Available online at www.ijournalse.org

**Emerging Science Journal**

(ISSN: 2610-9182)

Vol. 10, No. 1, February, 2026

# Data Governance Meets Generative Artificial Intelligence: Towards A Unified Organizational Framework

Bruno M. V. Bernardo [1*], Henrique S. Mamede [2], João M. P. Barroso [1], Mijail Naranjo-Zolotov [3], Vítor M. P. Duarte dos Santos [3]

[1] *INESC TEC and Universidade de Trás-os-Montes e Alto Douro, Portugal.*

[2] *INESC TEC and Universidade Aberta, Portugal.*

[3] *NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Portugal.*

**Abstract**

As technology continues to evolve, organizations face growing and complex challenges and opportunities that affect their ability to govern, manage and harness data as a key source of competitive advantage. Equally, data are considered a powerful and unique source of success for organizations, which in turn, can impact their decision-making capabilities and play a critical role in their success. Hence, this article aims to provide a detailed identification, analysis and discussion over the current data governance context and its existing frameworks, highlighting their commonalities, differences and gaps, including ones related to data governance relationship with Generative Artificial Intelligence (GenAI). This article conducts an extensive methodological and in-depth analysis over a set of sixteen data governance frameworks based on different key data governance attributes, denoting that although there are numerous frameworks, they hold weaknesses, limitations and challenges which prevent them from being capable of incorporating and governing the use and management of AI, particularly the demands originating from GenAI. Our findings provide and propose a new and enhanced data governance framework which integrates the best features and ideas from the existing ones and initiatives derived from the advancements and particularities of AI and GenAI models, systems, and overall usage.

**Keywords:**

Data Governance;
Data Governance Frameworks and Tools;
Artificial Intelligence;
Generative Artificial Intelligence;
Data Management and Quality;
Data Assurance.

**Article History:**

| | | | |
|---|---|---|---|
| **Received:** | 28 | September | 2025 |
| **Revised:** | 22 | January | 2026 |
| **Accepted:** | 28 | January | 2026 |
| **Published:** | 01 | February | 2026 |

## 1- Introduction

Throughout the years, data have been considered as one of the terms that outline our age, where data are at the center of almost every enterprise activity, being gradually recognized as a key asset and no longer just "bits of information". Data encompasses a tremendous amount of possible finalities, from allowing business services and development to establishing the key building blocks of both public and private organizations [1, 2]. Besides, it is clear that data governance is becoming highly relevant due to the ever-increasing daily amount of data that is generated both internally and externally through more and more complex IT systems and infrastructure [3]. Consequently, organizations' ability to make robust decisions and sustain the competitive markets is determined by its capacity to develop data governance and management activities [4]. Hence, literature has concluded that the high volume of different data is originating data inaccuracies and inconsistencies that, if not correctly and timely addressed, can jeopardize a firm's strategy [3].

Besides, organizations must acknowledge the potential consequences from defining and executing a robust and efficient data governance framework, which can positively impact their capacity to achieve success and make decisions

* **CONTACT**: al75188@alunos.utad.pt

**DOI:** http://dx.doi.org/10.28991/ESJ-2026-010-01-029

[5]. In this context, data governance's significance is gradually increasing within the scientific community, being recently considered an emergent topic, as organizations are acknowledging the need for more knowledge on this field [6]. Literature defines the data governance field as a complex function that maximizes the value of data and minimizes the data-related risks and harm [7]. This field encompasses the management of data (unstructured, semi-structured, and structured) and related assets through a cross-functional framework, enabling the organization to establish a strategy, vision, and plan, enforce policies, procedures, and standards, and define roles and responsibilities [8]. Even considering organizations' data algorithms to perform data analysis and operational processes, it is undeniable that this field is one of the most fundamental approaches to govern and monitor those analysis and their supporting algorithms [9].

Likewise, with the advancement of time, technology and data, data governance is one of the major areas where more fundamental and differentiated advancements have taken place [10]. These enhancements have significantly impacted the way data have been and should be governed to ensure that data's attributes (precious, completeness, timeliness, availability, integrity, and consistency, among others) are maintained over time. Although trends indicate that the motivation to adopt data governance has not been through self-organizational proactive initiatives, this field has been used as a means to comply with the legal paradigm and manage data as the business progresses. Nonetheless, it is clear that any data governance framework should prompt organizations to improve and monitor the results of such activities, enabling the governance of both transactional and data at rest [3, 10, 11]. Hence, data governance practices are considered relatively new which organizations have yet to fully incorporate into their context. Literature has also shown that data governance and technology should not be analyzed separately, as the first is a cornerstone of the Industry 4.0, described as the Fourth stage of the Industrial Revolution [12]. This revolution is characterized by a rapid shift in how companies' approach their data assets, characterized by real-time decisions, Industrial Internet of Things (IIoT), including cyber and physical systems. As a result, as data and their environment of people, processes, and systems become increasingly interconnected, there is a recognized need for organizations to consider their data as one of their primary assets [12].

Considering the Artificial Intelligence (AI) field, literature has concluded that the rapid integration of AI across industries has contributed to a greater need for effective data governance practices, namely to ensure that AI's transparency and accountability, is entirely covered by a given data governance framework [13]. Baum et al. [14] emphasize that research scope and efforts around AI must focus on the data governance and auditing aspects of this field, including considering how AI techniques can be managed and governed throughout a given AI lifecycle. Also, literature acknowledges the emerging topic of AI governance, which is widely recognized and analyzed, primarily in relation to the ethical and legal requirements of AI systems. Nonetheless, literature denotes a notable gap yet to be filled, which relates to how AI should be governed within an organization's data governance, where their frameworks are either forgetting to extend their data governance initiatives to the AI systems and their requirements or focusing on AI as an object of regulation and not as an active approach [15]. This gap becomes particularly more concerning because the majority of AI-driven models being developed and deployed are often unclear in regards to their training dataset sources, and their algorithmic outputs are not provided with a transparent and clear explainability [15].

