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

Patrícia Bento, Miguel Neto & Nadine Côrte-Real

This is the final, accepted version of the conference contribution published by *IEEE*

Bento, P., Neto, M., & Corte-Real, N. (2022). How data governance frameworks can leverage data-driven decision making: A sustainable approach for data governance in organizations. In A. Rocha, B. Bordel, F. G. Penalvo, & R. Goncalves (Eds.), 2022 17th Iberian Conference on Information Systems and Technologies (CISTI): Proceedings (pp. 1-5). (Iberian Conference on Information Systems and Technologies, CISTI). IEEE Computer Society. https://doi.org/10.23919/CISTI54924.2022.9866895

How data governance frameworks can leverage data-driven decision making

a sustainable approach for data governance in organizations

Patrícia Bento
NOVA Information Management School,
Universidade Nova de Lisboa
Lisboa, Portugal
M20190053@novaims.unl.pt

Miguel Neto
NOVA Information Management School,
Universidade Nova de Lisboa
Lisboa, Portugal
mneto@novaims.unl.pt

Nadine Côrte-Real
NOVA Information Management School,
Universidade Nova de Lisboa
Lisboa, Portugal
nreal@novaims.unl.pt

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 (DG) 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 (CSF) 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 CSF 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..

I. INTRODUCTION

The aim of this study is to provide an assessment tool, that delivers 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 (CSF) for data governance.

As competition arises, enterprises are pressured to make good decisions and deliver value for its customers. In this context, data rise 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 [1].

While working with data seems crucial to sustain market position, it has risen several challenges for organizations. Although data are 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 [2].

In this context, data governance has created the path 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.

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 [3]. Implementing a data governance initiative and thriving from it is very challenging due to its ambiguity and abstraction, and the lack of interest in this subject leads to an erroneous approach in what regards data governance processes. Furthermore, there is limited research on sustainable data governance frameworks [4], which minimizes the importance of the subject in the companies.

Accordingly, the value of this research is providing a tool that identifies the strenghts and improvement points on the existing frameworks, as well as the gaps in the literature concerning the CSF for data governance.

II. DATA GOVERNANCE

Data Governance is described as a working-term, as it has no formal definition. In some articles, it is based on other definitions, adapting them to their own contexts [5][6][7]; in other articles, it is advised that organizations adapt their definition in accordance to its culture and environment, avoiding resistance from the stakeholders involved [8].

Some authors inspire their definition of data governance in the definition of IT governance by Weill and Ross [9] as "the decision rights and accountability framework for encouraging desirable behaviors in the use of IT" [10][11]. Khatri and Brown [12], 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" [12].

Abraham et al. [5], on the other hand, 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.

III. 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 [13] 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.

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.

The selection of the data governance frameworks to include in this assessment matrix was based on the following criteria: 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 [14][6][15]. 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 (Table 1) were selected to be assessed against the CSFs identified previously, based on the different maturity levels.

Frameworks	***One Size Does Not Fit All [11]***	***DG Structure [16]***	***DG Matrix [12]***	***DGI Framework [8]***	***SAS DG Framework [17]***
Context	Academic	Academic	Academic	Practical	Practical
Year	2009	2007	2010	2006	2018(?)
Areas of strenght	Flexible & tailored. Reuse of resources.	Collaboration between IT & business. Detailed roles and responsibilities.	Important contribute to the literature. Basis to other articles. New perspective of data decision domains	Oriented for results. Detailing of the different levels of DG.	Focus on DG issues & sustainability. Holistic & pragmatic approaches with "quick-wins"
Areas to improve	May require an perception of resources and needs. Requires maturity in DG.	Lack of practical guidelines. Focused almost exclusively on roles and responsibilities.	Lack of practical guidelines.	Lack of tools and technologies.	Requires a maturity assessment. Requires maturity in DG.

TABLE 1. Frameworks comparision

IV. CSF AND MATURITY LEVELS

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. These practices are defined as critical success factors, which describe the areas of activity in data governance that require the most attention to have a successful data governance program.

For this research, were used eleven CSF based on the combination of three papers [15][18][6] (Table 2).

