

Processes and Procedures – In the 'Organization' layer, one of the main tasks is to define data management processes, which corresponds to the 'Initial' maturity level. In the contingency model presented in this framework, one of the contingencies is, in fact, process harmonization. Nonetheless, the authors state that "the framework does not specify how the processes should be applied in different contexts." (Weber et al., 2009), preventing the framework from going to the next maturity level. In the description of these layer it is also stated that "It includes the establishment of appropriate measurement and control systems in order to monitor and improve the performance of DQM." (Weber et al., 2009), which could take this framework to the 'Quantitatively Managed' maturity level if it would have been more explored and documented. Nonetheless, the maturity level is 'Initial'.

Policies – Policies are mentioned in this framework as a component of role definition, that is, an area that business data steward must define and detail for its area of responsibility. In fact, in the Discussion section of this article, the authors mention that "An analysis of the guidelines and policy aspect of data governance is recommended" (Weber et al., 2009), as a recommendation for further investigation. Consequently, in this area, its maturity level is 'Initial'.

Data Requirements – The 'Information Systems' layer combines the development of a logical data model, the design of data architecture, and the definition of system support. In this layer, main tasks are related to the clear definition and identification of data objects and its dimensions. Although not mentioned specifically, data requirements could be added to this layer in a more explicit manner. The maturity level in this area is 'Initial'.

Tools and technologies – Tools and technologies are mentioned superficially in this article, in the 'Information Systems' layer, through the statement "define the utilization of cleansing methodologies and tools." (Weber et al., 2009), which places this framework within the 'Initial' maturity level.

Assessment – In the 'Strategy' layer, the implementation of a maturity assessment is mentioned, but not developed. Its maturity level is 'Initial'.

Sustainability – As mentioned previously, this framework does not have in consideration the time dimension of data governance in an organization, stated by the authors as a means to improve. Therefore, its maturity level is 'Initial'.

4.4.2. Framework 2 – Data Governance Structure

Training – In this article, 10 critical success factors for data governance are identified, one of them being training and awareness. Here, it is stated that "Data stakeholders need to be aware of the value of data governance. The importance of data quality and the benefits of quality data need to be communicated to all data stakeholders in order to raise their awareness." (Cheong & Chang, 2007).

Moreover, one of the responsibilities of Data Stewards is to “train and educate data users” (Cheong & Chang, 2007). Although it is mentioned in this article, there is no mention to structured and documented plans on training, neither its review. The framework maturity level in this area is ‘Initial’.

Roles and Responsibilities – This framework, as mentioned in the article, “was developed based on (…) roles and responsibilities” (Cheong & Chang, 2007). These are described in the article, with detailed activities, with the focus to mitigate the issues faced by data governance through the development of formal data governance with clear roles and responsibilities. The authors state that “Roles and responsibilities are defined so that members within the Data Governance structure are held accountable for their actions.” (Cheong & Chang, 2007). As the article defines and establishes guidelines for roles and responsibilities according to the business strategy, it does not include a review of such plan. The Data Governance structure, itself, is a well characterized and standardized plan with procedures, tools, and methods, with the opportunity to scale it overtime. In conclusion, although it checks all the conditions on the ‘Defined’ maturity level, the lack of a review component in the plan leads to simply a ‘Managed’ maturity level.

Buy-in – Even though stakeholders buy-in is not explicitly documented and structured, stakeholders are included in the data governance structure and each role – except user groups – has the responsibility to resolve issues that stakeholders may have in the lower levels. Accordingly, its maturity level is ‘Initial’.

Communication – In the descriptions of each role in this article, there is the assigned responsibility of managing stakeholders on lower levels and engaging stakeholders from different groups, promoting communication. In Figure 2, it is depictable that communication is present, not only within the Data Governance Structure, but also with IT in strategic, tactical and operational levels. Moreover, it is mentioned that IT governance council and Data governance council should work collaboratively and, individually, with the Corporate governance council. In conclusion, collaboration and transparent decision-making processes are mentioned in the article and part of the Data Governance structure itself, but not with a review of this plan, placing this framework in the ‘Managed’ maturity level.

Data Strategy – Data strategy is not thoroughly mentioned in this article. In this line, among the responsibilities of the Data Governance Council, there is the strategic alignment of business and IT initiatives. Moreover, there is the recognition that a proper structure and framework can drive strategic projects to conform and maintain good data governance. Accordingly, this area is in the ‘Initial’ maturity level.

Processes and Procedures – As mentioned by the authors, “This paper underscores the importance of a Data Governance structure together with policies and procedures for managing data effectively.” (Cheong & Chang, 2007). Regardless, there is no specific mention to processes and procedures, neither its management. This places the framework in the ‘Initial’ maturity level.

Policies – In the same rationale as the Processes and Procedures area, this area is in the ‘Initial’ maturity level.

Data Requirements – Data stewards are responsible for data requirements, translating business requirements into technical ones to pass to the user groups, where, in turn, the stakeholders specify reporting requirements. Although not mentioned explicitly, it can be assumed that these requirements are properly described and defined in the user group meetings, as required for the ‘Managed’ maturity level.

Tools and technologies – Tools and technologies are mentioned specifically as a result of a clear structure, leading to the acquisition of a data profiling and metadata repository tool, contributing to the discovery of anomalies more efficiently and the capture of information about data, promoting the access to this data organization wide. Therefore, the maturity level is ‘Initial’.

Assessment – No assessment method or tool is mentioned in this article. Its maturity level is ‘Initial’.

Sustainability – The scalability of the Data Governance structure can be a point worth mentioning in this framework, but it is based on only one case study, which leaves to question its adaptability to other realities. Therefore, its maturity level is ‘Initial’.

4.4.3. Framework 3 – Data Governance matrix

Training – In the article, it is mentioned one role associated with training: the data quality trainer. However, there is no explicit mention to training programs or which capacities/competencies people that deal with data should have. The maturity level is ‘Initial’.

Roles and Responsibilities – One of the focus in this article is the *loccus of accountability*, the “who” in decision making. Accordingly, for each of the five decision domains, different roles are suggested to answer the questions in the domain decisions. Not only these are documented, but they are integral part of the data governance framework itself, with the opportunity to organizations to construct their own model. Nonetheless, there is no reference to the review of this plan. In this line, this maturity level is ‘Managed’.