Furthermore, the literature reflects on the data governance field as one that currently faces complex challenges, leading into studies that focus ultimately on identifying data governance challenges [16–19], among others, concluding that data governance challenges are rising, namely due to the fast-paced technology innovation, including the usage of AI and GenAI systems and the increase in the data flows, volume, and complexity. Table 1 illustrates the key results of the above-mentioned studies regarding data governance challenges:

**Table 1. Data Governance (DG) Main Challenges – Overview of Literature Findings**

| # | Challenge Description | References |
|---|---|---|
| 1 | Lack of solid data management frameworks, including GenAI requirements and initiates to address those | [20–22] |
| 2 | The absence of unified, robust, and comprehensive data governance frameworks | [8, 21, 23] |
| 3 | Insufficient basis of knowledge and maturity in data governance practices | [2, 3, 11] |
| 4 | Limited to non-existent awareness of the role of leadership in a data governance framework | [19, 11, 24] |
| 5 | Significant cultural and motivational resistance to change management among stakeholders and leadership | [6, 25] |
| 6 | Proliferation of data silos and isolated data points within Organizations' social networks | [23, 26] |
| 7 | Rapid technological advancements and innovations creating governance challenges | [25, 27] |
| 8 | Need for strong data protection measures and supportive legislation | [2, 28–30] |
| 9 | Increased interaction and usage within the IoT ecosystem and cloud services | [21, 27] |
| 10 | Lack of auditing culture, policies, and mechanisms | [1, 6, 25, 26] |
| 11 | Insufficient guidance on establishing performance metrics and indicators | [1, 6, 19, 25] |

Given these challenges, we conclude that data governance capacity to integrate the GenAI specifications currently faces strong limitations, and thus our research is a vital tool for organizations and practitioners, and that it should be performed extensively, as to fulfil the current gap that is yet to be addressed, namely regarding the inexistence of an up-to-date, robust, scientifically analyzed and validated data governance framework [2, 6, 27].

### *1-1- Related Studies*

In this section, we provide an overview of previous research on data governance-related literature, specifically those that focus on identifying and describing data governance concepts, challenges, and frameworks and that establish a bridge between data governance and GenAI. Through this research, a total of 17 articles were identified and analyzed, where i) the overall articles focused on presenting a general or context-specific data governance framework, i.e., an extension of the general framework to a specific area or data feature, for example, Big Data, SMEs, Digital Platforms, among others, ii) only one article, Al-Ruithe et al. [31], performed a comparison between two frameworks, and iii) none included the features and integration of GenAI within data governance. Nonetheless, the analysis of Al-Ruithe et al. [31] focused on performing a systematic literature review on data governance and cloud data governance, mainly on the IBM and on the Data Governance Institute (DGI) data governance frameworks [31]. Despite this article being the closest to our work, particularly in presenting frameworks and describing their challenges and factors, it only includes a total of two, whereas our paper includes sixteen. Also, it does not conduct a critical analysis of the key characteristics of frameworks, including how they distinguish themselves or how a specific framework can incorporate the features of GenAI and establish initiatives to govern those systems.

Moreover, in Table 2, we present the key indicators on the seventeen articles analyzed, identifying that we did not find any publication that expansively acknowledges all the domains analyzed within our paper, namely the identification of i) the general data governance concepts and the challenges, ii) and iii) the presentation and description of a data governance framework, including general and context-specific, iv) the presentation of a comparison between data governance frameworks, vi) the definition of the data roles and stewardship, vi) the identification of data practices, tools and applications and vii) the inclusion of initiatives to govern and manage the GenAI usage and its systems. Therefore, Table 2 shows that our article proposal is novel, aiming to present a validated unified data governance framework that has not been provided in any literature.

**Table 2. Paper analysis and framework identification resulting from the review of relevant literature on data governance**

| Research Article Area of Focus – Data Governance Frameworks | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Articles Included** | **DG Concepts & Challenges** | **One DG General Framework** | **One DG Context-Specific Framework** | **DG Frameworks Comparison** | **DG Roles & Stewardship** | **DG Tools & Applications** | **Generative AI Use & Integration** | **Period (Published up to)** |
| [32] | ✔ | ✔ | | | | | | 2019 |
| [33] | ✔ | ✔ | | | ✔ | | | 2016 |
| [31] | ✔ | ✔ | | ✔ | ✔ | ✔ | | 2019 |
| [10] | ✔ | | ✔ | | | | | 2016 |
| [34] | ✔ | ✔ | | | | ✔ | | 2015 |
| [23] | ✔ | | ✔ | | ✔ | | | 2020 |
| [35] | ✔ | ✔ | | | ✔ | | | 2010 |
| [8] | ✔ | | ✔ | | ✔ | | | 2023 |
| [36] | ✔ | ✔ | | | ✔ | | | 2018 |
| [37] | ✔ | | ✔ | | ✔ | | | 2021 |
| [11] | ✔ | | ✔ | | ✔ | | | 2019 |
| [38] | ✔ | | ✔ | | ✔ | | | 2021 |
| [12, 39] | ✔ | | ✔ | | | | | 2022 |
| [26] | ✔ | ✔ | | | | | | 2017 |
| [40] | ✔ | | ✔ | | | ✔ | | 2019 |
| [41] | ✔ | | ✔ | | | | | 2019 |
| **Present Research** | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ | **Up-to-date (2025)** |

