Association for Information Systems

AIS Electronic Library (AISeL)

MCIS 2024 Proceedings

Mediterranean Conference on Information Systems (MCIS)

10-3-2024

The impact of data quality orchestration in data ecosystems: quantitative evidence from the Brazilian Judiciary

Felipe Fonseca Salerno
Universidade Federal do Rio Grande do Sul (UFRGS), 00193140@ufrgs.br

Antonio Carlos Gastaud Maçada
Universidade Federal do Rio Grande do Sul (UFRGS), acgmacada@ea.ufrgs.br

Follow this and additional works at: https://aisel.aisnet.org/mcis2024

Recommended Citation

Salerno, Felipe Fonseca and Maçada, Antonio Carlos Gastaud, "The impact of data quality orchestration in data ecosystems: quantitative evidence from the Brazilian Judiciary" (2024). *MCIS 2024 Proceedings*. 10.
https://aisel.aisnet.org/mcis2024/10

This material is brought to you by the Mediterranean Conference on Information Systems (MCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in MCIS 2024 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

THE IMPACT OF DATA QUALITY ORCHESTRATION IN DATA ECOSYSTEMS: QUANTITATIVE EVIDENCE FROM THE BRAZILIAN JUDICIARY

Research full-length paper

Salerno, Felipe Fonseca, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil, 00193140@ufrgs.br

Maçada, Antônio Carlos Gastaud, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil, acgmacada@ea.ufrgs.br

Abstract

In data ecosystems, the intense exchange of data between organizations offers opportunities for collaboration and growth. However, poor data quality circulating within these ecosystems obstructs the advancement of data-driven initiatives and undermines confidence in the data. To address this issue, data quality orchestration seeks to organize and coordinate data-driven activities within the ecosystem to maintain and improve data quality. This exploratory study investigates the effects of data quality orchestration on overall data quality in data ecosystems. Data from the Brazilian Judiciary, collected over two years of data quality orchestration (2021-2023), is analyzed using repeated measures ANOVA. The findings indicate that data quality orchestration had a positive effect on overall data quality, regardless of the size of the organizations involved. This study provides valuable empirical and quantitative evidence concerning data quality orchestration within data ecosystems.

Keywords: Data Quality Orchestration; Data Ecosystems; Repeated Measures ANOVA; Judiciary.

1 Introduction

In an increasingly competitive market, organizations are structuring themselves into ecosystems—a group of interdependent actors collaborating to achieve a common goal (Mann et al., 2022; Wang, 2021). From the perspective of data ecosystems, this concept illustrates a flexible and open system where multiple actors engage in data exchange, providing opportunities for collaboration and growth (Geisler et al., 2021; Guggenberger et al., 2020). Cappiello et al. (2020) highlight the relevant role of data ecosystems as fundamental technological enablers in the digital economy, while Otto et al. (2019) argue that these ecosystems facilitate the acquisition of valuable insights about both partners and the business, fostering efficient management practices. However, the literature highlights challenges associated with poor data quality within data ecosystems, which hinders the progress of data-driven initiatives and affects financial performance (Altendeitering et al., 2022).

Consultancy McKinsey & Company (2023) indicates that data quality assurance is important for developing new business opportunities, especially considering the intensive use of external data to create scenarios and support strategic decisions. A recent survey by Monte Carlo (2023) with 200 data engineers estimated that poor data quality impacts $3 of every $10 of revenue, showing that as data becomes more valuable to businesses, poor data quality becomes more costly. According to Gartner (2021), poor data quality results in an average cost of $12.9 million annually for organizations and, beyond its immediate impact on revenue, fosters suboptimal decision-making in the long run. Vafaei-Zadeh et al. (2020) argue that data quality plays an important role in facilitating informational integration among different organizations, thereby directly influencing performance. Zhang et al. (2022) found that poor data quality presents a challenge to data sharing, hindering its effective use by those involved. To maximize the value extracted from data, a data ecosystem must prioritize maintaining exceptional data quality to foster confidence among actors and support well-informed decision-making (Hannila et al., 2022; Choi et al., 2021).

Another factor influencing data quality within data ecosystems is the varying degree of sophistication with which an organization collects, manages, utilizes, and values its data. This variability can potentially affect how data exchanges are conducted and consequently impact data quality (Al-Sai et al., 2023; Gupta and Cannon, 2020). The literature suggests that when engaging in data exchanges with partners, maintaining an acceptable standard of data quality is central to avoid fostering distrust due to poor quality data. The impact of an organization’s size and its contingencies on data management and governance is a relevant subject of research in the information systems literature. This examination is critical as it could influence the quality of data within the data ecosystem, highlighting the role of organizational characteristics in shaping data-related practices (Hannila et al., 2022; Hannah and Eisenhardt, 2015). Therefore, within data ecosystems, the relevance of data quality should not be understated as high-quality data cultivates trust among stakeholders, empowering them to utilize it to enhance their operations and decision-making. Conversely, poor data quality could restrict its utility and the potential benefits derived from data (Karkošková, 2023).

The lack of data quality also affects governmental institutions. Vetrò et al. (2016) mention that one of the main impacts is on decision-making, such that poor-quality data can lead to ineffective and inefficient decisions, impacting public policies. Additionally, the authors assert that low-quality data negatively affects public transparency, hindering society's understanding of the activities carried out by institutions. This view is complemented by Purwanto et al. (2020), who state that citizens perceive data quality and associate it with the quality and trustworthiness of public services. Finally, Omari et al. (2021) reinforce the importance of public institutions defining a strategy to ensure that data meets the desired quality. Among the strategic steps, the authors propose the definition of principles and quality standards, as well as constant data quality monitoring. Thus, it is evident that data quality is also important for public institutions and their data ecosystem, influencing transparency and accountability, and affecting society's trust and perception regarding service quality.