Buy-in – a There is no mention to employees buy-in in this article. Accordingly, its maturity level is ‘Initial’.

Communication – The communication component in this framework is represented in two main moments: in the Data Principles domain, through the decisions “What are the mechanisms for communicating business uses of data on an ongoing basis?” and “What is the program for establishing and communicating data quality?” (Khatri & Brown, 2010); and with the development of the data governance matrix itself, that “may be useful for communicating a given organization’s data governance approach.” (Khatri & Brown, 2010). With the absence of a review plan for these strategies, this maturity level is ‘Managed’.

Data Strategy – Data strategy in this framework is inside the domain Data Principles, where decisions based on strategies to clarify the role of data as an asset are made, by the potential roles. Here, decisions such as data usage, ongoing communication mechanisms, desirable behaviors, and other data related decisions are made. In the same line as the previous CSFs, this maturity level is ‘Managed’, due to the lack of a review strategy.

Processes and Procedures – Various data related processes and procedures are mentioned in the framework itself as domain decisions, across all data governance domains, such as “the mechanisms for communicating business uses of data”, “the program for establishing and communicating data quality”, “data be consistently defined and modeled”, “data access standards and procedures”, and “data definition, production, retention, and retirement for different types of data” (Khatri & Brown, 2010). All these previous statements refer to processes and procedures as being defined and implemented through this data governance framework. Moreover, statements that refer, for example, to decisions on how will the "assessment results be integrated with the overall compliance monitoring efforts" and "data quality as well as the associated program be evaluated" (Khatri & Brown, 2010), illustrate not only the review of such strategies, but also its improvement overtime. The plan is well characterized and defined in standards ("decision to standardize business processes"), methods, procedures, and tools. In conclusion, this maturity level is 'Defined'.

Policies – Policies are included in the Data Principle domain, where "data principles therefore establish the extent to which data is an enterprise wide asset, and thus what specific policies(…) are appropriate", as well as in the Data Access domain: "industry standards serve as a guide for the writing and updating of an organization's access policies(…)" (Khatri & Brown, 2010). The maturity level is 'Initial'.

Data Requirements – Data requirements are distributed between business data usage and compliance with regulation, incorporated in domain decisions in the following statements "What are the uses of data for the business", "How does the regulatory environment influence the business uses of data?", and "How do the compliance issues related to legislation affect data retention and archiving?" (Khatri & Brown, 2010). In this line, as these decisions are included in the data governance framework, being properly defined and documented according to the business strategy, this maturity level is 'Managed'.

Tools and technologies – Additionally to IT tools, which are mentioned in this article to "help surface quality issues for the business owners" (Khatri & Brown, 2010), other keywords are used to describe tools and technologies to be implemented, such as "mechanisms for communicating business uses of data" and "the program for backup and recovery" (Khatri & Brown, 2010). These domain decisions are built in a manner that lead decision makers to decide on which tools and technologies to use to cope with different data requirements. Accordingly, this maturity level is 'Managed'.

Assessment – The assessment methods described in this article are particular to data quality and risk, where Data Quality domain provides a roadmap to assess data; and risk assessment is recommended to be conducted on an ongoing basis. Its maturity level is 'Initial'.

Sustainability – This framework, as mentioned by the authors, "can be used by practitioners to design their data governance", that is, this constitutes a starting point to design a data governance framework or for researches to identify important data governance issues. Therefore, this framework is not considered to have sustainability features, as such, its maturity level is 'Initial'.

4.4.4. Framework 4 – DGI Data Governance Framework

Training – Training is perceived in this framework within the "Controls" component, which is where risks are prevented, detected, and corrected. In the article, it is mentioned that "Data Governance may also be asked to recommend ways that existing general controls ((…), training, SDLCs and Project Management, etc.) could be modified to support governance goals or enterprise goals." (Attard & Brennan, 2018)(Thomas, 2006). In this last statement, the author suggests that existing training programs could be adjusted to data governance, fulfilling the necessity to have employees with data competencies. Although training is mentioned in the framework, it is not possible to deeply understand its scope, as it can depend on the training program methods that the organizations use. There could be organizations with a training program documented, and quantitatively managed, or organizations with no structure at all. In this line, the maturity level is 'Initial'.

Roles and Responsibilities – Roles and responsibilities are mentioned priorly in the "Mission and Vision" component, where it is recommended that its definition in data governance bodies have in consideration the three-part mission suggested – reactive issue resolution, proactive rules, and ongoing services. Components "Data stakeholders" and "A Data Governance Office" define and establish the responsibilities and accountabilities for these roles. Similarly, the component "Data Stewards" is directly related with this CSF, where a Data Stewardship Council is recommended as "a set of Data Stakeholders who come together to make data-related decisions" (Thomas, 2006). Their responsibilities are described as setting policies and specify standards and define recommendations to a higher-level Data Governance board. Multiple levels of stewardship are also mentioned, as well as the example of Data Quality Stewards.

Finally, the component "Accountabilities" is also directed to this CSF, with the definition of accountabilities that can be integrated into everyday processes and the organization software development life cycle. As an example, the article focus this section in compliance, where individual managers may not be prepared to do the necessary steps and have responsibilities in identifying "all the tasks and integration points for designing and implementing controls, documentation, and auditable proof of compliance" (Thomas, 2006). In this sense, organizations may have a centralized group developing those requirements and disseminating them to stakeholders. As a final note, the author states "(…) most governance efforts involve cross-functional teams. Your governance coordinators will need to understand and follow your organization's protocols for engaging staff, assigning tasks, and providing status to management." (Thomas, 2006).

In light with all these, this framework has roles and responsibilities defined and with guidelines established, built according to business and data strategy, and, although it does not have a review plan, it is flexible enough for organizations to adapt and adjust to its own needs. Its maturity level is 'Managed'.

