Analysis of the Relationship Between Data Governance and Data-Driven Culture

Felipe F. Salerno and Antonio Carlos G. Maçada

October 24, 2022

ANALYSIS OF THE RELATIONSHIP BETWEEN DATA GOVERNANCE AND DATA-DRIVEN CULTURE

Felipe Fonseca Salerno [1] [0000-0003-1305-7957] and Antônio Carlos Gastaud Maçada [1] [0000-0002-8849-0117]

[1] Escola de Administração, Universidade Federal do Rio Grande do Sul – UFRGS, Porto Alegre, Brazil

00193140@ufrgs.br

Abstract. As organizations increase their use of data, among scholars there is growing interest in data governance and data-driven culture, and studies suggest investigating the relationship between these two phenomena would provide a better understanding of data behavior in organizations. Thus, this exploratory research investigates the relationship between data governance and data-driven culture using partial least squares structural equation modeling (PLS-SEM). The results show the relationship between data governance and data-driven culture is strong, and that it is mediated by data quality. Additionally, based on the Resource-Based View of the firm, our results indicate data governance and data-driven culture should be addressed jointly when evaluating their contribution as an organizational resource.

Keywords: Data governance, data-driven culture, partial least squares structural equation modeling (PLS-SEM).

1 Introduction

Data governance[1] is defined as the exercise of authority and control over data, in which data is considered a strategic organizational asset [1]. According to Fan [2], it enables better decision-making and protects stakeholders' needs concerning data, including planning and supervision activities aimed at ensuring and improving data quality. Liaw *et al.* [3] highlight the alignment between data governance and organizational strategy, arguing that data is used to sustain and promote the achievement of the organization's objectives.

Abraham *et al.* [4] and Magnusson *et al.* [5] suggest that when analyzing data governance, the structural, procedural, and relational mechanisms should be considered. The structural mechanisms refer to the roles and responsibilities within an organization, including the definition of authorities for decision-making concerning data. The procedural mechanisms include the data strategy, policies, rules, and

[1] Note: Throughout this article the term “data governance” is used as a synonym of “information governance” [17].

standards that guide the treatment, use, storage, and security of data. Finally, the relational mechanisms are those that collaborate in data-related communication initiatives, such as training and data awareness [4, 5].

It is argued that data governance and its mechanisms enhance data quality. Wende [6] claims data quality has a direct relationship with governance, since when defining norms, standards and authorities, the organization expects to improve the quality of its information. A similar view is shared by Brous *et al.* [7], who conducted a case study and found data governance collaborated to increase data quality, especially in organizations operating large volumes of information. In the authors' view, this gain in quality contributes towards improving organizational results.

In the same vein, Ladley [8] argues that data governance is essential for any organization that wants to improve data quality and become data-driven, i.e. use data as an organizational asset. Anderson [9] defines ‘being data-driven’ as "the construction of tools, skills and, most importantly, a culture that acts based on data". For the author, good quality data is a fundamental requirement for an organization to become data-driven, and key to reliable decision-making.

Regarding data-driven culture[2], Huppertz *et al.* [10] suggest it is closely related to digital transformation. For the authors, the routine and strategic use of data is fundamental for organizations operating in a technological ecosystem, while also being a vector of innovation and strategic orientation. Chatterjee *et al.* [11] point out that a data-driven culture impacts business performance, collaborating to disseminate familiarity with data and its related activities among managers and employees, such as the use of dashboards, artificial intelligence, and analytics. Thus, there is considerable evidence to show fostering a data-driven culture helps organizations achieve their strategic objectives.

One theory that facilitates the analysis of the role of data-driven culture is the Resource-Based View (RBV) of the firm, which argues that organizations are able to improve their competitiveness and performance through their resources [12]. Wade and Hulland [13] highlight the relevance of RBV in the context of information systems research, arguing the theory permits the analysis of the impact of a particular resource and its effect on the organization as a whole. Under this lens, data is seen an organizational resource, which, if used appropriately, would allow the organization to become more competitive. Chatterjee *et al.* [11] also identify a relationship between RBV and data-driven culture, stating that recognizing data as an organizational asset contributes towards the development of a data-driven culture.

Recent studies, such as those from Mikalef *et al.* [14-16] have investigated the wider relationship between data-driven culture and data governance, addressing issues such as big data, innovation, and performance, and indicating the need for further research to explore the practical implications and new theoretical lenses. From the RBV perspective, it is argued a strong relationship between data governance and data-driven culture could be considered a resource and assist organizations in achieving their strategic objectives. Therefore, the present study investigates that relationship using

[2] Note: Throughout this article the terms "data culture", "data-driven culture" and "data-oriented culture" are used as synonyms [18].

partial least squares structural equation modeling (PLS-SEM) from the perspective of the RBV, adapting previous research [14-16] to address the topics in question.

This study seeks to quantify the relationship between data governance and data-driven culture and, in so doing, detail and clarify its nature. Recent studies, such as Chatterjee *et al.* [11], highlight the importance of understanding this relationship for managers that want to improve organizational performance. The same authors also suggest the RBV provides a new interpretation of the topic, contributing towards the advance of scholarly research in the area.