An alternative proposed in the literature for enhancing data quality within data ecosystems is orchestration. This approach involves the coordination of assets and activities to achieve specific objectives, encompassing various aspects of interest to the involved stakeholders, such as security management and data quality (Autio, 2022; Linde et al., 2021). From a theoretical perspective, resource orchestration theory supports the notion that effectively managing and integrating both internal and external resources can generate organizational value. This implies that organizations must align and adapt their strategies in response to market dynamics, whether positive or negative changes (Autio, 2022; Cui et al., 2017). Based on these concepts, the term “data quality orchestration” represents the capability to organize and coordinate the ecosystem assets towards maintaining and improving its data quality.

Understanding the impact of data quality orchestration allows directing the data ecosystem in favor of improving its data quality and consequently fostering data-driven initiatives. Following the literature's suggestion to investigate how contingencies such as organizational size influence data-related initiatives, the following research questions (RQs) are formulated:

RQ1: *What is the effect of data quality orchestration on the overall data quality of a data ecosystem?*

RQ2: *Is the effect of data quality orchestration influenced by the size of the organization?*

To address these RQs, this study examines the official data quality indicator within the Brazilian Judiciary data ecosystem over two years of data quality orchestration (2021-2023), employing repeated measures ANOVA to evaluate its effects on the data ecosystem’s overall data quality. Additionally, to illustrate how contingency factors may affect data quality orchestration, the size of the actors is included in the analysis. This paper begins with a literature review on data quality orchestration, followed by a detailed examination of the research context. The findings are expected to contribute to the body of research in information systems by providing real-world quantitative evidence regarding the impact of data quality orchestration on the data ecosystem's overall data quality.

2 Data Quality Orchestration

Teece (2020) and Linde et al. (2021) argue that orchestration in ecosystems can be considered a dynamic capability, involving the coordination of assets and activities to achieve specific objectives, where the capabilities of sensing, seizing, and reconfiguring are essential to understand the business environment, seize opportunities, and manage threats and transformations. Orchestration also involves persuading behaviors to achieve common goals of the ecosystem, encouraging the definition of roles among the actors, and seeking to create value through the mobilization of assets in a relationship that goes beyond the simple structure of command and control (Autio, 2022). Finally, orchestration entails the continuous adaptation and evolution of strategies and tactics in response to changes in the ecosystem, fostering resilience and sustainability over time. Considering data as a strategic asset for organizations (Zhang et al., 2019), the concept of orchestration could be extended to data ecosystems.

According to Oliveira and Lóscio (2018), data ecosystems consist of actors, roles, relationships, and resources configured in a network, driven by the common interests of the organizations within it. Gelhaar et al. (2021) propose analyzing the taxonomy of data ecosystems from the control dimension, indicating two classifications: centralized, representing the presence of an actor controlling essential resources within the ecosystem; and decentralized, where data control is distributed among the actors. Regarding typology, Guggenberger et al. (2020) classify ecosystems into five main types based on their predominant features, one of which is the orchestrated ecosystem characterized by four predominant attributes. The first is centrality, indicating that relationships are centralized by a key actor within the ecosystem. The second is centralized power, emanating from the presence of a central actor exerting influence over others. The third is specialization, where each actor brings a unique value contribution to the ecosystem. The final attribute is collective intent, where actors work towards a common objective of collective interest. In summary, orchestrated ecosystems are “communities controlled by a central power and a central object used to orchestrate the individual specializations” (Guggenberger et al., 2020, p. 9).

In orchestrated ecosystems, there is a central actor with power and influence over others, capable of influencing, motivating, and promoting cooperation to achieve a goal. Teece (2020) contends that leadership involvement is indispensable for the success of orchestration initiatives, while simultaneously stressing the importance of fostering a collaborative culture among ecosystem participants to ensure effective coordination and alignment towards common goals. Given the prevalent challenge of poor data quality in ecosystems (Karkošková, 2023), it is proposed that the orchestrating organization could enhance data quality through effective data management and governance within the data ecosystem, particularly focusing on its data and metadata. Essentially, data quality orchestration embodies the endeavors of an actor aimed at improving data quality within its data ecosystem. The concept is closely intertwined with data governance and management, where orchestration aligns with governance policies, working in tandem with data resource management to effectively coordinate and harmonize efforts aimed at achieving data quality objectives, including topics such as data training, compliance, and workflow design (Najafabadi and Cronemberger, 2023; Mukhopadhyay and Bouwman, 2019).

From a theoretical standpoint, data quality orchestration could be examined through the lens of resource orchestration theory. This theory focuses on efficiently coordinating resources through orchestration capabilities, emphasizing relationships and interactions with partners (Lin et al., 2023; Schreieck et al., 2022). The theory comprises three processes: structuring, bundling, and leveraging (Sirmon et al., 2011). These processes entail activities aimed at organizing available resources, optimizing and complementarily grouping them, and leveraging them to generate value for the ecosystem.

In the context of data, the theory enables viewing data quality as a valuable resource that empowers ecosystem actors to enhance their data-related activities (Cui and Han, 2022). Thus, data quality orchestration expands the concept of orchestration by highlighting its effectiveness in driving data quality improvement within data ecosystems.

To quantitatively demonstrate the effects of data quality orchestration in a data ecosystem, an empirical analysis was conducted within the Brazilian Judiciary, contextualized in the following section.

3 Research Context

This section begins with a brief overview of the Brazilian Judiciary data ecosystem, highlighting the orchestrating role of the National Council of Justice (CNJ). The CNJ was established in 2005 under Article 103-B of the Brazilian Federal Constitution, and its responsibilities include overseeing the administrative and financial activities of the Judiciary. Additionally, it is tasked with producing statistical reports on the Judiciary's activities and, based on these reports, proposing measures and standardizing the operations of the courts (Brazil, 1988). In other words, it is an official government central body that orchestrates the administrative and financial activities of the courts with the aim of enhancing control and transparency, not being a jurisdictional body (CNJ, 2023a).

