Journal of Innovation & Knowledge 9 (2024) 100598

Journal of Innovation & Knowledge

https://www.journals.elsevier.com/journal-of-innovation-and-knowledge

Data governance & quality management—Innovation and breakthroughs across different fields

Bruno Miguel Vital Bernardo[a,*], Henrique São Mamede[b], João Manuel Pereira Barroso[a], Vítor Manuel Pereira Duarte dos Santos[c]

[a] INESC TEC and Universidade de Trás-os-Montes e Alto Douro, Portugal
[b] INESC TEC and Universidade Aberta, Portugal
[c] MagIC and NOVA IMS, Universidade Nova de Lisboa, Portugal

ARTICLE INFO

Article History:
Received 28 August 2023
Accepted 11 October 2024
Available online 30 October 2024

Keywords:
Data governance
Data quality
Data assurance
Digital forensics
PRISMA
Bibliometric mapping

JEL classification:
G20
G30
L00
M10
M42

ABSTRACT

In today's rapidly evolving digital landscape, the substantial advance and rapid growth of data presents companies and their operations with a set of opportunities from different sources that can profoundly impact their competitiveness and success. The literature suggests that data can be considered a hidden weapon that fosters decision-making while determining a company's success in a rapidly changing market. Data are also used to support most organizational activities and decisions. As a result, information, effective data governance, and technology utilization will play a significant role in controlling and maximizing the value of enterprises. This article conducts an extensive methodological and systematic review of the data governance field, covering its key concepts, frameworks, and maturity assessment models. Our goal is to establish the current baseline of knowledge in this field while providing differentiated and unique insights, namely by exploring the relationship between data governance, data assurance, and digital forensics. By analyzing the existing literature, we seek to identify critical practices, challenges, and opportunities for improvement within the data governance discipline while providing organizations, practitioners, and scientists with the necessary knowledge and tools to guide them in the practical definition and application of data governance initiatives.

Introduction

Background

The ever-expanding volume of data is a key asset for companies (Bennett, 2015). Data can be regarded as a "secret weapon" that determines an organization's capability to make informed decisions and maintain competitiveness in ever-changing and volatile markets, including the financial and industrial sectors (Ragan & Strasser, 2020, p. 1). To effectively manage their data, organizations must acknowledge and recognize the impact that robust and efficient data governance can have on achieving success and making well-founded decisions (Hoppszallern, 2015).

Throughout the years and as data and technology have advanced, the need to find solutions that ensure adequate data governance has grown significantly (Dutta, 2016). As part of the data governance procedure, strategy and actions, it is essential for data to be governed in a way that guarantees accuracy, integrity, validity, and completeness. However, many organizations are currently approaching data governance solely through the narrow lens of non-transactional data (data at rest). They are overlooking the risks associated with transactional data that may hinder their success and ability to be competitive in the market (Dutta, 2016). Hence, this narrow approach exposes organizations to significant risks, namely information errors. These errors can lead to increases in costs, reputational and financial losses, compliance risks, and others that may arise from the organizations' inability to define an adequate data governance and quality framework (Dutta, 2016).

This field is crucial for an organization's success and any activity performed by data governance and quality practitioners. Recent trends demonstrate that the world and its market are being characterized by an increase in compliance and regulatory requirements, advances and changes in the technology and data landscape, and a heightened focus on excellence, financial governance, and reputation (Dutta, 2016). These trends are compelling organizations to reassess

* Corresponding author.
E-mail address: bmvb1995@gmail.com (B.M.V. Bernardo).

https://doi.org/10.1016/j.jik.2024.100598

their data governance initiatives and to define and establish frameworks that address their specific data governance needs, providing management with a set of processes that allow for the governance and quality assurance of transactional and non-transaction data (Dutta, 2016). Nonetheless, since data governance corresponds to a complex and evolving field that can be considered one of an organization's most imperative and intricate aspects, there is still a gap to be filled by companies. This gap relates to the necessity for companies and their management to integrate and elevate data governance in their business operations to the highest level (Alhassan et al., 2019a; Bernardo et al., 2022; Johnston, 2016; Sifter, 2017).

According to Zorrilla and Yebenes (2022), data governance is not just a part but a cornerstone of digital transformation, particularly within the context of the Industrial Fourth Revolution (Industry 4.0, also known as I4.0). This revolution represents a significant shift in how organizations manage and control the value derived from the entire life cycle of a given product, or service. These authors emphasize that the digitalization of the industrial environment, achieved through the integration of operational and information technologies (OT and IT), is heavily influenced by effective data governance. This includes the use of cyber-physical systems, the Industrial Internet of Things (IIOT), and the application of real-time data generation for decision-making and insight-gathering.

Given the interconnected and interdependent nature of data, people, processes, services, and cyber-physical systems, the depth and complexity of digital transformation becomes clear. In line with this, for data to become an organization's competitive edge, it must be managed and governed like any other strategic asset, if not more so. Therefore, implementing a data governance framework is essential. This framework should define who has the authority and control to make decisions about data assets, facilitate shared and communicated decision-making, and establish the necessary capabilities within an organization to support these functions (Zorrilla & Yebenes, 2022).

Similarly, organizations must recognize that developing, designing, implementing, and continuously monitoring a data governance program and its initiatives requires a significant investment of resources, including personnel, time, and funds. Besides, building a robust data governance program—including its model, culture, and structure—takes time to fully address the complete needs of an organization, including business operations, risk management, compliance, and choices (Bernardo et al., 2022; Lancaster et al., 2019; Sifter, 2017). Bennett (2015) also emphasizes the vital need for companies to comprehend data governance, its components, standards, and the criteria that the framework must meet. Without this comprehension, enterprises may risk mismanaging data privacy and the information duties they hold and manage, which could lead to financial, operational, and reputational harm (Bernardo et al., 2022).

In a similar vein, Lee (2019) points out that while organizations and their boards are focused on improving cybersecurity, they often neglect to invest sufficient effort and resources in defining and developing a more robust and extensive data governance framework. Such a framework should address the accuracy, protection, availability, and usage of data. Although data governance intersects with cybersecurity, it is a broader field that encompasses additional components. These include concerns that should be considered as high-priority objectives for organizations, such as data management and governance, quality principles, the definition of data roles, responsibilities, processes, and compliance with data regulations and privacy laws (Bernardo et al., 2022; Lee, 2019). In reality, organizations have historically neglected and overlooked data governance, providing this field with minimal attention due to the high level of investment and complexity that it would require (Bernardo et al., 2022; Janssen et al., 2020). Similarly, corporations are refocusing and shifting their efforts to address data governance concerns, recognizing it as one of the top three factors that differentiate successful businesses from those that fail to extract value from their data. As a result, companies are developing their stakeholder positions and roles so that they can emphasize the value of technology, people, and processes within their data governance programs (Bernardo et al., 2022; Janssen et al., 2020).

Moreover, the literature confirms that many businesses across various industries and markets lack a sound data governance and management framework, structure, and plan that could shield them from potential harm, namely data disasters, losses, and system failures (Bernardo et al., 2022; Johnston, 2016; Zhang et al., 2016). Also, some authors emphasize how vital it is for enterprises to define robust auditing and assurance procedures to enhance their data governance and ensure that they are implemented in a productive, approach-oriented, continuous, and compliant manner (Bernardo et al., 2022; Johnston, 2016; Perrin, 2020). Furthermore, Cerrillo-Martínez and Casadesús-de-Mingo (2021) suggest that while this field has great potential, literature sources and guidelines on the subject are still "scarce and generally excessively theoretical."

The literature alerts the community to the fact that businesses are more focused on experimenting with and utilizing artificial intelligence than on ensuring the quality of the data life cycle. This neglect includes the processes of acquiring, collecting, managing, using, reporting, and safeguarding, or destroying data—steps that are crucial for establishing a solid foundation for artificial intelligence to be used (Bernardo et al., 2022; Janssen et al., 2020). Even though these stages of data quality require a tremendous amount of time, people, and effort, companies tend to give them minimal attention. In reality, companies should prioritize developing and delivering initiatives that would allow them to identify the critical data sets, understand their nature and sources, track their flow through people, processes, and systems, and enhance their knowledge on data governance and quality (Bernardo et al., 2022; Janssen et al., 2020).

In addition, data-dependent activities such as data assurance and digital forensics analysis are strongly affected by an organization's maturity in governing data. Thus, organizations should primarily build and define a solid data governance framework before conducting rigorous digital forensics and data assurance analysis. In fact, only after setting up this data governance framework should organizations engage in forensics and data assurance tasks, because doing so can potentially improve their daily activities and operations, increase and improve data quality, and strengthen data-dependent activities such as reporting and decision-making. Otherwise, without proper data governance, organizations risk making poor decisions based on inadequate or inaccurate data (Bernardo et al., 2022; Ragan & Strasser, 2020).

The challenges associated with the data governance field are growing daily, particularly due to the relentless increase in data and the ongoing race to improve efficiency and competitiveness in the market (Paredes, 2016). Companies are concentrating on identifying roles and responsibilities, such as the Chief Data/Digital Officer (CDO), to lead and govern their data governance frameworks (Bennett, 2015; Bernardo et al., 2022). Ragan and Strasser (2020) emphasize the need for organizations to nominate and design a Data Czar, or other role to oversee their data governance initiatives. However, many firms and stakeholders resist this change due to the complexity and high investment required for a data governance framework (Bernardo et al., 2022; Ragan & Strasser, 2020). Additionally, without a comprehensive understanding of the business and data flow, organizations will face challenges in defining leadership positions, roles, and responsibilities necessary to govern their data and ensure their strategy is effectively implemented. Nonetheless, this overall process is inherently complex because there is no single method, framework or approach to data governance that fits all organizations in a standardized manner (Bernardo et al., 2022; Paredes, 2016; Ragan & Strasser, 2020).

Related studies

This section aims to summarize previous publications related to literature reviews, including systematic ones, on data governance topics. Sixteen papers were analyzed, revealing that they all focused on defining data governance principles and foundations. The analysis showed that the majority address general aspects of this field, namely key principles (16/16 articles), challenges (10/16 articles), and roles and responsibilities (11/16 articles). However, there was limited exploration of more in-depth essential topics, including: i) the exploration of data governance frameworks and their comparisons (1/16 articles); ii) the data lifecycle and assurance (6/16 articles); iii) analysis on existing data governance tools (3/16 articles); and iv) exploration of data governance maturity models and their impact (2/16 articles).

Moreover, we found that only half of the articles presented a data governance framework, and only one paper, Al-Ruithe et al. (2019), included a comparison of different frameworks. This paper is the closest one related to our work, particularly in its presentation and comparison of data governance frameworks. However, it did not include the relationship between these frameworks with the field's key components that our work intends to explore, such as structures, responsibilities, existing tools, lifecycle, and maturity models.

Much of the research literature was provided and published before 2020. We verified that of the 16 papers reviewed, approximately 69 % were dated and included papers published by 2017, 12 % included papers released by 2019, and the remaining 19 % by 2020. Our research, which was conducted using a thorough and robust process, consists of the latest publications on the subject available at the time of writing. We employed an extensive research process, utilizing relevant databases and rigorous methodologies such as PRISMA and a bibliometric analysis of the data obtained, minimizing the likelihood of not including relevant publications.

Table 1 shows that while most literature reviews focus mainly on the theoretical aspects of data governance, some do address specific domains, such as the data lifecycle, roles and responsibilities, and framework presentation. However, we did not find any article that comprehensively covers all these domains and establishes relationships between them, such as the analysis of multiple frameworks alongside data maturity models. Therefore, Table 1 illustrates our proposal for a novel and comprehensive approach to understanding the critical domains of data governance. This inclusive and accessible approach is intended not only for data governance practitioners and organizations but also for anyone seeking a deeper understanding of the field. It covers a wide range of topics, including data governance concepts and principles, challenges, a framework presentation, a comparison and analysis, the data lifecycle and assurance, existing functions and structures within a data governance framework, data governance tools, and maturity models.

Study contributions

In this article, we address and justify the existing gap in characterizing the foundations of data governance, its evolution, and its maturity within enterprises and among its practitioners, which has been highlighted by several authors in the literature. One of the major challenges is the insufficient consideration and awareness of the importance of data governance and its key concepts, coupled with a lack of sufficient literature sources to guide and support organizations. To address this gap, we provide a detailed analysis of the data governance field, covering all previously described components and examining the relationship between data governance and other data-dependent fields, such as data assurance and digital forensics.

Throughout this article, we explore the relationship between data governance and two different but connected data-dependent fields: data assurance and digital forensics. Data assurance focuses on

Table 1
Analysis of papers focusing on the review of the various domains of Data Governance (DG).

Articles Included	Paper Focus								Period (Published up to)
	DG Principles	DG Challenges	DG framework presentation	DG framework Comparison	Data lifecycle and Assurance	Roles / Responsibilities	DG Tools	DG Maturity Models	
(Bennett, 2015)	✓	✓			✓	✓			2017
(Janssen et al., 2020)	✓	✓	✓			✓			2020
(Cerrillo-Martínez & Casadesús-de-Mingo, 2021)	✓	✓				✓			2020
(Al-Ruithe et al., 2019)	✓	✓	✓	✓		✓			2017
(Abraham et al., 2019)	✓	✓	✓			✓			2019
(Alhassan et al., 2016)	✓	✓	✓						2015
(Bordey, 2018)	✓	✓	✓						2016
(Chakravorty, 2020)	✓	✓	✓		✓	✓			2020
(Clarke, 2016)	✓				✓	✓	✓		2018
(Demarquet, 2016)	✓				✓	✓			2015
(George et al., 2017)	✓				✓				2017
(Hay, 2015)	✓		✓				✓		2013
(Koltay, 2016)	✓	✓				✓			2016
(Meyers, 2014)	✓								2012
(McDowall, 2017a, b)	✓					✓		✓	2017
This Research Article	✓	✓	✓	✓	✓	✓	✓	✓	**Up to date analysis (2023)**

ensuring the quality and reliability of data throughout the lifecycle, including collection, processing, storage, and reporting stages. Digital forensics is a relatively new discipline within forensic science that offers rigorous methodologies and procedures for data examination and evaluation. This field can be applied in various areas beyond judicial proceedings, providing valuable support for organizations. Our primary reason for this focus is based on the fact that these two areas are vital and represent some of the most novel and emerging areas for the success of data governance. Moreover, both fields, alongside data governance, share a common focus on maintaining data integrity, completeness, and security throughout the data management lifecycle. In addition, digital forensics relies heavily on robust data governance policies to ensure data integrity, transparency, and authenticity. Meanwhile, data assurance focuses on complementing the data governance field with a set of mechanisms to continuously assess, validate and safeguard data quality. This ensures organizations remain in compliance with internal policies, current laws and regulations, and international standards.

As Grobler (2010a) emphasizes, the field of digital forensics can support an organization's operations. It specifically helps management and data users achieve the organization's data strategy, increases their ability to use digital resources, and complements the role of technology and information within the business context. Grobler further states that the interoperability of information systems and the increasing frequency of incidents heighten the need for organizations to leverage digital forensics in establishing their data governance framework (Grobler, 2010a, b).

Similarly, the literature highlights that data governance effectiveness is not solely dependent on the quality of the data, but rather on the absence of well-established policies and procedures, the organization's inability to comply with these standards and the lack of monitoring mechanisms such as the Key Performance Indicators (KPIs) and Key Risk Indicators (KRIs). As a result, organizations must establish methods to ensure that their data assets and the data used within their specific context have their quality attributes assured (Cheong & Chang, 2007; Hikmawati et al., 2021). Likewise, data assurance is closely related to the data governance field because it provides features for analyzing, validating, and verifying data integrity, quality, and compliance. This process helps prevent errors and ensures adherence to policies, standards, and relevant legislation. Therefore, we believe that by exploring digital forensics and data assurance, organizations can achieve a more robust and effective data governance. This will enable companies to: i) implement precise data collection methods, including the identification of relevant data and sources while preserving data quality attributes such as accuracy, consistency, completeness, availability and relevance; ii) enhance their data extraction capabilities using tools, especially automated ones, to preserve data assets and ensure authenticity of the gathered data; and iii) conduct thorough analyses supported by documented data lineage and generate reports based on high-quality data (Martini & Choo, 2012).

It is important to note that data governance interacts with other relevant disciplines and complementary areas, such as data security, business analytics, and artificial intelligence. However, while these areas are related to data governance, our comprehensive analysis indicates that the literature covers data governance more extensively than the fields of digital forensics and assurance, which are considered more novel and emerging.

This study includes a systematic literature review and an in-depth bibliometric examination of the aforementioned topics. It identifies key issues, opportunities, and challenges within these fields and clearly defines the primary research question as follows:

"How to conduct an **extensive methodological systematic review** to define, design, and enhance a **data governance program** - breaking through the fundamentals of Data Governance, Data Assurance & Digital Forensics Sciences."

While seeking the application of the four different stages over the systematic literature review, the purpose of this study is to clearly answer the research questions (RQ) posed, namely:

- **RQ1**: What are the key challenges and opportunities in data governance, and what benefits and issues can be learnt from successful implementations?

In addressing Research Question 1, we focus on analyzing the number of benefits and challenges associated with the definition and implementation of a data governance framework, particularly through real case studies of organizations that have undergone this process. We believe that the benefits and opportunities in this field, outweigh the number of challenges currently faced. Nonetheless, it is crucial to understand what these challenges mean for an organization and what lessons can be learned from existing case studies.

- **RQ2**: What is the current level of maturity and background surrounding the topic of data governance?

For Research Question 2, we aim to analyze the current state of the data governance field and the literature around it to determine its maturity level. Consequently, we examined what the literature identifies as key aspects of the data governance framework that led to greater maturity levels, including existing frameworks, data stewardship roles, and consistent data management practices.

- **RQ3**: What are the current methodologies to support a data governance program and assess its maturity level?

For Research Question 3, we aim to identify the current and different methodologies that support the data governance field, including different frameworks and maturity assessment techniques. We believe that these practices will help organizations evaluate their current practices, identify gaps, and set objectives for improvement, ultimately leading to a more robust data governance environment.

- **RQ4**: What significant baselines from other fields can enhance data governance activity, structure, and archaeology?

Research Question 4 analyzes and explores the foundational principles of other data-related fields whose synergies can enhance an organization's data governance program, with a particular focus on data assurance and digital forensics. We believe that these connections, which have not yet been fully explored by the academic community, could lead to improved data quality management and more precise analytics techniques, offering deeper insights into strategic decision-making.

- **RQ5**: What are the main obstacles and constraints that the data governance discipline is currently facing and may face in the future?

Research Question 5 seeks to showcase the advantages of data governance through real-life case studies of successful data governance frameworks. It also aims to identify the obstacles and challenges organizations face in this field, such as resistance to change, the complexity of integrating diverse data systems, and a lack of specialized skills and management support. Additionally, it considers future challenges that may involve managing growing data volumes and ensuring data privacy and security in an increasingly digital and interconnected environment.

By answering these questions, this article aims to inform and demonstrate to the community how to break through the essential concepts of the data governance field, ultimately fostering its quality. At the same time, we seek to raise awareness and offer a different

approach and perspective to understanding the main challenges faced by organizations. We believe that defining these research questions will bring added value to these fields, benefiting both specialists and organizations. This will lead to the definition of four main objectives, which correspond to the following statements:

- **OB1**: Identify, characterize, and evaluate the current environment, namely the latest and most appropriate data governance tools and methodologies, to lay the foundation for OB2, OB3, and OB4;
- **OB2**: Identify the importance of data governance frameworks that provide a sustainable and robust foundation for data governance programs, enabling organizations to obtain a fully committed and up-to-date environment;
- **OB3:** Comprehend the available tools, maturity assessment procedures, and methods, and explore how organizations can leverage them;
- **OB4:** Acknowledge, classify, and compare the differences and similarities among existing tools and methodologies regarding data governance and data quality to find potential barriers and synergies for future exploitation.

These objectives add value to the data governance field by identifying and addressing existing gaps, raising awareness on this field to any researcher and practitioner, and supporting broader goals. Additionally, this work contributes to goal 9 of the UN Global Goals (Industry and Infrastructure) which aims to expand and increase scientific research, enhance the technological capabilities of industrial sectors worldwide, and promote innovation by 2030 (Denoncourt, 2020; United Nations, Washington, DC, 2020).

Phase I – Data, methods and planning

Given the complexity of the data governance field, a systematic review of the literature is essential for an in-depth analysis. To achieve this, we employed several methodologies to collect information and data that met pre-defined qualifying criteria, aiming to address the research questions and objectives effectively. As illustrated in Figs. 1 and 2, we structured the article and research into four stages: i) planning and definition, ii) execution of the systematic literature review and its techniques, iii) analysis and discussion of the findings, and iv) conclusions and recommendations for future work.

Likewise, the details for Phase II are presented in Fig. 2, which includes the application of the PRISMA methodology, full-text analysis, scrutiny methods, quality assessment, and bibliographic mapping of the data analyzed.

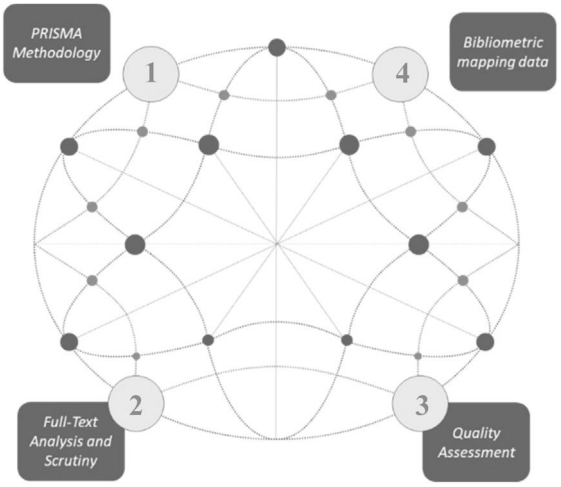

Fig. 2. Overview of the techniques applied in the systematic literature review.

The theoretical background was established using the well-established systematic literature review method known as PRISMA (Preferred Reporting Items for Systematics and Meta-Analyses). This approach was employed to identify and examine relevant literature in any form that could support this research. It also provided access to the most up-to-date scientific publications encompassing knowledge that is already available and validated by the community (Okoli & Schabram, 2010). The literature review process aims not only to summarize previous concepts and findings but also to present new evidence. This evidence is derived from consolidating all the activities in this article, including the PRISMA method, annotated bibliography, quality assessment, and bibliometric network analysis (Hong & Pluye, 2018).

Similarly, PRISMA is designed to provide scientific documentation on research (Moher et al., 2015). It involves a 27-item checklist containing four main steps that comprise its statements, with the primary goal of aiding researchers improve systematic literature reviews (Moher et al., 2009). In this context, the main goal is to compile and provide an overall analysis on data governance, data assurance and digital fields by scrutinizing the available literature (Palmatier et al., 2018; Snyder, 2019). By integrating different perspectives, we address the research questions and identify areas needing further investigation (Snyder, 2019). As a result, the theoretical background was developed to facilitate more comprehensive research in these fields. Moreover, this structure remains adaptable,

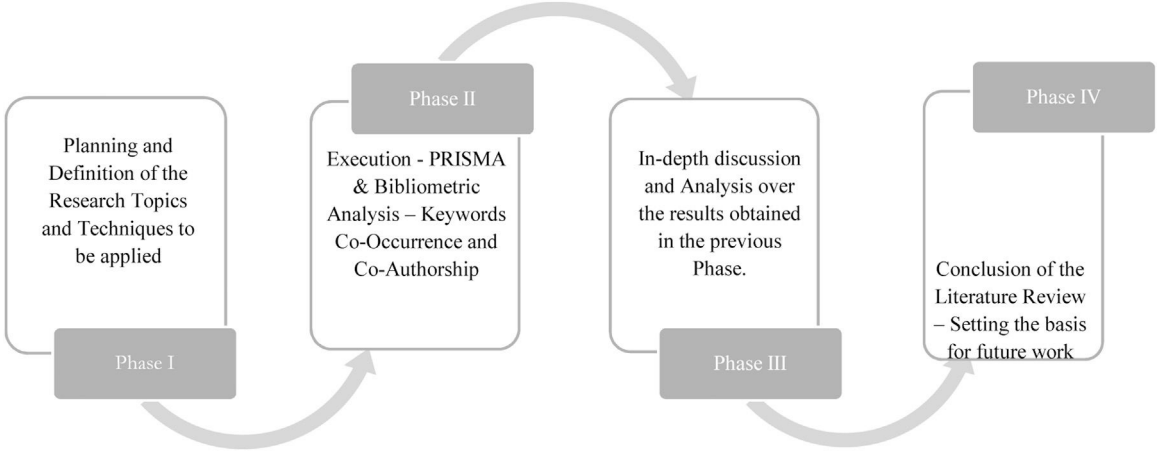

Fig. 1. Systematic literature review phases.

allowing for the inclusion of additional literature that may be appropriate for this research. Finally, the PRISMA application can guide the search process using different keyword queries across publication databases (Moher et al., 2015).

It is crucial to understand that the broad range of information available across various databases and levels of significance can sometimes mislead researchers. The literature indicates that with the rapid evolution of Information and Systems Technology, there are now an almost limitless number of open and closed-access publications, journals, search engines, and databases. This abundance necessitates a structured methodological literature review process to be able to define levels and eligibility criteria for articles to be included in the theoretical background review (Bannister & Janssen, 2019; Smallbone & Quinton, 2011).

Additionally, considering the initial state-of-art and background review, we recognize that rapid advances and adoption in the fast-paced IT industry have impacted these fields. This has introduced several numerous prospects and challenges that, although prominent in this field, are not exclusive to it. Therefore, an essential step in the literature review is to include existing studies and research and outline the roadmap for analysis (Snyder, 2019). The review roadmap focused on the topics presented in Fig. 3, namely:

Fig. 3. Roadmap of the literature review.