Among the benefits of this research are the insights it provides into the impact of data governance on data-driven culture by detailing the extent of their relationship. It also complements the quantitative models from Mikalef *et al.* [14-16] by providing new insights into one of the themes investigated by those authors. Finally, it also provides subsidies for the use of RBV, contributing towards the development of theory and an understanding of data governance and data-driven culture as resources capable of ensuring competitive advantage. Thus, the findings of this study are expected to help managers see data governance as a means of promoting the development of a data-driven culture.

2 Concepts and theory

2.1 Data Governance

With the onset of the digital transformation and the increasing volume of data held by organizations, data governance has become a popular topic in information systems research. Data governance is defined as the exercise of authority and control over data, where data is used as a strategic asset and collaborates to reduce data-related risks [1]. In a synthetic definition, Kremser and Brunauer [18] point out that it consists of "formal implementation and enforcement of the authority on data management and data-related assets". Thus, data governance consists of the definition of policies, standards, procedures, structures, and other initiatives related to data strategies, such as use, storage, security, and data sharing [1, 4, 18]. Frequently, data governance is presented through frameworks in which action dimensions and the main processes that interact with it are represented [4, 19, 20], which facilitates the analysis of data governance from different perspectives.

In addressing data governance, Wende and Otto [19] adopt the contingency lens, in which organizational characteristics, such as size, structure, and decision-making styles directly impact the implementation of data governance. For the authors, data governance concerns the definition of data-related decision structures, which should be designed in harmony with the organizational reality, whereby the principles and procedures for data should be consistent with the contingencies of the organization [19]. From this approach, the main purpose of data governance is to improve data quality, and thus permit better decision-making. For Khatri and Brown [20], one of the main contributions of data governance is to indicate which structure or person within the organization is responsible for data-related decision-making. In other words, data governance has an intimate relationship with decision-making, either by defining decision-making authorities or by supporting it with quality data [20].

From a theoretical perspective, Abraham *et al.* [4] refers to the importance of the structural, relational, and procedural mechanisms of data governance. The structural mechanisms comprise governance and accountability structures, such as defining roles, responsibilities, and authorities for data-related decision-making. The procedural mechanisms include data strategy, policy, standards, processes and procedures, contracts, compliance, problem management, and performance measurement. These mechanisms are mainly intended to ensure data is correctly and securely stored and effectively used. Finally, relational mechanisms consist of communication, training, and the coordination of decision-making to ensure alignment between functions. Therefore, for the authors, data governance can be defined in terms of the structural, procedural, and relational mechanisms of which it is constituted. [4].

Although the approaches adopted by the above-mentioned authors differ, they agree regarding the role of data governance in improving data quality, which is seen as being critical for strategic decision-making. Abraham *et al.* [4] defines data quality as "the ability of data to satisfy use requirements before a context", that is, data quality is not a static factor since it can be interpreted differently according to the intended use. Thus, data quality improvement is seen as a positive consequence of data governance.

In summary, data governance aids the organization by treating data as a strategic asset, improving data quality and consequently the decision-making process. From these two aspects – data quality and decision-making – the next chapter addresses data-driven culture and its relationship to data governance.

2.2 Data-driven Culture

Data-driven culture consists of a specific form of organizational culture that manifests itself through data orientation. For Kremser and Brunauer [18], it is a culture in which organizational decisions are preferably based on insights extracted from data, and access to and use of data is encouraged, so that knowledge and skills are continuously shared. Adding to this, Chatterjee *et al.* [11] point out that a data-driven culture implies a pattern of behavior in which data is seen as critical to the organization's success. A similar view is shared by Anderson [9], who cites the key characteristics of a data-driven organization: data access and sharing, data literacy, goals and indicators, an inquisitive and learning culture, and data leadership. These features are described in greater detail below.

Regarding data access and sharing, both Kremser and Brunauer [18] and Anderson [9] suggest an organization should stimulate broad access to data. Here, the role of data governance is particularly important as it defines policy regarding access, accountability, and confidentiality. Thus, for data-based decision-making to exist, it is a fundamental requirement that people have data available. Kremser and Brunauer [18] use the term "democratization", which conveys the idea that data should be present at all organizational levels, observing possible access limitations.

Data literacy consists of the ability of managers and people in the organization to understand and work with data and its tools. Anderson [9] notes that while it is not necessary to be an expert in statistics, people need to understand the basics of patterns, charts, and tools. While not using the term 'literacy', Berndtsson *et al.* [21] refers to analytical capacity, which suggests the ability of employees to understand and to operate data. Thus, data-driven culture embraces basic knowledge about data and the skills necessary for its use, which can be stimulated through training and learning activities. Here, a parallel can be drawn with the relational mechanisms of data governance that allow for the dissemination of data practices, skills and knowledge.