### *1-2- Study Objectives and Contributions*

Our article aims to address the solution to the existing gap that data governance is currently facing, emphasized by the analyzed articles, which reflects that more research is needed in data governance and that information on its frameworks comparison and key criteria is still scarce and needs to be further developed [8, 31, 32]. As a result, our work aims at making a comparison between data governance frameworks, identifying how they differ from each other and acknowledging the different criteria that organizations should consider so that a unified framework can be designated considering the best features of each of the existing data governance frameworks and including GenAI in their governance initiatives requirements. What's more, our analysis revealed that most articles in the literature discuss general topics on data governance, including its concepts (17/17 articles), primary challenges (12/17 articles), roles and stewardship (9/17 articles). Also, these articles presented a total of 16 frameworks, grouped into two categories: general frameworks (7/17 articles) and context-specific (10/17 articles). More in-depth essential topics, such as i) frameworks comparison (1/17 articles) and ii) tools and applications (3/17 articles), were not thoroughly explored. Excluding the outlier framework in the date context, presented in 2010 by Khatri & Brown [35], the remaining 16 articles, averaged a publication date of around 2018.9, where 68.75 % contains a publication date before or equal to 2020, and the remaining 31.25 % are dated after 2020. Even so, we verified that the five most recent articles presented one framework that is context-specific rather than general context, namely Big Data, Ethics, Small- and Medium-Enterprises (SMEs), and Digital Platforms, which highlighted the fact that literature is still yet to reach a data governance unified framework which can in turn be applied in different and specifics contexts [8, 37, 38].

Moreover, there is a dearth of study on data governance framework theories, practical methods, and tools, which, aligned with the continuous improvements, dependence and complexity of the technology and information assets, is considered to be generating low levels of data quality and trust, incapacitating data sharing activities and minimizing data value [27]. Hence, it can be concluded that the current knowledge on data governance is still in its initial stages [6]. Macfeely et al. [2] outlined that the "Dataverse" literature, which results from the continuously expansion of the global data ecosystem, driven by technology enhancements, data value, and volume, is currently acknowledging the urgent need for some new and enhanced data governance frameworks [2]. Although there are existing frameworks, there is still a lack of precise definitions, which limits the practical applicability in real-world applications and scenarios. Besides, literature highlights that there is a need for more analysis and evaluation of whether the data governance frameworks that currently exist are a reasonable basis for an enterprise application or if there is a need for an additional one [3]. Thus, this study aims to describe and compare existing frameworks and synthesize a new, improved model that integrates the best features of each, considering frameworks that were extended to fit a general and specific purpose.

In line, the literature considers that much work must be done within this field, namely regarding its integration of AI techniques and advancements, stating that "what is needed is not a pause in development, but governance" to ensure that AI models are employed in a way that is constructive, sustainable and positive [14]. For Lam et al. [22], it is vital that literature addresses data governance issues and provides solutions on how organizations can govern their digital transformation. Hence, data governance and its relation with AI must ensure that the overall impact of a given AI model is managed, known, and monitored and that its initiatives and risks are continuously assessed and mitigated [14]. Similarly, a lack of a data governance framework that does not cover the AI systems can heavily impact and influence the retrieval of value [42]. Hence, data governance and AI systems should have a framework founded on a human-centric approach that takes into consideration both data processing practices and human rights standards, ensuring data principles such as impartiality, security, accountability, interoperability, and explainability. Besides, for Chen [43] while GenAI in a given framework should have a relevant focus, the current data governance landscape still lacks controls and mechanisms to monitor the challenges of GenAI. Considering this, we aim to answer the following question:

*"How can a novel and enhanced data governance framework be developed through a comparative and integrative synthesis that combines the best strengths and features of existing frameworks and integrates Generative AI-specific requirements and complexities?"*

This main research question is supported by three sub-research questions (RQ), namely:

- **RQ1**: What are the primary domains, characteristics, and features of existing data governance frameworks, and how do they differ from one another?
- **RQ2**: What are the key challenges and risks associated with GenAI and its integration in a data government framework context?
- **RQ3**: What set of initiatives can a novel and enhanced data governance framework include to address the challenges, specifications and opportunities that GenAI presents?

In addition, our paper is organized through different but connected Sections, where Section 2 presents the theoretical background of data governance, artificial intelligence, and GenAI, while analyzing and discussing key concepts, benefits and challenges related to data governance. Section 3 details the research methodology and planning, outlining the phases of our study. Section 4 provides an identification and comparative analysis of existing data governance frameworks and initiatives, analyzing the specification of each of the frameworks in scope of our study. In Section 5, we present the Unified Data Governance Framework, its key attributes, specifications and how it defines a bridge between data governance and generative AI, and a set of GenAI initiatives which can be performed to enhance any data governance framework. Finally, Sections 6 and 7 summarize the main findings and contributions of the research and highlights the outcomes obtained through the focus group activity performed and directions for future work.

## 2- Theoretical Background

This section explores the literature conducted on data governance, its characteristics, benefits, challenges, and frameworks while examining its overall relationship with the field of AI, specifically Generative Artificial Intelligence.