Phase II – Execution

Systematic literature review - PRISMA application

Using PRISMA, we identified and collected publications that meet specific predefined criteria, represented by variables such as the search query, database reference, time frame, publication years, language, and other relevant factors. In this first step, scholarly studies and literature considered during the exploration and search process are retrieved from the general database "EBSCOhost Online" (the primary research base). Following PRISMA's guidelines, the SLR, illustrated in Fig. 4, encompassed four main steps: 1) Identification, 2) Screening, 3) Suitability, and 4) Inclusion.

Consequently, a flowchart was produced and adapted from the R package developed and presented by Haddaway et al. (2022) for producing PRISMA flowcharts. The search query and its expression were designed by considering publication characteristics, namely their abstract, title, and keywords. The expression was applied by including Boolean logical operators, i.e., "AND"/"OR," to establish logical relationships. For Stage 1, Identification, the subsequent search query was formalized and applied to target publications containing specific terms, or expressions in their abstract, title, and keywords, including the following:

("data governance" OR "Data Assurance and Governance" OR "Data Forensics and Governance" OR "Data Govern" OR "data

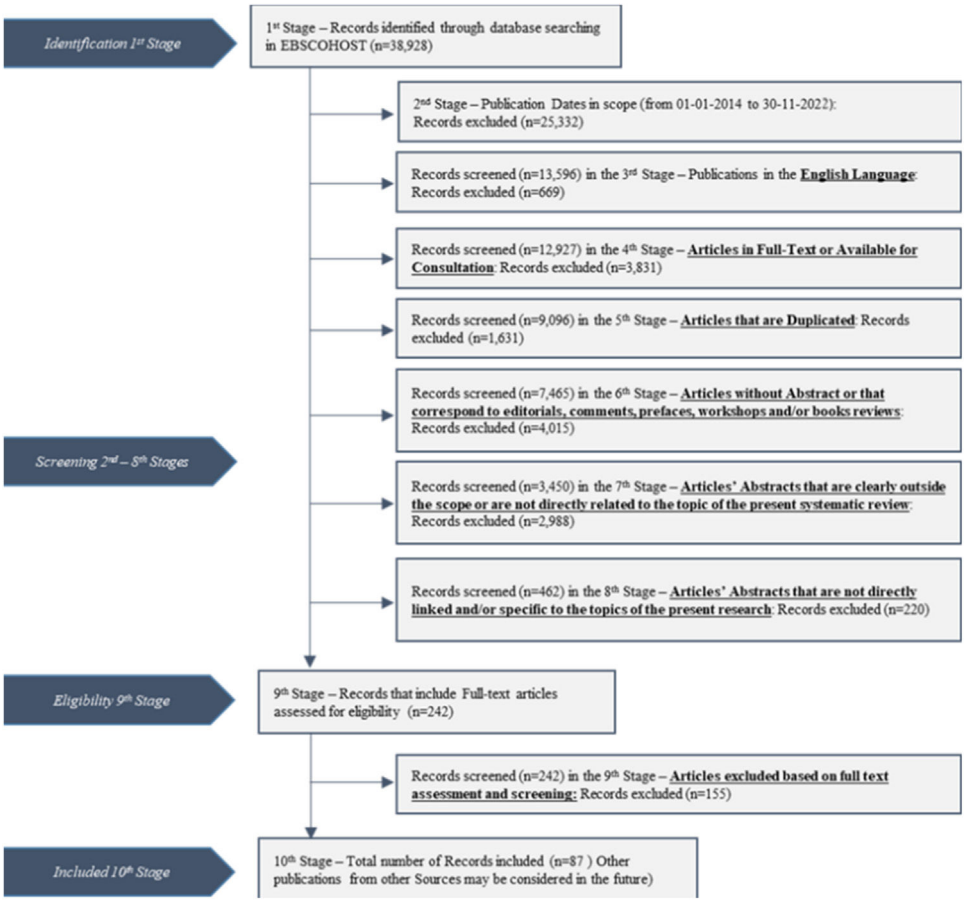

Fig. 4. PRISMA methodology flowchart.

governance & Quality" OR "data governance and Assurance" OR "data governance and Forensics" OR "data governance and Quality" OR "Data Quality & Governance" OR "Data Quality and Governance" OR "Govern of Data" OR "Digital Forensics" OR "Digital Forensic").

The search expression outlined above focused on the main topics and any derivations that can be obtained from these terms. Also, the search process was conducted in November 2022 using EBSCOhost Online and its advanced search engine as the primary database. Google Scholar was also used as a secondary database to identify additional relevant publications beyond what was found on EBSCOhost. The analysis of these sources was carried out in 2023.

The PRISMA application assisted in creating a flowchart that represents and includes each stage of the methodology: Stage 1 (Identification), Stage 2-8 (Screening and exclusion criteria), Stage 9 (Eligibility), and Stage 10 (Inclusion). In Stage 1, the search query identified a total of 38,928 records through database searches in EBSCOhost Online.

Given the total number of records, it was essential to define the screening and exclusion stages (Stages 2-8) to filter out studies that adhered to the pre-established exclusion criteria. These criteria included, in the early stage (Stage 2), the requirement that publications fall within the date range from 1 January 2014 to 30 November 2022, resulting in the exclusion of 25,332 publications that fell outside of this range. In Stage 3, a total of 13,596 articles were screened to retrieve only English-language publications, excluding 669 articles. Stage 4 focused on removing publications that were not accessible for analysis and in full text, addressing the 12,927 articles from the previous run and eliminating 3,831 publications. During this screening process, we excluded references that, despite appearing in the search query, lacked available documentation, or full-text resources for our full-text review. Despite these necessary exclusions, our literature review included what we believe are the most relevant articles on the topics. We considered it essential to filter these types of references to maintain the integrity and reliability of the literature review, ensuring that it only contains accessible and verifiable references for further consultation by researchers, though this approach may limit the scope of the review in some cases. Moreover, Stage 5 focused on analyzing the 9,096 articles retrieved so far. In this Stage, the primary task was to remove duplicate studies, i.e., instances where the exact same publication appeared more than once. This process led to the exclusion of 1,631 articles, leaving 7,465 articles to be analyzed in Stage 6. Here, the main goal was to filter and eliminate studies that did not have an abstract section, or were categorized as editorial columns, observations, preludes, book volumes, reviews, and workshops, resulting in the exclusion of 4,015 studies.

This process resulted in 3,450 articles obtained for the run in Stage 7, which focused on excluding publications with abstracts that were clearly outside the research topic, or not directly related to the analysis. In this Stage, 2,988 articles were removed, leaving 462 articles to be included in Stage 8. As a result, in Stage 8, the main objective was to exclude articles whose abstracts were not closely connected to, or explicit about the research subject, resulting in the exclusion of 220 records and leaving 242 for Stage 9, the eligibility phase. Consequently, these 242 articles were retrieved for a full-text, in-depth assessment of their content, relevance, and importance to the studied topics. After this full-text analysis in Stage 9, 155 articles were excluded, leaving 87 articles to be incorporated in the literature review. Additional records may be included if any article is later identified as a vital source of information for this paper.

The final 87 articles obtained through the PRISMA process were distributed relatively evenly across different publication dates. This ensures that the literature review encompasses both older and more recent articles, providing the researcher with a broader perspective on the topics under analysis.

Finally, after applying the PRISMA methodology at each step, presented in Fig. 4, 87 articles were deemed eligible and incorporated into the literature review.

To assist in managing the bibliography and references, the BibTeX information of these articles was imported to Zotero, an open-source reference management tool. Consequently, this tool played a crucial role in the quantitative and qualitative analysis of the literature and in managing the references used throughout this paper (Idri, 2015).

Bibliography quality assessment

In addition to applying PRISMA, we conducted further analysis by developing a quantitative quality assessment metric to support the literature review. The metric scored articles on a scale from 0.0 (Poor) to 0.5 (Mild) and 1.0 (Good), with a maximum score of 4.0 per article. Articles scoring above the cut-off score of 2.0 were considered more relevant for the literature review, while those scoring below 2.0 were excluded.

This assessment was fully supported by a detailed full-text analysis, with scores assigned based on the criteria presented in the framework shown in Fig. 5:

- **Objective:** Does this article contain topics relevant to my work?
- **Objective:** Does this article provide factual indicators relevant to my work?
- **Subjective**: Does the article contain valuable topics for me?
- **Subjective:** Does it represent an article acknowledging critical aspects of the topic being studied?

Onwuegbuzie and Frels (2015) state that the quality assessment and its metric will be used to address gaps identified in the literature review process and to generate a structure that reflects the researcher's value system. Winning and Beverley (2003) argue that a key concern for researchers when performing reviews should be the trustworthiness and authenticity of the methodologies applied in the theoretical background review process. Therefore, these assessment techniques were designed to enhance the thoroughness and transparency of this process. As Bowen (2009) suggests, the researcher, being the subjective interpreter of information and data in publications, should strive for an analysis process that is as accurate and transparent as possible.

Accordingly, the literature identifies several gaps that affect the literature review process, including: 1) insufficient description of the document analysis and the processes used to perform the literature review; 2) inadequate understanding of what is already known regarding a particular field, leading researchers to investigate research questions on fields and topics that have already been thoroughly analyzed by others; 3) lack of consideration of researcher bias, which can lead to the selection of studies that align with the researcher's perspective; and 4) poor understanding of existing methodological frameworks and techniques regarding how they were applied and how conclusions were derived (Bowen, 2009; Caldwell & Bennett, 2020; Deady, 2011; Levac et al., 2010).

Various authors emphasize the need for researchers to include a quantitative quality evaluation approach in their review process. This approach adds a robust and rigorous technique that extends beyond traditional qualitative assessments, particularly by uncovering and analyzing the publications' metadata and scientific information (Campos et al., 2018; Mackenzie & Knipe, 2006; Major, 2010; Niazi, 2015).

Consequently, we conducted a full-text analysis of the 87 articles, evaluating both objective and subjective perspectives. As a result, Table 2 presents the quality assessment of all 87 articles that reached this phase. As shown in Table 2, 12 articles were assessed and scored below the cut-off score of 2.0 and were therefore excluded because they were deemed irrelevant, or of no value for the topics under analysis. The quality assessment resulted in 75 articles being selected for

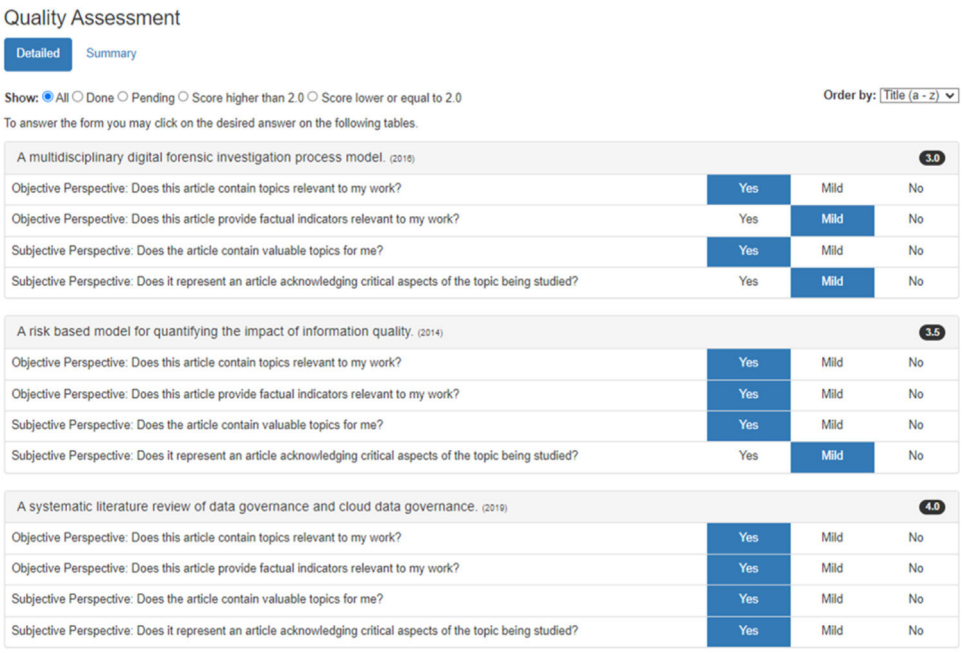

Fig. 5. Example of the quality assessment conducted.

Table 2
Bibliographic references – quality assessment.

#	Bibliographic Reference	Quality Score
1	Privacy Governance and the GDPR: How Are Organizations Taking Action to Comply with the New Privacy Regulations in Europe?	2.5
2	Managing cross-regulatory data challenges in practice.	4
3	Microsoft Releases Office 365 Security, data governance Tools.	0.5
4	Responsible data governance in Projects: Applying a Responsible Research and Innovation (RRI) Framework.	3.5
5	Seven Best Practices to Boost Big data governance Efforts.	4
6	Spotlight on a Discipline: Forensics.	3.5
7	TDWI Technology Survey: The State of data governance.	2.5
8	The centerpiece of data governance: Making information quality pay off.	4
9	The chequered past and risky future of digital forensics.	2.5
10	The Data Divide: Data ethics and data governance need to be part of every employee's onboarding, highlighting their responsibility along the supply chain.	2.5
11	Towards a Systemic Framework for Digital Forensic Readiness.	2.5
12	Understanding data governance, Part II.	4
13	Understanding data governance, Part I.	4
14	Visualizing Digital Forensic Datasets: A Proof of Concept.	3.5
15	We need to think about data governance for dementia research in a digital era.	3.5
16	A multidisciplinary digital forensic investigation process model.	3
17	A risk based model for quantifying the impact of information quality.	3.5
18	A triage framework for digital forensics.	3
19	Agile in data governance Design.	2.5
20	Are you prepared for your next data disaster?	3
21	Are You Creating a Data Swamp?	2.5
22	Artificial intelligence in the context of data governance.	3
23	Artificial intelligence and moral rights.	2.5
24	Beyond the Hype: Data Management and data governance.	3
25	What is information governance and how does it differ from data governance?	4
26	Challenges ahead on the digital forensics and audit trails.	2.5
27	Carefully communicate performance metrics.	1
28	Changes in roles, responsibilities and ownership in organizing master data management.	3.5
29	Challenges in digital forensics.	3
30	Common challenges of data governance.	4
31	Commanding data governance.	2.5
32	Critical Success Factors for data governance: A Theory Building Approach.	4
33	Critical success factors for data governance: a telecommunications case study.	4
34	Critical Factors in data governance for Learning Analytics.	0.5
35	What We See, What We Don't See: data governance, Archaeological Spatial Databases and the Rights of Indigenous Peoples in an Age of Big Data.	4
36	data governance 101: IR's Critical Role in data governance.	2.5
37	Data governance and protection.	2.5
38	data governance and Its Scientific Outlook In Indonesia: A Literature Review.	0.5
39	data governance Gamification.	3
40	Data governance for public transparency.	4

(*continued*)

Table 2 (*Continued*)

#	Bibliographic Reference	Quality Score
41	data governance in the Age of Programmatic Advertising.	2.5
42	Data governance: A conceptual framework, structured review, and research agenda.	3.5
43	data governance.	3.5
44	Data Vision vs. Data Strategy: Why Credit Unions Need Both: Learn why locating your CU's North Star - the guiding light that will inform its data analytics strategy - is so crucial.	3
45	Designing data governance that delivers value.	3.5
46	Digital Forensic Readiness: Are We There Yet?	3
47	Why data governance Should Be Part of Your Boardroom Conversations.	3
48	Effective data governance: From strategy through to implementation.	4
49	Establishing a data governance Center of Excellence Within Your Bank.	3.5
50	Establishing an Effective data governance System: Data governance is necessary for compliance with current regulatory expectations for data integrity in pharmaceutical R&D and manufacturing organizations.	3
51	Five Key Reasons Enterprise data governance Matters to Finance ... and Seven Best Practices to Get You There.	2.5
52	Governance of data and information management in smart distribution grids: Increase efficiency by balancing coordination and competition.	0.5
53	Guideline on data integrity.	3.5
54	How Data Management and Governance Can Enable Successful Self-Service BI.	3.5
55	How to ensure provision of accurate data to enhance decision-making.	4
56	Implementing Projects for Enhancing Records, data governance.	1
57	Information governance leadership: Controlling data and information to achieve strategic objectives.	0.5
58	Intimidated by AI?	1
59	Is It Too Soon for Unstructured data governance?	1
60	IT's Evolving Role in data governance.	2.5
61	Auditing Artificial Intelligence: Internal auditors can develop a framework for conducting AI engagements, despite a lack of standards and guidance.	3.5
62	The data governance Act and the EU's move towards facilitating data sharing.	3
63	Joining the dots: how to approach compliance and data governance.	2.5
64	Indicators for maturity and readiness for digital forensic investigation in era of industrial revolution 4.0.	3
65	Privacy and Security data governance: Surveillance Mechanisms and Resilience Risks of Smart City Technologies.	0
66	Structure Your data governance.	3
67	Taxonomy of Challenges for Digital Forensics.	4
68	The six steps to become a successful CDO.	3
69	The role of information governance in big data analytics driven innovation.	3
70	Towards a capability maturity model for digital forensic readiness.	3
71	Understanding the How and Why of CU data governance.	4
72	Unlocking data: Where is the key?	2.5
73	A systematic literature review of data governance and cloud data governance.	4
74	Considerations around the use of data.	1
75	What's Your Data Strategy?	4
76	Data classification – the foundation of information security.	2.5
77	data governance in the Dairy Industry.	0.5
78	Data governance case at KrauseMcMahon LLP in an era of self-service BI and Big Data.	3
79	Data governance activities: an analysis of the literature.	3.5
80	Data governance: going beyond compliance.	2.5
81	Digital forensics and investigations meet artificial intelligence.	3
82	Why information governance needs top-down leadership.	3
83	Ensuring the Quality of Data in Motion: The Missing Link in data governance.	4
84	Getting smarter with data: understanding tensions in the use of data in assurance and improvement-oriented performance management systems to improve their implementation.	2.5
85	Data governance: Organizing data for trustworthy Artificial Intelligence.	4
86	Reducing the impact of cyberthreats with robust data governance.	2.5
87	Data governance, data literacy and the management of data quality.	3.5

the literature review and for analysis in the bibliometric network analysis, which is detailed in Table 2.

As shown in Fig. 6, the 75 publications are distributed according to the defined period. This distribution provides a balance of mature and recent insights and knowledge, proposing a more innovative approach to the fields under research.

Bibliometric networks analysis

Alongside PRISMA and the in-depth quality assessment, we conducted a bibliometric analysis of the literature. This analysis examined the co-occurrence of keywords within publications, the correlation between authors, and the most cited articles. The VOSviewer application was used to produce and visualize bibliometric networks, allowing us to explore data through visualization tools and techniques, such as clustering co-authorship and keyword co-occurrence. Bibliometric activity is described in the literature as a statistical method for validating and evaluating scientific publications—including articles, books, chapters, conference papers, or any other scientific works—using statistics and metrics to effectively measure their impact on the scientific community, as well as the relationships between the studies and their bibliographic metadata (de Moya-Anegón et al., 2007; Donthu et al., 2021; Ellegaard & Wallin, 2015; Lo & Chai, 2012; Merigó & Yang, 2017). Ellegaard and Wallin (2015) note that bibliometric analysis and its methods are firmly established as a scientifically valid approach and a fundamental component of the methodology applied in research assessment. This approach introduces a quantitative dimension to the literature review, using accurate and precise metrics to assist the researcher in the literature review process. It enhances the subjective assessment of articles that may be considered eligible for review (Ellegaard & Wallin, 2015; Zupic & Čater, 2015).

As demonstrated by Hong and Pluye (2018), the development and use of literature reviews have been growing over the last 40 years, making them a major component of consolidated scientific work. With this rise, bibliometric analysis has also expanded in recent

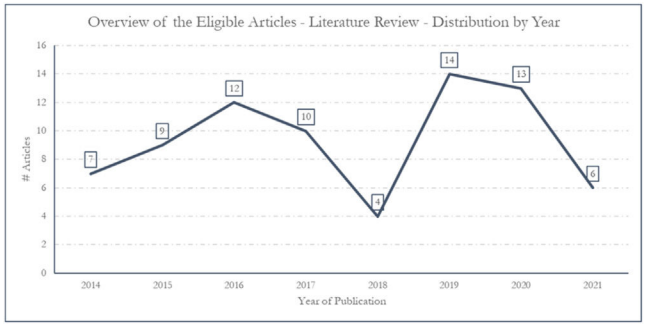

Fig. 6. Overview of the eligible articles - literature review - distribution by year.

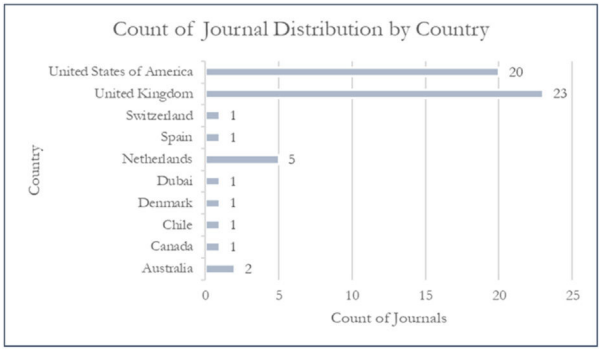

Fig. 7. Count of journal distribution by country.

years, impacting researchers' literature review and introducing new challenges to the review process. This includes assessing both the objective and subjective quality of the literature and its impact (Zupic & Čater, 2015).

The continuous growth of data, including published works, requires now more than ever a robust methodology for researchers to assess and evaluate the information and data they need from the vast amount available nowadays. As a result, bibliometric analysis and its methods are expected to help filter essential works for researchers (Ellegaard & Wallin, 2015; Zupic & Čater, 2015). In fact, bibliometric scrutiny is vital when analyzing substantial volumes of bibliographic data (Gaviria-Marin et al., 2018).

Consequently, we conducted the following analyses: i) Journal Analysis, ii) Authors Co-Authorship Examination, and iii) Keywords Co-Occurrences Investigation. While performing these analyses, we discussed the results and connected the bibliometric findings to the discussion of the review. The first analysis, i) Journal Analysis, involved summarizing and analyzing the metadata of the journals in which the publications appeared, focusing on characteristics such as Quartile ranking (if applicable), country, and research fields. We used data from the SCImago Journal and Country Rank (SJR) for this analysis. The second analysis, ii) Authors Co-Authorship Examination, explored interactions among authors within a research field to understand how scholars communicate and collaborate, and whether this collaboration was diverse, or isolated. Authors who collaborate amongst themselves form a network that Donthu et al. (2021) describe as "invisible colleges," where authors focus on developing knowledge within a specific field of interest. This analysis helps the researcher understand the dynamics of author collaboration over different periods of time (Donthu et al., 2021).

Additionally, the Co-Authorship evaluation helps perceive the level of collaboration among authors on a given topic. It provides insight into the field's social structure and allows the researcher to understand whether authors collaborate outside their immediate group (i.e., the group of authors who appear in each other's publications). Often, authors engage with their group and publish together frequently, but they may not engage with authors outside of their group, which limits the inclusion of different perspectives and knowledge in their research (Donthu et al., 2021; Zupic & Čater, 2015). This analysis was conducted using VOSViewer to perform the necessary text mining (Van Eck & Waltman, 2011, 2014).

For the third analysis, iii) Keywords Co-Occurrences Analysis, we also used VOSViewer for clustering visualization. This analysis identifies which keywords meet the predefined criteria (Van Eck & Waltman, 2011, 2014). It compares keywords from articles published within a specified time frame to identify those that appear in at least two, or more articles, although these criteria may vary depending on the researcher's choice (Donthu et al., 2021; Gaviria-Marin et al., 2018).

Thus, this analysis enables the examination of the most frequently used keywords within the eligible articles and serves as a longitudinal method to comprehend and track the progression of a given field over time (Walstrom & Leonard, 2000). It also involves connecting keywords using the bibliographic metadata of the publications. However, a central challenge for the researcher is that keywords can appear in various forms and may have different meanings depending on the context (Zupic & Čater, 2015).

i) Journal Analysis

The aim of this analysis is to summarize and examine the journals associated with each eligible article, focusing on characteristics such as Quartile ranking (if applicable), country, and research fields. To obtain this information, we used the SCImago Journal and Country Rank (SJR) to retrieve the main characteristics of each journal. For articles not listed in the SJR, we used the Google Scholar search engine. As stated by Mañana-Rodríguez (2015), it is important to have a quantitative assessment of different approaches, including theoretical and practical ones. Roldan-Valadez et al. (2019) argue that researchers must understand the impact of a publication using different types of bibliometric indices. One widely used metric is the SJR index, which is based on data from Scopus and incorporates centrality concepts from social networks. This metric is openly available and contains journal- and country-specific scientific measures and indicators, and is available on its website: www.scimagojr.com (Ali & Bano, 2021; Roldan-Valadez et al., 2019).

Considering this, our literature review comprised 75 articles from 56 different journals. The three most represented countries were the United Kingdom, the United States of America, and the Netherlands, illustrated in Figs. 7 and 8.

The 75 articles reviewed come from a total of 56 journals. Of these, 42 articles (56 %) were retrieved from quartile-ranked journals, while the remaining articles have not yet been classified by quartile ranking. Note that journal characteristics can change over time. This information is presented in Table 3.