For Anderson [9] the organizational objectives and strategic direction can be quantitatively represented by goals and indicators that encourage employees and managers to work toward their realization, providing transparency in relation to the expected results. Agyei-owusu *et al.* [22] comment on the relationship between data-driven culture and performance, highlighting the importance of metrics and indicators in verifying the results obtained by the organization. Thus, a data-driven organization uses metrics and indicators in a transparent and shared way between its levels, concepts aligned with data governance mechanisms [4, 9, 22].

In relation to the existence of an inquisitive and learning culture, Anderson [9] points out that in data-driven organizations decisions are often challenged based on numeric evidence. In other words, healthy debate is stimulated through the use of data, promoting discussions that enhance decision-making. Kremser and Brunauer [18] reinforce that aspect, pointing out that discussing decisions using data as an input stimulates the development of a data-driven culture, encouraging the organization to question its decisions based on data evidence.

Finally, the last aspect of a data-driven culture cited by Anderson [9] is data leaders. For the author, these key figures need to actively inspire and promote a data-driven culture, in a top-down approach, ranging from the collection stage to decision-making and organizational learning. Chatterjee *et al.* [11] state that "business leaders need to emphasize the usefulness of establishing a data-driven culture to achieve success." Once again, there is an intersection between data governance and data-driven culture, in which structural mechanisms collaborate in the identification of authorities and decision structures.

2.3 The Resource-Based View

According to the resource-based view (RBV) of the firm, 'resources' are all the assets, capabilities, processes, information, and other items that allow an organization to implement strategies to improve its efficiency and effectiveness, thus constituting a strength that collaborates with organizational strategy [12, 23]. The authors define three categories of resources: physical capital, human capital, and organizational capital. The physical capital consists of physical technologies, such as equipment and its geographical location. Human capital includes resources such as experience, training, relationship, and intelligence. Finally, organizational capital is understood to be resources related to structure, formal and informal planning, and relationships with other organizations, among others [12, 23]. Barney [12] argues that an organization can acquire a sustainable competitive advantage from efficient resource management, and Mahoney and Pandian [24] complement that by pointing out that RBV incorporates strategic perceptions of different organizational competencies and capacities, in which resources play a relevant role in defining business strategy.

According to Barney [12] and Penrose [23], a competitive advantage exists when a strategy of value creation is implemented in the organization, and a sustainable competitive advantage exists when, in addition to creating value, it cannot be copied by competitors. However, this does not mean that a sustainable competitive advantage would be eternal, as it is subject to the effect of major transformations, only that it is not susceptible to being copied by competitors.

Barney [12] goes on to describe four characteristics of resources, namely, that they are valuable, rare, imperfectly imitable, and with no near substitutes; which are fundamental to ensure a sustainable competitive advantage. Resources are valuable when they allow the company to implement strategies that enhance its efficiency and effectiveness. Rare resources are those that are difficult to obtain and seldom combined. Being imperfectly imitable involves three main aspects: historical conditions, causal ambiguity, and social complexity. Finally, resources should not have close substitutes, otherwise they lose their competitiveness [12].

In the field of information systems (IS), Wade and Hulland [13] suggest adopting the RBV enables the relationship between IS, strategy, and organizational performance to be addressed, especially from the perspective of resources and competitive advantages. Moreover, they argue it permits the assessment of different resources and add that, when using the RBV in IS research, as a basic step, it is necessary to specify and detail technological resources, as well as to evaluate their dynamism in the technological environment.

3 Model and hypotheses

3.1 Data Governance

As theorized within the RBV, the efficient management of resources could provide an organization with a sustainable competitive advantage. Accordingly, in the model proposed in this study, data governance and data-driven culture are considered organizational resources. Thus, as a first step, it is necessary to understand and detail the relationship between data governance and data-driven culture, since both could be considered resources capable of providing an organization with competitive advantages [9, 11, 15, 18]. Studies such as Faria *et al.* [25] suggest data governance is a factor associated with the generation of value for the banking industry, while Chatterjee *et al.* [11] identified an association between data-driven culture and organizational performance. That is, recent studies provide evidence that reinforces the potential of data governance and data-driven culture as resources that, according to the RBV, could support sustainable competitive advantages.

Thus, considering the suggestion from Wade and Hulland [13] when using RBV in IS research, the first step is to identify and the detail the resources analyzed under this theoretical lens. Therefore, investigating the relationship between data governance and data-driven culture is primordial, which motivated the establishment of four research hypotheses, according to Fig. 1. The model is exploratory in nature, being the first step towards the further investigation of other components, which explains its generalist approach, and is intended to guide and direct future research on the subject, specifically from the RBV perspective.