### *2-1- Data Governance*

While various definitions of data governance exist, there is a consensus that it encompasses the overarching practice of managing, controlling, and overseeing data through a network of interconnected policies, procedures, and authorities, guiding the different stakeholders and their data actions, from data collection, analysis, and sharing to security, storage, and reporting [42]. Similarly, Scholz et al. [3] defines the overall concept as the set of activities that support oversight and management of an organization's operations, business processes, data, and data assets. The Data Management Association, DAMA, defines data governance as the practice that encompasses the overall arrangement of responsibilities, governance and its rights, internal control, and robust decision-making processes that characterize the maturity of which organizations can manage, govern and monitor their data assets [32]. In simple terms, data governance aims at allowing organizations to leverage high-quality data to make robust decisions and to support the short- and long-term viability of their strategy while ensuring that data-related risks are residually mitigated [3]. To do so, data governance includes not only the people and processes but also the procedures and technologies that gather, explore, store, and report on data, being a never-ending process [40]. Commonly, one of data governance's key features is the definition of a policy, which is a cornerstone encompassing relevant domains such as data roles and responsibilities, data protection and quality. Aligned with this, is the need for a strategy, a document or a set that outlines how information is produced and managed to deliver the planned goals and objectives [40]. For these practices, it is necessary to outline them within a framework. Such framework is the design roadmap that supports the organization in the definition of data roles and responsibilities, a data driven culture, data governance strategy and policy, data quality standards, data lineage, training requirements, amongst others [40]. Meyers [44] stated that such framework should be a systematic approach that involves defining mechanisms to oversee and manage the data a given organization interacts with through formally supported processes and procedures compliant with internal and external guidelines and regulations. Literature believes that the more organizations rely on technology to support their operations and to manage and use data, the more critical role, a clearly defined framework can be in leading to operational efficiency and robust choices [5].

Accordingly, literature emphasizes important notions of data governance including the data lake concept, which serves as a central repository of information that enables firms to identify, store, and distribute large volumes of data, and accommodating contextual information and metadata of data, including details on their purpose, quality standards, formats, traceability, and system dependencies, among others [10]. Also, critical notion is the data lineage and its formalization, where Bordey [45] proposed that a firm should periodically perform a set of questions regarding its data quality maturity, namely i) does the firm know the data activities being performed and the roles and responsibilities established to manage them?; ii) Which conditions and obligations are required for an organization to plan and perform the data-related functions and activities effectively; iii) Where are the organization's data stored within its data-related assets? iv) What is the overall goal and result expected from the data utilization, and to whom should it be distributed?; and v) Which follow-up questions can take place after data processing, delivering, and storing?. Besides, literature identified three key concepts, including i) data visibility, the processes that organizations undertake to obtain a clear sight of what data, data-assets and activities exists, where data are located and if data are recoverable; ii) federated data, the standardization of data that is gathered from sources; and iii) data privacy and security, the compliance with data privacy and security standards, including, deletion of non-necessary, outdated, duplicated, erroneous and low value data, through a proper, regular, and adequate access management and system privileges [46, 47].

Although data governance is presented as a systematic approach, some literature considers that this reality is still in its infancy for several firms, mainly due to the lack of resources, budget constraints and the inability to focus on more than just data management [23]. In lines manner, Abraham et al. [32] emphasized that there is still no academic consensus regarding this field full scope and on a robust framework that can address the organization's entire needs. Likewise, data governance relevancy should be highlighted across all organizations, independent of their type (private or public), employee number, market segment and capitalization, profit or non-profile, and size (small and medium-sized companies and larger ones), among other. Therefore, existing frameworks must be able to adapt to the different features that distinguish organizations and established in a manner that aligns with their capacity, resources, and size, where it is expected that larger institutions take the lead and that ultimately, smaller ones begin, in a more phased process [38]. Consequently, literature emphasizes that current frameworks are usually defined in a way that suits larger firms, concluding that there is not yet a fully effective and efficient on-size-fits-all approach. Also, current frameworks do not encompass specific initiatives that ensure the management and monitorization of Generative AI systems and models usage, that organization perform on their daily basis [38]. Therefore, our paper's contribution, as illustrated in Figure 1, focuses on addressing the aforementioned gap.