Buy-in – One of the responsibilities of the Data Governance Office is to provide stakeholder care, which can be seen as a form of "buy-in". This is further described in the Challenges section, where DGO "understand stakeholders' information needs, preferred terminology, and special interests. (…) They employ stakeholder participation matrices to make sure stakeholders aren't overlooked and that the right people get their part of the message in the right sequence." (Thomas, 2006). This places the framework in the 'Managed' maturity level.

Communication – In the same line as described above, in the stakeholder care responsibility of Data Governance Office (DGO) is included communication. Moreover, communication is identified in the article as one of the main sources of data governance success. Here, a detailed role of DGO is described, with the development of "layers of communication pieces – elevator speeches, value statements, impact statements, presentations, and other documentation – so they can deliver the right versions of governance messages to the right people with the right level of detail." (Thomas, 2006). In this manner, the maturity level is 'Managed'.

Data Strategy – Data Strategy can be identified in two different components, “Mission and Vision” and “Goals, Governance Metrics / Success Measures, Funding Strategies”. The definition of the program’s mission and vision is related with data strategy, as well as its goals, metrics, and funding. In both components, there are techniques to define the approach that the program should follow to accomplish its objectives. Accordingly, this maturity level is ‘Managed’.

Processes and Procedures – The component “Proactive, Reactive, and Ongoing Data Governance Processes”, as stated, “describes the methods used to govern data.” (Thomas, 2006). The article itself recognizes that “Ideally, these processes should be standardized, documented, and repeatable. They should be crafted in such a way to support regulatory and compliance requirements for Data Management, Privacy, Security, and Access Management.” (Thomas, 2006). The level of structure and formality is left to each organization to decide, but the DGI has some recommendations (see in the chapter “Improvement Opportunities” of this work). These processes are described and well characterized in standards, procedures, tools, and methods. Accordingly, the maturity level is ‘Defined’.

Policies – Policies are in the “Data Rules and Definitions” component, which comprises of objectives as “Create new rules/definitions; Gather existing rules/definitions; Address gaps and overlaps; Align and prioritize conflicting rules/definitions; Establish or formalize rules for when certain definitions apply.” (Thomas, 2006). This encompasses the definition, description, and standardization of policies, as well as its review and improvement. As such, the maturity level is ‘Defined’.

Data Requirements – Data requirements are also mentioned in the “Data Rules and Definitions” component, as compliance requirements. A centralized group clarifies, implements, and disseminates these requirements to stakeholders, enabling a better understanding for everyone and aligning with the data strategy. In this sense, the maturity level is ‘Managed’.

Tools and technologies – Technologies are rarely mentioned in the article, however, in the focus on Privacy/Compliance/Security these are referenced to “locate sensitive data, to protect data, and/or to manage policies or controls.” (Thomas, 2006). In the “Data Warehouses and Business Intelligence (BI)” focus, there is the reference of tools as data warehouses, data marts, or BI tools. Nonetheless, it is at the ‘Initial’ maturity level.

Assessment – Although no assessment is explicitly mentioned, a seven-phase data governance life cycle is proposed to implement any data governance program. In this methodology, the first three phases – “Develop a value statement”, “Prepare a roadmap”, and “Plan and Fund” – contribute to the assessment of data governance state in the organization, as well as its gaps to further define the strategy. Despite this, there is no standardization nor characterization in methods and tools. It is at the ‘Managed’ maturity level.

Sustainability – An interesting component of sustainability can be identified in the data governance life cycle methodology, where this method allows for a strong structured program, with continuous improvement through monitoring, measuring and report. Even though it is not specifically introduced as such, it provides the basis for a sustainable, long-term solution. It is, however, not further explained and, as such, it is in the ‘Managed’ maturity level.

4.4.5. Framework 5 – SAS Data Governance Framework

Training – Training is mentioned briefly in the Process area, within the scope of communication infrastructure, instead of data-related manners. Nevertheless, in the People area, it is stated that “An organized structure of people who have the proper skills is essential for the success of the data governance program.” (SAS Institute Inc, 2018). For this matter, the maturity level is ‘Initial’.

Roles and Responsibilities – As mentioned above, it is stated that “A common component of successful data governance is getting the right business stakeholders involved in decisions about data and how it is managed.” (SAS Institute Inc, 2018). In the “Decision-Making Bodies” area there is a list of data governance constituents, in which is required that they “become entrenched in the decision-making process around data rules and processes.” (SAS Institute Inc, 2018). Further, in the “Decision Rights” area, the RACI tool is suggested to define the roles for data governance activities and responsibilities. Additionally to the roles in decision making bodies, “Data Stewardship” is other area that focuses on roles and responsibilities, specifically data stewards and data custodians. Both these areas have a clear definition of roles and responsibilities, lacking only the review of this plan to get to the next maturity level. Currently, it is on the ‘Managed’ maturity level.

Buy-in – In the “Corporate Drivers” area, the authors identify that the challenge to demonstrate the return of investment may cause difficulties in getting broad consensus and participation across business units. To overcome this, the suggestion is to align data governance activities to corporate drivers, which will align its “wins” with key business objectives. Other strategy mentioned is to include leaders from both business and IT in the data governance council to guarantee that the needs from both business units are met. Finally, the authors mention that “Once the data governance role is part of a people’s jobs, they are more likely to make better decisions about the role of data – and how it applies to the corporate mission.” (SAS Institute Inc, 2018). In spite of these strategies, the plan does not include a review component, which places the framework in the ‘Managed’ maturity level.

Communication – In the “Process” area, it is recognized that complete, concise communication is paramount to data governance success. For this matter, “Any communications framework should focus on reusability, broad acceptance and effectiveness. Along with this, effective training on all facets of the communication infrastructure can help integrate a strong communications effort throughout the program.” (SAS Institute Inc, 2018). As such, there is the recognition of a structured communication plan, with the intent of reusability, which, by lacking a review plan, places the maturity level in ‘Managed’.