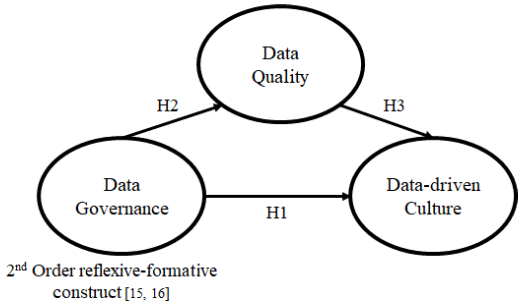

Fig. 1. Research hypotheses

Because data governance defines the standards related to data in the organization, promoting the formalization of data structures, responsibilities, and processes [8, 26], it is assumed to exert a positive effect on data-driven culture. Previous research, such as Anderson [9] and Kremser and Brunauer [18], has indicated a positive association between these constructs, but there is a lack of quantitative studies that explore the magnitude of this association. Theoretically, this positive effect would be due to the fact data governance contributes towards raising the awareness of decision-makers regarding their responsibilities and competencies. Thus, by organizing and mapping the processes involving data in the organization, governance allows managers to identify which data should be analyzed, thus overcoming one of the barriers cited by Anderson [9] for the development of a data-driven culture. In a similar vein, Huppertz *et al.* [10] argue that the establishment of functions and data-based decision-making promotes *a data-driven culture*, and these characteristics are directly related to benefits achieved though data governance. Hence, our first research hypothesis is:

H1: Data governance has a positive effect on data-driven culture.

In the literature, there is a broad consensus that improved data quality is among the most important benefits of data governance [4, 8, 19]. Studies such as Kim and Cho [26] and Otto [27] provide evidence of a positive relationship between the implementation of data governance and the quality of information. Nonetheless, studies that explore this relationship quantitatively from the perspective of data-driven culture could provide further insights into this subject, which is one reason it is included as a research hypothesis here. Furthermore, as detailed below, this hypothesis is necessary to assess the mediation hypothesis. Therefore, given the above, our hypothesis is:

H2: Data governance has a positive effect on data quality.

3.2 Data Quality

As highlighted by Anderson [9], having quality data encourages organizations to adopt a data-driven culture since it enables decisions to be made based on reliable data. Due to the reduced risk, other managers are more likely to use data in their decisions. Wook *et al.* [31] adds to that view, suggesting there is a positive effect between data quality and its use in big data applications. For Kremser and Brunauer [18], having access to high quality data makes managers more willing to trust and act based on data. Thus, the authors include data quality as a prerequisite for a data-driven culture. Hence, our hypothesis is:

H3: Data quality has a positive effect on data-driven culture.

Finally, based on the results reported by Mikalef *et al.* [15, 16], which demonstrate a mediating relationship between data governance and innovation capabilities in the context of big data, data quality is also assumed to be mediator between data governance and data-driven culture. In that perspective, the improvement of data quality is a benefit of data governance, which may vary to some degree. For example, two organizations, despite applying similar data governance mechanisms, might obtain different results in terms of data quality due to peculiarities related to the diversity of data in each organization. Therefore, our hypothesis is:

H4: Data quality has a mediating effect between data governance and data-driven culture.

4 Method

4.1 Partial least squares structural equation modeling (PLS-SEM)

According to Hair *et al.* [32], the PLS-SEM technique can be used to estimate complex models with constructs, indicators, and paths, being widely adopted by scholars in cases where there are no assumptions about the distribution of the analyzed data. It is also a consolidated technique for the investigation of the hypotheses in the social sciences, as proposed in this study (Fig. 1).

4.2 Development and validation of the questionnaire.

To carry out this research, a questionnaire composed of 20 items related to data governance, data quality, and data-driven culture was applied. As suggested by Mikalef *et al.* [15], data governance was considered a second-order formative construct, consisting of structural (03 items), procedural (04 items), and relational (02 items) constructs. The questionnaire was validated by a specialist in the area and was then tested with a pilot sample of 11 respondents through Google Forms. After the feedback from the pilot sample, adjustments were made in the wording of the items and the structure of the form, especially considering aspects concerning the translation from English to Brazilian Portuguese.

4.3 Sample and data collection

The target population of the research was people who work and use data in their work routines. Data collection was performed virtually, mainly through social networks (LinkedIn, Facebook, WhatsApp) through the Google Forms, using a Likert scale from 01 to 05. The data were collected between November 15 and 30, 2021. As suggested by Hair *et al.* [33], the transposed matrix of responses was analyzed to identify and exclude responses concentrated at one point of the scale. Thus, of the 149 responses received, 142 were considered valid. Table 1 demonstrates the characteristics of respondents. The sample is characterized as non-probabilistic and for convenience, and the sample size is higher than the minimum of 119 responders indicated by the G*Power software [34], adopting an effect size parameter of 0.15, alpha of 0.05, and power 0.95, as suggested by Cohen [35].

Table 1. Sample Characteristics

Dimension	Category	Percentage
Gender	Male	43,7%
	Female	55,6%
	Not declared	0,7%
Sector	Trading & Services	49,3%
	Government	33,8%
	Industry	4,9%
	Non-profit Organization	5,6%
	Other	6,3%
Number of Employees	Fewer than 09	9,9%
	Between 10 and 19	4,9%
	Between 20 and 49	7,0%
	Between 50 and 99	12,0%
	Between 100 and 499	9,9%
	More than 500	56,3%

4.4 Data Analysis

A descriptive analysis and exploratory factor analysis were performed in SPSS 18 to verify the correlation structure of the variables. The Single Harman Factor Test for Common Method Bias (CMB) was also performed, according to guidelines from Hair *et al.* [32]. Once the necessary validations and adjustments were completed, partial least squares structural equation modeling (PLS-SEM) was conducted using SMART PLS 3, estimating the measurement and structural models.