In the exercise of its constitutional responsibilities and to promote the adoption of new technologies to drive digital transformation in Brazilian courts, the CNJ launched the Justice 4.0 Program in January 2021, which seeks to structure more agile, effective, and accessible services in the courts (CNJ, 2022a). One of the program's areas of action concerns information management and judicial policies, aiming to promote the formulation, implementation, and monitoring of judicial policies through data and evidence. Specifically, the program strives to integrate cutting-edge technological solutions into court procedures to increase efficiency and accessibility. Concurrently, it prioritizes the discerning utilization of data-driven insights to inform public policy and operational strategies, leveraging data-based decision-making processes.

In pursuit of this objective, the CNJ developed a centralized database of Judiciary data and metadata named DATAJUD, gathering data on the activities of all 90 Brazilian courts except the Federal Supreme Court (CNJ, 2023b). This initiative was described as a significant advance in judicial management, enabling courts and society to track data related to the Judiciary's activities (CNJ, 2023a). By providing access to comprehensive and up-to-date information through the centralized database, CNJ aims to empower courts to enhance their decision-making processes, optimize resource allocation, and streamline judicial operations. Furthermore, the initiative is expected to increase transparency and accountability, thereby facilitating informed public policy regarding judicial affairs (CNJ, 2022a).

To implement the Justice 4.0 Program, the CNJ has published 24 regulations (as of September 2023) that determine and guide courts on the adoption of digital technologies and data management (CNJ, 2022a; 2023a). Additionally, training cycles in data science have been conducted, consisting of eight courses covering topics related to statistics, data science, Python and R programming, and machine learning for Judiciary servers and judges (CNJ, 2023a). Furthermore, webinars and various events have been held to promote the initiative to the courts and society. Therefore, it is evident that since the launch of the Justice 4.0 Program, the CNJ has acted as an orchestrator of the data ecosystem, fostering and disseminating best practices to the courts and facilitating the transition towards a more technological and data-centric judiciary.

As indicated by Karkošková (2023), one main challenge in data ecosystems is poor data quality. To overcome that challenge, the CNJ issued regulations standardizing data collection methods and offered data science training cycles, webinars, and events to promote data-driven initiatives, mobilizing the courts towards improving data quality (CNJ, 2023a; 2023b). Fundamentally, the CNJ organized and coordinated the data ecosystem assets towards the maintenance and improvement of its data quality, which is the essence of data quality orchestration. Before the beginning of data quality orchestration

(May 2021), the average quality indicator in the State Courts was 11%, meaning that only about 1 in 10 cases met the desired data quality criteria. Essentially, the data ecosystem had poor data quality, which hindered the creation of the centralized database and the use of data for court administrative management. After CNJ's data quality orchestration, the most recent measurement available (August 2023) indicates a quality indicator of 78%, suggesting that data quality orchestration was effective in enhancing the overall data quality of the data ecosystem (CNJ, 2023a; 2023b). Therefore, examining the effects of data quality orchestration within the judiciary's data ecosystem is interesting as the findings might be extended to other ecosystems with similar data quality issues.

The role of the CNJ as an orchestrator of the ecosystem is exemplified in Figure 1. In it, the institution coordinates activities involving court data and metadata, exercising data quality orchestration through governance and data management policies, training staff and technical personnel, designing data workflows, and ensuring constant supervision and monitoring through compliance initiatives and audits. With these initiatives, the CNJ manages to integrate the data into a database, which can be publicly visualized through thematic dashboards. As a result, both society and court managers have access to the data, contributing to public transparency and enhancing court management (CNJ, 2022a; 2023a; 2023b). Thus, data quality orchestration works to ensure that the data made available to stakeholders is of sufficient quality for secure utilization.

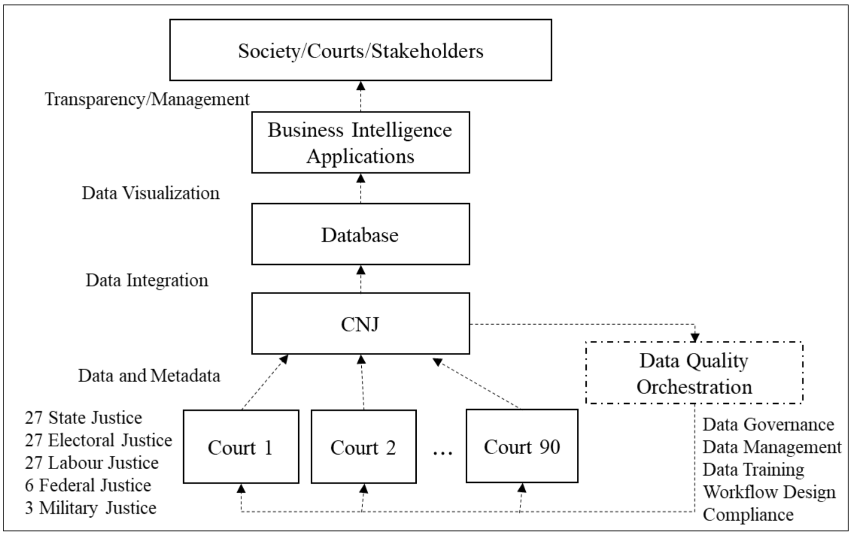

Figure 1. CNJ as a data quality orchestrator in the Brazilian Judiciary data ecosystem.

The CNJ's activities could also be identified within the processes defined by the resource orchestration theory (Sirmon et al., 2011). Concerning the structuring processes, these can be defined as actions aimed at organizing the technological ecosystem that enables the flow of data among stakeholders. This structure enables courts to regularly send data to the CNJ, ensuring the continuity of the data ecosystem. Regarding bundling actions, these can be observed through the consolidation of data from all courts by the CNJ, which analyzes and enriches the data. In other words, bundling activities involve combining data from different courts, generating insights that surpass those obtained individually. Finally, leveraging actions involve creating value through data quality orchestration, which directly manifests as an improvement in data quality and indirectly impacts decision-making and other data-driven initiatives (CNJ, 2022a; 2023a; 2023b). Thus, the CNJ's role as an orchestrator could be analyzed within the framework of the resource orchestration theory. To statistically assess the evolution of data quality within the data ecosystem over the two years of orchestration, the following section discusses the use of the repeated measures ANOVA technique.