Data Strategy – Data Strategy can be identified in the “Program Objectives” and “Guiding Principles” areas. As one of these guiding principles, it is mentioned as an example “Data will be managed as a shared asset to maximize business value and reduce risk.” (SAS Institute Inc, 2018). “Data Management” area can also be perceived as a part of data strategy, as its functions are designed to implement data governance policies and include multiple disciplines, depending on the data strategy. In this statement “While data management is fairly broad, not all of these disciplines must be included in the first phases of a governance program. Some programs focus more on business definitions (metadata) initially, while others may emphasize a single view of the customer (master data).” (SAS Institute Inc, 2018), it can be perceived that the data strategy is described and well characterized in standards, methods, and tools, with a review to shift the data management functions when accurate, as well as the establishment and improvement of these standards over time. Accordingly, this maturity level is ‘Defined’.

Processes and Procedures – The “Data Architecture” area defines the processes used in data management and includes the standardization of procedures as a duplication of effort preventive measure. Also “Data Quality” mentions the standards and procedures on the quality of data. “Data Administration” sets standards and procedures for managing quotidian operations, including monitoring procedures. At last, in the “Process” area is stated “Measurement allows organizations to maintain control over data governance processes.(…) A data governance program is a program of continuous improvement, so effective measurement is a basic component of any successful program.” (SAS Institute Inc, 2018), which places this framework within the ‘Quantitatively Managed’ maturity level.

Policies – Similar to Processes and Procedures, policies are mentioned various times in this framework. As an example, in “Data Architecture” area, one data architecture artifact is policy documents. Moreover, in “Data Security”, there is policies to determine access levels to data and suggestions to apply those policies. These are determined by the data governance council and executed by data stewards. As there is no review plan along the program, the maturity level is ‘Managed’.

Data Requirements – Data requirements are not explicitly mentioned in the framework, as it is built to encompass all data requirements in an organization. Rather, the organization must adjust the data governance program and, as such, its data requirements to its needs. These are defined and documented within each of the correspondent areas, according to the data requirements that are included in that state. As mentioned, this framework allows for a flexible, incremental data strategy, which directly impacts data requirements. Accordingly, as in the Data Strategy CSF, the maturity level is ‘Defined’.

Tools and technologies – Unlike other frameworks, this encompasses a strong component of tools and technologies. “Data Quality”, for example, includes automated tools to “(…) cleanse and enrich data in both batch and real-time modes.” (SAS Institute Inc, 2018). As for technology, it “is used in a standalone fashion and integrated with transactional systems for ultimate flexibility” (SAS Institute Inc, 2018). In “Data Warehousing, Business Intelligence and Analytics” list of tools and techniques is, for example, data movement tools. Finally, in the “Technology” area, automated tools to perform tasks as data profiling and monitoring are considered the current best practices, allowing “the organization to scale to the largest volumes of data processing, maintaining the same rules and processes for any application.” (SAS Institute Inc, 2018). Nonetheless, it is at the ‘Managed’ maturity level, as there is no review plan.

Assessment – Although no assessment is explicitly mentioned, the framework itself can help the organization to understand how the existing facets can be implemented into the initiative. However, it is at the ‘Initial’ maturity level.

Sustainability – Sustainability is one of the bases on this framework, focusing on short-term needs while assuring long-term governance capability. The framework itself is seen as a tool to achieve a sustainable data governance program, by leveraging the individual parts that may already exist in the organization and use them as part of the data governance framework. As the framework itself is structured to be reviewed, incremental, and measured against organization goals, the maturity level is 'Quantitatively Managed'.

4.5. Assessment Matrix

CSF / Frameworks	*1 - One Size Does Not Fit All*	*2 - Data Governance Structure*	*3 - Data Governance Matrix*	*4 - DGI Data Governance Framework*	*5 - SAS Data Governance Framework*
Training	Initial	Initial	Initial	Initial	Initial
Roles and Responsibilities	Managed	Managed	Managed	Managed	Managed
Buy-in	Initial	Initial	Initial	Managed	Managed
Communication	Initial	Managed	Managed	Managed	Managed
Data Strategy	Defined	Initial	Managed	Managed	Defined
Processes and Procedures	Initial	Initial	Defined	Defined	Quantitatively Managed
Policies	Initial	Initial	Initial	Defined	Managed
Data Requirements	Initial	Managed	Managed	Managed	Defined
Tools and Technologies	Initial	Initial	Managed	Initial	Managed
Assessment	Initial	Initial	Initial	Managed	Initial
Sustainability	Initial	Initial	Initial	Managed	Quantitatively Managed

Table 4 - Assessment Matrix

Level 1 – Initial; Level 2 – Managed; Level 3 – Defined; Level 4 – Quantitatively Managed; Level 5 – Optimizing;

5. DISCUSSION

5.1. Improvement Opportunities

This section works as a guide on what can be useful to add when implementing these frameworks, whether the organization has already initiated a data governance program with such framework or because it resonates better with its components.

5.1.1. Framework 1 – One Size Does Not Fit All

This framework dates 2009 and, however it contributed significantly to the literature through individualizing data governance initiatives to organizations in order to accomplish their goals, nowadays it is expressed through the literature the importance of training and employee expertise. In this sense, adding a training plan/recommendation to this framework will fill the gap created by assuming the stakeholders already possess knowledge, as mentioned in the article: "the underlying assumption of DQM is that business experts possess knowledge about their data requirements and, hence, about data quality." (Weber et al., 2009).

In the Roles and Responsibilities area, this framework has a very strong component, lacking only the continuity that may benefit the organizations, through reviewing the plan as a means to evaluate and improve it, and adding time dimension. This along would place this framework in the 'Defined' maturity level. As identified by the authors: "It is also important to elaborate more precisely on the third factor in this model: how to define and measure DQM success" (Weber et al., 2009), it is lacking a measuring system in order to evaluate, improve and predict success for this framework to scale in to the 'Quantitatively Managed' maturity level. This is even more relevant as the statements in 'Optimizing' maturity level are already implemented, with the establishment of a committee and data ownership, mentioned in the 'Organization' layer: "identify and assign ownership of corporate data objects" (Weber et al., 2009). Although this framework has a strong component on people, it is highly directed to accountabilities, leaving the stakeholders' communication and agreement behind. This can be a great improvement opportunity for this framework, aligning its other strength points with communication and stakeholders buy-in, as well as training, it completes the CSFs on the people area.