5 Findings

5.1 Exploratory factor analysis and biases

This analysis tested the interrelationship between the variables collected and the corresponding theoretical dimensions. Initially, to verify whether the appropriateness of the factor analysis, the Kaiser-Meyer-Olkin (KMO) criteria were observed and found to be within the established parameters (0.884), while the Bartlett's test was significant (p-value <0.001) [36, 37]. Thus, factor analysis was performed, in which three factors were extracted, whose variables were grouped as expected, and explained 61.1% of the total variance. As recommended by Podsakoff *et al.* [38], to verify possible common method bias (CMB), the Harman test was performed using SPSS 18. By reducing 43.1% of the variance by a single factor, it was impossible to explain most of the variance, indicating it is unlikely the sample was affected by common method bias.

5.2 Measurement Model – 1st Order

The measurement model was tested in SmartPLS 3 following the guidelines from Hair *et al.* [32]. Since this is a type II reflective-formative second-order model, the measurement was conducted in two stages, as suggested by Becker *et al.* [39] and Hair *et al.* [32]. Accordingly, the first level analysis, including all the variables of the model, was performed. When checking the outer loadings, the variables QUA04 and CDD02 presented values below 0.708, and were excluded from the model, in accordance with Hair *et al.* [32]. Based on analysis of the variance inflation factor (VIF), multicollinearity was identified in the procedural governance construct, which motivated the exclusion of the variables DGP04 and DPG05 [32]. After these adjustments, the outer loadings and VIF were within the reliability standards (>0.708 and <3, respectively) [32]. Thus, the measurement model was attained, as shown in Fig. 2.

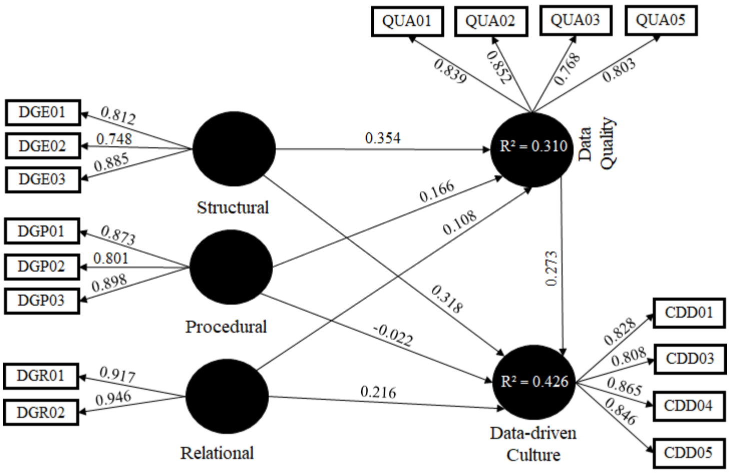

Fig. 2. Analysis of the 1st order measurement model (reflective)

To measure the validity of the constructs, their internal consistency was analyzed. The Cronbach's Alpha (CA) of the constructs was above the recommended value of 0.7 [40] and the composite reliability (CR) was within the parameters defined by Hair *et al.* [32], indicating that internal consistency is satisfactory, considering the sample analyzed. Convergent validity was also confirmed through the average variance extracted (AVE), where all values were above 0.5, indicating that the constructs converge with their indicators [33].

In sequence, the Fornell and Larcker criteria [41] were used to verify the square root of the AVE of the constructs and its relationship with the correlation between them. All values on the main diagonal, referring to the AVE square root, were superior to the correlation between the constructs (Table 2). As suggested by Henseler *et al.* [42], the Heterotrait-Monotrait (HTMT) ratio test was also performed, in which all the values were below 0.85, a conservative measure proposed by the authors (Table 3). Thus, the validity of the discriminant between the constructs was confirmed.

Table 2. Convergent Validity and Discriminant Validity

Construct	CA	CR	AVE	Discriminant Validity				
				Data-driven Culture	Structural	Procedural	Relational	Data Quality
Data-driven Culture	0.858	0.903	0.701	**0,837**				
Structural	0.753	0.857	0.667	0,576	**0,817**			
Procedural	0.823	0.893	0.736	0,452	0,671	**0,858**		
Relational	0.850	0.929	0.868	0,506	0,592	0,608	**0,932**	
Data Quality	0.832	0.888	0.666	0,522	0,529	0,469	0,418	**0,816**

CA: Cronbach’s Alpha; CR: Composite Reliability; AVE: Average Variance Extracted

Table 3. HTMT Ratio

	Data-driven Culture	Structural	Procedural	Relational	Data Quality
Data-driven Culture					
Structural	0,699				
Procedural	0,525	0,840			
Relational	0,581	0,737	0,729		
Data Quality	0,614	0,640	0,548	0,494	

Thus, following the conclusion of the measurement model steps recommended by Hair *et al.* [32], we were able to move on to the 2nd order model [39] in which data governance is a formative construct of the structural, procedural, and relational constructs [15].