4 Methodology

To investigate the research questions (RQs), this study utilized the official indicator of data quality within the Brazilian Judiciary data ecosystem. This indicator, calculated by the National Council of Justice, evaluates adherence to data transmission quality and recording standards, primarily under National Resolutions No. 76/2009 and No. 46/2007 (CNJ, 2023a). For example, one criterion stipulates that the subject registration within a process should be detailed to the most specific level of the hierarchy, with a minimum requirement of level 3 detailing. As shown in Figure 2, a process labeled with the subject "Partition" ought to be categorized with code 14923, representing level 5 specificity. However, it is permissible to report the code at level 3 (5808) for the sake of data quality. This hierarchical logic also applies to procedural classes and actions, necessitating a minimum of level 3 specification. In addition to these criteria, various others exist, such as the completion of basic data (document numbers, active and passive parties, etc.) and adherence to interoperability standards. Therefore, the quality indicator stands as an official measure used by the CNJ, consolidating several quality criteria and measured as a percentage, where a higher percentage indicates higher data quality. Although in IS literature data quality is a multifaceted concept (Najafabadi and Cronemberger, 2023; Zhang et al., 2019), this research considered the official data quality indicator as it is the national reference on the matter.

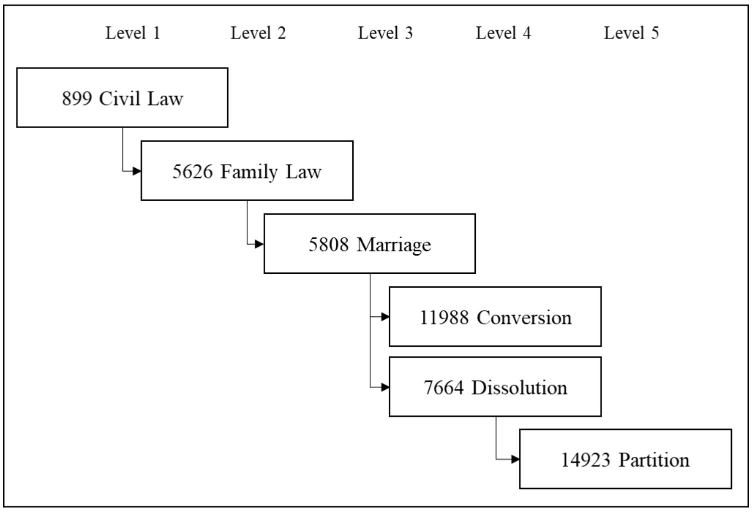

Figure 2. Example of data hierarchy criteria structure.

Historical data was available for all 27 State Courts of Justice, which collectively accounted for 72.9% of case filings in Brazil, equivalent to roughly 23 million cases in 2022 (CNJ, 2023a). The data is publicly accessible and was retrieved from the CNJ website. The data quality indicator is measured as a percentage, where 0% indicates no quality at all while 100% indicates that all quality criteria established by the CNJ have been achieved. It was available every month from May 2021, shortly after the beginning of the data quality orchestration, through August 2023. However, there were gaps in the data for four specific months (Nov/Dec 2022, Feb/May 2023). To address these gaps, measurements were standardized at three-month intervals, commencing in July 2021 and concluding in July 2023 (CNJ, 2023a; 2023b).

As suggested by Park et al. (2009), when working with data following a standardized measurement structure over time, one of the analysis techniques employed is ANOVA for repeated measures. The objective of this technique is to investigate differences in mean scores across various time intervals and differences in mean scores among groups. In this study, to elucidate the influence of contingencies on data initiatives as proposed by Sambamurthy and Zmud (1999) and Weber et al. (2009), the categorization is organized according to the size of the court (small, medium, large). This classification is determined by the CNJ itself using a principal component analysis technique, aiming to depict variations in resources and workload among the courts. Table 1 presents data from three courts of different sizes, highlighting substantial differences in court structure.

Indicator	Small (TJPI)	Medium (TJCE)	Large (TJRJ)
Total Expenses	$ 8.6 million	$ 1.6 billion	$ 7.3 billion
New Cases	261522	480540	2100621
Pending Cases	595629	1159546	7426744
Judges	178	505	908
Employees	3634	8582	24147

Table 1. *Example of indicators according to the court's size.*

Through the use of repeated measures ANOVA, this research aims to assess whether the effect of data quality orchestration over time (within-subjects) contributed to an overall improvement in the ecosystem's data quality; and if the size of the courts influenced this outcome (between-subjects). This method proves especially valuable given the strong correlation caused by the repeated measures over time, necessitating a thorough examination of the model's assumptions with particular emphasis on assessing sphericity (Muhammad, 2023). Regarding the RQs, this technique allows for a statistical evaluation of the effects of data quality orchestration on the data ecosystem's overall data quality over the years and any potential effects of the court's size. The following section presents the obtained results.

5 Results

An initial examination of quarterly data displayed in Table 2 reveals a notable enhancement in the average data quality after the beginning of the data quality orchestration. The overall average data quality, which accounts for all state courts, surged from 34.50% in July 2021 to 69.45% in July 2023. This represents a substantial boost in data quality after the beginning of data quality orchestration by the CNJ.