Another clear improvement opportunity would be on the processes and procedures area whereas, if they were clearly defined and standardized, along with predictive methods, since other components as review, improvement over time, and measurement are already met, it would allow this framework to step in to the 'Quantitatively Managed' maturity level in this area. On a lower scale, as data strategy is described as one of the layers in this data governance model, being in the 'Defined' maturity level, it could be important to take the next step through the constant improvement over time. On the other hand, policies, tools and technologies, and sustainability have a big gap in this framework, and would require the most effort to get to higher maturity levels. Further development on a maturity assessment component within the organizations, to identify the level of data governance maturity in the organization, would complement this framework objective of individualization of data governance models according to each organization.

5.1.2. Framework 2 – Data Governance Structure

Since training is perceived as one of the critical success factors for data governance in this article, and it has attributed responsibility to the Data Steward, it could be further explored in order to achieve the succeeding maturity levels. Rather than simply attributing the responsibility to one group of stakeholders, it could be documented and structured, preferably in standards that encompass processes, tools, and methods, and after, quantitatively measured. In line with the Roles and Responsibilities maturity level assignment to this framework, the inclusion of a review component in the structured plan could move this framework from 'Managed' to 'Defined' maturity level. From there, the application of measurement and predictive methods could bring it to 'Quantitatively Managed' level.

On the other critical success factors, this framework has to come a long way to achieve other frameworks maturity levels. Nonetheless, it is relevant to mention that this article dates 2007 and, as such, Data Governance and its frameworks' presence in the literature were not as developed as they are today. Even with these gaps, this framework brought a solution to one of the most mentioned difficulties in data governance, the collaboration between data governance and IT, as well as its "separation". In organizations where this is the main bottleneck for data governance it can be a good starting point to mitigate those issues.

5.1.3. Framework 3 – Data Governance Matrix

This framework was very important in the development of further frameworks and in the discussion of data governance in the literature. The objectives of the authors with this framework were accomplished, as it does is used by both practitioners and researchers, both to identify data governance problematics and to design other data governance frameworks. Overall, this framework has a strong base to a data governance framework, with all the CSFs in the descriptions of roles and responsibilities. Passing those guidelines to a structured document and elaborating a formal data governance plan, this framework would have much higher levels of maturity, possibly being between the most matured ones.

5.1.4. Framework 4 – DGI Data Governance Framework

In a general perspective, this framework is very complete, with references to all CSFs. However, mentioning them might not be enough to all organizations, especially the ones that do not have a strong structure to effectively implement these CSFs. As a final statement, the article includes the recommendations of DGI on formal, documented, and repeatable procedures for: Aligning Policies, Requirements, and Controls; Establishing Decision Rights; Establishing Accountability; Performing Stewardship; Managing Change; Defining Data; Resolving Issues; Specifying Data Quality Requirements; Building Governance Into Technology; Stakeholder Care; Communications; and Measuring and Reporting Value (Thomas, 2006). Within this scope, there are all the CSFs for data governance, which would lead this framework to the ultimate maturity level: Optimizing. Nonetheless, these are not explained in detail in the framework, which leads to lower maturity levels on each CSF. With a deeper development of these points, this framework would have the most matured level and, according to this assessment, represent a better solution to implement a data governance initiative.

5.1.5. Framework 5 – SAS Data Governance Framework

The training CSF is an improvement opportunity for this framework, where the right skills in the right people – as the framework identifies as essential – could be assured in the framework itself, through the design of a structured plan of data-related training. In this manner, organizations that do not have this type of knowledge can still leverage data governance at its full extent. As for the other CSFs, some have the highest maturity levels in this scope of frameworks, with special mention to sustainability. Through a further development of such components and its constant improvement, this framework could reach the highest maturity level.

5.2. VERTICAL/FRAMEWORK SCORE

This vertical analysis, or within framework, accomplishes the main objective of this assessment matrix: to evaluate each framework and its maturity levels according to the CSFs for data governance. The metric used will be mode, as this is a qualitative variable.

5.2.1. Framework 1 – One Size Does Not Fit All

Throughout this work it has been reinforced that this framework has a very strong people component, namely on accountabilities. This is depictable as its vertical score on the "people" area of the matrix is fairly higher than in all the other areas. On the other hand, tools and technologies, policies, and sustainability are the biggest gaps in this framework. Overall, its score is Initial.

5.2.2. Framework 2 – Data Governance Structure

This framework has a similar score as the one before, only with a more mature level in data requirements. However, as described above, this framework could have important improvements to elevate to higher levels of maturity. In general, its score is Initial.

5.2.3. Framework 3 – Data Governance Matrix

The Data Governance Matrix framework has very strong basis on most of the CSF, only lacking the implementation of a formal and documented structured of each area to step into higher maturity levels. The maturity levels are tied between Initial and Managed, which will lead to a score of Managed, as there is one maturity level as Defined.

5.2.4. Framework 4 – DGI Data Governance Framework

The DGI framework is fairly complete, with mentions and actions in all CSF areas and their explanation, however it does not have described the manner that organizations should implement such efforts in order to be more matured in data governance. Its score is Managed.

5.2.5. Framework 5 – SAS Data Governance Framework

This framework has the higher maturity levels in most CSF due to its completeness and detailed explanations, through the use of examples and guidelines that are easy to follow. Its score is Managed.

Frameworks	*1 - One Size Does Not Fit All*	*2 - Data Governance Structure*	*3 - Data Governance Matrix*	*4 - DGI Data Governance Framework*	*5 - SAS Data Governance Framework*
Vertical score	Initial	Initial	Defined	Managed	Managed

Table 5 - Vertical score (Frameworks)

5.3. HORIZONTAL/CSF SCORE

In addition to a vertical analysis, within framework, the horizontal analysis, within CSF, can also bring important insights, especially when looking at a specific area of data governance. This can identify which framework, from the ones in the matrix, provides the highest maturity level in each CSF. On one hand, it can help organizations to choose which CSF are more important or they need the most according to its paradigm and, on the other hand, contribute to the literature and further framework development as a basis of objectives to accomplish. Moreover, this analysis identifies the most common gaps in data governance frameworks and, thus, the areas that need further research.