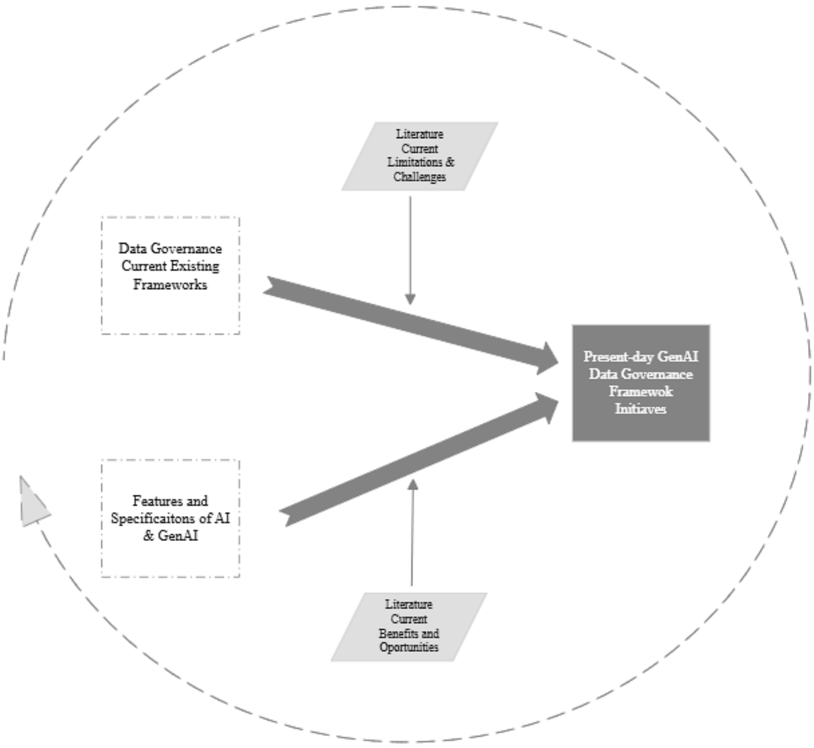


**Figure 1. Paper Contribution to Data Governance and GenAI Practices**

### *2-2- Artificial Intelligence (AI) and Generative AI*

Literature defines Artificial Intelligence (AI) as the feature of information systems that advance human intelligence through the usage of computational power and its enhancements that allow individuals and organizations to have systems that are capable of learning, interpreting, and generating data and novel knowledge, based on a set of already existing data, achieving their desired outcome through continuous customization of AI models [48]. For Liza [49], AI encompasses various layers of computational methods deployed to perform specific tasks that mimic human thinking and behavior, such as generating content, supporting decision-making, and predicting data patterns or behaviors. By replicating the human cognitive functions, AI aims at powering the human brain through a machine capable of processing large and complex data through neural networks [49]. Considering the two distinct AI systems, Discriminative and Generative AI (GenAI), literature has defined these as i) Discriminative, which corresponds to a given AI system that generates predictions based on data that is inputted into its model, in which by defining inputs the system is capable of generating the looked-for outputs, e.g., classification and pattern recognition of a given data; ii) Generative, which excels at producing new data and knowledge outputs through actively learned distributions and its probabilities allied with training patterns that were given to the system [13]. Besides this difference, Generative AI can be considered the most recent type of AI, being considered the "cutting-edge branch", which mainly focuses on producing novel and original knowledge based on its strong capacity to learn and gather knowledge from the data it is inputted with [50].

The new data and knowledge produced can be shaped in many formats, from pictures, audio, and video to text, speech, software, and product programming and design [51]. Literature denotes that in 2023, about 5% of organizations were using GenAI applications and methods, where it is estimated that by 2026, this number will essentially increase to about 80%, contrasting with traditional Discriminative AI, which is expected to focus on pattern recognition, classification and producing decisions based on pre-existing sets of data [50]. Even so, a study conveyed by BlackBerry in 2023 concluded that 75 % of organizations in the survey thought of banning GenAI, whereas, in 2024, a survey conducted by Capgemini concluded that 97 of the organizations allowed their human resources to leverage and use the GenAI, exemplifying the current struggle that firms are being faced with. GenAI can be employed in almost every process, including the governance of policies and procedures, software coding, and supporting processes and decisions [52]. Besides, with the increasing affluence and diverse availability of language models, already prepared and trained transformers, and open-source AI, data users are now able to autonomously build AI-driven models and systems. Firms' enterprise resource planning (ERM) software also contain embedded AI, e.g., SAP, Oracle, and AWS services [50].

Accordingly, the GenAI branch encompasses different techniques, including i) GAN, Generative Adversarial Network, which corresponds to a deep learning technique that produces new data that resembles the training one; ii) Transformer-based models, used for the processing of natural language tasks; iii) VAE, Variational Autoencoder, used as a generative unsupervised learning model that allows users to reduce noise from a given data, identify anomalies and generate knowledge, through the identification of key elements from the training dataset (encoding), the recreating of the original data (decoding behavior), the addition of probabilistic inferences (generative mechanism) and the re-encode of those outputs into a latent space; iv) Autoregressive techniques, which use probability distribution as an approach to produce new data; v) Boltzmann machines, a learning unsupervised model, based on probability distribution from a specific dataset, that uses this knowledge to conclude on data that is yet to be explored; and vi) Flow-based techniques, a model to approach large datasets with a complex dimension, capturing complex probability distributions and that can be employed to generate unobserved but realistic samples or fill incomplete ones and forecast events [51].

Similarly, the application of GenAI is seen across industries, independent of their business, maturity, and size, where the most recognizable forms resides in OpenAI (ChatGPT), Microsoft (Copilot) and Google (Bard) [48, 50]. Accordingly, the overall GenAI lifecycle is composed of a workflow of three different but inter-related stages, namely: i) data collection, where organizations seek to gather a representative amount of data as input to the AI model; ii) post-training, where the AI method through input data starts to produce novel knowledge; and, iii) the refining stage, where the user analyses its results and performs fine-tunes. Besides, GenAI continue to rise as technological advances and AI systems become more accessible, generating benefits namely i) democratization of data, which is emphasized by the simplification that GenAI provides data users with, allowing non-technical experts to access, utilize, and interpret AI models and their outcomes; ii) identify and treat data patterns and problems; iii) use and apply data analytics and programming over data that is unstructured or contain complex structures; and, iv) allow for the performance of real-time data analytics.

In contrast, literature reflects on the downside of GenAI, namely as challenges to ensure its safety, transparency, and accountability impact the efficacy and efficiency of a proper data governance framework, including i) inaccuracy of the training data of an AI system, which is dependent on the collection, quality and bias of training data [13, 53]; ii) incomplete, biased, or incorrect natural language texts and prompts that support the definition of the AI system's behavior and its task, as prompt engineering can involve commands that ensure the AI system performs as the user desires, rather than the opposite. Proprietary AI systems, such as ChatGPT, DeepSeek, Copilot, and Bard, among others, offer organizations a fast and accessible way to implement AI tools. However, there is growing apprehension about the bias that comes from the training data used to develop these models, which can influence results [13, 50, 53]; iii) fraudulent and malicious actions due to the lack of preventive and detective controls and responsibilities over the AI-generated data [43]; iv) inadequate or not defined policies, procedures, and guidelines to support the usage of AI systems and to ensure it remains transparent, legal and ethical [13]; v) Generative AI models usage as an undoubtful source of accurate and precise information, and not as a complementary and supportive source; vii) AI techniques that disregard the ethics principles and standards that set the bar between the right and wrong and that guides in defining the moral conduct [22, 54]; viii) current data governance frameworks are failing to adequately address the GenAI's complexities, predominantly their nature of being self-reproductive and autonomous, as they tend to focus on regulatory compliance, setting a static structure for responsibilities and data risk mitigation, while not accommodating the particular properties of GenAI systems [52]; ix) although data privacy regulations are mature standards, across the global perspective, with legislation such as the European GDPR (General Data Protection Regulation), Brazil's "Lei Geral de Proteção de Dados", Australia's "Privacy Amendment Act 2017", and the CCPA, 2018 (The California Consumer Privacy Act), they do not extensively consider the data and their detailed content that models provided by GenAI, leading into uncertainties and risks for organization that deploy these and strive to have responsible AI techniques [28, 40, 55].