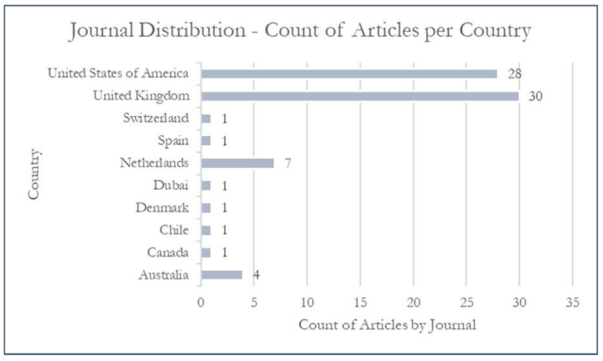

Fig. 8. Count of articles per country - journal distribution.

Table 3
Journal details description.

#	Journal Name	ISSN	# Articles SLR	Quartile	Country	Relevant Fields of the Publisher for this Research
1	ABA Banking Journal	1945947	1	n.d.	United States	Banking & Technology
2	AI & Society	9515666	1	Q2	United Kingdom	Artificial Intelligence & Human-Computer Interaction
3	Alzheimer's Research & Therapy	17589193	1	Q1	United Kingdom	Cognitive Neuroscience
4	Annals of Mathematics & Artificial Intelligence	10122443	1	Q3	Netherlands	Artificial Intelligence & Applied Mathematics
5	Australian Journal of Forensic Sciences	450618	1	Q3	United Kingdom	Pathology and Forensic Medicine
6	Banker Middle East	n.d.	1	n.d.	Dubai	Banking & Technology
7	Best's Review	15275914	1	n.d.	United States	Insurance & Technology
8	Bioethics	2699702	1	Q2	United Kingdom	Heath (social science)
9	Business Horizons	76813	1	Q1	United Kingdom	Business and Internacional Management
10	Business Intelligence Journal	15472825	5	n.d.	United States	Business Intelligence
11	Business Officer	0147877X	1	n.d.	United States	Business
12	CIO (13284045)	13284045	1	n.d.	Australia	Information Systems & Technology
13	Computer Fraud & Security	13613723	3	Q3	United Kingdom	Computer Science
14	Computer Weekly	104787	2	n.d.	United Kingdom	IT Management
15	Computers & Security	1674048	1	Q1	United Kingdom	Computer Science
16	Computers in Industry	1663615	1	Q1	Netherlands	Engineering & Computer Science
17	Credit Union Times	10587764	2	n.d.	United States	Banking & Technology
18	El Profesional de la Información	13866710	1	Q1	Spain	Information Sciences & Systems
19	Feliciter	149802	1	n.d.	Canada	Data Management
20	Governance Directions	22034749	3	n.d.	Australia	Governance & Management
21	Government Information Quarterly	0740624X	1	Q1	United Kingdom	Social Sciences
22	Harvard Business Review	178012	1	Q1	United States	Business, Management and Accounting
23	Health Research Policy & Systems	14784505	1	Q1	United Kingdom	Heath Policy
24	IFLA Journal	3400352	1	Q2	United Kingdom	Social Sciences
25	Information & Management	3787206	1	Q1	Netherlands	Business, Management and Accounting
26	Information Systems Management	10580530	1	Q2	United Kingdom	Computer & Social Sciences
27	Information Today	87556286	1	n.d.	United States	Information Technology
28	Internal Auditor	205745	1	n.d.	United States	Internal & IT Audit
29	International Journal of Information Management	2684012	2	Q1	United Kingdom	Business, Management and Accounting & Computer Science
30	International Review of Law, Computers & Technology	13600869	1	Q3	United Kingdom	Computer & Social Sciences
31	International Social Science Review	2782308	1	n.d.	United States	Medicine & Social Sciences
32	Journal of Accounting Education	7485751	1	Q1	United Kingdom	Business, Management and Accounting & Social Sciences
33	Journal of Computer Information Systems	8874417	1	Q2	United Kingdom	Computer & Social Sciences
34	Journal of Corporate Accounting & Finance (Wiley)	10448136	1	Q2	United States	Business, Management and Accounting & Economics, Econometrics and Finance
35	Journal of Decision Systems	12460125	2	Q3	United Kingdom	Business, Management and Accounting & Computer Science
36	Journal of Field Archaeology	934690	1	Q1	United Kingdom	Social Sciences
37	Journal of Forensic Sciences	221198	2	Q2	United States	Pathology and Forensic Medicine
38	Journal of International Commercial Law & Technology	19018401	1	Q4	Denmark	Computer Science
39	Journal of Securities Operations & Custody	17531802	3	n.d.	United Kingdom	Security Operations & Custody
40	Journal of Technology Management & Innovation	7182724	1	Q3	Chile	Business, Management and Accounting
41	KM World	10998284	2	n.d.	United States	Knowledge Management
42	McKinsey Insights	n.d.	1	n.d.	United States	Strategy & Technology Consulting
43	Molecular Systems Biology	17444292	1	Q1	United States	Information Systems
44	NACD Directorship	1934279	1	n.d.	United States	Biotechnology & Bioengineering
45	Network Security	13534858	3	Q3	Netherlands	Computer Network & Informations Systems and Management
46	NetworkWorld Asia	8877661	1	n.d.	United States	Enterprise Technologies
47	New Directions for Institutional Research	2710579	1	n.d.	United States	Resource Coordination & Information Analysis
48	New Hampshire Business Review	1648152	1	n.d.	United Kingdom	Business & Management
49	New Jersey Banker	0028-5536	1	n.d.	United States	Banking & Technology
50	Personal & Ubiquitous Computing	16174909	1	Q2	United Kingdom	Computer & Social Sciences
51	Pharmaceutical Technology Europe	17537967	1	Q4	United States	Pharmaceutical Technology
52	Police Practice & Research	15614263	1	Q1	United Kingdom	Social Sciences
53	Proceedings of the European Conference on Management, Leadership & Governance	20489021	1	n.d.	United Kingdom	Business & Management
54	Spectroscopy	8876703	2	Q4	United States	Analytical Chemistry
55	WHO Drug Information	10109609	1	Q4	Switzerland	Medicine & Public Heath
56	Wireless Networks (10220038)	10220038	1	Q2	Netherlands	Computer Science & Engineering

Similarly, we categorized each eligible publication into four major areas, outlined in Table 4: data governance and quality, data assurance, artificial intelligence, and digital forensics.

ii) Authors Co – authorship analysis

In this analysis, we utilized VOSviewer, which provided solid and robust text mining techniques for the researcher. One such technique involves a complete counting analysis of co-authorship, where the unit of analysis is the authors of each of the 75 eligible publications. To avoid restricting this analysis, we did not limit the maximum number of authors per publication, and a minor threshold was set at 2 articles per author. Consequently, the co-authorship analysis was applied to 126 authors of which five met the defined threshold, shown in Fig. 9.

Similarly, we conducted an additional iteration of this analysis, setting the minimum threshold to 1 while allowing an unlimited

Table 4
Primary research areas of the literature review.

Dimension Category	Bibliographic Reference
# Data Governance	(Abraham et al., 2019; Alhassan et al., 2016, 2019a, b; Al-Ruithe et al., 2019; Bennett, 2017; Bindley, 2019; Bordey, 2018; Burniston, 2015; Cerrillo-Martínez & Casadesús-de-Mingo, 2021; Chakravorty, 2020; Clarke, 2016; Demarquet, 2016, 2016; Dighe, 2014; George et al., 2017; Gupta et al., 2020; Hassan & Chindamo, 2017; Hay, 2015; Hubbard et al., 2020; Janssen et al., 2020; Jiya, 2021; Johnston, 2016; Koltay, 2016; Lancaster et al., 2019; Lee, 2019; Mansfield-Devine, 2017; McDowall, 2017a, b; Mcintyre, 2016; Milne & Brayne, 2020; Perrin, 2020; Petzold et al., 2020; Riggins & Klamm, 2017; Russom, 2015; Seerden et al., 2018; Shabani, 2021; Sifter, 2017; Smith, 2016);
# Data Quality	(Dallemule & Davenport, 2017; Dutta, 2016; Harper, 2020; Meyers, 2014; Mikalef et al., 2020; Paredes, 2016; Ragan & Strasser, 2020; Rivett, 2015; Sucha, 2014)
# Data Assurance	(Bennett, 2015; Borek et al., 2014; N. Clarke, 2019; Gardner et al., 2018; Johnson, 2015; Manus, 2019; Saporito, 2019; Swoyer, 2016; Tankard, 2015; Vilminko-Heikkinen & Pekkola, 2019);
# Artificial Intelligence	(Bone, 2020; Sánchez & Sarría–Santamera, 2019; Fischer & Piskorz-Ryń, 2021; Kopp, 2020; Miernicki & Ng (Huang Ying), 2021);
# Digital Forensics	(Ariffin & Ahmad, 2021; Bashir & Khan, 2015; Casey, 2019; Costantini et al., 2019; Elyas et al., 2014; Englbrecht et al., 2020; Gold, 2014; Karie & Venter, 2015; Lutui, 2016; Mouhtaropoulos et al., 2014; Tassone et al., 2017; Valdez, 2018; Vincze, 2016).

number of authors per publication and maintaining a minor threshold of 2. This iteration included a total of 126 authors, resulting in the aggregation of 71 distinct clusters, with the largest cluster containing five authors. The top five clusters, ranked by the number of authors per document, are presented in Fig. 10 and Table 5.

By analyzing the data, we observed that, although there are a total of 71 distinct clusters, each containing one or more authors, none of these researchers are connected to others from different articles or clusters. This lack of connection may suggest a deficiency in collaboration between clusters, indicating that authors did not engage in collaborative work outside their respective clusters. Additionally, clusters containing more than one author may suggest that the authors within each cluster collaborate on the same publication. According to Kraljić et al. (2014), researchers' key contributions should focus on both existing and future scientific knowledge and the sharing of this knowledge. Researchers should therefore aim to foster collaborative work and networks to integrate diverse perspectives and approaches.

iii) Keywords co-occurrences analysis

For this analysis, we used the VOSviewer tool to assess 349 keywords, setting the minimum occurrence threshold at 3. Consequently, 30 out of the 349 keywords met this threshold and were selected for analysis. VOSviewer grouped these 30 keywords into four distinct clusters, which are presented in Table 6 and Fig. 11: Cluster 1 (red) – data management; Cluster 2 (green) – computer crimes/forensic sciences; Cluster 3 (blue) – data quality; and Cluster 4 (yellow) – data governance.

Additionally, the top five keywords identified were data governance (17 occurrences, 30 total link strength), data management (15 occurrences, 32 total link strength), big data (9 occurrences, 23 total link strength), computer crimes (9 occurrences, 23 total link strength) and criminal investigation (9 occurrences, 23 total link strength). The analysis, which encompassed four clusters, included a total of 30 items, 117 links, and an overall link strength of 183, shown in Fig. 11.

The overlay visualization, which represents the Co-Occurrence analysis, includes an additional variable represented by the average year of the keywords' publications. As shown in Fig. 12, the most interconnected terms in this network correspond to articles published between 2016 and 2019. Thus, within Cluster 1 (red) – data management – the average publication year of the keywords is between 2017 and 2018. In Cluster 2 (green) – computer crimes/forensic Sciences – the average year is between 2016 and 2017. For Cluster 3 (blue) – data quality – the time frame falls between 2017 and 2018. Finally, Cluster 4 (yellow) – data governance – stands out from the remaining groups with a more recent average publication year between 2018 and 2019.

Phase III - Results and discussion

This section discusses each Research Question (RQ) and details the results obtained to answer them. Each subsection corresponds to a specific Research Question, labeled RQ1 through RQ5.

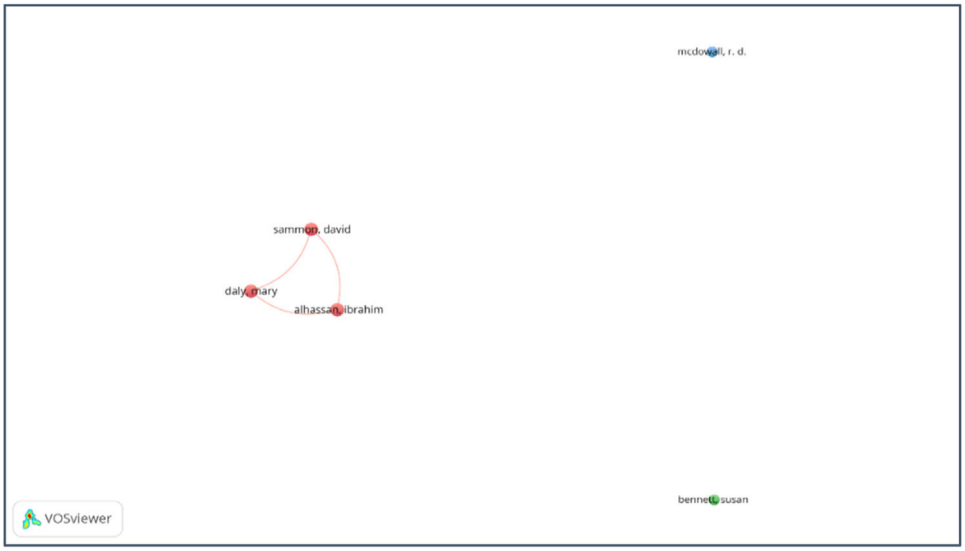

Fig. 9. Co-authorship occurrence network (VOSviewer - Threshold = 2).

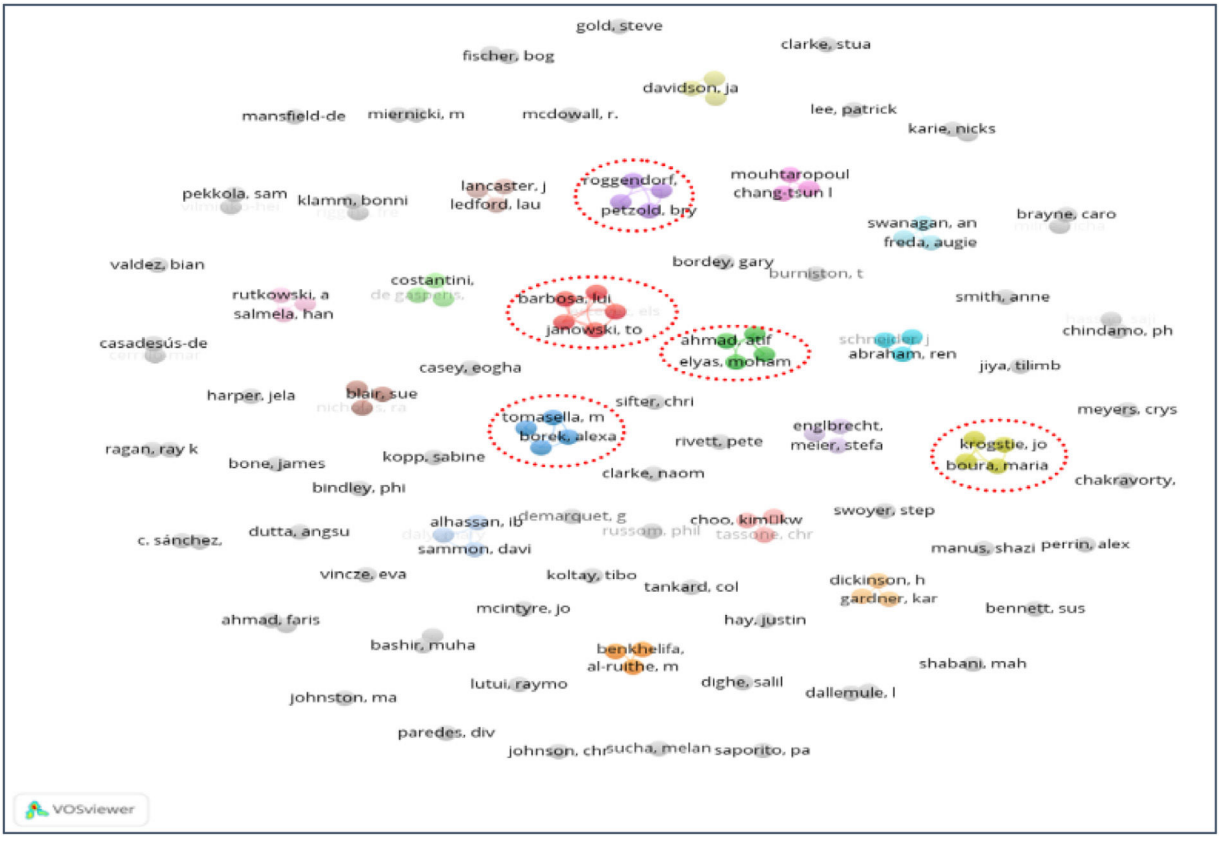

Fig. 10. Co-authorship occurrence network (VOSviewer - Threshold = 1).

Data governance — theoretical background

Theoretical background

In response to **Research Question 1 (RQ1)**, it is important to describe the theoretical context, foundations, and key concepts of the data governance field. According to the Data Management Association (DAMA), a significant reference in this area, data governance involves a combination of responsibilities, control environments, governance, and decision-making related to an organization's data assets (Al-Ruithe et al., 2019). This concept is distinct from data management, which DAMA characterizes as the process of defining data elements, acquiring, managing, and storing them within an organization, and overseeing how data flows through IT systems. Data management also includes the policies, procedures, and plans that support this workflow (Al-Ruithe et al., 2019).

Given this, we define data governance as the planning, management, and governance of data and related activities. According to Sifter (2017), data governance can be viewed as a combination of highly managed and structured collections of resources and assets, which are critical to an organization's business and decision-making processes (Sifter, 2017, p. 24). The main objective of data governance is to enhance a company's capability, effectiveness, and sustainability by supporting its decision-making processes with robust, high-quality data, leading to more precise and informed choices (Sifter, 2017).

Ragan and Strasser (2020) similarly view this field as one that enables organizations to leverage data to communicate their needs, while also aligning business strategy with available data. This alignment leads to better supported, more organized processes and more informed decisions.

Moreover, the success of a data governance program is closely tied to an organization's culture, as well as the dedication and engagement of its administration and human resources (Ragan & Strasser, 2020). Hoppszallern (2015) highlights the importance of clear governance in information technology, noting the possible negative effects of unclear governance on an organization and the value that can be created through effective implementation. Similarly, Sifter (2017) argues that strong data governance may address obstacles related to an organization's processes, people, and technology, enabling it to tackle data management problems and uphold a long-term commitment to integrity in data and asset governance. Meyers (2014) shares

Table 5
Co-authorship analysis - top 5 ranked by clusters of author.

# Cluster	Reference	# Of Authors per Cluster	Link Strength	Links between authors in different clusters
1	(Janssen et al., 2020)	5	4	0
2	(Elyas et al., 2014)	4	3	0
3	(Borek et al., 2014)	4	3	0
4	(Mikalef et al., 2020)	4	3	0
5	(Petzold et al., 2020)	4	3	0

Table 6
Keywords co-occurrences ranked by the occurrences - descending.

Keyword Label	# Occurrences	Total Link (Strength)	Average Publication Year
data governance	17	30	2018.5
data management	15	32	2016.9
big data	9	23	2018.2
computer crimes	9	23	2015.7
criminal investigation	9	23	2016.0
corporate governance	8	15	2016.5
information resources management	8	19	2016.1
data analysis	7	17	2017.1
data quality	6	16	2018.2
data security	6	11	2018.2
forensic sciences	6	15	2017.3
artificial intelligence	5	5	2020.2
Data	5	10	2018.6
data protection	5	9	2017.6
database management	5	9	2018.0
acquisition of data	4	5	2018.0
business intelligence	4	9	2015.5
business planning	4	9	2016.3
digital forensics	4	10	2017.5
information technology	4	9	2015.8
information technology security	4	10	2016.0
business databases	3	7	2016.7
business enterprises	3	5	2015.3
chief information officers	3	5	2017.3
computer security	3	9	2014.7
Cyberterrorism	3	7	2015.0
data integration	3	4	2018.7
data warehousing	3	8	2015.7
information sharing	3	6	2019.3
open coding	3	6	2018.0

this perspective, recognizing data governance as the application and oversight of data management rules, processes, and formal procedures throughout an organization's operations.

Furthermore, the literature identifies data governance as one of the most crucial areas for firms to incorporate into their operations. Indeed, it can be seen as a critical asset for the success of their business strategy because data are seen as the future and a competitive edge that can differentiate successful businesses from those that fall short (Ragan & Strasser, 2020, p. 10). Rivett (2015) argues that data governance represents a significant gap that businesses need to address, especially as it becomes increasingly complex over time. However, organizations should remember that data governance involves more than just managing data. It also encompasses how data are collected, produced, utilized, controlled, and reported, as well as who is responsible for ensuring quality throughout the lifecycle (Janssen et al., 2020).

Moreover, Riggins and Klamm (2017) describe this field as an exercise in process and leadership focused on data-based decisions and related matters. It comprises a system of decisions and responsibilities defined by an agreed-upon model, which outlines how an organization can take action, what type of actions can be pursued, and when those actions should occur in relation to data and information. Dutta (2016) describes this field as the combination and definition of standards, processes, activities, and controls over enterprise data to ensure that the data are available, integrated, complete, usable, and secure within the organization. Organizations are striving to define and implement resolutions that manage the quality and master data, ensuring truthfulness, quality, and completeness—specifically non-transactional data, data at rest, while often disregarding transactional data, data in motion (Dutta, 2016).

As highlighted by Abraham et al. (2019), while data governance is increasingly recognized as being critical to the success of organizations, there is still no unanimous agreement within the scientific community regarding its scope. Current literature and other relevant sources often focus on specific aspects of the discipline, such as data protection, security, and the data lifecycle. Alternatively, some sources provide limited reviews that focus on theoretical approaches without offering frameworks and tools to improve and strengthen data governance and quality. Therefore, our approach to this topic ensures that this discipline is analyzed in a way that integrates both theoretical and practical concepts, tools, and frameworks, aiming to improve the overall effectiveness and efficiency of this field and its impact on organizations.

Archaeology and relevant concepts

Research Question 1 (RQ1) focuses on the recognition that, for a given data governance framework and strategy, there are several concepts that organizations need to understand and embed in their operations. According to Gupta et al. (2020), a data governance framework should encompass strategies, knowledge, decision-making processes related to data management, relevant regulations,

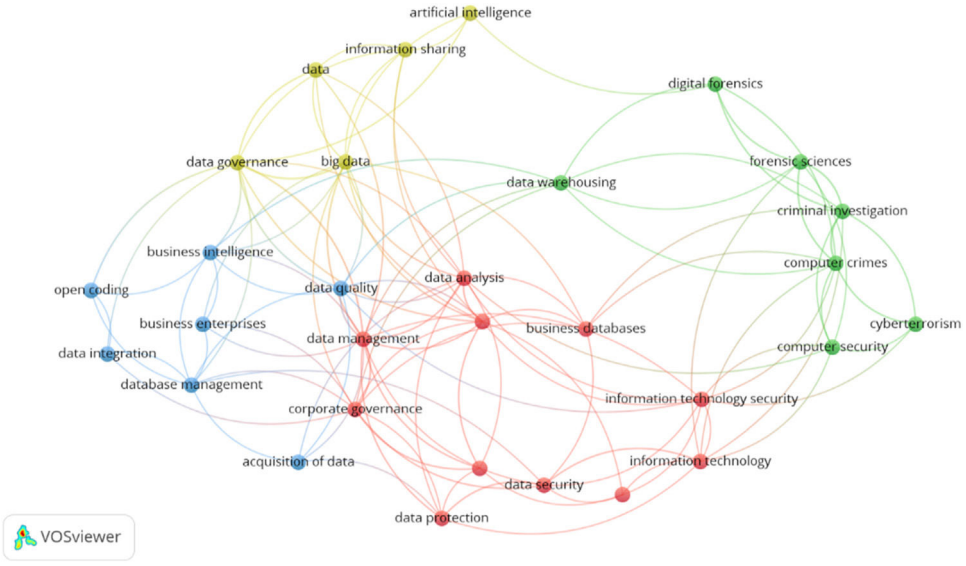

Fig. 11. Keywords co-occurrence analysis (VOSviewer).

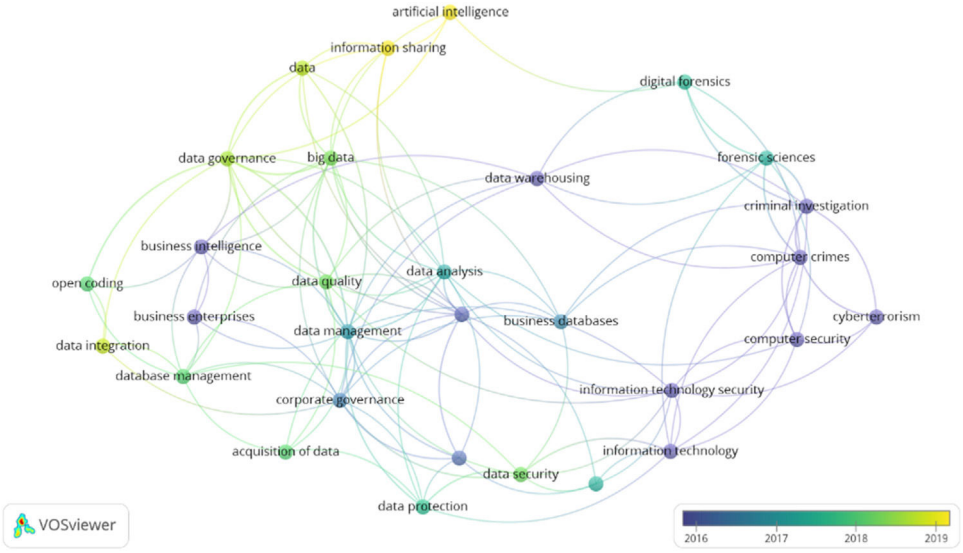

Fig. 12. Keywords co-occurrence analysis – year overlay (VOSviewer).

policies, data proprietorship, and data protection. One important concept is the notion of data lake, which is considered a centralized repository that allows an organization to collect and distribute large amounts of data. In a data lake, data can be both structured and unstructured, often with limited contextual information, such as the data's purpose, its timeliness, its owner, and how, or whether it has been used (Rivett, 2015).