5.3.1. CSF 1 – Training

This CSF is fairly unexplored in frameworks, with a few mentions but not structured or formalized. Nevertheless, in most frameworks, there is the identified need for experience, with the right skilled stakeholders to the success of data governance. Indeed, it is important that this is stated in the framework so that organizations understand that, however roles and responsibilities are a fairly important component of data governance, without guidance on how to achieve this, specially through the available resources to contribute for the initiatives sustainability, it can be overwhelming or even dispendious. Accordingly, this CSF area should be more explored by both practitioners and researchers, preferably with guidelines on what capabilities are needed to each role and responsibility, and how to design a program that can deliver them. The overall score is Initial.

5.3.2. CSF 2 – Roles and Responsibilities

Roles and responsibilities are one of the common areas within most frameworks. In the literature, in particular, this topic is highly recognized as one of the most important when establishing a data governance initiative, especially after the compliance regulations started to appear. The overall score on this CSF is Managed, being common to all frameworks. The biggest gap in these frameworks, which stops them from having a Defined maturity level or even higher in this CSF, is the lack of a review plan according to the performance of the initial roles and responsibilities. Nonetheless, this area is highly explored by both practitioners and researchers, only lacking a deeper research on the continuity and revision of this plan according to performance metrics.

5.3.3. CSF 3 – Buy-in

Although this CSF is being more explored in more practice-focused frameworks, its overall score is Initial. This is an underexplored area that can have such strong effect on how the data governance initiative is accepted and taken by the employees, which can have major effects on their performance in this program. Accordingly, this area should be further recognized and explored by researchers and practitioners in order for the initiative to go smoothly and contribute to its long-term success. Its overall score is Initial.

5.3.4. CSF 4 – Communication

Transparent and clear communication is important in many aspects of the organizations, which is no different in data governance. Being an underexplored area, data governance can sometimes be difficult to understand and perceive its value, especially because its measurement is fairly complicated to demonstrate. In this sense, by having clear communication, stakeholders are aware of their roles and responsibilities, avoiding mistakes and imperception. Its overall score is Managed, which leads to assume that these frameworks do not have in consideration that the communication strategy may need to be adjusted according to multiple factors, such as the state of the program, shifts in the organization culture and strategy, and underperformance. It is recommended that this area is further explored in this sense, elevating its maturity level.

5.3.5. CSF 5 – Data Strategy

Data Strategy score is Managed, which suggests a fairly acknowledge of its importance in most frameworks. Nevertheless, as with all strategies, this can suffer alterations along the time, whether it is due to change of paradigms, for example, or to adjust according to the program performance. It is important to have such changes in mind and build *a priori* a data governance program that can encompass them without failure risk.

5.3.6. CSF 6 – Processes and Procedures

This CSF score is Defined, as one of the frameworks has a higher maturity level than this. People-focused frameworks, as it is the example of the first two, with Initial maturity levels, reference this CSF as part of the responsibilities of some roles in data governance, not further exploring them individually. The other three frameworks have higher maturity levels because this is one of its main focus, not only as stakeholder's responsibility, but in a technical and technological manner as well.

5.3.7. CSF 7 – Policies

Policies are the guidelines that stakeholders must follow in order for the program to run smoothly, guaranteeing consistency and trustability in data governance outcomes. In this line, it is very important in an initiative as data governance, to ensure its success and prosperity. The overall score is Initial, which is concernable because it can have major impacts on the initiative results. It is recommended that this area is fairly more explored in data governance frameworks, with a special attention to its documentation and formalization, which is the first step towards a higher maturity level.

5.3.8. CSF 8 – Data Requirements

In order to make decisions based on data, it is important to have the proper data for such decisions, which initiates with defining what does "proper data" mean for each decision. This helps building the data requirements, which define which type of data, in what manner, and how it should be handled. The people involved in the flow of data between its extraction and its delivery for decision making should have a clear understanding of these requirements in order to deliver what is necessary. This CSF score is Managed, as the review component of this plan is not yet contemplated in most frameworks. As other factors, these can change according to the organization state and these changes should be part of this plan to avoid failure risk.

5.3.9. CSF 9 – Tools and Technologies

This is an underexplored area, with only an Initial score. The lack of recognition on how important it is to have automated, embedded data governance processes can held the organizations from achieving its data objectives, delaying the initiative success. Not only these allow for greater amounts of data and reducing processing times, but it also has a great part in other CSF, boosting its performance and pace, and enabling to go further. Having a tools and technologies plan can boost the data governance initiative success and sustainability within organizations.

5.3.10. CSF 10 – Assessment

Prior to initiate a data governance initiative, it is important to assess its presence in the organization in order to define the strategy according to what the organization already has. Tools as an assessment matrix for the organization are recommended before the design of the data governance initiative. The overall score in this CSF is Initial.

5.3.11. CSF 11 – Sustainability

The score on this CSF is Initial, which was already expected, since it is one of the identified gaps in this work. Data governance should not be a short-term project, but rather a continuous program to ensure that data is treated as an asset in the organization and leveraging its results as part of organizational growth. Accordingly, sustainability should be one of the bases in a data governance program, to ensure its durability and long-term success.

CSF	*Horizontal score*
1 – Training	Initial
2 – Roles and Responsibilities	Managed
3 – Buy-in	Initial
4 – Communication	Managed
5 – Data Strategy	Defined
6 – Processes and Procedures	Defined
7 – Policies	Initial
8 – Data Requirements	Managed
9 – Tools and Technologies	Initial
10 – Assessment	Initial
11 – Sustainability	Initial

Table 6 - Horizontal score (CSF)

6. CONCLUSION AND FUTURE WORK

In this work, some of the main gaps in a sample of data governance frameworks were identified through the design and implementation of an assessment matrix. This assessment matrix is focused on the maturity levels of five data governance frameworks within the critical success factors of a data governance initiative identified in the literature. Through the analysis of the results on this assessment, it is depictable that some areas of data governance require deeper exploration in both data governance literature and the design and implementation of data governance frameworks.