Size	Jul/21	Oct/21	Jan/22	Apr/22	Jul/22	Oct/22	Jan/23	Apr/23	Jul/23
Small	30.84	47.14	53.52	56.36	52.74	50.70	56.70	57.78	64.42
Medium	37.50	52.02	53.92	51.99	51.52	64.78	68.67	66.69	70.71
Large	33.53	49.30	54.78	54.90	55.33	64.93	64.22	70.41	70.50
Overall	34.50	49.91	54.23	54.09	53.44	62.24	64.47	66.69	69.45

Table 2. *Average data quality based on court size*

Following Parker et al.'s (2009) recommendations, Wilks' Lambda is employed as the criterion for the ANOVA analysis, interpreted as the proportion of generalized variance in the dependent variables explained by the predictors, striking a balance between statistical power and underlying assumptions (Table 3). Regarding the assumption of sphericity in the covariance structure (Table 4), Mauchly's test yielded a significant result ($p < 0.05$), leading to the rejection of the sphericity hypothesis (Muhammad, 2023). Consequently, given that epsilon (ε) is below 0.750, the Greenhouse-Geisser (G-G) correction criteria is used to adjust the p-values. It does so based on F-values with reduced degrees of freedom using the Box method (Vittinghoff et al., 2011). Table 5 presents the tests of within-subjects effects, which assess the significance of changes within the same courts across different months, and

Table 6 presents the tests of between-subjects effects, evaluating the differences among sizes over time.

Results show that the p-value for the *month* variable was statistically significant ($p < 0.05$) in the within-subjects effects. This suggests that there is a significant difference in data quality across the analyzed period. Specifically, the observed significance indicates a positive improvement in data quality over time within the data ecosystem. In other words, this confirms that there was an improvement in the data ecosystem's overall data quality after the beginning of the data quality orchestration. However, the analysis revealed a non-significant interaction effect between month and size (*month*size*) ($p > 0.05$), suggesting that there is insufficient evidence to support the claim that the court's size influenced the improvement of data quality over time. This finding is further supported by the non-significant result ($p > 0.05$) for the between-subjects effect of size, indicating that there is no significant difference in data quality improvement based on the court's size.

Effect		Value	F	Hypothesis df	Error df	Sig.	η^2 partial
Month	Wilks' Lambda	0.418	2.965	8.000	17.000	0.028	0.582
Month * Size	Wilks' Lambda	0.450	1.043	16.000	34.000	0.440	0.329

Table 3. Multivariate tests

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser
Month	0.001	258.507	35	0.001	0.284

Table 4. Mauchly's Test of Sphericity

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	η^2 partial
Month	G-G	19918.634	2.275	8756.342	8.785	0.001	0.268
Month * Size	G-G	1408.955	4.550	309.692	0.311	0.890	0.025
Error (Month)	G-G	54414.753	54.594	996.709			

Table 5. Tests of Within-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	η^2 partial
Intercept	657307.163	1	657307.163	81.539	0.001	0.773
Size	1027.814	2	513.907	0.064	0.938	0.005
Error	193470.525	24	8061.272			

Table 6. Tests of Between-Subjects Effects

Expanding on the analysis, Figures 3 and 4 depict the graphs of mean effects for data quality and its breakdown by court size, respectively. As previously indicated, a substantial enhancement in data quality over time is shown in Figure 3. Concerning size, there are several intersections among the lines in Figure 4, visually confirming the absence of evidence to support the claim that size influenced the improvement in data quality during the orchestration period. It is also noteworthy that a prominent inflection point occurred in July 2022 within the historical series, resulting in a more substantial increase in data quality among medium and small-sized courts, while larger courts did not exhibit the same level of improvement. This indicates the necessity for further research to identify potential explanations for this behavior.