Another vital concept corresponds to data lineage, which represents the traceability of a given data within an organization. Data lineage allows organizations to visualize and understand the impact of data throughout the organization's system and databases, as well as to trace data back to its primary source, including the stage where it is stored (Dutta, 2016). The literature highlights the importance of embedding data lineage in an organization's end-to-end processes to ensure accurate and complete traceability throughout the data lifecycle, including the systems and processes the data pass through (Dutta, 2016).

Bordey (2018) emphasizes the importance of understanding data lineage and how a firm can control and monitor the data life cycle. To assess a firm's data quality capability, Bordey proposes a set of questions that any firm should consider: i) What activities are being performed, and who is responsible for executing and monitoring them?; ii) What are the conditions and requirements for the planned and performed activities?; iii) Where can the necessary data be retrieved, and what exact phases are required to manage and use that data?; iv) What will be the result of this activity, and for whom is it being produced?; and v) What are the next steps after the result is delivered? (Bordey, 2018).

Moreover, to build and develop a robust and effective data governance strategy and framework, it is essential to establish foundational pillars that should be employed to provide a solid baseline for this field. We recognize that organizations remain vulnerable to data disasters and are hesitant to invest in prevention measures, such as robust and secure infrastructure designed to guard against potential system outages and data disasters, namely through the loss of data (Johnston, 2016).

Johnston (2016) identifies three fundamental topics for managing and maintaining a solid data governance strategy: data visibility, federated data for better governance, and data security. The author emphasizes that defining the baseline for a data governance strategy and framework is crucial. According to this author, the first pillar should be data visibility, which encompasses the activity an organization should go through to understand "what you have, where it is, and how to recover it, when necessary" (Johnston, 2016, p. 52). In addition, Johnston describes the principles organizations should comply with to secure their data: i) deleting low-value data that that has little to no business value, such as outdated, or duplicated data, or data that has exceeded its retention period; ii) organizing valuable data by storing it in controlled repositories using information technology tools; iii) enforcing rigorous data security procedures and mechanisms to ensure data privacy and the protection of financial, personal, and other types of data, while also ensuring it is secure and properly backed-up; and finally, iv) maintaining and controlling access, user privileges and their management (Clarke, 2016). Similarly, Mcintyre (2016) suggests that every organization should implement data redundancy measures to ensure an organization is prepared for a disaster and recovery. This includes setting Recovery Point Objectives (RPOs), which determine how much updated and current data the firm is willing to lose, and Recovery Time Objectives (RTOs), which define how long the firm can go without access to its data assets.

Similarly, Rivett (2015) describes that context, or cataloging of data, is a layer that can answer critical questions and provides the organization with greater visibility over their data and information. This visibility allows organizations to perceive their data, determine their usage, identify ownership roles, and know who to contact in case of data errors. According to Rivett, a business can only define and implement an effective data governance strategy—one that includes appropriate data management procedures to ensure proper governance and compliance—if they have complete visibility of the data within the organization's infrastructure.

In addition to data visibility, Johnston (2016) emphasizes the need for organizations to focus on federated data to improve organizational governance. The process of federating data involves retrieving data from different databases or sources and standardizing it into a single, unified source for front-end analysis and visualization. This process allows companies to analyze and access data from a single source without needing to modify, move, or recover data across different infrastructures (Johnston, 2016). Johnston also identifies data security as the third pillar, which includes protecting and securing data, defining extensive policies, and applying encryption and authentication mechanisms to ensure that only authorized individuals can access the data (Johnston, 2016). Similarly, Rivett (2015) examines the importance of context and governance in implementing a data governance strategy. This author stresses that organizations need to be able to address questions related to data

responsibilities, the up-to-date status of data, data meaning and usefulness, and the technologies used for management and storage.

Similarly, Sifter (2017) argues that for a data governance framework to achieve excellence, it must be supported by five essential attributes. The first attribute is related to data accuracy, ensuring that all data flows are managed, controlled, audited, and monitored across the organization's system landscape. Sifter also highlights the importance of data consistency and availability, which are essential for enabling business stakeholders to access data whenever necessary. The other three key attributes, according to Sifter (2017), include: i) the ability of organizations to maintain a single repository for essential data storage, ii) a commitment to maintaining data quality and availability across all processes, and iii) effective planning activities that address data requests by business stakeholders with suitable swiftness and repeatedly whenever necessary. Additionally, it is essential to ensure that data governance aligns with the organization's Master Data Management (MDM). MDM encompasses a set of practices and techniques whose primary goal is to increase and sustain data quality and usage throughout the organization's environment and processes (Vilminko-Heikkinen & Pekkola, 2019).

Benefits and added value

In response to **Research Question 2 (RQ2)**, and because organizations increasingly focus on fostering robust data governance practices to support their personnel, business processes activities, and technology infrastructure, they are likely to achieve positive outcomes that create value for the overall business and its stakeholders. Sifter (2017) explains that data governance can enhance an organization's human capital by providing more trained, equipped, and knowledgeable personnel, particularly in roles that require ownership and accountability in governance and data management. This approach is not about restricting access to data for certain privileged users, but rather about enabling all users within the organization to access and utilize data governed by appropriate controls (Riggins & Klamm, 2017).

Similarly, this author suggests that an organization's processes will also improve because they are more controlled and monitored for consistency and efficiency due to the strong support of established policies, procedures, and practices. As a result, data governance provides organizations with a set of tools and directives aimed at ensuring that the correct data are accessed and analyzed by the appropriate people whenever and wherever decisions need to be made (Riggins & Klamm, 2017). Additionally, an organization's technology infrastructure will be strengthened as companies strive to optimize and update their systems across the entire environment, rather than just within individual business units (Sifter, 2017). In fact, with better use and understanding of the data they manage, companies are more likely to make well-supported decisions, thereby unlocking greater potential from their data (Burniston, 2015). Consequently, accountability and reliance on data are enhanced when organizations take responsibility for their data and systems, leading to better-designed and controlled decision-making processes (Sánchez & Sarría–Santamera, 2019). Therefore, robust data management and governance can significantly improve organizational transparency (Cerrillo-Martínez & Casadesús-de-Mingo, 2021)

Moreover, Riggins and Klamm (2017) highlighted several benefits that organizations can gain from data governance including: i) the creation of a mutual, company-wide approach to data processes, which helps generate a single source of truth for the organization's data; ii) the definition and recognition of data controls and policies, which when aligned with organizational needs, foster data quality, accessibility, and security; iii) the development of a central repository for data management, aimed at defining a common terminology and taxonomy within the organization; iv) the establishment of real-time reports and standardized reporting techniques; and v) the enhancement of decision-making processes through the use of trusted data.

Furthermore, the importance of addressing data governance and quality is becoming increasingly prominent in both organizational practices and among regulatory bodies. Shabani (2021) notes that the European Union recently announced a proposal for a new Data Governance Act aimed at facilitating data sharing across various fields while establishing a bridge with the GDPR (General Data Protection Regulation). According to Shabani, robust data governance mechanisms and secure data-sharing practices should be adopted to guarantee compliance with rules and directives from regulatory bodies, such as the GDPR (Shabani, 2021). The author argues that providing organizations with a strong regulatory framework for data governance, which addresses all core elements of this field, can empower both organizations and individuals to better control and enhance their data usage and sharing (Shabani, 2021). Similarly, Janssen et al. (2020) propose that organizations should adopt a framework to regulate and control data sharing, stating that organizations must include requirements and standards for data sharing, formalize agreements, contracts, and service level agreements (SLAs), define authorization levels and approval workflows, and implement audit mechanisms to ensure compliance with relevant legislation and agreements. Kopp (2020) adds that when organizations recur to outsourcing activities, they must clearly describe and characterize the activities being performed as outlined in the formal contract.

Data governance frameworks

Areas of application

Concerning **Research Question 2 (RQ2)**, it is important to recognize that a given data governance strategy and framework can be applied across a wide range of areas and situations. These applications can serve as benchmarks for developing a solid and robust data governance program (Hassan & Chindamo, 2017). For instance, Perrin (2020) describes the recent impact of data governance in the advertising industry, where data fuel decisions made by marketers and brand managers to guide decision-making processes and create brand experiences for consumers. Perrin highlights the significance of current data management and usage regulations, such as the CCPA (California Consumer Privacy Act) and the GDPR (General Data Protection Regulation). Given that advertising relies on vast amounts of data for decision-making and brand promotion, organizations should consider benchmarking their data governance strategies against those in other industries and areas, whether similar, or different from their own. This approach can help them identify key learning points and best practices (Dencik et al., 2019).

Similarly, Perrin (2020) stated that data governance in advertising must encompass data efficacy, transparency, and consumer control. While data can be considered a new currency, increasingly stringent regulations are imposing stricter restrictions on data management and usage. Therefore, organizations must implement a data governance framework and strategy that aligns with the applicable regulations. This includes promoting internal processes to understand the different types of data within the organization, how they are collected, and how they are audited for efficacy and compliance. In the advertising industry, it is crucial for companies to integrate data governance into their operations, because failing to comply with regulations, or jeopardizing consumer trust could make it "nearly impossible to rebuild" that trust (Perrin, 2020, p. 35).

Sifter (2017) describes the importance of data governance in financial institutions, such as insurance companies, financial services, and banks. According to the author, a successful data governance program must address people, processes, and technologies. Sifter notes that these organizations often struggle with managing data quality and access to both data and technology. The author views data governance as a crucial tool for preventing data corruption and issues, maintaining long-term data integrity, and producing vital business

data and assets that support operations and business management (Sifter, 2017).

Existing strategies, tools and frameworks

Regarding **RQ2**, Sifter (2017) argues that for a data governance initiative to achieve excellence, it requires support from five key components: foundational elements, data portfolio management, implementation management, engineering and architecture, and operations and support. For the foundational elements, Sifter states that it is important for an organization to establish a charter and vision that can be controlled and assessed over time. This should include clearly defined ownership roles and responsibilities regarding data governance.

For the data portfolio management component, Sifter (2017) argues that organizations must be capable of designing, building, defining, and maintaining an up-to-date data inventory. This inventory should provide a comprehensive overview of how data flows within the organization's technologies, including how data are secured, preserved, and managed. For instance, a data inventory, also known as a data dictionary, or glossary, acts as a catalog of all the data used within the organization. It serves as a reference point for data and assigns data ownership. It resembles a data repository, with detailed metadata, including format, type, technical specifications (e.g., character limits, possible blanks, among others), and accountability and quality criteria and metrics (Clarke, 2019). The author highlights the importance of defining and regularly updating a data inventory to improve an organization's ability to maintain a centralized repository of data definitions, structures, and relationships, which is essential for consistency and clarity across the organization (Clarke, 2019).

Regarding implementation management, Sifter (2017) considers it essential to determine and define methods related to project, change, and access management within an organization. These methods should also facilitate the optimized management of human capital and training programs. Additionally, Sifter (2017) argues that engineering and architecture should support a data governance framework by establishing policies, procedures, and mechanisms across various aspects of the organization. This includes software management, quality assurance and auditing, change and security management, and the management of the technology landscape that characterizes the organization's operations and activities.

Lastly, Sifter (2017) emphasizes the need for organizations to have robust operations and support mechanisms, allowing their personnel to have clear communication channels whenever there is a need for assistance, or in the case of an incident. Additionally, Dutta (2016) proposes a framework to ensure the quality of data in motion, or transactional data. We believe that each of the steps outlined in this framework can be adapted to enhance an organization's data governance framework. Dutta presents a different approach to data governance, which is illustrated in Fig. 13. This approach is structured around the design of a conceptual framework consisting of six interrelated dimensions:

In Stage 1, Discover, the organization must identify and acknowledge all critical information and how it flows within the technological and operational architecture. This involves defining and developing metrics and their foundations. The organization needs to inventory all data across existing systems, databases, and data warehouses, understanding, registering, and documenting data lineage, including sources from both internal systems and external systems and providers (Dutta, 2016). In this stage, it is also important for the organization to establish metric standards and implement techniques to improve data quality through collaboration with data owners.

In Stage 2, the organization should focus on identifying and defining the risks associated with data quality, including data domains and both data at rest and in motion. This involves formalizing issues related to data quality features, as well as recognizing the risks,

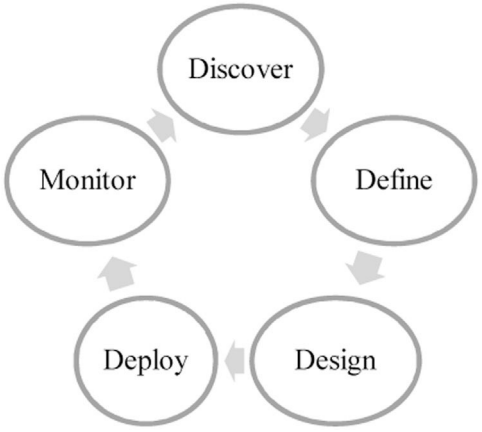

Fig. 13. Framework for data quality of data in motion. Adapted from Dutta (2016).

potential gaps, and pain points that characterize the organization's current environment (Dutta, 2016). By identifying these risks, the organization will be in a position to develop a clear approach to evaluate, mitigate, respond to, and treat them properly.

Data transparency, reliability, and usability can be considered part of the data observability feature in a data governance program. This concept involves all the mechanisms and processes that allow employees to effectively monitor, assess, and produce insights from their data (Schmuck, 2024). According to Schmuck, data observability serves as a solution within a data governance program and encompasses characteristics such as data modeling and design, data quality and its architecture, metadata registry, data warehousing, and business intelligence, data integration and interoperability, and document management (Schmuck, 2024).

In Stage 3, Design, organizations must develop their information analysis techniques and establish processes for managing exceptions, errors, and incidents. To improve these processes, organizations should use automated techniques rather than traditional approaches like sample sizing. They should also implement controls designed to monitor real-time data and processes, such as the success of batch routines (Dutta, 2016). By designing these approaches, organizations should prepare for Stage 4, which involves deploying a set of actions and activities based on the prioritized risks, namely those that are most important and impactful to the environment. The implementation of these actions must involve various components of the organization, including technology, people, and processes, to ensure the effective deployment of these initiatives.

The last Stage, Stage 4, corresponds to Monitoring, where the organization's primary goal is to observe and review the framework that was applied to its context. This includes evaluating data quality metrics, incident management process, risk controls, and any activities that may influence and jeopardize the success of the defined framework (Dutta, 2016).

Furthermore, Al-Ruithe et al. (2019) identify three key components that should be ensured in data governance: roles related to data quality assurance, areas of decision-making, and overall responsibilities. They suggest that an effective data governance strategy should involve defining, monitoring, and implementing a comprehensive organization-wide data warehouse (Al-Ruithe et al., 2019). Despite the lack of recent literature on different methodologies and frameworks in this field, several existing data governance frameworks should be analyzed and compared to develop a robust framework that encompasses the best critical aspects of each. For example, the Data Governance Institute (DGI) framework, described by Al-Ruithe et al. (2019), includes ten components that an organization should ensure: 1) Mission and Vision involves defining and setting rules and boundaries, presenting the mission, strategy, and vision,

and aligning them with data stakeholders while safeguarding compliance and regulatory requirements. In this step, the DGI suggests that organizations define data governance groups to provide an overview and ensure understanding of how data travel through human and technological assets; 2) Goals, Governance Metrics, and Funding strategies define clear goals and metrics following the SMART criteria (specific, measurable, actionable, relevant, and timely). It ensures that employees understand successful data governance and how to monitor and assess it continuously. In this Stage, it is important to have funding strategies to support roles such as data governance officers and other data stakeholders; 3) Rules include defining and aligning manuals, policies, and procedures to support data-related activities, including compliance standards, business rules, data definitions, lineage, existing conflicts, and areas for improvement, among others; 4) Decision rights clearly define decision-making processes and responsibilities, ensuring that employees understand their roles in the data governance framework and the creation of metadata for data-related decisions; 5) Accountabilities ensure that the organization can perform and prove compliance with its operations and business activities, maintaining control and documenting each step. 6) Control refers to the idea that any existing data within an organization's environment is inherently at risk, whether in terms of quality, or security. Therefore, organizations must ensure that their framework encompasses a variety of controls—both preventive and detective, whether manual, IT-dependent, or automatic. These data-related controls allow for a more comprehensive overview and monitoring of all operational and business activities. Additionally, these controls can aid any external audit and compliance requirements; 7) A data stakeholder refers to an individual, or a group of individuals who could either be impacted by or influence any data and related decisions within the existing framework; 8) A data governance officer is a role focused on facilitating and defining data governance activities. This includes articulating and aligning roles and responsibilities with the organization's critical processes and formal procedures, as well as deploying data governance initiatives and programs, among other activities; 9) Data Stewards are data users responsible for ensuring that data follows particular pre-defined criteria: completeness, accuracy, and integrity, as well as reporting to a business role, or a data quality team. According to the DGI, one of the best approaches an organization could take is to establish a Data Stewardship Council, bringing together all data stewards across the organization to define, monitor, and enhance the data governance framework and its current procedures. McDowall (2017a, b) mentions that alongside the data steward, there should be a Technology Steward—an individual responsible for managing the technological IT requirements of the data owner, or user. This role is essential to ensure the segregation of duties and the administration of systems and data warehouses; Lastly, the DGI refers to 10) Proactive and Ongoing Data Governance Processes. This emphasizes that the data governance framework must not be static. Instead, it must be continuously maintained, monitored, and improved while establishing mechanisms and metrics to guarantee continuous data governance and quality.

According to Al-Ruithe et al. (2019), another important framework is the IBM data governance framework, which is composed of 14 elements derived from a software provider perspective. Those components are: Required – 1) Definition of a business problem; 2) Sourcing executive sponsorship and funding; 3) Evaluating the maturity status; 4) Developing a roadmap in order to increase its current maturity robustness; 5) Defining the organizational blueprint and data flows; 6) and 7) Defining and implementing a data dictionary, while acknowledging existing data; 8) Designing and implementing a repository to store details on existing metadata; 9) Building metrics to measure operational success. Bordey (2018) suggests that organizations use stories related to user operations to describe operational measures, i.e., numbers, or values that can be provided with context and address a given performance metric to ensure the operation's success; 10) Designating data stewards, managing data quality, and implementing a management strategy for master data. Optional elements – 11) Using analytics to support governance; 12) and 13) Managing and monitoring privacy and security elements, and how data flows within the organization's resources, including human and technological ones; 14) Measuring the results of each data governance program and its initiatives.

Equally important, Abraham et al. (2019) present a conceptual framework for this field, describing six dimensions it should encompass: i) governance mechanisms, which represent the structure, procedures, and relational activities within an organization. At this level, it is vital to define, implement, and align roles, responsibilities, and procedural mechanisms. Additionally, organizations must establish essential and robust tools such as a data strategy, policies, manuals, standards, processes, key performance indicators, monitoring mechanisms, and issue management methods to ensure compliance with applicable legislation. Harper (2020) suggests that the most robust quality metrics for data and information must encompass the following attributes: completeness (all information and fields are filled with no unjustified, or applicable missing elements); timeliness (data accessed is the most updated and covers the required period for the user); uniqueness (no duplicates, or wrong data exist within the data assets); accuracy (information in the data assets is correct and precise); consistency (data are presented in a manner that matches user expectations, including expected content and formats, such as dates in a date format, among others); validity (specific requirements, rules and standards are met); and lineage (the data's journey throughout the data assets and data stakeholders is documented from the initial acquisition of the data to final reporting and destruction). Additionally, the framework includes: ii) organizational scope, where the organization defines the extent of the data governance program; iii) data scope, identifying which data assets the organization intends to govern and monitor; iv) domain scope, which relates to the organization's capability to explore the data decision process, including the quality, security, lifecycle and storage of data used to support those decisions; v) antecedents, which comprise the contingency factors, both internal and external, that affect the organization's ability to establish a robust data governance framework; and, vi) consequences, which represent the effects generated by having this framework, including those related to performance, controls and risk assessments.

Similarly, Alhassan et al. (2016) describe a data governance framework in which five related concepts are considered: i) data principles, where organizations define the direction for the decision domain by setting requirements and internal rules for managing their data assets. In this Stage, it is important for organizations to recognize that there is no one-size-fits-all solution. As a result, the data governance principles and scope will depend on various factors, including dynamic data, system dependencies, and personnel, among other elements (Bordey, 2018); ii) data quality, which involves refining the foundations to ensure that data are acquired, transformed, stored and reported in a controlled, effective, and efficient manner; iii) metadata, and iv) data access, are domains in which organizations seek to establish data quality standards to ensure robust management. Bindley (2019) highlights the importance of maintaining data integrity, specifically consistency, accuracy, and completeness throughout the entire cycle, from the systems used to manage the data to the transformations and processes applied to the data. Bordey (2018) considers data profiling and quality essential for users to understand the data being used. This includes performing informative summaries and aggregations, as well as implementing controls and processes relevant to the subject matter. Similarly, metadata plays a crucial role in an organization's success in governing data because it corresponds to the information that the user has on the data, including business and technical documentation (Chakravorty, 2020). Finally, v) the data life cycle involves managing the entire

journey of a data point—from acquisition, analysis, and production to storage, reporting, and eventual destruction—within an organization's data management workflow.

Beyond the frameworks discussed above, Demarquet (2016) suggests that any data governance program should include a set of mandatory topics, such as: i) a change management program and workflow capabilities, which ensure that changes to the system, processes, and data follow a pre-defined process that encompasses the segregation of duties, testing, functional analysis and evidence regarding the approvals throughout the process; and ii) versioning and audit trails, which are also essential for tracking the history of creations, modifications, and deletions, including identifying the changes and the individuals responsible for those actions. Audit trails should be configured and active at all times to prevent users from disabling this configuration (George et al., 2017; Koltay, 2016).

Additionally, Hay (2015) proposes a different approach to data governance by gamifying the framework—using game mechanisms to empower data governance initiatives. The author highlights several mechanisms on which a data governance program can be based, namely: i) fast feedback, where users receive immediate feedback from their activities and actions, such as earning points for contributing to data definitions, or stewarding data, and being penalized if they do not provide any contribution; ii) transparency, where data management is made available for everyone to understand, functioning like an open community; iii) goals, where data users and other data roles recognize both short- and long-term objectives, including individual and group goals, to foster data quality metrics, business rules, and other essential governance aspects; and iv) badges, which publicly showcase the achievements of data users and other data roles, rewarding individuals for reaching different levels of compliance within data governance. For instance, Hay (2015) considers that badges and levels allow users to demonstrate their achievements and level up both individually and as teams. This system of scoring and earning badges can effectively foster and ensure that organizational goals are met; v) onboarding, which ensures that whenever a business user, or any data role joins the organization's governance initiative, the individual is introduced to it and efficiently integrated from day one; vi) competition and collaboration among data users and other data roles are encouraged by allowing them to work towards common objectives while tracking their standings on a leaderboard. This approach continuously enhances the maturity of data governance within the organization (Hay, 2015).

Correspondingly, the literature consistently highlights the critical importance of data governance for organizations of all sizes, industries, and types, whether large or small and medium-sized enterprises (SMEs), and whether private or public (Okoro, 2021). Okoro points out that while large organizations often lead the way in data governance, SMEs are beginning to understand the value of data-driven strategies but have not yet fully embraced data governance practices (Okoro, 2021). Consequently, Okoro suggests that while SMEs tend to prioritize their data assets, the data governance community has largely overlooked these organizations. As a result, there is a significant need for a data governance framework tailored to the specific requirements of SMEs because a standard universal approach that works for larger organizations may not be suitable for all. Similarly, the size of an organization can present unique challenges regarding data governance practices because SMEs often face more significant constraints in terms of resources and expertise compared to larger companies. Nonetheless, data governance practices and a well-established framework can support organizations in managing their assets more effectively, preserving data quality, and complying with regulatory requirements (Okoro, 2021).

Moreover, recognizing this challenge, Okoro (2021) proposes a data governance framework tailored for SMEs. The key components of this framework include: i) data security and privacy, which involves prioritizing data privacy and security by ensuring that data collection is limited to the business operations requirements and disposing of it when it is no longer needed. To achieve this, the author recommends that organizations implement robust authentication mechanisms, controls, and classification schemes for data, and monitor their data architecture, including related procedures and policies; ii) data impact assessment, which involves assessing the degree of data anonymization to reduce risks and protect confidentiality, especially regarding the identification of a person. Additionally, assessments of the impact of data protection, such as Privacy Impact Assessments (PIAs) and General Data Protection Regulation (GDPR), can help organizations ensure the safety, compliance, and confidentiality of their data processing activities; iii) designation of a data review board, which involves establishing a board that includes experts in data science, legal compliance, and security. This board should address and analyze data governance subjects such as data storage, collections, and assets; iv) avoidance of unfairness or bias, where it is vital for organizations and data users to make conscious efforts to prevent bias and ensure fairness when collecting, analyzing, and using data to make decisions. The author suggests that data users should guarantee that the data collection process represents diverse populations and is free from any inherent biases that could potentially impact results and their implications; v) data governance, ethics, and GDPR, where organizations should align their data governance strategy with ethical guidelines and any relevant data protection regulation that impacts the context, ensuring that data processing activities are precise, fair and transparent. Organizations must set strong data protection measures and manage their data responsibilities to adhere to ethical and data governance standards (Okoro, 2021). Within this framework, we believe it provides a robust ground for SMEs to fully harness the potential of their data, focusing on components related to data protection, data quality, and data strategy.