CSF	*Description*	*Suggested actions*	*Reference*
Employees data competencies	Capability and awareness to handle data manners.	Continuous training in DG. Employee's awareness to critical and sensible data.	[15]
Clear data processes and procedures	If not guaranteed, leads to lack of trusting in data.	Embed and optimize processes and procedures into the system.	[15]
Flexible data tools and technologies	Software and hardware that affect data.	Appropriate IT infrastructure and integrated data, with automatation and testing.	[15]
Standardized easy-to-follow data policies	Implementation and clarity. Unifying, simplifying and renovating these policies.	Tools and technologies should also be implemented to monitor and update data policies.	[15]
Established data roles and responsibilities	Unclarity on what are the roles and what responsibilities each role has.	Establishing a committee and identify data owners.	[15]
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.	[15]
Focused and tangible data strategies	Planning for DG, with an orientation for achieving short-term and long-term objectives.	Establishing an executive team to encourage organization-wide structure for data governance.	[15]
Obtaining stakeholder buy-in'	Involving and reaching agreements, contributing to increase interest and commitment.	Debates and feedback.	[18]
Effective and strategic communications	Importance of communication and alignment with the organization goals.	Disseminating communication aligned with organizational goals.	[18]
Assess data governance situation	Assessing the current state of DG in the organization before implementing an initiative.	Data governance maturity level assessment.	[6]
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.	[6]

TABLE 2. Critical Success Factors

These CSF were used to evaluate the DG Frameworks, based on the maturity levels of CMMI Maturity Models. These levels are used 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" [19].

For this research, the maturity levels were ajusted to fit the CSF reality (Table 3).

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
Optimized	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 3. Maturity Levels

V. 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.

As mentioned before, the sustainability of a DG 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.

The assessment matrix (Table 4) for these framworks was designed as follows: for each framework, direct or indirect mentions to each CSF were identified, then the respective maturity level was attributed, in accordance to its presence in the framework.

CSF / Frameworks	*One Size Does Not Fit All [11]*	*Data Governance Structure [16]*	*Data Governance Matrix [12]*	*DGI Data Governance Framework [8]*	*SAS Data Governance Framework [17]*
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 1. Assessment Matrix

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

The assessment matrix depicts a prominence of lower levels, namely, Initial and Managed. This suggests a large room for improvement in what regards these frameworks and its maturity levels in this CSF.

Further analysing this assessment matrix, it is possible to measure vertical and horizontal scores, based on mode of the maturity levels depicted. This analysis provides a more precise look on the state of each CSF and each framework.

On a vertical perspective, the score identifies the mode for each framework, summarizing its maturity level in the CSF. In this line, *One Size Does Not Fit All* [11] and *Data Governance Structure* [16] have an overall maturity level of Initial, *DGI Data Governance Framework* [8] and *SAS Data Governance Framework* [17] are on the Managed maturity level, and *Data Governance Matrix* [12] has the highest score, with Defined maturity level. These results depict the need for further development of the frameworks, particularly in what regards the use of quantitative measures, and establishing a review and improvement of the defined plan.

On the other hand, the horizontal score identifies the mode on each CSF, summarizing its maturity level in the five frameworks. On the Initial maturity level there are the CSF: 'Training', 'Buy-in', 'Policies', 'Tools and technologies', 'Assessment' and 'Sustainability'; the Managed maturity level includes the CSF: 'Roles and Responsibilities', 'Communication' and 'Data Requirements'; finally, on the Defined maturity level: 'Data Strategy' and 'Processes and Procedures'. As for the vertical score, this enlightens the need to further develop these CSF within the frameworks and the literature.

Both vertically and horizontally it is depicted a prominence of lower levels, reaching only the third level – Defined. Although other factors can influence the sustainability of a data governance initiative, these can contribute to a new perspective on how to attain the required sustainability for the success of such initiative.

VI. 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.

REFERENCES

[1] 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

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

[3] 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

[4] Abueed, R. A. I., & Aga, M. (2019). Sustainable Knowledge Creation and Corporate Outcomes: Does Corporate Data Governance Matter? *Sustainability*, *11*. https://doi.org/10.3390/su11205575

[5] 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

[6] 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

[7] 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

[8] 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

[9] Weill, P., & Ross, J. W. (2004). *IT Governance on One Page* (No. 349). Retrieved from http://web.mit.edu/cisr/www

[10] Otto, B. (2011). A morphology of the organisation of data governance. ECIS 2011 Proceedings, (272).

[11] 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

[12] 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

[13] 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

[14] 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

[15] 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

[16] 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.

[17] 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/sasdata-governance-framework-107325.pdf

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

[19] 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