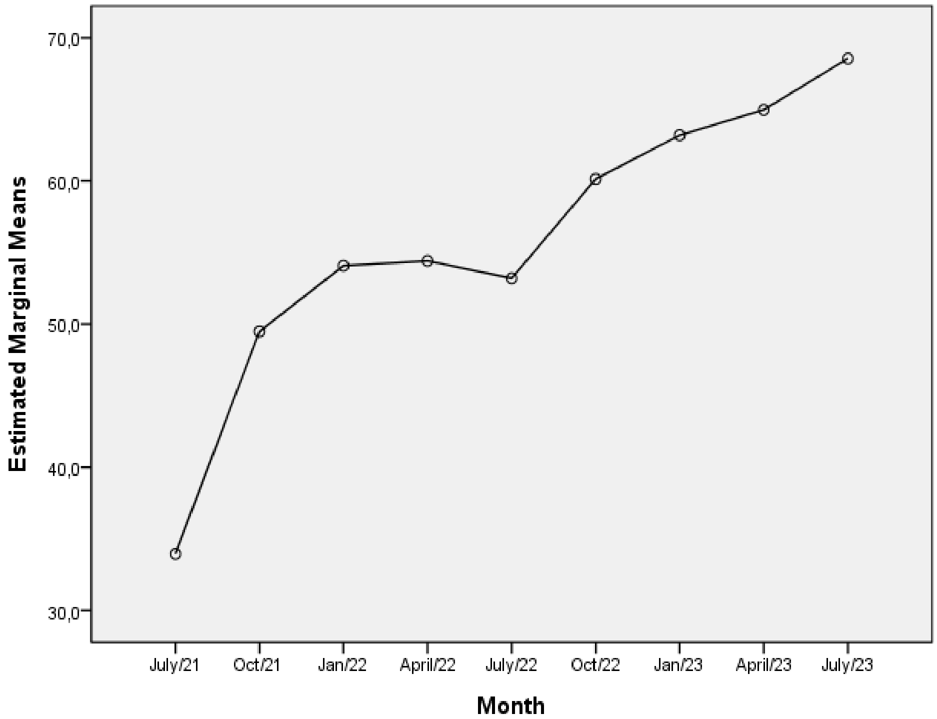

Figure 3. *Estimated Marginal Means of Data Quality*

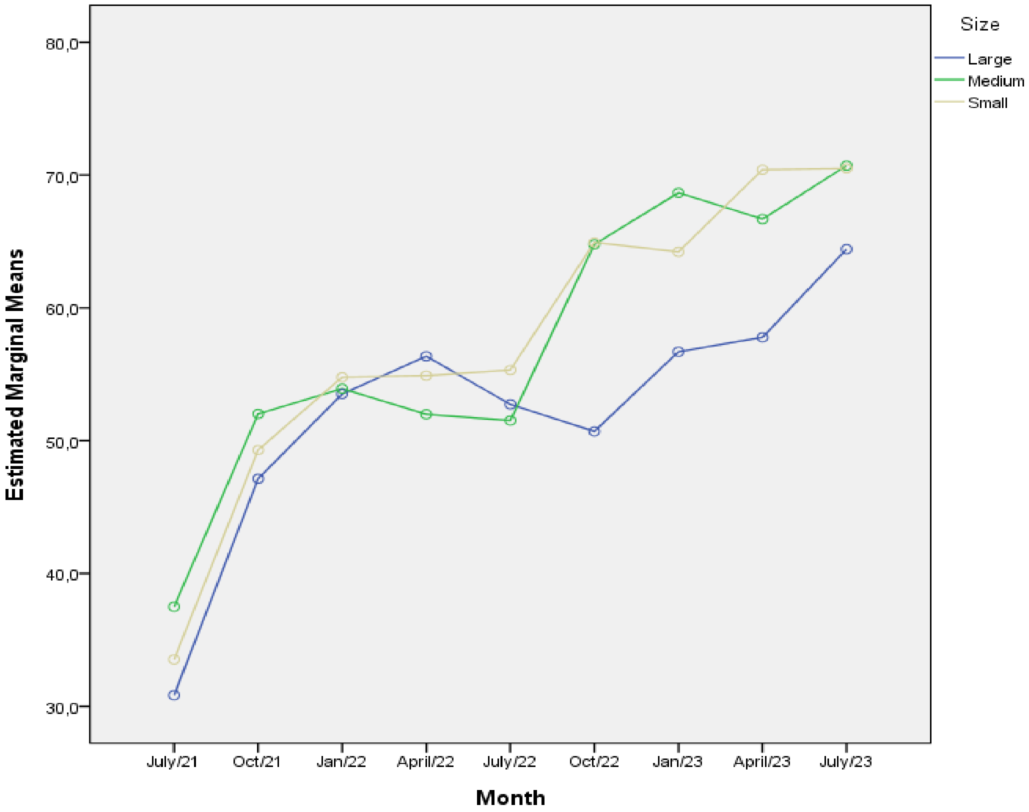

Figure 4. *Estimated Marginal Means of Data Quality - by Size*

6 Discussion

This research investigated the effects of data quality orchestration on the overall data quality of a data ecosystem through quantitative analysis. To do so, it analyzed the Brazilian Judiciary data ecosystem, utilizing data from 27 state courts over two years of data quality orchestration. Data was analyzed using repeated measures ANOVA, a technique that allowed for the examination of changes in data quality over time, capturing trends and patterns in the data ecosystem. The results provided valuable insights that contribute to the information systems literature, shedding light on how data quality orchestration can foster improvements in the overall data quality of data ecosystems.

Initially, following the introduction of data quality orchestration activities by the ecosystem orchestrator (CNJ), there was an improvement in data quality, encompassing the establishment of data ecosystem norms, capacity-building initiatives for stakeholders, and the promotion of data utilization. These findings reinforce arguments in the orchestration literature concerning the role played by leadership and the pursuit of common data ecosystem objectives, whereby the orchestrating organization assumes an important part in improving overall data quality in ecosystems (Autio, 2022; Teece, 2020; Linde et al., 2021). As proposed by Guggenberger et al. (2020), the results constitute a real-world example of an orchestrated ecosystem, underscoring the need for the orchestrating actor to exercise authority and coordination in seeking to enhance data quality in the data ecosystem.

This scenario can also be examined through the lens of the Resource Orchestration Theory. The structuring, bundling, and leveraging processes carried out by the CNJ during the establishment of the data ecosystem contributed to enhancing data quality, thereby adding value to the data ecosystem through improved decision-making and accountability to society. Furthermore, the perspective of a central orchestrator, exemplified by the CNJ, extends the theory by emphasizing the significance of cooperation among stakeholders and centralized coordination. This empirical demonstration of defined roles related to orchestration indicates the relevance of a central coordinating entity. In essence, this case study facilitates the identification of practical instances of structuring, bundling, and leveraging processes, thereby expanding the perception of data as resources through the lens of the Resource Orchestration Theory (Lin et al., 2023; Schreieck et al., 2022; Cui and Han, 2022).

Another noteworthy contribution to the information systems literature pertains to contingencies, which are represented by the size of the courts. Interestingly, this study found that court size had no statistically significant effect on the process of improving data quality. This presents a counterpoint to the notion that contingencies exert a substantial influence on data management and governance (Sambamurthy and Zmud, 1999; Weber et al., 2009), thus motivating further research to deepen the understanding of this phenomenon within the context of orchestrated ecosystems. Building upon Miao et al. (2017), one plausible assumption is that data quality orchestration might mitigate the influence of contingencies by centralizing the coordination of resources within the data ecosystem, potentially reducing the constraints actors might face. Additionally, the results reveal the dynamic nature of data quality, which can fluctuate both positively and negatively over time. This highlights the need for continuous monitoring within the data ecosystem and the periodic review of data governance policies (Karkošková, 2023).

7 Conclusion

This study provided a detailed analysis of the effects of data quality orchestration in the Brazilian Judiciary data ecosystem. By examining data from 27 state courts over a two-year period of orchestration by the National Council of Justice (CNJ), the findings show an improvement in data quality following the introduction of orchestration. The results align with existing literature on orchestration, highlighting the role of a central coordinating body in managing and improving data quality within data ecosystems (Guggenberger et al., 2020). The CNJ's efforts, including standardization, training, and promotion of data initiatives, significantly enhanced data quality from an average of 34.5% to 69.45% over the two-year period.

Using resource orchestration theory, the study demonstrated that the CNJ's actions in structuring, bundling, and leveraging resources effectively improved data quality (Lin et al., 2023; Schreieck et al., 2022; Cui and Han, 2022). This work highlights the relevance of stakeholder collaboration and centralized management to support data quality orchestration. Remarkably, the study found that the size of the courts did not significantly influence the improvement in data quality during orchestration, suggesting that orchestration efforts can be effective across different organizational sizes.