As the main bottleneck for the achievement of higher maturity levels within the data governance frameworks, the concepts of review and constant improvement of the plans in the specific areas, along with its effective measurement, constitutes the one of biggest gaps found in this work. Accordingly, it is recommended that these techniques are further explored and included in data governance frameworks on all critical success factors, in order to allow organizations to attain higher data governance maturity through the use of such initiatives.

The implementation of the assessment matrix, even though it was only applied to five data governance frameworks, provided essential insights on data governance and data governance frameworks gaps, and possible explanations on the lack of success of data governance initiatives in organizations. It is, however, recommended that this tool is further explored through a case study to reevaluate the maturity level descriptions based on experts and stakeholders' opinions, as well as its implementation on both more data governance frameworks and organizational contexts.

More important than comparing the frameworks, is to understand that its design and scope can prevent organizations from thriving from data governance initiatives. This is depicted as most frameworks, when evaluated through the formal maturity levels, performed more poorly than what was previously expected. Moreover, the comparison should not be made based on solely its maturity, but have also in consideration many other factors, such as the date it was developed, the scope it is designed, its objectives and the gaps it was made to fulfill, as well as recognition from experts. In addition, organization's needs, resources, and objectives are often disparate, as each data governance framework can fit those in different manners. Rather than trying to find the "best" solution, it should be the goal to both critically analyze and improve the available solutions, and fitting the organization with the right solution for their own context.

Due to its broad spectrum of definition of maturity levels, this tool can be implemented in other disciplines, programs, or projects, adapting the critical success factors and frameworks. It can also serve as an inspiration to use the available tools – namely assessment matrixes – in different manners, bringing new paradigms and perspectives to the literature. In addition, these and other frameworks, data governance related or not, could be evaluated in the scope of other comparison methods available in the literature.

7. BIBLIOGRAPHY

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, *49*, 424–438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008

Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2016). A Conceptual Framework for Designing Data Governance for Cloud Computing. *Procedia Computer Science*, *94*, 160–167. https://doi.org/10.1016/j.procs.2016.08.025

Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2019). A systematic literature review of data governance and cloud data governance. *Personal and Ubiquitous Computing*, *23*(5–6), 839–859. https://doi.org/10.1007/s00779-017-1104-3

Alhassan, I., Sammon, D., & Daly, M. (2018). Data governance activities: a comparison between scientific and practice-oriented literature. *Journal of Enterprise Information Management*, *31*(2), 300–316. https://doi.org/10.1108/JEIM-01-2017-0007

Alhassan, I., Sammon, D., & Daly, M. (2019). Critical Success Factors for Data Governance: A Theory Building Approach. *Information Systems Management*, *36*(2), 98–110. https://doi.org/10.1080/10580530.2019.1589670

Attard, J., & Brennan, R. (2018). Challenges in value-driven data governance. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11230 LNCS*, 546–554. https://doi.org/10.1007/978-3-030-02671-4_33

Benfeldt, O., Persson, J. S., & Madsen, S. (2019). *Why Governing Data Is Difficult : Findings from Danish Local Government*. (January). https://doi.org/10.1007/978-3-030-04315-5

Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data Governance as a Collective Action Problem. *Information Systems Frontiers*, *22*(2), 299–313. https://doi.org/10.1007/s10796-019-09923-z

Berndtsson, M., Forsberg, D., Stein, D., & Svahn, T. (2018). Becoming a data-driven organisation. *Twenty-Sixth European Conference on Information Systems (ECIS2018)*, (January). Portsmouth, UK.

Bhansali, N. (2014). *DATA GOVERNANCE Creating Value from Information Assets* (N. Bhansali, ed.). Retrieved from http://docplayer.net/2788186-Data-governance-creating-value-from-information-assets.html

Brous, P., Janssen, M., & Vilminko-Heikkinen, R. (2016). Coordinating decision-making in data management activities: A systematic review of data governance principles. *IFIP International Federation for Information Processing*, *9820 LNCS*, 115–125. https://doi.org/10.1007/978-3-319-44421-5_9

Cave, A. (2017). *Exploring Strategies for Implementing Data Governance Practices* (Walden University). Retrieved from https://scholarworks.waldenu.edu/dissertations

Cheong, L. K., & Chang, V. (2007). The need for data governance: A case study. *ACIS 2007 Proceedings - 18th Australasian Conference on Information Systems*, (June), 999–1008.

CMMI Product Team. (2006). *CMMI ® for Development, Version 1.2 Improving processes for better products*. Retrieved from http://www.sei.cmu.edu/publications/pubweb.html

DalleMule, L., & Davenport, T. H. (2017). What’s Your Data Strategy? Retrieved January 18, 2021, from Harvard Business Review website: https://hbr.org/2017/05/whats-your-data-strategy

Diván, M. J. (2017). Data-Driven Decision Making. *2017 International Conference on Infocom Technologies and Unmanned Systems (ICTUS'2017)*. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8285973

European Comission. (n.d.). Data protection in the EU | European Commission. Retrieved December 2, 2020, from https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

Fu, X., Wojak, A., Neagu, D., Ridley, M., & Kim, T. (2011). Data governance in predictive toxicology: A review. *Journal of Cheminformatics*, *3*(24), 1–16. https://doi.org/10.1186/1758-2946-3-24

Gupta, U., & Cannon, S. (2020). *A Practitioner's Guide to Data Governance: A Case-Based Approach* (1st ed.). Retrieved from https://books.google.pt/books?id=KwLuDwAAQBAJ&pg=PA228&lpg=PA228&dq=Foundations+of+Data+Governance+uma+gupta+pdf&source=bl&ots=wuz-PXMkw7&sig=ACfU3U3821Z406OqcH9-RDN_zAlYxueSTA&hl=pt-PT&sa=X&ved=2ahUKEwiL5I_skarsAhUqAWMBHUyCCHwQ6AEwB3oECAEQAg#v=onepage&q=Foundations of Data Governance uma gupta pdf&f=false

Haneem, F., Kama, N., Taskin, N., Pauleen, D., & Abu Bakar, N. A. (2019). Determinants of master data management adoption by local government organizations: An empirical study. *International Journal of Information Management*, *45*, 25–43. https://doi.org/10.1016/j.ijinfomgt.2018.10.007