Government institutions, both public and private, operate in a unique landscape when it comes to data governance. Unlike larger firms and SMEs, these institutions face distinct challenges that existing practices and frameworks do not fully address. The strategic value placed on data in government institutions often does not fully align with these frameworks, which are primarily designed for business operations and goals. This mismatch results in ineffective controls over crucial data governance processes, such as data collection, processing, analysis, and reporting (Mao et al., 2021). To address this, the authors explored and developed a new data governance concept for these institutions, called the Data Middle Platform concept. This concept consists of a framework designed to address the complex context of government institutions by centralizing data management and supporting cross-departmental integrations. The concept is presented with key components including: i) service empowerment objectives, which focus on defining the data governance framework in a public context by protecting and enhancing data management capacities, especially for sensitive data, to develop robust public service capabilities. To achieve this, these institutions need to form strategies based on their public service goals and explore their ability to manage cross-relationships, namely across different levels, systems, and departments; ii) defining roles through structure, which pertains to how institutions should delineate roles and responsibilities across their levels, including the foreground, the middle platform, and the background. The author considers the foreground to represent public service access points used as dynamic interfaces where citizens interact with the governmental institution, requiring constant updates to meet public needs. In contrast, the background relates to the administrative departments and the layer encompassing decision-makers, including data governance leaders and councils, who are responsible for planning, strategy development, and resource coordination.

To connect these two components, a middle platform serves as a bridge between the foreground and the background, enabling collaboration between them. This platform also includes key roles within

data governance practices, such as: data stewardship, which is responsible for ensuring data quality and compliance with policies; a Chief Data Officer (CDO), who oversees the overall data governance framework and strategy; a Chief Information Officer (CIO), who is responsible for the IT infrastructure; and a Chief Service Officer (CSO), who ensures that data-related initiatives positively impact public services. The platform is also responsible for: iii) defining and implementing procedures, including policies and standards. These procedures include mechanisms that support the decision-making process, data ownership lists, data governance and quality policies, procedures and standards, security protection guidelines and techniques to appraise and assess the effectiveness of the data governance framework and strategy; iv) openness through communication and sharing, which involves constructing and communicating a shared awareness of the data governance framework among stakeholders. It also ensures the facilitation of interagency data sharing (IDS) for collaboration and cooperation, as well as open government data (OGD) to promote the public use of data and thus enhance transparency; v) monitoring and control to ensure compliance, through which institutions must define and implement data security protection mechanisms, including personal data privacy protection, security audits, and supervision. Additionally, institutions must assess data quality, including accuracy, integrity, consistency and timeliness, using indicators and metrics as outlined by the International Organization for Standardization (2008); vi) execution of the data management life cycle, where the data governance framework must oversee and govern the entire data management life cycle. This includes activities such as data collection (the process of gathering data from various departments, citizens, and other internal and external sources), processing (activities to transform, clean, join, and plan data to ensure its quality), analysis (the process that allows for the visualization and reporting of data to make it insightful and accessible), and capitalization, i.e., activities performed to manage metadata and continuously monitor data flows (Mao et al., 2021).

Additionally, considering the importance of Industry 4.0, it is crucial to examine the potential benefits of a data governance framework that addresses precise specifications, functionalities, and requirements. This framework should particularly focus on challenges unique to the industrial environment, such as its distinct context and characteristics, inter-company collaboration and its silos, compliance with laws and regulations, and adherence to third-party service levels agreements (SLAs). With this in mind, Zorrilla and Yebenes (2022) propose a data governance system framework where data and data governance form the core of the architecture. This framework is based on three key principles: i) Data-as-a-Service (DaaS), ii) Monitoring-as-a-Service (MaaS) and iii) Platform-as-a-Service (PaaS). The data governance framework for Industry 4.0, proposed by the authors, includes several components: a maturity model, a method for architecture development, a standards information base, a reference architecture, a content metamodel, and a reference model establishing a set of Architecture Building Blocks (ABBs). In line with the framework's reference architecture, the authors draw on international standards and their iterations to define the Reference Model ABBs. For example, they incorporate i) an iteration of the TOGAF® Standard v9.2 framework, in which the authors introduced a new entity called "Policy." This standard provides a framework for enterprise architecture, offering a structured approach for the organization and governance of technology, business processes, and data. It also incorporates the ISO/IEC/IEEE 42010:2011 standard, which establishes guidelines for the characterization and description of systems and software architectures, ensuring consistency and comprehensiveness; iii) an iteration of the Reusable Asset Specification standard, which defines each of the Architecture Building Blocks into different classes. These classes include: the asset (Name, ID, and Description); the profile (asset type and its lineage): classification (allowing the blocks to be managed in a repository); usage (detailing the set of activities performed within the asset); related-asset (relationships with other blocks); solution (Solution Building Blocks to be deployed within the governance framework); requirement; and architecture-description.

Considering the standards utilized, the authors developed two distinct ABBs. The first ABB is related to architecture principles, where data governance principles are defined, listed in a repository, and managed. This ABB encompasses essential organizational aspects, including requirements, mission, vision, strategy, goals, data governance bodies, and roles. The second ABB relates to policies and standards, including governance and stewardship. It encompasses the organization's critical data governance activities, such as compliance with standards, regulations, and established guidelines, as well as the definition of performance monitoring indicators.

International standards and guidelines

Considering the frameworks discussed above, it is evident that organizations must align their data governance framework with comprehensive international standards that provide essential guidelines for data governance and quality. The International Organization for Standardization (ISO) has produced key principles that organizations can integrate into their data governance practices. Among these, three particularly relevant and complementary standards are: i) ISO/IEC Standard 25012:2008; ii) ISO/IEC 38505-1:2017; and iii) ISO/IEC TR 38505-2:2018. These standards establish critical components for effective data governance and quality management. ISO/IEC 25012 defines a detailed model for data quality, emphasizing the key attributes and features that management should analyze to improve data quality. ISO/IEC 38505-1 provides a framework for organizations to ensure effective data governance and emphasizes the need for a data accountability map and its key attributes. Finally, ISO/IEC TR 38505-2 discusses the implications for governing bodies to consider data accountability and management requirements. It details how governing bodies should develop a set of policies and procedures for data use, ensuring alignment with the organization's overall vision, mission, and objectives (International Organization for Standardization, 2008, 2017, 2018).

In line with this, ISO/IEC 25012, recognized by the International Organization for Standardization, plays a pivotal role in data governance, particularly concerning data quality. This standard provides a detailed model and principles for evaluating data quality, leveraging the data life cycle, which is often longer than the software life cycle (International Organization for Standardization, 2008; Rafique et al., 2012). This model focuses on data stored in computer systems in a structured format and aims to label and assess data quality requirements throughout the processes of producing, acquiring, and integrating data. It also helps identify criteria to ensure data quality, with a focus on assessment and continuous improvement. Additionally, it validates data compliance with current laws, regulations, and data quality requirements (International Organization for Standardization, 2008).

This standard asserts that data quality can only be assessed using a predefined data quality model, designed to guide the evaluation and continuous improvement of data quality while ensuring compliance with current laws, regulations, and requirements. Accordingly, ISO/IEC 25012 provides a model composed of fifteen different characteristics, including accuracy, consistency, timeliness, and completeness. This model is divided into two perspectives: i) an inherent perspective, representing the extent to which the intrinsic features of data meet stated and implied needs under predefined conditions. It considers factors such as the data domain, its relationships, restrictions, and metadata; and ii) a system-dependent perspective, corresponding to how effectively data quality is achieved and maintained within a computer system environment. It includes aspects such as hardware technology (ensuring data availability and required precision), a computer system software (providing backup for correct data

recovery) and other software that ensures data migration and portability (International Organization for Standardization, 2008; Rafique et al., 2012).

Moreover, this standard identifies five specific inherent data quality characteristics, which include the following: i) accuracy, described as the degree to which data values correctly represent the actual value of a given concept. This characteristic measures the difference between the correct value and the value used; ii) completeness, associated with the data quality feature that ensures an organization has values for all expected attributes; iii) consistency, where data are expected to be free from any illogicality and must be coherent within their context of use, iv) credibility, which pertains to the authenticity and believability of data in their context of use; and v) currentness, where data correspond to the appropriate age and intended period of use (Haug, 2021; International Organization for Standardization, 2008). Similarly, the standard identifies three specific data quality characteristics related to the system-dependent point of view, including: i) availability, which ensures that data are only retrieved by appropriate and authorized users within their context of use; ii) portability, which refers to the ability of data to be shared, substituted, or integrated between different systems while maintaining quality; and iii) recoverability, which corresponds to the capacity to recover and preserve data and quality in the event of failure, data loss, or corruption.

ISO/IEC 25012 outlines a total of seven additional data quality characteristics that are common to both points of view: i) accessibility, which relates to the ease with which data can be retrieved and accessed by users; ii) compliance, where organizations ensure that their data adhere to relevant standards, policies, or regulations that are currently in place, iii) confidentiality, which ensures that data are protected from unauthorized and inappropriate access, maintaining privacy; iv) efficiency, which measures how well data processing optimizes the use of necessary resources; v) precision, which refers to the exactness of data, including the ability to be distinguished in their context of use, vi) traceability, where an organization's data contains audit trail characteristics to track the data's history, origin and lineage; and vii) understandability, which refers to the ability of users to easily read, interpret, and comprehend the meaning of data in their context of use (Calabrese et al., 2020; International Organization for Standardization, 2008; Rafique et al., 2012).

Based on the principles established by ISO/IEC 25012, the International Organization for Standardization recognized the urgent need for organizations to define the requirements and responsibilities of their governing bodies to ensure that their flood of data is continuously directed, evaluated, and monitored (International Organization for Standardization, 2017). Consequently, the ISO/IEC 38505-1:2017 standard was established in 2017. It provides principles and guidelines to support an organization's governing body—including data owners, directors, partners and managers, and others—in managing and governing data to maximize value (Sothilingam et al., 2021).

Accordingly, this standard stipulates that an organization's governing body must ensure its responsibilities regarding data governance. This body is accountable for managing data and ensuring compliance with obligations to both internal and external stakeholders. As a result, the standard emphasizes that the governing body should adopt a model based on three principles: Evaluate, Direct, and Supervise. These principles encompass the following activities: i) oversee and define a data governance framework that aligns with the organization's level of data dependency and allows for the evaluation of current and future data use; ii) clearly define the importance of data in achieving the organization's objectives and overall strategy, including the potential benefits and risks associated with its data; iii) direct the preparation and establishment of strategies, policies, and procedures to ensure that the data meet business objectives and establish a subcommittee to support the governing body in its overall activities, particularly from a strategic point of view; and iv) monitor and assess the overall effectiveness of the data governance and management framework, including the pursuit of maturity and conduct independent audits (International Organization for Standardization, 2017; Serrano & Zorrilla, 2021).

Additionally, ISO/IEC 38505-1:2017 defines a model for supporting an organization's data accountability through a mapping diagram. This model links the governing body's activities with the data management life cycle, including data creation, storage, processing, archiving, and deletion. The data accountability model encompasses the following activities: 1) collect, which focuses on acquiring, gathering, and creating data. This includes data entry into an Enterprise Resource Planning (ERP) system, transactions between and from other systems, sensors, reports, and subscriptions to data feeds. The process should also consider past decisions and the internal and external context. In this model, the governing body should determine how the organization will use or monetize data to achieve its strategic objectives; ii) store, which ensures the proper location and storage of data, whether physical or logical. In the store activity, the governing body should approve policies and procedures for allocating resources for data storage and subscriptions; iii) report, where organizations oversee the extraction, examination, and reporting of their data, either manually, or automatically, to support decision-making. In the report step, the governing body should guide managers in using the appropriate tools to maximize the total value of data; iv) decide, which involves making decisions based on the analysis and investigations of reports by both people and automated systems. In the decide step, the governing body should ensure that a data-driven culture is established and aligned with the organization's data strategy and expected behaviors; v) distribute, which refers to how the reported activity flows within an organization's internal and external environments. In this step, the governing body should define and implement a data distribution policy aligned with the organization's strategic plan; and vi) dispose, which focuses on identifying data that is ready for disposal during the reporting activity, followed by permanent destruction. In this activity, the governing body must ensure that approved policies and procedures for data disposal are followed for data that is no longer valid or maintainable (International Organization for Standardization, 2017).

After defining the data accountability map and the role of the governing body in managing it, the International Organization for Standardization introduced a technical report standard, known as ISO/IEC TR 38505-2. This standard provides detailed guidance on how the governing body should develop policies and procedures for data use, including activities within the data management lifecycle. These policies should be aligned with the organization's overall vision, mission, and objectives (International Organization for Standardization, 2018). This standard also connects with ISO/IEC 38505-1:2017 by emphasizing that, while the governing body is responsible for defining the data strategy and aligning it with the organization's overall structure and strategy, the management team is held accountable, within their delegated authority, for establishing and implementing these policies.

Furthermore, ISO/IEC TR 38505-2 defines a cascade mechanism for connecting business strategy to data management. In this mechanism, the governing body produces an informing policy in collaboration with management. This policy guides and influences the strategy, policies, processes, and controls to ensure alignment with the data strategy. It also ensures that reports are effective and that alerts are produced by controls.

Similarly, this standard establishes key areas that management should address within a data governance framework, including: i) the identification and definition of business activities and their needs, regulatory requirements, current gaps, and areas for improvement. This includes defining the roles of sponsors and stakeholders, as well as methodologies and approval processes for policy development; ii) the creation, revision, approval, and distribution of policies related to

data and data governance; iii) the implementation of a mechanism to present and communicate policies, including the development of instructional activities to support the defined policies and ensure their operating efficacy and efficiency; iv) compliance monitoring with established policies and making necessary improvements, which includes conducting annual reviews and pursuing continuous improvements (International Organization for Standardization, 2018).

Responsibilities and roles

In response to **Research Question 2** (RQ2) regarding the structure of a data governance model and its strategy, it is crucial that both roles and responsibilities related to data use and management are clearly defined, distinguished, and communicated within the organization (Jiya, 2021). In fact, Jiya (2021) notes that data ownership is often not straightforward, potentially involving shared or disputed responsibilities. According to Abraham et al. (2019), those in data governance roles must manage and supervise the framework or program on a day-to-day basis while ensuring coordination among the roles and responsibilities defined at the organizational level, including data owners, stewards, and users. Bennett (2017) expresses the growing prevalence of the Chief Data Officer (CDO) role in organizations. The CDO, the leader of data governance, is responsible for establishing and managing the data governance framework, ensuring that the firm's data assets are reliable, and that data quality is maintained throughout the data life cycle.

Moreover, Abraham et al. (2019) emphasize that roles such as data leadership, sponsor, data user, steward, and owner should be defined within an organization's structure, enabling reporting structures to facilitate and govern existing data assets. Understanding not just the existing roles within a data governance framework but also how they interact with daily operations is crucial. Data owners are typically operational executives responsible for managing the organization's data assets. They are accountable for reporting data requirements, risks, and existing controls (Abraham et al., 2019). In a RACI Matrix (Responsible, Accountable, Consulted, Informed), they are usually assigned the "Accountable" role, meaning they oversee and are answerable for specific activities (Clarke, 2019). The data steward role, on the other hand, usually represents the operational leader of a given operation who is considered an expert in a particular data asset. This role translates and maps business requirements of the data assets into technical specifications, including database design (Abraham et al., 2019; Clarke, 2019). In the RACI matrix, data stewards are typically "Responsible," ensuring that rules and standards are enforced on the data and associated data sets (Clarke, 2019). Therefore, organizations must recognize that governance processes combining accountability with local empowerment can effectively support business operations and services (Gardner et al., 2018).

Moreover, at the lowest operational level of data governance, data producers and data users play crucial roles. Data producers are individuals who create, aggregate, and transform data generated by others. Data users, on the other hand, are the primary consumers of a specific data asset, typically using it for data-related reports (Abraham et al., 2019). In a RACI matrix, data producers are usually categorized as "Responsible," often working under the direction of data owners or stewards. In contrast, data users are generally "Consulted" or "Informed" since they consume data content as customers and define what data they need and how they intend to use them (Clarke, 2019). Given these existing roles, organizations should establish committees or councils that facilitate hierarchical and cross-functional interaction, engaging data stewards with owners and users.

Additionally, Jiya (2021) emphasizes that the role of a Data Protection Officer (DPO) should be closely integrated with data governance, namely by providing expert knowledge on data protection issues, assessing incidents, and contributing to maintaining a robust data management system. Recent legislation, particularly the GDPR, has increased the focus on the importance of storing, using, and managing personal data, making it a major priority and concern in an organization's strategy and environment (Abraham et al., 2019). Furthermore, Dighe (2014) highlights that the growing demands for data governance are rapidly expanding, imposing challenges across various area and industries, including the financial and banking industries, where regulation has grown substantially.

Benefits and opportunities

Regarding the benefits of data governance **(RQ2)**, Bennett (2015) argues that if an organization is capable of promoting and implementing a proper and effective governance framework, it can significantly enhance data quality and information across various areas of an organization. This improvement, in turn, ensures that data and information are available to the appropriate functions, particularly in supporting the decision-making process.

Similarly, enhanced data analytics and management can lead to increased revenue by developing new products or services and by creating synergies and efficiencies that may reduce costs. Improved data governance also allows organizations to better manage existing data, including more efficient retention criteria. According to Sifter (2017), the leading objective of data governance is to improve the sustainability and efficiency of an organization's operations by capitalizing on and delivering crucial business data while improving and strengthening the decision-making process.

In a survey examining the state of data governance across three different periods—based on responses from 91 respondents in 2014, 117 in 2011, and 119 in 2007—Russom (2015) found that by 2014, more organizations had implemented a data governance program than in 2007. The study also showed that data governance has continuously guided data-intense business initiatives, especially in areas like business intelligence, data privacy, and compliance, becoming increasingly important to organizations over time.

Main challenges

In response to **Research Question 5 (RQ5)**, the literature highlights that different and intricate challenges in the field of data governance are becoming increasingly complex. These challenges include the companies' ability to adapt to and sustain various realities and advances, as well as the difficulty leaders face in understanding the significance of this field and the risks associated with inadequate or unsuccessful practices (Bennett, 2015). Additionally, enterprises are increasingly focused on using technologies and their features as a service and tool to enhance competitiveness and support business operations (Bennett, 2015).

Given this context, companies are experiencing an increase in data points and silos—repositories of data that are held and managed by specific individuals or groups and are not readily accessible to others within the same company. This isolation of data can have a negative impact on businesses, making the implementation of a sound data governance program more crucial than ever (Johnston, 2016). Sifter (2017) states that corporations, particularly banks and financial institutions with larger employee bases, often develop data silos where different units or departments rely on and use separate data warehouses, data marts, sources, and systems. We believe this issue can be solved by defining and establishing a solid data governance program that integrates the company's culture, people, processes, and systems. To encourage value generation and manage risk and mitigation effectively, enterprises are urged to define and implement an efficient governance strategy and structure (Bennett, 2015; Chakravorty, 2020; Hoppszallern, 2015).

Similarly, challenges in data governance are becoming increasingly prevalent, especially given the continuous growth of data and the competitive pressure to utilize and manage data more effectively for businesses to gain a competitive edge (Paredes, 2016; Stratton, 2014). Table 7 presents the key challenges identified in the various

Table 7
Overview of data governance challenges.

#	Challenge Description	References
1	Lack of solid data management frameworks	(Gardner et al., 2018; Haug, 2021; Hikmawati et al., 2021; Janssen et al., 2020)
2	Lack of awareness regarding the importance of the role of leadership in addressing data governance	(Cheong & Chang, 2007; Hikmawati et al., 2021; Paredes, 2016)
3	Technology innovation and fast improvements	(Bennett, 2015; Chakravorty, 2020)
4	Robust data protection procedures and legislation	(Sánchez & Sarría-Santamera, 2019; Coche et al., 2024; Milne & Brayne, 2020)
5	Higher usage and interaction within the IoT ecosystem and Cloud services	(Hikmawati et al., 2021)
6	Enterprises are facing data points and silos	(Janssen et al., 2020; Johnston, 2016; Sifter, 2017)
7	Strong and resilient amount of resistance towards change displayed by stakeholders/ leadership	(Chakravorty, 2020; Hikmawati et al., 2021)
8	Lack of DG knowledge and maturity	(Chakravorty, 2020)
9	Lack of guidance in establishing performance metrics and indicators	(Paredes, 2016; Stratton, 2014)
10	Lack of auditing policies to ensure data governance and quality	(Bennett, 2015; Chakravorty, 2020)

studies analyzed related to data governance and quality, including the following:

In addition to these general challenges, Riggins and Klamm (2017) present specific issues related to data quality that organizations often face, including: i) duplicate data across systems and databases; ii) lack of timeliness, quality, appropriateness, completeness and integration of data; iii) non-existence and insufficient controls over data and data security; iv) data that is used but generates less important and valuable reports for the organization; and, v) data collection processes that are error-prone and inefficient (Chakravorty, 2020; Mansfield-Devine, 2017; Manus, 2019). In fact, innovation in data collection through various technologies and methods has driven the need for more robust data governance mechanisms (Milne & Brayne, 2020).

In addition to defining data governance, Bennett (2015) examines and reflects on the importance of information governance (IG) in an organization's overall business strategy. The author defines information governance as the combined application of processes, activities, and technologies that are implemented within an organization to maximize the value of existing information while mitigating associated risks and costs. Bennett (2015) claims that the unceasing rise of data and technological disruptions facing the world and its companies must be the key drivers of a robust information governance strategy and framework. Furthermore, the author claims that the effective and structured leadership of information governance and its framework are key to ensuring that appropriate strategies, procedures, policies, and processes are effectively embedded within an organization.

Bennett (2015) argues that the increasing volume of data that each organization is facing amplifies the risks that a proper information governance framework should address, such as privacy breaches in IT and communication systems, the costs associated with storage and retention policies regarding the ever-increasing data, legal and compliance risks, and cyber incidents. The author further argues that promoting and implementing an effective information governance framework will foster and improve data management and collaboration. This will, in turn, equip organizational leadership with quality data to support its decision-making and strategic planning (Bennett, 2015). Additionally, organizations will improve existing data management practices, including better and more efficient retention criteria.

Data assurance

Relevant concepts

In response to **Research Question 4 (RQ4)**, understanding the concept of data assurance, its foundations, and its potential impact on the success of an enterprise's data governance is crucial for evaluating the importance and effectiveness of a data governance framework or program. This discipline addresses critical features of the framework for data management, including data quality, security, reliability, and monitoring (Dutta, 2016).

The literature indicates that organizations strive to enhance their governance, processes, and operations by leveraging artificial intelligence and its methods. Consequently, it is vital to recognize the different data domains within an organization. Typically, organizations contain two types of data domains, each presenting its own set of challenges and quality issues that may arise throughout the process flow (Dutta, 2016).

The first data domain is represented by data at rest, which refers to non-transactional data stored within a particular system, database, or data warehouse, not actively flowing through the organization. Examples of these domains include data within a CRM (Customer Relationship Management) system or an administration system, which serves as a primary source of input for other systems and databases (Dutta, 2016). Hassan and Chindamo (2017) describe data warehouses as technological tools that support business and reporting operations by acting as repositories for critical data assets and information, thus separating core operational and transactional systems from the processes of analyzing and reporting data. Common data warehouse technologies include Microsoft SQL, Oracle Database, and IBM dashDb (Hassan & Chindamo, 2017). However, Hay (2015) states that setting up a data warehouse is a complex and challenging process for any organization, regardless of its size and characteristics. This is mainly due to the organization's difficulty in ensuring their data assets are complete, accurate, and accessible (Hay, 2015). In fact, Mikalef et al. (2020) report that organizations struggle with data management due to siloed business units, making the data inaccessible and resulting in unclear definitions of the data assets, among other issues.

The second data domain corresponds to data in motion, which refers to data that is actively flowing and being exchanged or processed between two or more systems, databases, or data warehouses. This is commonly known as transactional data (Dutta, 2016).

When considering business analytics—a set of tools and techniques that organizations and individuals use to examine and analyze data—three types of business analytics techniques and components can be identified. The first corresponds to descriptive analytics, which aims to answer the question of what has happened or is currently happening (Riggins & Klamm, 2017). Tools such as business reporting, dashboards, scorecards, and data warehousing are commonly used in this approach. The second technique is predictive analytics, where the focus is on understanding and forecasting what will happen and why by using tools and methods like data mining, web and media mining, text mining, and forecasting (Riggins & Klamm, 2017). The third approach is prescriptive analytics, which seeks to determine what should be done and why, using activities such as optimization, business and decision modeling, and simulations (Riggins & Klamm, 2017). The summary of this taxonomy can be seen in Fig. 14.

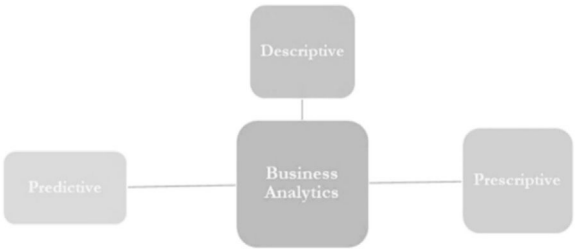

Fig. 14. Business analytics taxonomy. Adapted from Riggins and Klamm (2017).

Furthermore, data assurance involves various dimensions, from data quality and reliability to lifecycle and security. Understanding the implications of these areas in the data assurance field is crucial because they can significantly impact the effectiveness of a data governance program. Wang and Strong (1996) identified key data features—accuracy, consistency, completeness, cut-off (timeliness), availability, and relevance—as essential to overall data quality and that companies should address.