This study contributes meaningfully to the field of information systems by providing empirical evidence on data quality orchestration. It is important to acknowledge the limitations of this research, as it primarily serves as an exploratory study on data quality orchestration within the Brazilian Judiciary. One notable limitation is the lack of detailed information about the data quality orchestration activities, including the exact start date of each action. Such detailed information could provide more precise statistics on the effects of each activity on data quality. Additionally, there was no detailed information about how the official data quality indicator was calculated by the CNJ, which limited its analysis.

Future research should explore whether data quality orchestration has influenced the performance of the courts, particularly concerning core operational activities. Another suggestion is to investigate how data quality orchestration integrates with other ecosystem initiatives. In addition, the graphical analysis suggested an inflection point in July 2022 within the historical series, which requires further research to identify potential explanations for this behavior. Lastly, it is recommended to develop a framework for data quality orchestration capabilities to aid organizations in orchestrating efforts aimed at improving data quality within the data ecosystem.

References

Al-Sai, Z. A., et al. (2023). Big data maturity assessment models: A systematic literature review. *Big Data and Cognitive Computing*, 7(2). https://doi.org/10.3390/bdcc7010002

Altendeitering, M., Dübler, S., & Guggenberger, T. (2022). Data quality in data ecosystems: Towards a design theory. *Proceedings of the Twenty-Eighth Americas Conference on Information Systems*, Minneapolis, MN. https://aisel.aisnet.org/amcis2022/DataEcoSys/DataEcoSys/3/

Autio, E. (2022). Orchestrating ecosystems: A multi-layered framework. *Innovation*, 24(1), 96-109. https://doi.org/10.1080/14479338.2021.1919120

Brazil. (1988). *Constituição da República Federativa do Brasil de 1988*. Brasília, DF: Senado Federal. https://planalto.gov.br/ccivil_03/constituicao/constituicao.htm

Cappiello, C., Gal, A., Jarke, M., & Rehof, J. (2020). Data ecosystems: Sovereign data exchange among organizations (Dagstuhl Seminar 19391). *Dagstuhl Reports*, 9(9), 66-134. https://doi.org/10.4230/DagRep.9.9.66

Choi, Y., et al. (2021). Towards data-driven decision-making in government: Identifying opportunities and challenges for data use and analytics. *Proceedings of the 54th Hawaii International Conference on System Sciences*. https://aisel.aisnet.org/hicss-54/dg/digital_transformation/7/

Conselho Nacional de Justiça (CNJ). (2022a). *Relatório de diagnóstico dos tribunais nas atividades de saneamento de dados do Datajud*. https://bibliotecadigital.cnj.jus.br/jspui/bitstream/123456789/546/1/pnud-relatorio-v2-2022-06-14.pdf

Conselho Nacional de Justiça (CNJ). (2022b). *Relatório final Gestão Ministro Luiz Fux - Programa Justiça 4.0*. https://www.cnj.jus.br/wp-content/uploads/2022/09/af-pnud-relatorio-v3-web.pdf

Conselho Nacional de Justiça (CNJ). (2023a). *Relatório de entregas do Programa Justiça 4.0 - Gestão da ministra Rosa Weber*. https://www.cnj.jus.br/wp-content/uploads/2023/09/relatorio-de-entregas-gestao-ministra-rosa-weber-26-9-23.pdf

Conselho Nacional de Justiça (CNJ). (2023b). *Acompanhamento Datajud*. https://acompanhamento-datajud.cloud.cnj.jus.br/acompanhamento_datajud/

Cui, M., Pan, S. L., Newell, S., & Cui, L. (2017). Strategy, resource orchestration and e-commerce enabled social innovation in rural China. *The Journal of Strategic Information Systems*, 26(1), 3-21. https://doi.org/10.1016/j.jsis.2016.10.001

Cui, Z., & Han, Y. (2022). Resource orchestration in the ecosystem strategy for sustainability: A Chinese case study. *Sustainable Computing: Informatics and Systems*, 36. https://doi.org/10.1016/j.suscom.2022.100796

Gartner. (2021). *How to improve your data quality*. https://www.gartner.com/smarterwithgartner/how-to-improve-your-data-quality

Geisler, S., et al. (2021). Knowledge-driven data ecosystems toward data transparency. *ACM Journal of Data and Information Quality*, 14(1). https://dl.acm.org/doi/pdf/10.1145/3467022

Gelhaar, J., Groß, T., & Otto, B. (2021). A taxonomy for data ecosystems. *Proceedings of the 54th Hawaii International Conference on System Sciences*. https://hdl.handle.net/10125/71359

Guggenberger, T. M., et al. (2020). Ecosystem types in information systems. *Proceedings of the Twenty-Eighth European Conference on Information Systems (ECIS2020)*, Marrakesh, Morocco. https://www.researchgate.net/publication/341188637

Gupta, U., & Cannon, S. (2020). Data governance maturity models. In *A practitioner's guide to data governance* (pp. 143-165). Emerald Publishing Limited. https://doi.org/10.1108/978-1-78973-567-320201007

Hannah, D. P., & Eisenhardt, K. M. (2018). How firms navigate cooperation and competition in nascent ecosystems. *Strategic Management Journal*, 39, 3163-3192. https://doi.org/10.1002/smj.2750

Hannila, H., Silvola, R., Harkonen, J., & Haapasalo, H. (2022). Data driven begins with data; Potential of data assets. *Journal of Computer Information Systems*, 62(1), 29-38. https://doi.org/10.1080/08874417.2019.1683782

Karkošková, S. (2023). Data governance model to enhance data quality in financial institutions. *Information Systems Management*, 40(1), 90-110. https://doi.org/10.1080/10580530.2022.2042628

Lin, J., et al. (2023). How to build supply chain resilience: The role of fit mechanisms between digitally-driven business capability and supply chain governance. *Information & Management*, 60(2). https://doi.org/10.1016/j.im.2022.103747