Hannila, H., Silvola, R., Harkonen, J., & Haapasalo, H. (2019). Data-driven Begins with DATA; Potential of Data Assets. *Journal of Computer Information Systems*. https://doi.org/10.1080/08874417.2019.1683782

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, *37*(3). https://doi.org/10.1016/j.giq.2020.101493

Khatri, V., & Brown, C. V. (2010). Designing Data Governance. *Communications of the ACM*, *53*(1), 148–152. Retrieved from https://dl.acm.org/doi/pdf/10.1145/1629175.1629210

Kim, H. Y., & Cho, J.-S. (2018). Data governance framework for big data implementation with NPS Case Analysis in Korea. *Journal of the Academy of Business and Retail Management Research (JBRMR)*, *12*(3), 36–46. Retrieved from www.jbrmr.com

Kiron, D., & Ransbotham, S. (2017). Analytics as a Source of Business Innovation. *MIT Sloan Management Review*. Retrieved from https://sloanreview.mit.edu/projects/analytics-as-a-source-of-business-innovation/

Kohlegger, M., Maier, R., & Thalmann, S. (2009). *Understanding Maturity Models Results of a Structured Content Analysis*. Retrieved from http://www.sei.cmu.edu/cmmi/

Koltay, T. (2016). Data governance, data literacy and the management of data quality. *Internal Federation of Library Associations and Institutions*, *42*(4), 303–312. https://doi.org/10.1177/0340035216672238

Korhonen, J. J., Melleri, I., Hiekkanen, K., & Helenius, M. (2013). Designing Data Governance Structure: An Organizational Perspective. *GSTF Journal on Computing (JOC)*, *2*(4), 11–17. https://doi.org/10.5176/2251-3043

Ladley, J. (2020). *Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program* (2nd ed.; M. Conner & P. Llewellyn, Eds.). Retrieved from https://books.google.pt/books?hl=pt-PT&lr=&id=AkW9DwAAQBAJ&oi=fnd&pg=PP1&dq=data+governance+&ots=OO-QPPDNvK&sig=VhChCIMZKkKtXKIvWWhrON0rhms&redir_esc=y#v=onepage&q&f=false

Lee, Y. W., Madnick, S. E., Wang, R. Y., Wang, F. L., & Zhang, H. (2014). A cubic framework for the chief data officer: Succeeding in a world of big data. *MIS Quarterly Executive*, *13*(1), 1–13.

Mullon, P. A., & Ngoepe, M. (2019). An integrated framework to elevate information governance to a national level in South Africa. *Records Management Journal*, *29*(1), 103–116. https://doi.org/10.1108/RMJ-09-2018-0030

Nielsen, O. B. (2017). A Comprehensive Review of Data Governance Literature. *Selected Papers of the IRIS*, (8). Retrieved from http://aisel.aisnet.org/iris2017/3

Otto, B. (2011a). A morphology of the organisation of data governance. *ECIS 2011 Proceedings*, (272).

Otto, B. (2011b). Data governance. *Business and Information Systems Engineering*, *3*(4), 241–244. https://doi.org/10.1007/s12599-011-0162-8

Otto, B. (2011c). Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers. *Communications of the Association for Information Systems*, *29*(3), 45–66. https://doi.org/10.17705/1CAIS.02903

Otto, B., & Lis, D. (2020). Data Governance in Data Ecosystems-Insights from Organizations. *AMCIS 2020 Proceedings*. Retrieved from https://aisel.aisnet.org/amcis2020/strategic_uses_it/strategic_uses_it/12

Panian, Z. (2010). Some Pratical Experiences in Data Governance. *World Academy of Science, Engineering and Technology Management*, *62*, 939–946.

Paulk, M. C., Weber, C. V, Garcia, S. M., Chrissis, M. B., & Bush, M. (1993). *Key Practices of the Capability Maturity Model, Version 1.1*.

Ransbotham, S., & Kiron, D. (2017). Analytics as a Source of Business Innovation. *MIT Sloan Management Review*. Retrieved from https://sloanreview.mit.edu/projects/analytics-as-a-source-of-business-innovation/

Rivera, S., Loarte, N., Raymundo, C., & Dominguez, F. (2017). *Data Governance Maturity Model for Micro Financial Organizations in Peru*. https://doi.org/10.5220/0006149202030214

SAS Institute Inc. (2018). *The SAS ® Data Governance Framework: A Blueprint for Success [White Paper]*. Retrieved from https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/sas-data-governance-framework-107325.pdf

Smallwood, R. F. (2014). *Information Governance Concept Strategies and Best Practices*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Thomas, G. (2006). *The DGI data governance framework*. Retrieved from http://www.datagovernance.com/%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:DGI+Data+Governance+Framework#0

Thompson, N., Ravindran, R., & Nicosia, S. (2015). Government data does not mean data governance: Lessons learned from a public sector application audit. *Government Information Quarterly*, *32*(3), 316–322.

Vial, G. (2020, October 7). Data Governance in the 21st-Century Organization. Retrieved January 18, 2021, from MIT Sloan Management Review website: https://sloanreview.mit.edu/article/data-governance-in-the-21st-century-organization/

Watson, H. J. (2016). Creating a Fact-Based Decision-Making Culture. *Business Intelligence Journal*, *21*(2), 5–9.

Weber, K., Otto, B., & Österle, H. (2009). One Size Does Not Fit All—A Contingency Approach to Data Governance. *ACM Journal of Data and Information Quality*, *1*(1), 27. https://doi.org/10.1145/1515693.1515696.http

Were, V., & Moturi, C. (2017). Toward a data governance model for the Kenya health professional regulatory authorities. *TQM Journal*, *29*(4), 579–589. https://doi.org/10.1108/TQM-10-2016-0092

Yallop, A. C., & Aliasghar, O. (2020). No business as usual: a case for data ethics and data governance in the age of coronavirus. *Online Information Review*, *44*(6), 1217–1221. https://doi.org/10.1108/OIR-06-2020-0257

Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology*, *58*(4), 479–493. https://doi.org/10.1002/asi.20508