5.3 Measurement Model - 2nd Order

To analyze the 2nd order measurement model, the factors of the latent variables of the constructs that compose data governance (structural, procedural, relational) were extracted, as recommended by Becker *et al.* [39]. In this model, bootstrapping was performed with 5,000 iterations. Initially, the validity of the second-order construct was verified through outer weights, which were significant at 0.05, while the VIF also presented acceptable values [32], as shown in Table 4.

Table 4. Analysis of the 2nd order measurement model

2nd Order	1st Order	Outer Weight	T Statistic	Outer Loadings	VIF
Data Governance	Structural	0.498***	3.623	0.893	1.809
	Procedural	0.320*	2.081	0.841	1.924
	Relational	0.352**	2.686	0.812	1.654

*Note: *** p < 0.001, ** p < 0.01, * p < 0.05*

Since the measurement model was validated in terms of confidence and validity for both the 1st Order and 2nd Order constructs, we can present the Structural Model.

5.4 Structural Model

Following the steps suggested by Becker *et al.* [39], the 2nd order reflective-formative structural model is analyzed using SMART PLS 3. Collinearity was analyzed in the 2nd order measurement model, as reported in the previous section. Thus, Fig. 3 and Table 5 present the results of the structural model. Both determination coefficients were significant (p-value < 0.001) and the results obtained from the model are tested against the research hypotheses.

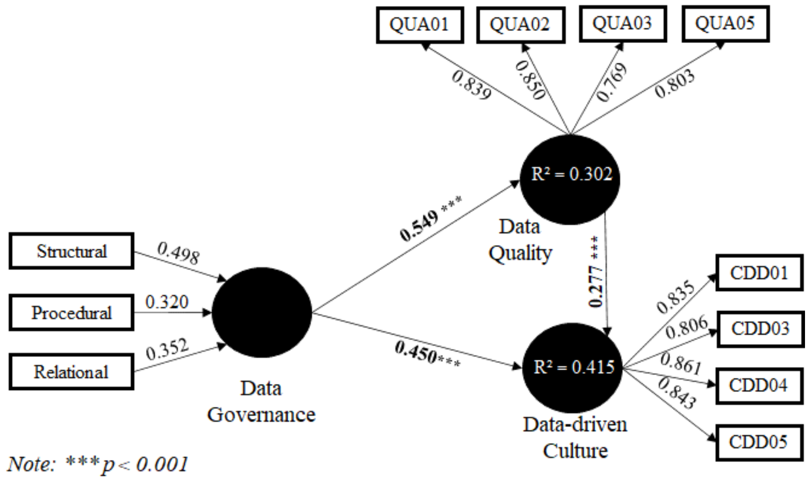

Fig. 3. Determination coefficient (R^2) and path coefficients.

Table 5. Summary of results

Path	Path Effect	Standard Deviation	T Statistic	Hypothesis
Data Governance → Data-driven Culture	0.450	0.085	5.285	Supported***
Data Governance → Data Quality	0.549	0.060	9.118	Supported***
Data Quality → Data-driven Culture	0.277	0.085	3.263	Supported***

*Note: *** p < 0.001*

Concerning the mediation hypothesis H4, data quality had a mediating effect between data governance and data-driven culture. The total effect of data governance on data-driven culture was 0.602, and even with the inclusion of data quality as a mediator, the direct effect of data governance in data-driven culture was significant (0.450). According to Hair et al. [32], as all effects are positive and significant, there is complementary partial mediation. Table 6 presents these results. The analysis also demonstrates the existence of predictive relevance (Q^2) for data-driven culture (Q^2= 0.273) and data quality (Q^2=0.192) [33]. The q^2 value of 0.413 indicates a great predictive relevance [33] of data quality for the data-driven culture construct.

Table 6. Effects Analysis - Data governance and Data-driven culture

Total Effect	Direct Effect	Indirect Effect
0.602* **	0.450***	0.152**

*Note: *** p < 0.001, ** p < 0.01*

5.5 Discussion

The results instigated a series of new developments on the subject. All the hypotheses tested in the research were supported, validating and complementing previous theory and framework propositions such as those from Mikalef *et al*. [15, 16]. For instance, the positive effect of data governance on data-driven culture is supported by new evidence, which leads to an estimation of effect sizes and supports further research to expand the knowledge on this subject. It is noteworthy, according to Cohen [44], that effects above 0.35 are considered large, indicating the relevance of the results. Our findings also provide strong evidence to support the role of data quality in mediating the relationship between data governance and data-driven culture. While the literature has identified an intimate relationship with these constructs [4, 8, 19], our study found a complementary partial mediation which might assist further research Hence, this study provides an introductory model that can be used as a basis for further investigations by expanding the PLS-SEM model to include additional constructs as a mean of integrating multiple research approaches.