Linde, L., Sjödin, D., Parida, V., & Wincent, J. (2021). Dynamic capabilities for ecosystem orchestration: A capability-based framework for smart city innovation initiatives. *Technological Forecasting and Social Change*, 166. https://doi.org/10.1016/j.techfore.2021.120614

Mann, G., Karanasios, S., & Breidbach, C. F. (2022). Orchestrating the digital transformation of a business ecosystem. *The Journal of Strategic Information Systems*, 31(3). https://doi.org/10.1016/j.jsis.2022.101733

McKinsey & Company. (2023). *Real-world data quality: What are the opportunities and challenges*? https://www.mckinsey.com/industries/life-sciences/our-insights/real-world-data-quality-what-are-the-opportunities-and-challenges#/

Miao, C., Coombs, J., Qian, S., & Sirmon, D. (2017). The mediating role of entrepreneurial orientation: A meta-analysis of resource orchestration and cultural contingencies. *Journal of Business Research*, 77. https://doi.org/10.1016/j.jbusres.2017.03.016

Monte Carlo. (2023). *Survey: The state of data quality 2023*. https://www.montecarlodata.com/blog-data-quality-survey

Muhammad, L. N. (2023). Guidelines for repeated measures statistical analysis approaches with basic science research considerations. *Journal of Clinical Investigation*, 133(11). https://doi.org/10.1172/JCI171058

Mukhopadhyay, S., & Bouwman, H. (2019). Orchestration and governance in digital platform ecosystems: A literature review and trends. *Digital Policy, Regulation and Governance*, 21(4), 329-351. https://doi.org/10.1108/DPRG-11-2018-0067

Najafabadi, M. M., & Cronemberger, F. A. (2023). Systemic effects of an open government program on data quality: The case of the New York State's Food Protection program area. *Transforming Government: People, Process and Policy*, 17(2), 192-203. https://doi.org/10.1108/TG-11-2021-0194

Oliveira, M. I. S., & Lóscio, B. F. (2018). What is a data ecosystem? *Proceedings of the 19th Annual International Conference on Digital Government Research*, Delft, Netherlands. https://doi.org/10.1145/3209281.3209335

Omari, H., Barham, S., & Qusef, A. (2021). Data strategy and its impact on open government data quality. *Proceedings of the 2021 International Conference on Information Technology (ICIT)*, Amman, Jordan, 648-653. https://doi.org/10.1109/ICIT52682.2021.9491766

Otto, B., et al. (2019). Data ecosystems: Conceptual foundations, constituents and recommendations for action. *Fraunhofer-Gesellschaft*. https://doi.org/10.24406/publica-fhg-300110

Park, E., Cho, M., & Ki, C. S. (2009). Correct use of repeated measures analysis of variance. *The Korean Journal of Laboratory Medicine*, 29(1), 1-9. https://doi.org/10.3343/kjlm.2009.29.1.1

Purwanto, A., Zuiderwijk, A., & Janssen, M. (2020). Citizens' trust in open government data: A quantitative study about the effects of data quality, system quality and service quality. *Proceedings of the 21st Annual International Conference on Digital Government Research (dg.o '20)*. Association for Computing Machinery, New York, NY, USA, 310-318. https://doi.org/10.1145/3396956.3396958

Sambamurthy, V., & Zmud, R. W. (1999). Arrangement for information technology governance: A theory of multiple contingencies. *MIS Quarterly*, 23(2), 261-290. https://doi.org/10.2307/249754

Schreieck, M., Manuel, W., & Helmut, K. (2022). From product platform ecosystem to innovation platform ecosystem: An institutional perspective on the governance of ecosystem transformations. *Journal of the Association for Information Systems*, 23(6), 1354-1385. https://doi.org/10.17705/1jais.00764

Sirmon, D. G., et al. (2011). Resource orchestration to create competitive advantage: Breadth, depth, and life cycle effects. *Journal of Management*, 37(5), 1390-1412. https://doi.org/10.1177/0149206310385695

Teece, D. J. (2020). Hand in glove: Open innovation and the dynamic capabilities framework. *Journal of Software Maintenance and Evolution: Research and Practice*, 1, 233-253. http://dx.doi.org/10.1561/111.00000010

Vafaei-Zadeh, A., Ramayah, T., Hanifah, H., Kurnia, S., & Mahmud, I. (2020). Supply chain information integration and its impact on the operational performance of manufacturing firms in Malaysia. *Information & Management*, 57(8). https://doi.org/10.1016/j.im.2020.103386

Vetrò, A., et al. (2016). Open data quality measurement framework: Definition and application to open government data. *Government Information Quarterly*, 33. https://doi.org/10.1016/j.giq.2016.02.001

Vittinghoff, E., McCulloch, C. E., Glidden, D. V., & Shiboski, S. C. (2011). Linear and non-linear regression methods in epidemiology and biostatistics. In C. R. Rao, J. P. Miller, & D. C. Rao

(Eds.), *Essential statistical methods for medical statistics* (pp. 87-113). North-Holland. https://doi.org/10.1016/B978-0-444-53737-9.50006-2

Wang, P. (2021). Connecting the parts with the whole: Toward an information ecology theory of digital innovation ecosystems. *MIS Quarterly*, 45(1). https://aisel.aisnet.org/misq/vol45/iss1/15/

Weber, K., Otto, B., & Österle, H. (2009). One size does not fit all – A contingency approach to data governance. *Journal of Data and Information Quality*, 1(1), 1-27. https://doi.org/10.1145/1515693.15156967

Zhang, D., et al. (2022). Orchestrating big data analytics capability for sustainability: A study of air pollution management in China. *Information & Management*, 59. https://doi.org/10.1016/j.im.2019.103231

Zhang, R., Sadiq, S., & Indulska, M. (2019). Discovering data quality problems. *Business & Information Systems Engineering*, 61(5), 575-593. https://aisel.aisnet.org/bise/vol61/iss5/3