## 3- Research Methodology and Planning

To achieve the objectives of this article, this section outlines and details the systematic approach to be employed in this research, with the primary objective of developing and providing innovative and practical knowledge to data governance and GenAI. The methodology applied corresponds to an adaptation of the stepwise approach of the Design Science Research (DSR) Methodology presented by Hevner et al. [56] and Ostrowski et al. [57]. The literature considers the DSR approach to be one of the most robust methodologies that enable researchers to produce artifacts, including toolboxes, structured processes, and frameworks, based on a systematically detailed approach. The traditional DSR methodology follows a set of six different phases, including: i) identifying and stating the problem and the study's main motivations; ii) defining the research objectives that the study and its artifact should address; iii) designing and production the prototype of the artifact; iv) demonstration and disclosure of the artifact; v) assessment of the utility, veracity, and validity of the artifact proposed; and vi) the communication of the artifact and its evaluation results. Below, in Figure 2, we provide a detailed description of our methodology structure and planning into a stepwise process flow approach, namely through a set of three Phases: Phase I – Exploration Phase; Phase II – Conceptual Phase; Phase III – Conclusive Phase.

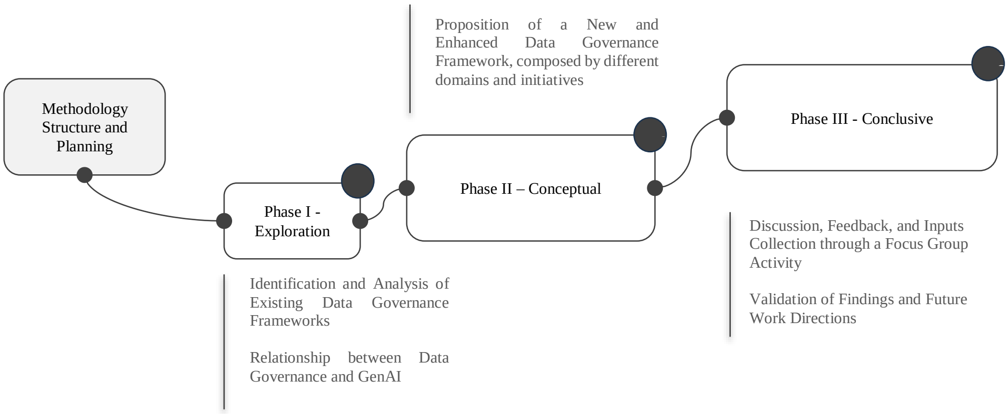


**Figure 2. Methodology Structure and Planning**

### *3-1- Phase I – Exploration Phase*

In the first phase of our methodology, the ground basis for the research, we focus on identifying, exploring, and clarifying the main issues (problem statement) and potential opportunities (study motivations and contributions), as well as on the formulation of the primary research objectives and its overall question [56–58]. Throughout our Background section, our goal is to identify and explicitly state the research gaps and problems identified throughout the literature, defining feasible and valuable objectives that our work aims to address. Consequently, our research focuses on the Phase I – Exploration Phase, where we perform a revision and thorough analysis of the literature surrounding the data governance field and its frameworks, focusing on the key concepts, characteristics, significant challenges to stabilize the problem statement and main potential opportunities and benefits to define the study objectives and motivations that will support our proposition of designing and defining an enhanced data governance framework. Also, our research will encompass and acknowledge the current research findings in GenAI, specifically its relationship with data governance, to gain a deeper understanding. Hence, we collected and reviewed key literature, research articles, and studies.

### *3-2- Phase II – Conceptual Phase*

In this phase, we intend to define a new, structured and enhanced data governance framework with actionable initiatives that establish a bridge between GenAI specifications, while integrating the best features of numerous data governance frameworks. Within the data collection and analysis step, we aim to gather, examine, and derive insights from our extensive background analysis and to evaluate the potential integration of GenAI initiatives into an existing framework. Moreover, we will present a systematic comparison between numerous frameworks, identifying how they differ from each other and acknowledging the different and best criteria that organizations should take into account so that a new, unified and improved framework can be designated, including GenAI in their initiatives. Also, we will present the key baseline for our framework presenting clearly how it addresses the current challenges and gaps.

### *3-3- Phase III – Conclusive Phase*

The final phase involves a focus group, which is conducted with a group of experts in data governance and GenAI to obtain explicit and specialized knowledge while integrating those insights into our proposition to provide an enhanced version of it. To this intent, we will evaluate the findings and insights obtained by producing transcripts and analyzing their content to compare these insights with our literature review findings. Thus, addressing significant challenges through an evidence-based set of initiatives that support a GenAI-focused framework. Likewise, this Phase provides research-validated findings, formulating GenAI initiatives within a new and enhanced framework, and summarizing the key research conclusions and contributions to both data governance and GenAI. Besides, all sources of information and insights are critically and thoroughly analyzed, compared, and discussed to ensure their validity and applicability to our work and our fields of interest. To conclude our paper, we will present a synthesis of our data governance framework and its actionable GenAI strategies to address the gaps in existing frameworks, which will enhance the transparency, responsibility, and compliance of organizations within these matters, and present the future direction topics.