Recent literature demonstrates that, despite the dependence of data quality on the internal and external quality of its sources, the fundamental concepts of data quality have remained consistent over time. Data quality is defined as the ability of data to meet the implied needs of the user whenever required (Taleb et al., 2021). According to these authors, these dimensions and assurance assertions can be categorized as follows: 1) Accuracy—the degree to which data correctly represents the aspect that it aims to describe and characterize; 2) Consistency—the uniformity of the data across its lifecycle, systems, and sources, from the moment it enters the organization until it is stored, reported, or destroyed; 3) Completeness—the alignment and reconciliation of data with all required data elements for a given purpose; 4) Cut-off/Timeliness and Availability—ensuring data are available whenever required and up-to-date for the specific time it is being used; 5) Relevance—ensuring that the data being used and required are applicable and useful for a given context and intent (Taleb et al., 2021; Wang & Strong, 1996).

Data assurance aims to ensure that data are continuously monitored, maintained, and improved in quality within the data governance framework by leveraging different practices such as: i) data profiling, which is the process of scrutinizing data from various sources, both internal and external, to gain a thorough understanding of the data, and quality features, such as completeness, timeliness, and consistency. This involves techniques like statistical analysis, data visualization, and data mining to identify anomalies, patterns, and trends (Naumann, 2014); ii) data cleansing, also known as data scrubbing or cleaning, corresponds to a technique that detects and solves any issues or errors, inconsistencies, and inaccuracies in data. It also ensures that data formats are standardized, and duplicates and blanks are analyzed and treated using techniques such as statistical and descriptive statistics, algorithms, regular expressions, and machine learning models. This process is crucial for accurate analysis and reporting, ensuring that correct and standardized taxonomies are applied (Ridzuan & Wan Zainon, 2019); iii) data validation is a method certifying that data are precise, correct, and reliable throughout the lifecycle, while adhering to predefined business rules and standards. The goal is to identify and correct errors or imprecisions that may occur in data before being analyzed and reported.

Various techniques can be employed in the verification and validation of data to enhance quality and the success of a data governance program. These techniques include: i) field and format, which is the process of validating necessary fields against their expected data type, range of values, types and formats (e.g., ensuring dates are in date format and numbers in numeric format); ii) cross-field validation, ensuring that the relationships between different data, including data from different systems and repositories, are consistent and match with reality; iii) validation of data reference and source, which involves mapping data to pre-established business rules and standards, such as comparing a fiscal number or postal code to the country's valid taxonomy; iv) range and pattern analysis, which entails analyzing data to check whether it falls within defined thresholds, or follows expected patterns (Sohrabi et al., 2022); and v) data monitoring, which involves continuously evaluating, measuring, and reporting on data quality and errors. This process monitors adherence to defined quality standards, business rules, and objectives, aiming to identify any issues as early as possible for a timely and proactive response.

For this dimension, techniques such as those described can allow for effective data monitoring, profiling, cleaning, and validation. However, there are additional techniques, including data auditing to ensure that repositories where the data are stored align with business and regulatory rules and standards. Statistical analysis, which includes pattern and trend identification as well as outlier analysis, is also crucial. Other important techniques include data lineage analysis, where organizations formalize and describe the traceability of a data point from its origin (the moment data enters an organization) to its final stage (when data are destroyed, stored and reported), identifying all the steps that the data point passes through within the organization's systems. Lastly, data visualization techniques allow organizations to visually structure the data they intend to represent, enabling traceability, trend detection, and identification of potential issues (Ehrlinger & Wöß, 2022).

Considering this, we can highlight data assurance as an integral part of the data governance field, ensuring data quality and overall data management while ultimately driving better decision-making and reporting processes.

Relevant trends

It is essential to consider **Research Question 4 (RQ4)** when discussing the importance of data assurance and data governance. Consequently, identifying potential trends in these areas where enterprises can play a role is vital. Firstly, there is a potential relationship between data assurance, artificial intelligence, and machine learning techniques. According to Gandomi and Haider (2015), these two techniques are increasingly being integrated into data governance frameworks, helping organizations automate their data assurance processes, including data profiling, cleansing, validation, and monitoring. They also assist in identifying events, patterns, and trends that may positively or negatively impact downstream processes. Secondly, with data assurance comes the importance of defining and implementing stewardship regarding data ownership, management, and its ethical and responsible use (Borgman, 2018). Data assurance can potentially foster an organization's data stewardship by ensuring that clear roles and responsibilities for governing data are defined and implemented. It also raises awareness of the ethical, accountable, and transparent handling of their data and assets. Thirdly, data assurance not only strengthens a more robust management of data quality but also enables more controlled, transparent, and secure sharing of data among an organization's sectors, stakeholders, subsidiaries, and external partners. These practices aim to ensure that data validation and monitoring are effectively implemented, thereby maintaining the quality standards described earlier (Gandomi & Haider, 2015). Lastly, a key trend closely related to data assurance and governance is the FAIR data principles, which emphasize the importance of ensuring that an organization's data are Fair: (F) findable, (A) accessible, (I) interoperable and (R) reusable (Wilkinson et al., 2016). By adopting these principles, organizations can create a more reliable and high-quality data environment, which in turn supports better decision-making, enhances collaboration, and improves reporting standards.

The first principle, Findable, ensures that data are always easily identifiable for any user who needs them. To achieve this, data should have a unique and permanent identifier and be accompanied by metadata to inform the user of the context, definition, characteristics, content, and source (Wilkinson et al., 2016). The second principle focuses on data availability and accessibility, ensuring that data can be accessed immediately when required. Organizations must ensure that their storage capacity is sufficient to accommodate all their data and that it is available in a way that is both accessible and system readable. The third principle emphasizes date interoperability, reinforcing the need for data to be integrated into tools and systems and used alongside other data. To facilitate this, organizations must store data in formats that adhere to business rules and standards, while ensuring that appropriate documentation and metadata are defined. Finally, the fourth principle focuses on the reusability of data, ensuring that data are available for reuse and can be analyzed multiple times as needed. Therefore, organizations must ensure that their resources are adequate, readily available, and appropriate to meet users' demands (Wilkinson et al., 2016).

All things considered, it is evident that the trends affecting data assurance significantly impact an organization's ability to lead and manage a successful governance program. As data volumes and complexity increase, ensuring effective data assurance becomes a top priority remains a significant challenge. According to Dama (2017), emerging trends in data assurance and governance reflect the ongoing evolution and growth in these fields. By effectively addressing these trends, organizations can enhance data quality, leading to better decision-making, improved management of data assets, and greater transparency among stakeholders.

Data governance maturity models

In response to **Research Question 3 (RQ3)**, after recognizing the importance of trends in data assurance, it is crucial to address one of the biggest challenges in data governance and assurance: assessing the maturity level of a firm's data governance environment. This assessment involves evaluating the company's performance across key data governance dimensions, including policies, procedures, standards, roles and responsibilities, and other relevant aspects. As noted by R et al. (2024), a data governance program should not be treated as a one-time project but as an ongoing, continuous practice that requires sustained effort and commitment from organizations in defining, implementing, and regularly reviewing their data governance maturity level. Many organizations lack a clear and comprehensive data governance framework, particularly since there is no universal framework standard. Even those that do have one often fail to perform continuous monitoring to ensure that its maturity aligns with the expected level (Dama, 2017).

Similarly, with the recent growth in the importance and impact of data governance, this discipline has become a crucial element in determining an organization's success and its ability to meet or exceed market demands. This is especially true as data evolves into an asset that plays a key role in decision-making, modernization, and gaining a competitive advantage (Weber et al., 2012).

To address this gap, the literature has presented several data governance frameworks to help manage data as an organizational resource. These frameworks define data strategies, policies, and procedures, and promote the establishment of data stewardship within the organization's processes (Weber et al., 2009). However, it is not enough for organizations to simply define and implement a data governance model. They must view this process as an ongoing cycle of continuous monitoring and maturity assessment to foster organizational improvement and evolution. Consequently, maturity assessment models have been developed to evaluate the effectiveness and current status of these frameworks, helping organizations understand where they were, where they are, and where they aim to be in the future. These models enable organizations to stay aligned with best practices and maintain a continuous and effective data governance framework (Otto, 2011).

Subsequently, it is crucial to examine the relationship between frameworks and maturity models, as well as analyze and compare different maturity models for data governance, including their foundations, applicability, and challenges. In fact, Otto (2011) states that many data governance maturity models in the literature (also presented in this paper), are essentially "checklists" that have no practical application and lack support for diverse business contexts (Otto, 2011, p. 49). Similarly, Belghith et al. (2021) argue that no single model can meet all the requirements and characteristics of a given organization. Therefore, it is important for the community to understand the different models and their key components.

All things considered, it is fundamental to establish a data governance framework, monitor it, and periodically assess its maturity level. McDowall (2017a, b) presents a data maturity model composed of five different levels. These levels are: 1) Performed (lowest maturity), 2) Managed, 3) Defined, 4) Measured, and 5) Optimized (highest maturity). At the Performed level, the company's processes are ad-hoc and reactive, with no formal data governance initiatives or maintenance. At the Managed level, processes are more planned and executed according to documented procedures, with some level of process monitoring and involvement of relevant stakeholders. At the Defined level, the organization's processes are consistently employed in alignment with robust, established procedures. The organization has a mature understanding of managing data and perceives it as a unique asset. At the Measured level, organizations can measure their performance using metrics related to processes and data, conduct analyses and predictions, and safeguard the quality of the data lifecycle. At the highest level, Optimized, organizations continuously improve their processes, create synergies, and share best practices across their industry. Here, data are viewed as critical for surviving in a dynamic and competitive market (McDowall, 2017a, b).

In 2007, IBM introduced a maturity model named the Data Governance Maturity Model (DGMM), designed to provide organizations with a technique to assess their data governance capabilities and status while identifying critical areas for improvement (Belghith et al., 2021; Firican, 2018). This framework outlines five different levels of data governance maturity, including: i) Initial, ii) Managed, iii) Defined, iv) Quantitatively Managed, and iv) Optimizing. Each level builds on the previous one, aiming to establish a sustainable and continuous data governance framework that is integrated into the organization's processes and decision-making culture. The first level, Initial, indicates an informal, reactive, and inconsistent data governance status. At this level, data ownership is typically unstructured, and data quality is often poor or not assured. It also represents a level where the organization may lack formalization or documentation, with little understanding of its data assets and management. At the second level, Managed, organizations demonstrate significant improvement from their initial stage. Regarding their data governance framework, some practices are documented, and data ownership is defined, though some processes may still be reactive. At this level, the organization recognizes the value of data, and data projects may be in development or implementation. There is also a modest degree of automation, such as batch processing or job routines.

The third level, Defined, marks a cultural shift within the organization where data governance is managed proactively rather than reactively. At this level, data governance procedures and practices are formalized, focusing on compliance and risk management through the organization's processes. Data stewards are appointed and identified, even if their roles may not yet be fully developed, and data quality risk assessments are in place. At the fourth level, Quantitatively Managed, the data governance program is fully integrated into the organization's processes and driven by strategic business objectives while supporting the decision-making process. At this level, the organization focuses on continuous, systematic monitoring

and measurement of its data governance effectiveness. The emphasis is on identifying and resolving data issues before they impact business processes and activities, including the use of predictive models to analyze and prevent future data-related events. Lastly, the final level of the IBM framework is Optimizing, where data governance is viewed as a competitive advantage. The organization's culture is data-driven, and data are integral to decision-making processes and achieving strategic goals. Organizations at this level allocate resources to continuously seek improvements and leverage data to drive innovation and growth (Belghith et al., 2021; Firican, 2018). This framework helps organizations evaluate their data governance capabilities and develop a roadmap to enhance the current state of their maturity. As a result, it is particularly useful for organizations with more mature data governance programs (Belghith et al., 2021; Firican, 2018).

Moreover, Zorrilla and Yebenes (2022) propose a data governance maturity model to help organizations assess their current state, capabilities, and framework. This model provides a roadmap for continuous improvement and enhanced maturity in this area. It includes five levels of maturity, each with a detailed description of its meaning and impact. Also, it offers a perspective on positioning data governance maturity within the organizational environment.

According to the maturity model provided by Zorrilla and Yebenes (2022), organizations can be categorized into five different maturity levels: Level 1) Initial (Ad Hoc) is where organizations are considered to be at a foundational stage with primitive, disorganized, and uncoordinated data governance practices. Data governance is not applied consistently across the organization, horizontally and vertically, and its functions are often limited to individual projects with ad-hoc ownership and stewardship responsibilities; Level 2) Managed, is where data governance practices are planned and carried out according to company policies. Processes are tracked and assessed, and there is growing recognition of the importance of data governance. Some degree of automation is required, and awareness is increasing. Level 3) Defined, is where data governance standards and practices are applied uniformly across the organization and are customized to meet specific requirements using integrated technologies. This allows for the automation of data management activities; Level 4) Measured, is where metrics are established and applied to assess data governance processes. Statistical and data mining tools are used to process performance and lifespan; Level 5) Optimized, is the highest level of maturity that involves constant use of performance indicators to evaluate and improve the company's processes. The organization has implemented and shared the industry's best practices, and data are vital to maintaining a competitive edge.

Digital forensics

Definition and context

This science corresponds to a branch of forensic science and represents the part of this field that focuses on the digital domain. According to Valdez (2018), forensic science applies scientific principles and legal procedures to solve queries or crimes. It is an emerging field with numerous possibilities for application. For instance, forensic science encompasses a range of fields and applications, including digital forensics, accounting, toxicology, odontology, and criminal investigation. It also provides valuable methodologies and techniques that can be applied beyond the field of forensic science itself (Valdez, 2018). Moreover, digital forensics, formerly known as computer forensics, involves the analysis of all electronic devices, including computers, cell phones, printers, and cloud-based systems, as well as documents, email, and malware. This field encompasses the process of analyzing digital evidence to draw conclusions, or make inferences that can support analyses, findings, or investigations. The process involves gathering, preserving, analyzing, and reporting on digital evidence to achieve specific goals. While traditionally used to assist in legal investigations, digital forensics also supports organizational analysis, data assurance, and governance practices (Casino et al., 2022).

Nonetheless, forensic science, including digital forensics, was traditionally perceived as a set of methodologies designed to prove someone's innocence or guilt. However, the primary purpose of digital forensics is to present factual evidence obtained through the analysis of electronic devices (Valdez, 2018). In digital forensics, it is crucial for investigators to properly guard, label and document each piece of evidence, ensuring a clear chain of custody. This concept involves logging and tracking the whereabouts of all the evidence obtained to ensure that legal requirements are met (Valdez, 2018). Additionally, in digital forensics, investigators must not work with the original evidence directly. Instead, they must create and work with a forensic copy, which is an exact replica of the original file (Valdez, 2018).

Existing models and methodologies

Within the field of digital forensics, various models and methodologies are used to conduct investigations. We believe that understanding this science and putting it into practice can help enhance data governance and quality. As Valdez (2018) notes, forensic professionals never stop learning as technology continues to innovate and evolve, and the same applies to data governance and quality, which evolve with each organization's unique context. One methodology in digital forensics involves eight steps, illustrated in Fig. 15 (Valdez, 2018). This approach mainly focuses on applying digital forensics to solve a given crime but can also be extrapolated to other activities and methodologies. Step 1 involves responding to an event that requires digital forensics expertise, such as a crime, and necessitates a search warrant to legally support the investigator's presence at the crime scene (Valdez, 2018). Step 2, already at the crime scene, the digital investigator focuses on securing the scene and protecting any evidence present. To ensure proper documentation, the investigator must photograph, or videotape all the evidence, label it, and assign a number to ensure that the records are accurately and safely registered and maintained. In Step 3, the digital investigator identifies and extracts digital devices, creating forensic replicas of the original data on the devices. These images are then validated using hash values

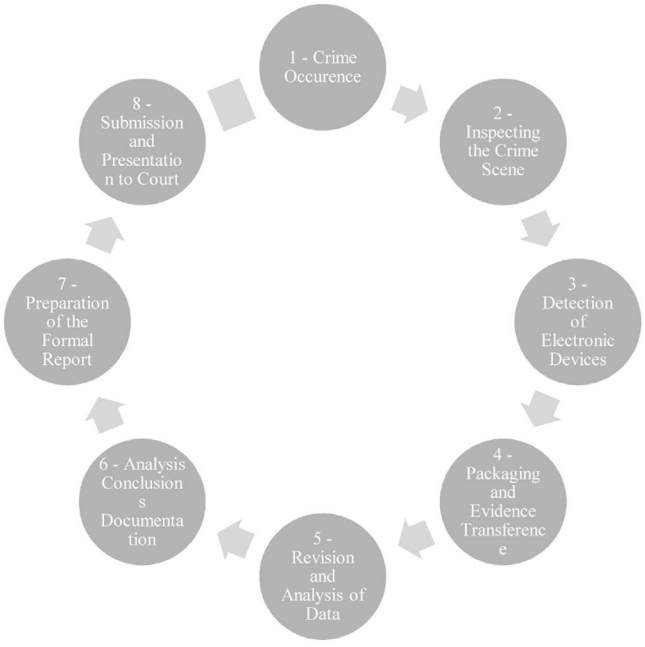

Fig. 15. Digital forensics methodology. Adapted from Valdez (2018).

(Valdez, 2018). Step 4 involves preparing and packaging the evidence for transfer to the investigator's laboratory or a forensics-prepared site. In Step 5, the investigator selects appropriate methodologies and techniques for the forensic analysis, examines the forensic images, documents each step of the analysis, and records all the results obtained (Valdez, 2018). In Step 6, the investigator ensures that all analyses and data are formally documented to support the conclusions drawn in Step 7, which involves preparing a report with these conclusions. The final step is the submission and presentation of the findings to legal bodies, such as the court, if necessary (Valdez, 2018).

Moreover, Casino et al. (2022) outline the main steps that are typically performed within a digital forensics model. The process begins with the Identification phase, where the analyst establishes the context and objectives of the analysis. During this phase, the analyst also determines all the required resources, including relevant policies, procedures, areas, and personnel. Following this, the analyst proceeds to the Collection and Acquisition stage. The primary goal here is to gather digital evidence while ensuring its safety and integrity so that it cannot be tampered with. Once the evidence is preserved and safeguarded, the analyst moves on to the Analysis phase. In this phase, various techniques, tools, and methods are employed to obtain the desired outcomes. Stages 4 and 5 involve Reporting, Discovery, and Disposal. At this point, the analyst consolidates the analysis and its outcomes, thoroughly documenting the procedures and steps taken. A comprehensive presentation of the outcomes is then prepared, which can be disclosed to various parties, including internal stakeholders, external parties, regulators, or the courts. Finally, while presenting the results, the analyst must ensure that all the work performed is properly stored for future reference, or inspections (Casino et al., 2022).

Additionally, the National Institute of Standards and Technology (NIST) recognizes forensic science as a process consisting of four primary phases: i) Collection, where the analyst identifies, labels, registers, and stores all relevant digital evidence from various sources, whether internal or external. Throughout this process it is crucial to ensure that the data remain unchanged and preserved at every stage; ii) Examination, where the analyst processes the data acquired in the previous stage. This involves performing automated, semi-automated and manual extractions, and assessing the quality of these extractions; iii) Analysis, where the analyst applies various methods to retrieve findings or outcomes that address the organization's concerns. It is essential that all methods used are legally compliant and properly performed; and iv) Reporting, where the analyst compiles the outcomes of the analysis into a presentation or report. This report details the findings and outlines all the steps taken in the previous stages, ensuring the traceability and auditability of the work (Kent et al., 2006).

Accordingly, these authors emphasize the importance of organizations ensuring that the data, models, and techniques used in the digital forensics process are consistent, traceable, authenticated, verifiable, and secure. Similar to the objectives of data assurance, these features aim to ensure that data are properly controlled, managed and accurately used in any analysis, decision-making, and reporting process, thereby enhancing the effectiveness and quality of a governance framework.

This is achieved by ensuring that the data and techniques applied are: i) consistent, where organizations ensure that data have not been modified during collection, so the original state is accurately represented; ii) traceable, where organizations maintain a record of all systems and flows through which data have passed, understanding all transformations and events related to that specific data, from entry into the organization until they are destroyed, reported and stored; iii) authenticated, verifiable and secure, where organizations ensure that the data provided are securely available, have not been manipulated, and are unique to the relevant business rules and processes.

Furthermore, they guarantee that only authorized individuals can access the data (Casino et al., 2022; Kent et al., 2006; Valdez, 2018).

Phase IV. Conclusion

Research conclusions

As highlighted in the literature, there is a consistent view among the authors analyzed that data governance is one of the most critical areas for firms to integrate into their businesses and operations. In other words, data governance may be seen as the key ingredient for a company's financial success because it is widely recognized that the future will be governed by data and data management. Therefore, this understanding underscores why data governance is a necessity (Ragan & Strasser, 2020, p. 10).

Consequently, Rivett (2015) argues that data governance is a gap that companies must fill, emphasizing the significance of this subject and its growing complexity. Corporations are increasingly striving to establish data governance frameworks to define and implement conjoined policies, best practices, manuals, and procedures to control and monitor data governance and its quality. Throughout this study, we can conclude that data governance is becoming a focal point for companies as they seek to meet compliance and regulatory demands from regulators and supervisory institutions (Coche et al., 2024). In this context, the author highlights that regulatory compliance is, in fact, a crucial aspect of a data governance program, stressing the importance for companies to understand and adhere to compliance requirements. For example: i) The General Data Protection Regulation (GDPR), which pioneered data treatment, collection and privacy regulations, particularly for personal data. The GDPR has influenced laws worldwide, such as Brazil's "Lei Geral de Proteção de Dados" and Australia's "Privacy Amendment Act 2017"; ii) the California Consumer Privacy Act (CCPA), which requires businesses to provide consumers with certain information about the personal data collected and its intended uses; and iii) the BCBS 239 principles from the Basel Committee on Banking Supervision, which focus on improving risk data aggregation and reporting procedures in banks, highlighting principles regarding data accuracy, consistency and traceability (Coche et al., 2024).

Nonetheless, organizations should remember that while data governance relates to data, it is not focused solely on data. It also considers the entire data lifecycle, including how data are collected, acquired, managed, monitored, utilized, reported, stored and destroyed. Additionally, data governance involves the roles of people, the processes, and the systems that support its governance and quality assurance (Janssen et al., 2020).

As a result, we believe that organizations must fully recognize the importance of data governance and its implications because it is one of the most multifaceted and emerging disciplines, offering companies numerous benefits and opportunities. However, it also introduces new and complex challenges, representing a gap that many businesses have yet to fully address and comprehend (Alhassan et al., 2019a; Johnston, 2016; Sifter, 2017).

Similarly, organizations must understand that planning, developing, defining, implementing, and maintaining a data governance framework aligned with a business strategy requires a significant investment in resources such as human capital, time, and assets. It also necessitates patience and a commitment to building robust foundations for data governance roles, responsibilities, frameworks, and culture. This includes implementing key risks and controls to govern data effectively, thereby enhancing and optimizing operations, administration, and compliance processes (Lancaster et al., 2019; Sifter, 2017). In summary, there are clear opportunities to improve and advance the field of data governance, including its synergies and relationships with other areas such as digital forensics and data

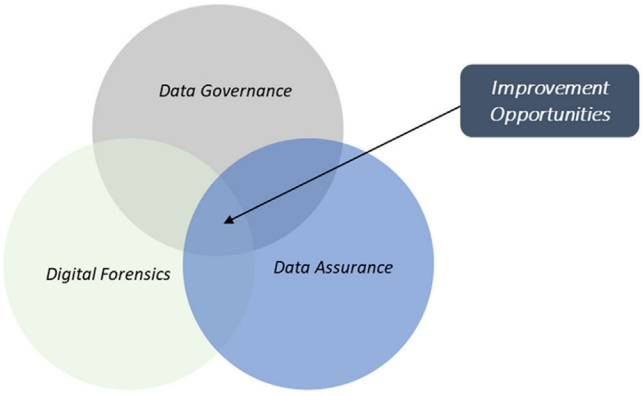

Fig. 16. Data governance and its synergies with other fields.

assurance. To help visualize this, we have drafted Fig. 16 to illustrate the objectives researchers should aim for in data governance.

Similarly, several authors consider data quality a critical aspect of any data governance framework and strategy because it directly impacts an organization's reliability, precision, and usability (Calabrese et al., 2020; Haug, 2021; Rafique et al., 2012). These authors argue that by adopting the comprehensive framework and model of data quality characteristics provided by ISO/IEC 25012, organizations are more likely to ensure that data meet their quality standards and requirements in their context of use, aligning with their data management and governance practices. We believe that, within the context of data governance, the model presented by ISO/IEC 25012 can significantly enhance the ability to achieve and maintain high data quality. By integrating both inherent and system-dependent characteristics, organizations can adopt a holistic approach to data quality management, leading to better decision-making and optimized operational efficiency.

Moreover, the guidelines provided by ISO/IEC 38500-1 and ISO/IEC TR 38505-2 are considered essential for ensuring effective data governance and management within an organization. These guidelines help define the roles and responsibilities of the governing body and management, incorporate data accountability requirements, and establish policies that align with the organization's business and data strategy. Additionally, they emphasize the need for organizations to ensure that their data governance policies are fully aligned with their business needs, strategy, and vision. The guidelines also stress the importance of continuous monitoring and assessment to maintain this alignment over time. Furthermore, they highlight various industry success stories where their implementation has brought tremendous value to organizations.