From the RBV perspective, as proposed by Wade and Hulland [13], our findings highlight the importance for managers to consider data governance and data-driven culture as organizational resources in the drive to achieve competitive advantage. Similarly, the results complement the findings from Chatterjee *et al*. [11] regarding the RBV as they indicate the magnitude of the effects between the constructs, permitting a more in-depth analysis. Finally, the hypothesis tested provided a basis for further development of the RBV theory and provide initial guidance for future research.

6 Conclusion

This exploratory research investigated the relationship between data governance and data-driven culture from four research hypotheses developed from the IS literature. In order to investigate the hypotheses, partial least squares structural equation modeling was conducted to identify the existence and the magnitude of any effects between the constructs. Following previous studies on the matter, data quality was also included in the model since it is closely related to data governance and data-driven culture [4, 9, 18]. In addition, following the suggestions of Wade and Hulland [13], the results were discussed under the Resource-based View lens, in which the constructs are considered organizational resources capable of impacting the development of competitive advantage.

The results show the four research hypotheses were supported, that is, data governance has a positive effect on data-driven culture, which is partially mediated by data quality. From the RBV perspective, the results complement the findings in the literature [11, 15, 16], namely data governance and data-driven culture are seen to be related resources, suggesting they should be addressed jointly when evaluating their contribution towards achieving competitive advantage [12, 13]. Those findings would potentially be useful to managers who need to understand the importance of the role of data governance in developing data-driven culture. Furthermore, the results complement previous research and provide an introductory basis that could support further development of PLS-SEM models.

Therefore, this research achieved its objectives in investigating the relationship between data governance and data-driven culture. Regarding the study limitations, it is exploratory research, and as such does not represent an exhaustive analysis of the research topic. Concerning the PLS-SEM modeling, the number of respondents was 142, which, although sufficient according to G*Power [34], is a limited sample size. For future studies, the research model could be enlarged to include more dimensions and investigate other forms of relationships between the constructs considering, for example, different levels of organizational maturity [10, 21].

References

1. DAMA-DMBOK. Guide to the Data Management Body of Knowledge (DAMADMBOK v2) (2017).
2. FAN, Z. Context-Based Roles and Competencies of Data Curators in Supporting Research Data Lifecycle Management: Multi-Case Study in China. Libri, v. 69, n. 2, p. 127-137 (2019).
3. LIAW, S. *et al.* An integrated organisation-wide data quality management and information governance framework: Theoretical underpinnings. Informatics in Primary Care, v. 21, no. Four, p. 199-206 (2014).
4. ABRAHAM, R. SCHENEIDER, J. VOM BROCKE, J. Data governance: A conceptual framework, structured review, and research agenda. International Journal of Information Management, v. 49, p. 424- 19 438 (2019).
5. MAGNUSSON, J., KOUTSIKOURI, D., PÄIVÄRINTA T. Efficiency creep and shadow innovation: enacting ambidextrous IT Governance in the public sector. European Journal of Information Systems, v. 29, n. 4, p. 329-349 (2020).
6. WENDE, K. A Model for Data Governance - Organising Accountabilities for Data Quality Management. ACIS 2007 Proceedings (2007).
7. BROUS, P.; JANSSEN, M.; KRANS, R. Data Governance as Success Factor for Data Science. IFIP International Federation for Information Processing 2020, p. 431-442 (2020).
8. LADLEY, J. Data Governance: How to design, deploy and sustain an effective data governance program. London: Academic Press (2019).
9. ANDERSON, C. Creating a data-driven Organization. Sevastopol, CA: O'Reilly Media (2015).
10. HUPPERZ, M. GÜR, I. MÖLLER, F. OTTO, B. What is a Data-Driven Organization? Twenty-Seventh Americas Conference on Information Systems (AMCIS) (2021).
11. CHATTERJEE, S.; CHAUDHURI, R.; VRONTIS, D. Does data-driven culture impact innovation and performance of a firm? An empirical examination. Annals of Operations Research (2021).
12. BARNEY, J.B. Firm resources and sustained competitive advantage. Journal of Management, v. 17, p. 99-120 (1991).
13. WADE, M.; HULLAND, J. Review: the resource-based view and information systems research: review, extension, and suggestions for future research. MIS Quarterly, v. 28, No. 1, pp. 107-142, March (2004).
14. MIKALEF, P. *et al.* Information Governance in the Big Data Era: Aligning Organizational Capabilities. Proceedings of the 51st Hawaii International Conference on System Sciences (2018).
15. MIKALEF, P. *et al.* The role of information governance in big data analytics driven innovation. Information & Management, v. 57 (2020).