## 4- Identification and Analysis of Existing Data Governance Frameworks

### *4-1- Data Government Frameworks and Comparison*

Kurniawan et al. [59] considered that data governance practices relate to an individual or entity's ability to define and implement a framework set to design, monitor, and oversee their data and related assets, including the information technology infrastructure. As a result, the main goal of a data governance framework is to guide and establish all data-related functions and processes, ensuring that data are presented in a precise, adequate, and secure manner. However, as

Joshi et al. [7] denotes that a framework takes into account more than just its technological side. Thus, in order to ensure that a data governance framework is defined and implemented without endangering the organization's productivity and innovation levels, the entity must understand the many existing frameworks and how they address each data-related domain [7]. In Table 3 we present the different frameworks analyzed, namely:

**Table 3. Data Governance (DG) Frameworks Identification and Characterization based on Literature Findings**

| Data Governance analyzed Frameworks and its Main Characteristics | | | | | |
|---|---|---|---|---|---|
| **#** | **Title** | **Main Domains or Phases** | **Scope** | **Date** | **Ref.** |
| A | The Framework for Data Decision Domains | i) Data governance domains; ii) Data quality; iii) Metadata; iv) Data access; and v) Data lifecycle. | General | 2010 | [35, 33] |
| B | The IBM DG Framework | i) Definition of a business problem; ii) Sourcing executive sponsorship and funding; iii) Evaluating the maturity status; iv) Developing a roadmap for maturity robustness; v) Definition of the organizational blueprint and data flows; vi) and vii) definition of a data dictionary; viii) Design and implementation of a metadata repository; ix) Design of metrics to measure operational success; x) Designation of data stewards; xi) Analytics to support governance; xii) and xiii) Manage security and privacy elements, and xiv) Measuring the Outcomes of data governance initiatives. | General | 2010 | [31] |
| C | The DG Institute (DGI) Framework | i) Mission and vision; ii) Goals, governance metrics and funding strategies; iii) Rules; iv) Decision Rights; v) Accountabilities; vi) Controls; vii) Data stakeholders; viii) Data Governance Officer; ix) Data Stewards; and x) Proactive and Ongoing Data Governance Processes. | General | 2014 | [31] |
| D | The Big Data Algorithms Systems (BDAS) DG Framework | i) evaluation of data quality and data bias; ii) analysis and detection of patterns; iii) organization and its data owners, users, and partners; iv) a bug bounty approach; v) fostering transparency; vi) data categories; vii) empowerment of data owners and users; viii) data gathering activities; ix) authorizations and accesses to data; x) distributed data storage; xi) data stewardship; xii) segregation of duties and conflicting roles; and xiii) data as a valuable asset. | Specific – Extension to Big Data | 2020 | [23] |
| E | The Analytics Governance Framework | i) Roles and responsibilities; ii) Policy and standard; iii) Analytics governance processes; and iv) Analytics governance control. | Specific – Extension to Data Analytics | 2023 | [8] |
| F | The Conceptual Framework: Data Governance | i) Governance mechanisms; ii) Organizational scope; iii) data scope; iv) domain scope; v) antecedents; and vi) consequences. | General | 2019 | [32] |
| G | The Analytical Framework for DG in Digital Platforms | i) Data ownership; ii) Original data quality; iii) Data user and value; and iv) Data stewardship. | Specific – Extension to Digital Platforms | 2019 | [11] |
| H | The Big Data Governance Framework | i) Enablers; ii) Governance domain; iii) Guiding principles; and iv) Goals. | Specific – Extension to Big Data | 2019 | [40] |
| I | The DG Framework for SMEs | i) data security and privacy; ii) data impact assessment; iii) designation of a data review board; iv) avoidance of unfairness or bias; and v) data governance and GDPR. | Specific – Extension to SMEs | 2010 | [38] |
| J | The Data Governance Center of Excellence | i) foundational elements; ii) data portfolio management; iii) implementation management; iv) engineering and architecture; and v) Processes, operations, and controls environment. | General | 2017 | [26] |
| K | The Quality of Data in Motion Framework | i) Discover; ii) Define; iii) Design; iv) Deploy; and v) Monitor. | Specific – Extension to Data Quality | 2016 | [10] |
| L | The DG Gamified Framework | i) Fast feedback; ii) Transparency; iii) Goals; iv) Badges; v) Onboarding; and vi) Competition and collaboration. | General | 2015 | [34] |
| M | The Data Middle Platform Framework | i) Service empowerment objectives; ii) Definition of roles and structures; iii) Design and implementation of policies and procedures; iv) openness through communication and data sharing; v) monitor compliance and control; and vi) data management life cycle. | Specific – Extension to Data Middle Platforms | 2021 | [37] |
| N | The Industry 4.0 DG Framework | i) Data-as-a-Service (DAAs); ii) Monitoring-as-a-Service; and iii) Platform-as-a-Service. | Specific – Extension to the Industry 4.0 | 2022 | [12, 39] |
| O | The Data Governance Framework for Big Data | i) Standards; ii) Policies and procedures; iii) Organization; and iv) Data integration infrastructure. | General | 2018 | [36] |
| P | The Responsible Data Governance Framework | i) Anticipation; ii) Reflection; iii) Engagement; and iv) Action or responsiveness. | Specific – Extension to Ethics and Responsibility | 2019 | [41] |

Moreover, in the selection process of the data governance frameworks for comparison and analysis, we ensured a balanced cross-section across sectors, including public sector organizations, private enterprises, SMEs, and NGOs. Through the background review performed, we retrieved widely recognized frameworks relevant to each context. For the public sector, we included The Analytics Governance Framework (Kanying et al. [8]); for private enterprises, The Data Middle Platform Framework (Mao et al. [37]); for public organizations and corporate contexts, The Conceptual Framework (Abraham et al. [32]) and The Industry 4.0 Data Governance Framework (Serrano & Zorrilla [12]; Zorrilla & Yebenes [39]); and for SMEs, The Data Governance Framework for SMEs (Okoro [38]). The remaining frameworks provide an integrated perspective across sectors, ensuring that it captures diverse organizational practices and challenges.