A common theme in these case studies is the importance of developing and managing data policies that align with overall business strategies, as well as creating a data accountability map. This approach enables organizations to identify key areas and methodologies where their data governance could be applied. Consequently, these studies show how vital proper data governance is to an organization's success and its ability to achieve broader goals. In this context, we believe that the standards analyzed—ISO/IEC Standard 25012:2008, 38505-1:2017, and 38505-2:2018—provide organizations with a set of clear and structured principles for defining a robust data governance framework, establishing the roles and responsibilities for governing such a framework, and adhering to data quality principles.

Furthermore, based on the findings of this paper, we outline below the conclusions drawn from the research questions posed in this study:

RQ1) What are the key challenges and opportunities in the area of data governance, and what benefits and issues can be learnt from successful implementations?

As seen throughout the literature, data governance involves integrating policies, manuals, formal procedures, structures (functions and responsibilities), and processes that characterize a firm's data assets. Essentially, it is the governance of a "business-critical asset" like data, encompassing the planning, management, and control of data and the associated processes. Studies demonstrate that data governance has become increasingly relevant to organizations. Therefore, we encourage readers to recognize the key topics in this field and incorporate them into their organization's core processes and strategy.

Given this, it is important to understand that a data governance framework comprises practices that include strategies, policies, procedures, roles and responsibilities, data stewardship, and control over the management of the data lifecycle. Organizations must also recognize that data quality encompasses various dimensions, such as accuracy, reliability, availability, consistency, and completeness. Data governance and quality impact every stage of the data lifecycle, from acquisition, or creation, through analysis, to storage, reporting, and eventual destruction. Establishing clear roles and responsibilities, such as data stewardship, is also essential. This may involve assigning specific roles for data management and accountability, such as a Chief Digital Officer who oversees the data governance program. Additionally, organizations should embed data lineage and a data dictionary into their processes. Data lineage ensures the traceability of data points, tracking their flow within the organization and its systems, from origin to storage, reporting, and destruction. Meanwhile, a data dictionary or inventory provides a centralized catalog of all critical data and its characteristics, including structure, meaning, taxonomy, relationships, sources, formats and the systems and databases where the data are used or available.

As highlighted in sections "3.2.5. Benefits and Opportunities" and "3.2.6. Main Challenges," the literature acknowledges that data governance can significantly influence an organization's ability to manage its data assets. Data governance ensures that the correct data are accessed and analyzed by the appropriate people within an accurate and complete context of use. It also fosters and enhances organizational technologies, leading to more optimized, controlled, monitored, and updated systems across various levels, departments, and executive lines. Additionally, it promotes accountability by establishing pre-defined roles and responsibilities for data stewardship. A well-established data governance framework allows organizations to produce a single version and source of truth for their data, providing greater accessibility and security, and supporting real-time reporting that improves decision-making and data quality.

In contrast, several challenges affect the effectiveness of data governance within an organization. Studies show that while several data governance frameworks exist, they often lack solid case studies and testing. These frameworks tend to be standard and do not always account for the specific needs of different contexts. There is also a lack of awareness regarding data governance, the importance of leadership in this role, and the need for a well-established data governance structure with clear roles and responsibilities. Additionally, organizations face issues due to inadequate data quality mechanisms and controls, including duplicate data across their data assets and the lack of accurate, complete, and timely data. These issues can impair decision-making processes and increase the likelihood of errors. Despite these challenges, it is evident that organizations can still improve their operations, manage and utilize their data assets more effectively, and support a more robust decision-making process to achieve their strategic goals.

As a result, we believe that by integrating these key data governance topics into their organizational environment, organizations can potentially unlock greater efficiency in governing data. This integration, in turn, can lead to improved decision-making and reporting processes while also ensuring compliance with regulatory requirements and maintaining trust among stakeholders and clients. The

potential benefits of such integration are significant, including increased operational efficiency, better decision-making, and enhanced stakeholder trust.

Similarly, although there are few documented and analyzed case studies of successful data governance implementations, we have verified that the standard ISO/IEC 38505-2 provides informative examples through a set of case studies illustrating the application of its principles and guidelines in various scenarios. For instance: i) A travel service company needed to improve customer satisfaction indicators through personalized suggestions. To achieve this, the company developed a governance structure driven by this need and leveraged defined data governance policies to support robust data analytics procedures. The result was a significant improvement in customer feedback, with the company continuously learning about and assessing diverse customer types and their travel experiences, thereby enhancing travel recommendations; ii) In China's financial industry, a case study highlights the importance of establishing data policies and tools to manage and analyze the vast volume of daily transactional data flowing through the Chinese Stock Exchange. This case underscores the governance controls necessary for processing such data, with a focus on critical data quality features like accuracy and timeliness; iii) An air transport company employed a data strategy using ISO/IEC 38505. This strategy helped the company identify the strategic value, risks and limitations of the various types of data circulating through its systems, ensuring alignment with its core mission and strategic objectives while optimizing data use and enhancing operational efficiency. As a result, the company facilitated the processing of over a billion airline passengers annually (International Organization for Standardization, 2018).

Additionally, the literature highlights another successful case study focused on implementing a data governance framework in Hainan Province through the conceptualized framework of the Government Data Middle Platform proposed by Mao et al. (2021). This case study shows that Hainan Province launched the Smart Hainan Master Plan for the period from 2020 until 2025, which focused on integrating various data sources and promoting cross-departmental and cross-level interactions. The plan encompassed the following key actions: i) the data resource center, which facilitated the creation of a data lake that integrated multiple data sources, gathering both structure and unstructured data from different departments; ii) the six platforms, which included activities such as data collection (gathering data from different sources), convergence and analysis (exploring data correlations and ensuring data quality attributes), data management (handling and storing metadata), service provision (delivering data-related services to the public), trusted exchange (safeguarding data flows and exchanges), and social platforms (mechanisms for broader data sharing support); iii) the four systems, which included policies, standards and procedures (aligning with the organization's data strategy and stakeholders), a basic capabilities system (supporting each step of the data middle platform), a system for managing operations and maintenance, and a system for managing security; and iv) the unified data portal, which provided users with a single point of access to data, fostering accessibility and delivery (Mao et al., 2021).

As a result, Hainan Province improved the delivery and quality of its public services, enhanced administrative efficiency and effectiveness, and strengthened the security of its data resources. This success underscores the effectiveness of both the Smart Hainan Master Plan and the Government Data Middle Platform.

RQ2) What is the current level of maturity and background surrounding the topic of data governance?

Our research indicates that the literature perceives data governance as a field that has evolved significantly in recent years, primarily to keep pace with the growth of data. Data governance can be considered a complex topic focused on defining and implementing formal procedures and processes to manage data effectively throughout the data lifecycle within organizations and their systems. Additionally, this field supports various purposes across different industries and sectors, including financial and banking markets, advertising and marketing, data privacy, and regulation.

Furthermore, while various frameworks for data governance exist in the literature, the scientific community highlights a lack of guidance on understanding and applying these frameworks. There is also a noted deficiency in comparative analysis of frameworks regarding their features and tools (Jiang et al., 2024). Additionally, there is no universal standardized framework that "fits" all organizations. Therefore, it is important to recognize the existing frameworks and the specific topics they address and apply. The literature also emphasizes the need to refine and enhance these frameworks to accommodate new features and demands arising from the ever-evolving technological landscape, including different tools, specifications, complex data sources, data flows, regulatory compliance standards, and other key aspects that may impact the effectiveness of data governance frameworks (Jiang et al., 2024).

Consequently, organizations can adopt one or more frameworks, or a hybrid approach, combining critical features from different methodologies to best suit their specific needs and business objectives. For instance, the following methodologies can be considered: i) The Data Quality of Data in Motion framework by Dutta (2016), which includes five stages: discovering, defining, designing, deploying, and monitoring; ii) The Data Governance Model proposed by Al-Ruithe et al. (2019), which focuses on data quality roles, decision areas, and responsibilities; iii) the DGI framework from the Data Governance Institute, which covers 10 dimensions, including mission and vision, governance metrics and funding tactics, rules, rights, accountabilities in decision-making, controls, leadership over data, data governance officers, data stewards, and proactive, ongoing data governance processes; iv) the IBM Data Governance framework, comprising 14 elements (10 required and four optional). Required elements include defining a business problem, sourcing an executive sponsor, evaluating maturity, developing a roadmap, creating an organizational blueprint, managing data flows, maintaining a data dictionary and metadata repository, building success metrics, appointing data stewards, managing data quality, and implementing a master data management strategy. Optional elements include analytics for governance support, privacy management, data flow monitoring, and measuring governance program and initiative results; v) The Conceptual framework by Abraham et al. (2019), which encompasses six dimensions: governance instruments, corporation, data, and domain scopes, antecedents, and consequences; vi) The Data Governance framework by Alhassan et al. (2016), which incorporates five concepts: principles, data quality, metadata, data accessibility, and the data lifecycle; vii) The Gamified Approach to Data Governance by Hay (2015), which includes elements such as steady feedback, clarity, ambitions, badges, onboarding experience, competition, and cooperation.

As a result, we can conclude that data governance is of paramount importance given that various models and frameworks are available to assist organizations in defining and implementing their data governance strategy. These models can be tailored to specific needs by combining features from different approaches. Ultimately, by establishing clear policies, procedures, processes, and roles, organizations can improve their data governance and quality, thereby enhancing their decision-making and reporting processes.

RQ3) What are the current methodologies to support a data governance program and assess its maturity level?

This paper acknowledges the importance of organizations defining and establishing a model to continuously assess and monitor the maturity of their data governance program. These maturity models are crucial tools in supporting an organization's strategy because they enable effective management and evaluation of the data governance framework. They help identify gaps and areas needing

improvement, allowing organizations to develop a roadmap to address these findings. Moreover, by assessing how well organizations acknowledge, manage, and structure their data and value, these models ensure compliance with data governance regulatory requirements and facilitate the optimization and continuous improvement of decision-making processes. Organizations can also use these models to understand their current position and trajectory within the data governance field, evaluating their past performance, present status, and future goals in terms of desired maturity levels. Typically, these models follow a five-level maturity scale with a bottom-up approach, ranging from the lowest maturity level, often termed "initial" or "inexistent," to the highest, commonly described as "optimized" or "continuous improvement." They thoroughly examine an organization's data governance framework, including policies, procedures, practices, roles and responsibilities, systems, and overall data quality and metrics.

Moreover, while these models may not be specifically tailored to an organization's specific characteristics and needs, and while new models may emerge as awareness and regulatory requirements in the data governance field evolve, they still provide a robust foundation for maturity assessment. Organizations can also leverage these models to conduct a thorough analysis of their maturity.

Consequently, it is essential to highlight the practical applications these models offer organizations. Specifically, these models can: i) focus efforts on the right areas by identifying gaps or weaknesses through comparisons of current practices with best practices, thus pinpointing areas needing improvement, such as policies, procedures, roles and mechanisms; ii) design and develop a roadmap for improvement, prioritizing initiatives and monitoring results to enhance data governance maturity; iii) benchmark data governance maturity and practices against industry standards, best practices, and regulatory requirements, and learn from the best practices of other organizations or regulators; iv) ensure adherence to regulatory compliance requirements, especially for organizations in the financial industry. This includes providing an overview of the maturity of information flows managed by their data governance framework; and v) promote a data-driven culture that encourages reliance on data, ownership, and integration within business processes, thereby enhancing data-driven decision-making.

Furthermore, a maturity assessment model for data governance is highly relevant and applicable to any organization that uses data, regardless of whether it is at a lower maturity level or has achieved an optimized one. These models are suitable for any organization seeking a comprehensive overview of its data governance framework, allowing firms to clearly position themselves and define a tailored roadmap of actions based on their size and strategy.

While valuable models exist for assessing the maturity of data governance frameworks, there are still gaps and challenges in the literature that need to be addressed. Specifically, more empirical evidence is needed to demonstrate the practical applicability and effectiveness of these models. Additionally, the literature suggests that the steps for implementing each model are not yet fully clear to organizations and their users. There is also a lack of consensus on how different dimensions or process areas should be measured, leading to variations in how different models evaluate the same areas.

RQ4) What significant baselines from other fields can enhance data governance activities, structure, and archaeology?

Our paper provides a general overview of two fields—Digital Forensics and Data Assurance—that we consider crucial and interrelated with data governance. These fields play a significant role in improving and enhancing the efficiency of the data governance framework. From our analysis, it is evident that while digital forensics and data governance are essential and distinctive fields, they share a common focus on managing, governing, and protecting digital and physical data. Despite their differences, these fields are interrelated and complementary, both being critical for ensuring that an organization's compliance and asset protection, particularly regarding data, align with best practices and remain effective throughout the data lifecycle, including the decision-making process.

In this context, we believe that the relationship and synergies between these fields have enormous potential, particularly in the following areas of Digital Forensics:

i) Legal and regulatory compliance, where both fields require strict adherence to legal and ethical recommendations and procedures. In Digital Forensics, this involves ensuring that the collection, preservation, analysis, reporting, and documentation of evidence—especially electronic evidence—follow robust and formalized procedures that respect privacy rights and data integrity. Similarly, data governance focuses on defining and implementing policies and procedures to ensure that the organization complies with data regulations such as GDPR, CCPA, and best practices and standards like those outlined in BCBS 239 (a Basel Committee guideline for risk management in data reporting and aggregation in banks and financial institutions). By working together, these disciplines ensure that an organization is aligned with and adheres to the legal and ethical standards in place;
ii) Data events and remediation, where data governance leads by establishing the policies, procedures, roles and responsibilities, processes, and resources necessary for effectively managing data. This structured approach positions the organization to quickly identify and respond to any event, including data breaches, cyber incidents, system failures, or errors. Data governance facilitates incident and event management, including response and remediation when necessary. At this stage, Digital Forensics becomes essential, providing the methods and tools to investigate the root cause, impact, and context of the event. This collaboration enables organizations to take quicker proactive and corrective actions, thereby strengthening their data governance and management processes;
iii) Reliability of the organization's data and data quality is achieved when organizations effectively and efficiently integrate digital forensics analysis and conclusions into their data governance framework, demonstrating a strong commitment to safeguarding their assets. This not only reinforces their internal culture but also communicates to investors, stakeholders, industry peers, and regulators that data governance and asset protection are top priorities;
iv) Policy development and implementation requires that both fields create and enforce well-defined policies, procedures, guidelines, and mechanisms to ensure effective data governance and incident management. By working together, these areas can monitor and enforce policy compliance, ensuring alignment with each other and with best practices;
v) Foster joint training, awareness, and proactive initiatives. As both disciplines work with data, organizations must ensure that there is ongoing training and awareness programs. This helps align their culture and personnel with current industry standards, regulatory requirements (if applicable), and the latest tools and techniques;
vi) Continuous monitoring, optimization, and improvement, where both fields can mutually benefit from each other's actions, insights, and findings. Digital forensics can identify and reveal vulnerabilities that may impact an organization's data governance framework. By collaborating and sharing knowledge, these fields can promote the continuous assessment, improvement, and optimization of a firm's ability to govern and use its data in the safest, most reliable, and accurate way possible.

Considering this, it is evident that although data governance, digital forensics, and data assurance are distinct disciplines, they are

interdependent and collaborative. Each plays a crucial role in promoting sustainable data governance for an organization's data and assets. We encourage readers to recognize that these three disciplines extend beyond their individual roles within the organization. By working together and promoting synergies, they enhance the firm's ability to manage and govern its data and respond to any events that may affect data quality. Data assurance, in particular, plays a vital role in this process, providing reassurance and confidence in the effectiveness of the data governance framework.

Additionally, data assurance can play a vital role in enhancing the effectiveness of the data governance framework, particularly through the following baselines:

i) Complementary synergies and objectives, where data assurance and data governance are distinct in focus, but work towards complementary objectives. Data assurance is concerned with verifying the accuracy, reliability, quality, and security of data, while data governance establishes the policies, procedures, and standards for how data are governed, acquired, analyzed, stored, and shared within the organization. Both disciplines aim to enhance data quality, improve decision-making and reporting processes, and ensure the security, availability, and consistency of data across the organization;
ii) Regulatory compliance requirements, where both disciplines are crucial for organizations to comply with current legislation and standards. Data assurance guarantees that data are of high quality—accurate, complete, secure, consistent, and reliable—while data governance provides the necessary framework and processes to manage data and meet regulatory requirements and best practice guidelines;
iii) Risk management procedures, in which both disciplines are crucial for managing data-related risks. Data assurance helps identify and mitigate risks associated with data quality, such as data manipulation, corruption, external data issues, or unauthorized access. Data governance, on the other hand, addresses risks related to regulatory compliance, privacy, data sharing, and reporting procedures. By integrating both disciplines, organizations can develop a comprehensive and robust data risk management strategy;
iv) Relationship with stakeholders and external providers, in which organizations that align their data assurance techniques with data governance practices may see a significant increase in trust and reputation among stakeholders. Both disciplines should focus on maintaining and managing the governance and quality of the organization's data, ensuring it is accurate, reliable, and consistent in data use;
v) Continuous improvement and monitoring, where both disciplines can mutually benefit from each other's processes and findings. Data assurance can help identity events and weaknesses, while data governance ensures that the organization classifies these events and plans for their response and management;
vi) Data lifecycle management, in which both fields should be involved throughout the entire data lifecycle—from the moment data enter the organization, including their creation, acquisition, or collection, to the analysis, storage, reporting, and eventual disposal. This involvement ensures that data remain accurate, consistent and available at all stages;
vii) Policy enforcement and system integration, because both fields require organizations to focus resources on developing and establishing policies, manuals, and formal procedures that align with their processes and systems. In the documentation process of policies and procedures, collaboration between these fields is crucial, because many procedures may overlap. Additionally, organizations will benefit from insights provided by both fields, including industry standards, best practices, and regulatory compliance requirements. As technology advances and organizations increasingly rely on systems to manage and analyze data, the relationship between data assurance and data governance becomes even more significant.

We believe that although data governance and data assurance are distinct areas, they closely interact and complement each other to enhance an organization's ability to manage the quality of its data assets. Both are crucial in ensuring that data remain consistent, accurate, secure, available, and reliable throughout their lifecycle, while also complying with legal and regulatory requirements and industry best practices. We recommend that organizations design their data governance strategies and frameworks to effectively integrate data assurance and digital forensics. Leveraging the key features and methods from these areas can improve the maturity of data governance, management, and quality. With technological advancements and the continuous growth of data, it is essential for organizations to adopt a versatile approach to data governance. This approach can ultimately maximize the value derived from data, which can be considered a key competitive asset in today's market.

RQ5) What are the main obstacles and constraints that the data governance discipline is currently facing and may face in the future?

While data governance offers many opportunities, it is crucial for organizations to understand the key challenges and constraints facing this field. The literature has highlighted several significant challenges that threaten the effectiveness of data governance in business and processes, including: i) the absence of a universal and standard data governance framework because there is often a lack of knowledge and guidance on existing frameworks and maturity levels; ii) a lack of awareness, with many organizations not fully recognizing the concept and importance of data governance, or how to leverage its principles effectively; iii) an unclear definition and implementation of leadership roles to address and manage data governance frameworks; iv) technological advances and increasingly stringent regulatory requirements, which pose significant challenges; v) the growing number of isolated data points within organizations (i.e., data silos); vi) the absence of monitoring mechanisms and metrics to ensure the success of data governance initiatives; and vii) complex technological infrastructures, which introduce challenges in maintaining data quality and consistency across company systems, including legacy systems.

Given these challenges, it is evident that organizations must address them when defining and implementing their data governance frameworks. They need to invest in and allocate sufficient resources to overcome these obstacles and effectively manage their data assets.

Research managerial implications

This article illustrates the significance of data governance and explores existing frameworks, along with their connections and synergies with other domains such as data assurance and digital forensics. Through a systematic literature review, we have identified key insights that can impact organizations, including their management, data governance leaders and structures, and policymakers. Our findings provide a deeper understanding of data governance and offer valuable guidance for improving day-to-day operations and governance practices. By leveraging this research, organizations will be better equipped to make informed decisions and take effective actions, particularly in processes that rely on data.

Research theoretical implications

Our study emphasizes the connection between data governance and an organization's success and decision-making. We found that while data are highly critical for organizations, data governance

initiatives are often still in their early stages, with limited awareness of available frameworks, tools and their relationship with other data-related fields. By providing organizations with potential frameworks, tools, and insights into synergies with related areas, we believe that this research will positively impact organizations and enhance their data governance maturity.

Limitations and future research

For future research, we recommend ongoing and exhaustive studies on data governance to keep up with technological advances in the field. Continuous research is essential for the continuous improvement of these areas, enabling both researchers and organizations to adapt to the rapidly evolving environment of data governance. Additionally, future studies should explore the potential relationships and synergies between data governance and other fields, such as data assurance, digital forensics, artificial intelligence, internal audit and control, risk management, and other data-related sciences.

Similarly, despite the value of the frameworks and maturity assessment models studied, several important questions remain unanswered within the scientific community and are crucial for organizations. Notably, there is a lack of empirical research on the application and effectiveness of these models. Moreover, there is still limited guidance on how organizations can advance their data governance frameworks and maturity assessment models.

Furthermore, we recommend that researchers interested in this field explore key areas that could impact data governance and its frameworks. These include the potential creation of a universal and standardized data governance framework and maturity assessment models, the adoption of advanced technologies such as blockchain, machine learning mechanisms, and data visualization techniques, as well as the exploration of detailed baselines for defining data stewardship.

Consequently, we believe that pursuing these future research initiatives in data governance will greatly benefit the field, as well as the organizations and practitioners that leverage this body of knowledge. This progress will enhance business operations, maximize asset value, improve decision-making and reporting processes, and foster transparency and stronger relationship with stakeholders.

Moreover, the study's methodology and the topics under analysis, which include current data governance frameworks and maturity assessment models, may be subject to selection biases that could influence the findings. The frameworks and models chosen for analysis were selected based on their availability and prominence in the literature, which may have led to the exclusion of lesser-known but potentially important approaches, especially those specific to particular industries, company sizes, or types. Additionally, the varying quality and scope of the literature could lead to gaps in the coverage of specific frameworks, or models, introducing biases from the perspectives of the publication and their authors.

Similarly, we acknowledge that subjective judgment and criticism may have influenced the criteria used to assess the maturity of the frameworks and models. However, we conducted meticulous research to ensure that we captured the most relevant frameworks and models available in the literature to date. By recognizing these limitations and potential biases, our goal is to offer the academic community and organizations a balanced, comprehensive, and transparent analysis. We also aim to highlight areas where further research and more comprehensive data are needed to fully understand the current and future state of the art in the field of data governance.

In line with this and considering the importance of the standards and guidelines provided by the International Standards Organization in the context of data governance and quality, we believe that future research should focus on evaluating their efficacy. This research should explore how these standards can be integrated and employed together to create a more robust and adaptive data governance framework and data quality principles. Additionally, we suggest that researchers analyze whether artificial intelligence-driven tools for data analytics can be applied to improve, monitor, and optimize data quality, and investigate the potential values these mechanisms could offer organizations.

Additionally, we believe that future research should examine the standards ISO/IEC 38505-1 and 38505-2 and their impact on data governance across various industries. This research should explore real-world examples of their practical application and benefits in different organizational contexts, including small, medium and large organizations. Researchers should also investigate how these guidelines can be used together to develop and monitor critical performance and risk indicators, ensuring continuous assessment and improvement of data governance effectiveness over time. Although these standards provide clear guidelines for data governance, data quality management, and accountability, it remains unclear how these principles can be integrated with existing data governance frameworks and how they compare in terms of effectiveness and implementation complexity across different industries. Therefore, we recommend that researchers conduct empirical studies to assess how these guidelines influence the operational effectiveness and efficacy of an organization's data governance framework and how they relate to other frameworks discussed in our study.

Contributions

Our study makes a significant and differentiated contribution to the growing body of knowledge in Data Governance, Data Assurance, and Digital Forensics. By analyzing critical features and establishing connections with other crucial data-related areas that have not yet been explored in the literature, our findings offer a unique perspective on this field. As a result, our work provides valuable insights and concepts that are applicable to organizations at any level of maturity in these areas. This study also lays the foundation for future research by suggesting baselines that can help foster and advance the body of knowledge in these data-related fields. Therefore, we encourage both the academic community and organizations to further explore and study these fields, as they hold considerable potential for growth and development.

Funding

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project UIDP/50014/2020. DOI 10.54499/UIDP/50014/2020 | https://doi.org/10.54499/UIDP/50014/2020.

Data availability statement

Data included in article/supp. material/referenced in article.

Additional information

No additional information is available for this paper.

Declaration of interest's statement

The authors declare no conflict of interest.

CRediT authorship contribution statement

Bruno Miguel Vital Bernardo: Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Henrique São Mamede:** Writing – review & editing, Validation,

Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **João Manuel Pereira Barroso:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Vítor Manuel Pereira Duarte dos Santos:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Formal analysis, Data curation, Conceptualization.

Acknowledgement

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project UIDB/50014/2020. DOI 10.54499/UIDB/50014/2020 | https://doi.org/10.54499/uidb/50014/2020.

References

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

Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: An analysis of the literature. *Journal of Decision Systems*, *25*, 64–75. doi:10.1080/12460125.2016.1187397.

Alhassan, I., Sammon, D., & Daly, M. (2019). Critical success factors for data governance: A theory building approach. *Information Systems Management*, *36*(2), 98–110. doi:10.1080/10580530.2019.1589670.

Alhassan, I., Sammon, D., & Daly, M. (2019). Critical success factors for data governance: A telecommunications case study. *Journal of Decision Systems*, *28*(1), 41–61. doi:10.1080/12460125.2019.1633226.

Ali, S., & Bano, S. (2021). Visualization of Journal ranking using Scimago: An analytical tool. *Library Philosophy and Practice*, 1–12.