16. MIKALEF, P.; KROSGSTIE, J. Examining the interplay between big data analytics and contextual factors in driving process innovation capabilities. European Journal of Information Systems, v.29, i. 3, p. 260-287 (2020).
17. LAJARA, T.T., Dull, A.C.G. Information governance framework: The defense manufacturing case study. 19th Americas Conference on Information Systems, AMCIS 2013 - Hyperconnected World: Anything, Anywhere, Anytime, p. 1984 (2013).
18. KREMSER, W.; BRUNAUER, R. Do we have a Data Culture? Springer (2019).
19. WENDE, K.; OTTO, B. A Contingency Approach to Data Governance. 12th International Conference on Information Quality (ICIQ-07), Cambridge, USA (2007).
20. KHATRI, Vijay; BROWN, Carol V. Designing Data Governance. Communications of the ACM, January, Vol. 53 No. 1, Pages 148-152 (2010).
21. BERNDTSSON, M. *et al.* Becoming a data-driven organisation. 26th European Conference on Information Systems (ECIS2018), Portsmouth, United Kingdom (2018).
22. AGYEI-OWUSU, B.; AMEDOFU, M.K.; ASAMOAH, D.; KUMI, C.A. The Effect of Data-driven Culture on Customer Development and Firm Performance: The Role of Supply Chain Information Sharing and Supply Chain Information Quality. In: DENNEHY, D.; GRIVA, A.; POULOUDI, N.; DWIVEDI, Y.K.; PAPPAS, I.; MÄNTYMÄKI, M. (eds) Responsible AI and Analytics for an Ethical and Inclusive Digitized Society. I3E, Springer, Cham (2021).
23. PENROSE, E. T. The Theory of the Growth of the Firm. New York: Wiley (1959).
24. MAHONEY, J.T.; PANDIAN, J. R. The resource-based view within the conversation of strategic management. Strategic Management Journal, v. 13, p. 363-380 (1992).
25. FARIA, F.; DULL, A.; KUMAR, K. Structural model of information governance for banks. RAE-Revista de Administração de Empresas FGV EAESP, São Paulo, v. 57, n. 1, Jan-Feb, p. 79-95 (2017).
26. KIM, H.Y.; CHO, J. Data Governance Framework for Big Data Implementation with a Case of Korea. IEEE 6th International Congress on Big Data, Honolulu, p. 384-391 (2017).
27. OTTO, B. Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers. Communications of the Association for information systems,v.29 (2011).
28. NAGLE, T.; REDMAN, T.; SAMMON, D. Assessing data quality: A managerial call to action. Business Horizons, v. 63, i. 3 (2020).
29. SHAMALA, P. *et al.* Integrating information quality dimensions into information security risk management (ISRM). Journal of Information Security and Applications, v. 36 (2017).
30. CAI, L.; ZHU, Y. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Science Journal, v. 14, no. 2 (2015).
31. WOOK, M.; HASBULLAH, N.A.; ZAINUDIN, N.M. *et al.* Exploring big data traits and data quality dimensions for big data analytics application using partial least squares structural equation modelling. J Big data, v.8, n. 49 (2021).
32. HAIR, J., RISHER, J., SARSTEDT, M. AND RINGLE, C. (2019), "When to use and how to report the results of PLS-SEM", European Business Review, Vol. 31 No. 1, pp. 2-24 (2019).
33. HAIR, J., HULT, G., RINGLE, C. AND SARSTEDT, M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). Sage: Los Angeles (2017).
34. FAUL, F.; ERDFELDER, E.; BUCHNER, A.; LANG, A. Statistical power analyses using G* Power 3.1: tests for correlation and regression analyses. Behavior Research Methods, v. 41 n. Four, p. 1149-1160 (2009).
35. COHEN, J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum, 2nd edition (1988).

36. KAISER, M. Kaiser-Meyer-Olkin measure for identity correlation matrix. Journal of the Royal Statistical Society, v. 52, p. 296-298 (1974).
37. TOBIAS, S.; CARLSON, J. Brief report: Bartlett's test of sphericity and chance findings in factor analysis. Multivariate Behavioral Research, v. 4, n. 3, p. 375-377 (1969).
38. PODSAKOFF, P.M.; MACKENZIE, S.B., LEE, J. Y.; PODSAKOFF, N.P. Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology, v. 88, n. 5, p. 879–903 (2003).
39. BECKER, J.M.; KLEIN, K.; WETZELS, M. Hierarchical latent variable models in PLS-SEM: guidelines for using reflective-formative type models. Long Range Planning, v. 45, i. 5–6, p. 359-394 (2012).
40. CRONBACH, L.; MEEHL, P. Construct validity in psychological tests. Psychological Bulletin, v. 52, no. Four, p. 281-302 (1955).
41. FORNELL, C.; LARCKER, D. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, v. 18, n. 1, p. 39-50 (1981).
42. HENSELER, J., RINGLE, C., SARSTEDT, M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science, v. 43, n. 1, p. 115-135 (2015).
43. ZHAO, X.; LYNCH, J.G.; CHEN, Q. Reconsidering Baron and Kenny: Myths and Truths About Mediation Analysis. Journal of Consumer Research, v. 37, n. 2, p. 197– 206 (2010).
44. COHEN, J. The power primer. Psychological Bulletin, v. 112(1), p. 155-159 (1992).
45. WANG, P. Connecting the Parts with the Whole: Toward an Information Ecology Theory of Digital Innovation Ecosystems. MIS Quarterly, v.45, n. One, p. 397-422 (2021).