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

Ariffin, K. A. Z., & Ahmad, F. H. (2021). Indicators for maturity and readiness for digital forensic investigation in era of industrial revolution 4.0. *Computers & Security*, *105*. doi:10.1016/j.cose.2021.102237 N.PAG-N.PAG.

Bannister, F., & Janssen, M. (2019). The art of scholarly reviewing: Principles and practices. *Government Information Quarterly*, *36*(1), 1–4. doi:10.1016/j.giq.2018.12.002.

Bashir, M. S., & Khan, M. N. A. (2015). A triage framework for digital forensics. *Computer Fraud & Security*, *2015*(3), 8–18. doi:10.1016/S1361-3723(15)30018-X.

Belghith, O., Skhiri, S., Zitoun, C., & Ferjaoui, S. (2021). *A survey of maturity models in data management*. 298–309. https://doi.org/10.1109/ICMIMT52186.2021.9476197

Bennett, S. (2015). Why information governance needs top-down leadership. *Governance Directions*, *67*(4), 207–212.

Bennett, S. (2017). What is information governance and how does it differ from data governance? *Governance Directions*, *69*(8), 462–467.

Bernardo, B., Barroso, J., & Santos, V. (2022). *Artificial intelligence and digital forensics on data governance breaking through its importance to organizations and its operations*. 3.

Bindley, P. (2019). Joining the dots: How to approach compliance and data governance. *Network Security*, *2019*(2), 14–16. doi:10.1016/S1353-4858(19)30023-6.

Bone, J. (2020). Auditing artificial intelligence: Internal auditors can develop a framework for conducting AI engagements, despite a lack of standards and guidance. *Internal Auditor*, *77*(5), 20–21.

Bordey, G. (2018). Agile in data governance design. *Business Intelligence Journal*, *23*(2), 23–32.

Borek, A., Parlikad, A. K., Woodall, P., & Tomasella, M. (2014). A risk based model for quantifying the impact of information quality. *Computers in Industry*, *65*(2), 354–366. doi:10.1016/j.compind.2013.12.004.

Borgman, C. L. (2018). Open data, grey data, and stewardship: Universities at the privacy frontier. *Berkeley Technology Law Journal*, *33*(2), 365–412. doi:10.15779/Z38B56D489.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*.

Burniston, T. R. (2015). Data governance. *ABA Banking Journal*, *107*(4), 56–57.

Calabrese, J., Esponda, S., & Pesado, P. M. (2020). Framework for data quality evaluation based on ISO/IEC 25012 and ISO/IEC 25024. *VIII conference on cloud computing, big data & emerging topics*. http://sedici.unlp.edu.ar/handle/10915/104778.

Caldwell, P. H., & Bennett, T. (2020). Easy guide to conducting a systematic review. *Journal of Paediatrics and Child Health*, *56*(6). doi:10.1111/jpc.14853 Article 6.

Casey, E. (2019). The chequered past and risky future of digital forensics. *Australian Journal of Forensic Sciences*, *51*(6), 649–664. doi:10.1080/00450618.2018.1554090.

Casino, F., Dasaklis, T., Spathoulas, G., Anagnostopoulos, M., Ghosal, A., Borocz, I., Solanas, A., Conti, M., & Patsakis, C. (2022). Research trends, challenges, and emerging topics in digital forensics: A review of reviews. *IEEE Access*, *10*. doi:10.1109/ACCESS.2022.3154059 1–1.

Cerrillo-Martínez, A., & Casadesús-de-Mingo, A. (2021). Data governance for public transparency. *El Profesional de La Información*, *30*(4), 1–13. doi:10.3145/epi.2021.jul.02.

Chakravorty, R. (2020). Common challenges of data governance. *Journal of Securities Operations & Custody*, *13*(1), 23–43.

Cheong, L., & Chang, V. (2007). The need for data governance: A case study. *ACIS 2007 proceedings - 18th Australasian conference on information systems*.

Clarke, N. (2019). How to ensure provision of accurate data to enhance decision-making. *Journal of Securities Operations & Custody*, *11*(2), 112–127.

Clarke, S. (2016). Reducing the impact of cyberthreats with robust data governance. *Computer Fraud & Security*, *2016*(7), 12–15. doi:10.1016/S1361-3723(16)30053-7.

Coche, E., Kolk, A., & Ocelík, V. (2024). Unravelling cross-country regulatory intricacies of data governance: The relevance of legal insights for digitalization and international business. *Journal of International Business Policy*, *7*(1), 112–127. doi:10.1057/s42214-023-00172-1.

Costantini, S., De Gasperis, G., & Olivieri, R. (2019). Digital forensics and investigations meet artificial intelligence. *Annals of Mathematics & Artificial Intelligence*, *86*(1–3), 193–229. doi:10.1007/s10472-019-09632-y.

Dallemule, L., & Davenport, T. H. (2017). What's your data strategy? *Harvard Business Review*, *95*(3) Article 3.

Dama, I. (2017). *DAMA-DMBOK: data management body of knowledge* (2nd Edition). Technics Publications, LLC.

Deady, R. (2011). Reading with methodological perspective bias: A journey into classic grounded theory. *Grounded Theory Review*, *10*(1) Article 1.

de Campos, E. A. R., Pagani, R. N., Resende, L. M., & Pontes, J. (2018). Construction and qualitative assessment of a bibliographic portfolio using the methodology Methodi Ordinatio. *Scientometrics*, *116*(2). doi:10.1007/s11192-018-2798-3 Article 2.

Demarquet, G. (2016). Five key reasons enterprise data governance matters to finance ... and seven best practices to get you there. *Journal of Corporate Accounting & Finance (Wiley)*, *27*(2), 47–51. doi:10.1002/jcaf.22121.

de Moya-Anegón, F., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Corera-Álvarez, E., Muñoz-Fernández, F., González-Molina, A., & Herrero-Solana, V. (2007). Coverage analysis of Scopus: A journal metric approach. *Scientometrics*, *73*(1), 53–78.

Dencik, L., Hintz, A., Redden, J., & Warne, H. (2019). Data scores as governance. *Information Polity: The International Journal of Government & Democracy in the Information Age*, *24*(1), 111–114. doi:10.3233/IP-190002.

Denoncourt, J. (2020). Companies and UN 2030 sustainable development goal 9 industry, innovation and infrastructure. *Journal of Corporate Law Studies*, *20*(1), 199–235. doi:10.1080/14735970.2019.1652027.

Dighe, S. (2014). Commanding data governance. *Banker Middle East*, *163*, 68–70.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285–296. doi:10.1016/j.jbusres.2021.04.070.

Dutta, A. (2016). Ensuring the quality of data in motion: The missing link in data governance. *Computer Weekly*, 1–4.

Ehrlinger, L., & Wöß, W. (2022). A survey of data quality measurement and monitoring tools. *Frontiers in Big Data*, *5*. https://www.frontiersin.org/articles/10.3389/fdata.2022.850611.

Ellegaard, O., & Wallin, J. A. (2015). The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics*, *105*(3), 1809–1831. doi:10.1007/s11192-015-1645-z.

Elyas, M., Maynard, S. B., Ahmad, A., & Lonie, A. (2014). Towards a systemic framework for digital forensic readiness. *Journal of Computer Information Systems*, *54*(3), 97–105. doi:10.1080/08874417.2014.11645708.

Englbrecht, L., Meier, S., & Pernul, G. (2020). Towards a capability maturity model for digital forensic readiness. *Wireless Networks (10220038)*, *26*(7), 4895–4907. doi:10.1007/s11276-018-01920-5.

Firican, G. (2018). IBM data governance maturity model. *LightsOnData. https://www.lightsondata.com/data-governance-maturity-models-ibm/*.

Fischer, B., & Piskorz-Ryń, A. (2021). Artificial intelligence in the context of data governance. *International Review of Law, Computers & Technology*, 1–10. doi:10.1080/13600869.2021.1950925.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. doi:10.1016/j.ijinfomgt.2014.10.007.

Gardner, K., Olney, S., & Dickinson, H. (2018). Getting smarter with data: Understanding tensions in the use of data in assurance and improvement-oriented performance management systems to improve their implementation. *Health Research Policy & Systems*, *16*(1). doi:10.1186/s12961-018-0401-2 N.PAG-N.PAG.

Gaviria-Marin, M., Merigo, J. M., & Popa, S. (2018). Twenty years of the journal of knowledge management: A bibliometric analysis. *Journal of Knowledge Management*, *22*(8), 1655–1687. doi:10.1108/JKM-10-2017-0497.

George, R., Truong, T., & Davidson, J. (2017). Establishing an effective data governance system: Data governance is necessary for compliance with current regulatory expectations for data integrity in pharmaceutical R&D and manufacturing organizations. *Pharmaceutical Technology Europe*, *29*(11), 40–43.

Gold, S. (2014). Challenges ahead on the digital forensics and audit trails. *Network Security*, *2014*(6), 12–17. doi:10.1016/S1353-4858(14)70060-1.

Grobler, M. (2010, March 16). *Developing digital forensic governance*.

Grobler, M. (2010). Managing digital evidence—The governance of digital forensics. *Journal of Contemporary Management*, *7*.

Gupta, N., Blair, S., & Nicholas, R. (2020). What we see, what we don't see: Data governance, archaeological spatial databases and the rights of indigenous peoples in an age of big data. *Journal of Field Archaeology*, *45*, S39–S50. doi:10.1080/00934690.2020.1713969.

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. *Campbell Systematic Reviews*, *18*(2), e1230. doi:10.1002/cl2.1230.

Harper, J. (2020). The centerpiece of data governance: Making information quality pay off. *KM World*, *29*(4), 6–8.

Hassan, S., & Chindamo, P. (2017). Effective data governance: From strategy through to implementation. *Governance Directions*, *69*(4), 207–210.

Haug, A. (2021). Understanding the differences across data quality classifications: A literature review and guidelines for future research. *Industrial Management and Data Systems*, *121*(12), 2651–2671. doi:10.1108/IMDS-12-2020-0756.

Hay, J. (2015). Data governance gamification. *Business Intelligence Journal*, 21–26.

Hikmawati, S., Santosa, P., & Hidayah, I. (2021). Improving data quality and data governance using master data management: A review. *IJITEE (International Journal of Information Technology and Electrical Engineering)*, *5*, 90. doi:10.22146/ijitee.66307.

Hong, Q. N., & Pluye, P. (2018). Systematic reviews: A brief historical overview. *Education for Information*, *34*(4). doi:10.3233/EFI-180219 Article 4.

Hoppszallern, S. (2015). Governance strategies can determine success of IT projects. *H&HN: Hospitals & Health Networks*, *89*(4) 20–20.

Hubbard, D., Freda, A., & Swanagan, A. (2020). Data governance 101: IR's critical role in data governance. *New Directions for Institutional Research*, *185–186*, 51–65.

Idri, N. (2015). Zotero software: A means of bibliographic research and data organisation; teaching bibliographic research. *SSRN Electronic Journal*. doi:10.2139/ssrn.2843984.

International Organization for Standardization. (2008). *ISO/IEC 25012:2008: software engineering—software product quality requirements and evaluation (SQuaRE)—data quality model*. Geneva: International Organization for Standardization.

International Organization for Standardization. (2017). *ISO/IEC 38505-1:2017—information technology—governance of IT — governance of data—Part 1: Application of ISO/IEC 38500 to the governance of data*.

International Organization for Standardization. (2018). *ISO/IEC TR 38505-2:2018—information technology—governance of IT — governance of data—Part 2: Implications of ISO/IEC 38505-1 for data management*.

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

Jiang, Z., Zhu, Z., & Pan, S. (2024). Data governance in multimodal behavioral research. *Multimodal Technologies and Interaction*, *8*(7). doi:10.3390/mti8070055 Article 7.

Jiya, T. (2021). Responsible data governance in projects: Applying a responsible research and innovation (RRI) framework. *Journal of Technology Management & Innovation*, *16*(1), 31–36. doi:10.4067/s0718-27242021000100031.

Johnson, C. (2015). Managing cross-regulatory data challenges in practice. *Journal of Securities Operations & Custody*, *7*(4), 284–295.

Johnston, M. (2016). Are you prepared for your next data disaster? *NetworkWorld Asia*, *13*(2) 52–52.

Karie, N. M., & Venter, H. S. (2015). Taxonomy of challenges for digital forensics. *Journal of Forensic Sciences*, *60*(4), 885–893. doi:10.1111/1556-4029.12809.

Kent, K., Chevalier, S., Grance, T., & Dang, H. (2006). Guide to integrating forensic techniques into incident response *(NIST special publication (SP) 800-86)*. National Institute of Standards and Technology. doi:10.6028/NIST.SP.800-86.

Koltay, T. (2016). Data governance, data literacy and the management of data quality. *IFLA Journal*, *42*(4), 303–312.

Kopp, S. (2020). Guideline on data integrity. *WHO Drug Information*, *34*(2), 231–254.

Kraljić, A., Kraljić, T., Poels, G., & Devos, J. (2014). ERP implementation methodologies and frameworks: A literature review. In *Proceedings of the European conference on information management & evaluation* (pp. 309–316).

Lancaster, J., Ledford, L., & Stephens, J. (2019). Structure your data governance. *Business Officer*, *52* 19–19.

Lee, P. A. (2019). Why data governance should be part of your boardroom conversations. *NACD Directorship*, *45*(6) 68–68.

Levac, D., Colquhoun, H., & O'Brien, K. K (2010). Scoping studies: Advancing the methodology. *Implementation Science*, *5*(1) Article 1.

Lo, Q.-Q., & Chai, K.-H. (2012). Quantitative analysis of quality management literature published in total quality management and business excellence (1996–2010). *Total Quality Management & Business Excellence*, *23*(5/6). doi:10.1080/14783363.2012.669553 Article 5/6.

Lutui, R. (2016). A multidisciplinary digital forensic investigation process model. *Business Horizons*, *59*(6), 593–604. doi:10.1016/j.bushor.2016.08.001.

Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods and methodology. *Issues in Educational Research*, *16*(2) Article 2.

Major, L. (2010). *Systematic literature review protocol: Teaching novices programming using robots*. 11.

Mañana-Rodríguez, J. (2015). A critical review of SCImago journal & country rank. *Research Evaluation*, *24*(4), 343–354. doi:10.1093/reseval/rvu008.

Mansfield-Devine, S. (2017). Data governance: Going beyond compliance. *Computer Fraud & Security*, *2017*(6), 12–15. doi:10.1016/S1361-3723(17)30052-0.

Manus, S. (2019). Data vision vs. data strategy: Why credit unions need both: Learn why locating your CU's North Star - the guiding light that will inform its data analytics strategy - is so crucial. *Credit Union Times*, *30*(43), 1–4.

Mao, Z., Wu, J., Qiao, Y., & Yao, H. (2021). Government data governance framework based on a data middle platform. *Aslib Journal of Information Management*, *74*(2), 289–310. doi:10.1108/AJIM-03-2021-0068.

Martini, B., & Choo, K.-K. R. (2012). An integrated conceptual digital forensic framework for cloud computing. *Digital Investigation*, *9*(2), 71–80. doi:10.1016/j.diin.2012.07.001.

McDowall, R. D. (2017). Understanding data governance, Part II. *Spectroscopy*, *32*(4), 12–18.

McDowall, R. D. (2017). Understanding data governance, Part I. *Spectroscopy*, *32*(2), 32–38.

Mcintyre, J. (2016). Data governance and protection. *New Hampshire Business Review*, *38*(2) 24–24.

Merigó, J. M., & Yang, J.-B. (2017). A bibliometric analysis of operations research and management science. *Omega*, *73*, 37–48. doi:10.1016/j.omega.2016.12.004.

Meyers, C. (2014). How data management and governance can enable successful self-service BI. *Business Intelligence Journal*, *19*(4), 23–27.

Miernicki, M., & Ng (Huang Ying), I. (2021). Artificial intelligence and moral rights. *AI & Society*, *36*(1), 319–329. doi:10.1007/s00146-020-01027-6.

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2020). The role of information governance in big data analytics driven innovation. *Information & Management*, *57*(7). doi:10.1016/j.im.2020.103361 N.PAG-N.PAG.

Milne, R., & Brayne, C. (2020). We need to think about data governance for dementia research in a digital era. *Alzheimer's Research & Therapy*, *12*(1), 1–3. doi:10.1186/s13195-020-0584-y.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group*, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264–269.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., , & PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*(1), 1. doi:10.1186/2046-4053-4-1.

Mouhtaropoulos, A., Chang-Tsun Li, & Grobler, M. (2014). Digital forensic readiness: Are we there yet? *Journal of International Commercial Law & Technology*, *9*(3), 173–179.

Naumann, F. (2014). Data profiling revisited. *ACM SIGMOD Record*, *42*, 40–49. doi:10.1145/2590989.2590995.

Niazi, M. (2015). Do systematic literature reviews outperform informal literature reviews in the software engineering domain? an initial case study. *Arabian Journal for Science & Engineering (Springer Science & Business Media B.V.)*, *40*(3). doi:10.1007/s13369-015-1586-0 Article 3.

Okoli, C., & Schabram, K. (2010). *A guide to conducting a systematic literature review of information systems research*.

Okoro, R. (2021). Proposed data governance framework for small and medium scale enterprises (SMEs). *All graduate theses, dissertations, and other capstone projects*. https://cornerstone.lib.mnsu.edu/etds/1126.

Onwuegbuzie, A. J., & Frels, R. K. (2015). Using Q methodology in the literature review process: A mixed research approach. *Journal of Educational Issues*, *1*(2) Article 2.

Otto, B. (2011). Organizing data governance: Findings from the telecommunications industry and consequences for large service providers. *Communications of the Association for Information Systems*, *29*(1). doi:10.17705/1CAIS.02903.

Palmatier, R. W., Houston, M. B., & Hulland, J. (2018). Review articles: Purpose, process, and structure. *Journal of the Academy of Marketing Science*, *46*(1), 1–5. doi:10.1007/s11747-017-0563-4.

Paredes, D. (2016). The six steps to become a successful CDO. *CIO (13284045)*, 4–4.

Perrin, A. (2020). Data governance in the age of programmatic advertising. *Information Today*, *37*(2) 35–35.

Petzold, B., Roggendorf, M., Rowshankish, K., & Sporleder, C. (2020). Designing data governance that delivers value. *McKinsey Insights* N.PAG-N.PAG.

R, S. O., Handayani, P. W., & Hidayanto, A. N. (2024). Health organization challenges in health data governance implementation: A systematic review. *Journal of Infrastructure, Policy and Development*, *8*(6). doi:10.24294/jipd.v8i6.3892 Article 6.

Rafique, I., Lew, P., Abbasi, M. Q., & Li, Z. (2012). Information quality evaluation framework: extending ISO 25012 data quality model. *World Academy of Science, Engineering and Technology - International Journal of Computer and Information Engineering*, *6*(5), 568–573. doi:10.5281/zenodo.1072956.

Ragan, R. K., & Strasser, T.'Buck' (2020). Understanding the how and why of CU data governance. *Credit Union Times*, *31*(10) 10–10.

Ridzuan, F., & Wan Zainon, W. M. N. (2019). A review on data cleansing methods for big data. *Procedia Computer Science*, *161*, 731–738. doi:10.1016/j.procs.2019.11.177.

Riggins, F. J., & Klamm, B. K. (2017). Data governance case at KrauseMcMahon LLP in an era of self-service BI and Big Data. *Journal of Accounting Education*, *38*, 23–36. doi:10.1016/j.jaccedu.2016.12.002.

Rivett, P. (2015). Are you creating a data swamp? *KM World*, *24*(3) S5–S5.

Roldan-Valadez, E., Salazar-Ruiz, S. Y., Ibarra-Contreras, R., & Rios, C. (2019). Current concepts on bibliometrics: A brief review about impact factor, Eigenfactor score, CiteScore, SCImago Journal Rank, Source-Normalised Impact per Paper, H-index, and alternative metrics. *Irish Journal of Medical Science (1971 -)*, *188*(3), 939–951. doi:10.1007/s11845-018-1936-5.

Russom, P. (2015). TDWI technology survey: The state of data governance. *Business Intelligence Journal*, *20*(2) 56–56.

Sánchez, M. C., & Sarría-Santamera, A. (2019). Unlocking data: Where is the key? *Bioethics*, *33*(3), 367–376. doi:10.1111/bioe.12565.

Saporito, P. (2019). The data divide: Data ethics and data governance need to be part of every employee's onboarding, highlighting their responsibility along the supply chain. *Best's Review*, *120*(4) 23–23.

Schmuck, M. (2024). Cultivating data observability as the next frontier of data engineering: A path to enhanced data quality, transparency, and data governance in the digital age. *Journal of Public Administration, Finance and Law*, *30*, 212–224. doi:10.47743/jopafl-2023-30-19.

Seerden, X., Salmela, H., & Rutkowski, A.-F. (2018). Privacy governance and the GDPR: How are organizations taking action to comply with the new privacy regulations

in Europe? In *Proceedings of the European conference on management, leadership & governance* (pp. 371–378).
Serrano, J. Y., & Zorrilla, M. (2021). A Data governance framework for industry 4.0. *IEEE Latin America Transactions*, *19*(12) Article 12.
Shabani, M. (2021). The data governance act and the EU's move towards facilitating data sharing. *Molecular Systems Biology*, *17*(3), 1–3. doi:10.15252/msb.202110229.
Sifter, C. J. (2017). Establishing a data governance center of excellence within your bank. *New Jersey Banker*, 24–25.
Smallbone, T., & Quinton, S. (2011). *A three-stage framework for teaching literature reviews: A new approach*. https://doi.org/10.3794/IJME.94.337
Smith, A. M. (2016). Seven best practices to boost big data governance efforts. *Computer Weekly*, 3–5.
Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. doi:10.1016/j.jbusres.2019.07.039.
Sohrabi, A., Sabahi, A., Garavand, A., & Ahmadian, L. (2022). Data validation techniques used in admission discharge and transfer systems: Necessity of use and effect on data quality. *Informatics in Medicine Unlocked*, *34*, 101122. doi:10.1016/j.imu.2022.101122.
Sothilingam, R., Pant, V., Shahrin, N., & Yu, E. S. K. (2021). Towards a goal-oriented modeling approach for data governance. In Proceedings of the forum at practice of enterprise modeling 2021 (PoEM-forum 2021) (PoEM 2021), Riga, Latvia, November 24-26, 2021, 3045 (pp. 69–77).
Stratton, J. (2014). Carefully communicate performance metrics. *Board & Administrator: For Administrators Only*, *31*(4), 1–4.
Sucha, M. (2014). Beyond the hype: Data management and data governance. *Feliciter*, *60*(2) Article 2.
Swoyer, S. (2016). IT's evolving role in data governance. *Business Intelligence Journal*, *21*(1), 49–55.
Taleb, I., Serhani, M. A., Bouhaddioui, C., & Dssouli, R. (2021). Big data quality framework: A holistic approach to continuous quality management. *Journal of Big Data*, *8*(1), 76. doi:10.1186/s40537-021-00468-0.
Tankard, C. (2015). Data classification – the foundation of information security. *Network Security*, *2015*(5), 8–11. doi:10.1016/S1353-4858(15)30038-6.
Tassone, C. F. R., Martini, B., & Choo, K. R. (2017). Visualizing digital forensic datasets: A proof of concept. *Journal of Forensic Sciences*, *62*(5), 1197–1204. doi:10.1111/1556-4029.13431.
United Nations, Washington, DC. (2020). *Industry, innovation and infrastructure: Why it matters?* https://www.un.org/sustainabledevelopment/wp-content/uploads/2019/07/9_Why-It-Matters-2020.pdf
Valdez, B. (2018). Spotlight on a discipline: Forensics. *International Social Science Review*, *94*(2), 1–6.
Van Eck, N. J., & Waltman, L. (2011). Text mining and visualization using VOSviewer. *ISSI Newsletter*, *7*(3), 50–54.
Van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. *Measuring scholarly impact* (pp. 285–320). Springer.
Vilminko-Heikkinen, R., & Pekkola, S. (2019). Changes in roles, responsibilities and ownership in organizing master data management. *International Journal of Information Management*, *47*, 76–87.
Vincze, E. A. (2016). Challenges in digital forensics. *Police Practice & Research*, *17*(2), 183–194. doi:10.1080/15614263.2015.1128163.
Walstrom, K. A., & Leonard, L. N. K. (2000). Citation classics from the information systems literature. *Information & Management*, *38*(2). doi:10.1016/S0378-7206(00)00054-9 Article 2.
Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5–33. doi:10.1080/07421222.1996.11518099.
Weber, K., Otto, B., & Oesterle, H. (2009). One size does not fit all—a contingency approach to data governance. *ACM Journal of Data and Information Quality*, *1*. doi:10.1145/1515693.1515696 Article 4.
Weber, S., Beck, R., & Gregory, R. W. (2012). Combining design science and design research perspectives—findings of three prototyping projects. *2012 45th Hawaii international conference on system sciences* (pp. 4092–4101). doi:10.1109/HICSS.2012.163.
Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1). doi:10.1038/sdata.2016.18 Article 1.
Winning, M. A., & Beverley, C. A. (2003). Clinical librarianship: A systematic review of the literature. *Health Information & Libraries Journal*, *20*, 10–21. doi:10.1046/j.1365-2532.20.s1.2.x.
Zhang, P., Zhao, K., & Kumar, R. L. (2016). Impact of IT governance and IT capability on firm performance. *Information Systems Management*, *33*(4), 357–373. doi:10.1080/10580530.2016.1220218.
Zorrilla, M., & Yebenes, J. (2022). A reference framework for the implementation of data governance systems for industry 4.0. *Computer Standards & Interfaces*, *81*, 103595. doi:10.1016/j.csi.2021.103595.
Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, *18*(3), 429–472. doi:10.1177/1094428114562629.