### *A. The Framework for Data Decision Domains*

In a paper published regarding the analysis of the overall important activities of data governance, Alhassan et al. [33] presented a data governance framework originally proposed and designed by Khatri & Brown [35], which included a total of five interrelated domains. Firstly, these authors presented the domain of data principles, which relates to the definition of the strategic roadmap and direction for governing and managing the organization's data through a set of internal requirements, standards, and rules. Although Alhassan et al. [33] reflected on the need for a universal approach to data governance and management, addressing all relevant data characteristics and needs, the literature reflects that data governance principles and their implementation are dependent on different aspects, from the volume, variety, and velocity features of data, to IT systems and data storage, human resources, among other elements [33, 35, 45].

The second domain focuses on the importance of defining and implementing key data quality standards that address each of the key steps of the data life cycle, from the data's acquisition stage to transformation, analysis, report, and archival [33, 35]. Examples of different data quality approaches can be in the form of data profiling, through the examination of a given dataset's structure, content, and context, analyzing topics such as missing values, trends, incorrect or duplicate values, data relationships, summaries, and aggregations, among others [45]. In this context, it is highlighted that data integrity and its assertions (precision, consistency, completeness, timeliness, among others) should be controlled and monitored throughout the entire cycle, including its archive and deletion stage [60]. In line with this, the third and fourth principles focus on two key concepts that directly impact the quality of data, namely, metadata and data access. Metadata is considered to play a significant role in the success of a given organization's data governance framework, as it corresponds to the description of the organizations' data, usually recognized as data about data, including data's business definition, classification, structure, format, source systems, relationships, transformation, and lineage applied through its lifecycle, supporting policies and procedures, applicable reports where data is used, among other features [25]. The last domain refers to the data life cycle, which is acknowledged as the governance of the entire journey and data flux of a given data point through the organization's processes, people, and systems, from its acquisition and processing to its reporting and deletion.

### *B. The IBM Data Governance Framework*

IBM's data governance framework corresponds to one of the earliest frameworks. Presented and analyzed by Al-Ruithe et al. [31], this framework methodology is characterized by being grouped into several distinct elements, comprising a total of 14 perspectives. Ten of these are mandatory, and four are optional as best practices for enhancing the framework. These mandatory elements can be described as i) business problem designation, ii) obtaining and managing executive funds and sponsorship incentives, iii) and iv) performing maturity assessments to analyze the current data governance status and establish a strategic roadmap of activities to increase the current level of maturity and robust the overall data governance program; v) project and outline the blueprint of the organization's architecture, including data, its streams and its supporting assets (IT infrastructure); vi) and vii) create a data dictionary, including the details of the organizations internal and external data; viii) store the organization's details in its metadata through a repository of information; ix) operationalize performance metrics and indicators to assess the effectiveness and success of business activities. In line, Bordey [45] denotes that companies can draw and employ stories related to the data users' operations to describe their activities' performance, namely through indicators such as numbers that, with a specific context, can yield a timely available status on organizations' performance and, x) nominate and define data stewards, individuals that are responsible for the data quality management and technical knowledge and for implementing a strategy regarding the management of master data. Besides these elements, the IBM framework includes a set of four optional measures that are thought to improve the data governance robustness of a given organization from xi) applying business analytics, xii) and xiii) monitor privacy and data security, including data sharing, usage, and its flows within the organization's human resources and technological architecture, to manage and assess the output result of each data governance initiative.

### *C. The Data Governance Institute (DGI) Framework*

The Data Governance framework proposed by the DG Institute (DGI), illustrated in Figure 3, and detailed by Al-Ruithe et al. [31] is composed of ten connected features, namely i) Vision and Organization Vision, which includes stating the strategic direction and boundaries for the organization data governance framework, ensuring that its vision, mission and overall objectives are aligned with the data stakeholders data usage. To this end, the DGI emphasizes that organizations can form groups around their framework to oversee how data goes across both human and technological data assets; ii) governance metrics, funding strategies, and objectives, by leveraging on the SMART principles (specific, measurable, actionable, relevant and timely), organizations are in a better position to define clear goals and metrics for their data governance framework and any data-related activity. In line, organizations' leaders should define their funding and budgeting strategies towards the framework, its leadership and data stewardship roles, namely because the lack of resources leads into an inadequate framework; iii) policies, procedures and governance rules, where organizations should promote the definition and alignment of data-related guidelines and policies, data definitions and lineage; iv) decision rights defined on the structure and responsibilities, to ensure that all employees acknowledge their data role.

Furthermore, the following component, v) relates to the responsibilities and accountabilities, where organizations strive to adhere to the currently applicable regulations throughout their operations, guaranteeing that internal controls and documentation around each business activity is formally defined; vi) Mechanism for data monitoring and control, where organizations strive to control, mitigate and monitor the risks that are inherent to its activity, namely those that result from data-related assets and activities. To do this, organizations should list, characterize, and implement internal controls, including manual, semi-automatic, or automatic detective and preventive ones, and set initiatives of external audit and compliance assessments using an independent third party; vii) data stakeholders, viii) a Chief Data Officer or Data Governance Officer, and ix) Data Stewardship and technology roles, where organizations should define both individual-level, council-level, and team-level roles and responsibilities. A higher-level position, such as a Chief Data Officer or Data Governance Officer, should lead these structures. Within this structure, Data stewards and custodians should be defined as those responsible for ensuring that the data flows and assets meet data quality standards, including completeness, accuracy, validity, and integrity, among others. They should also continuously report metrics on these activities to both business users and data quality teams. In this regard, Al-Ruithe et al. [31] denoted that it can be beneficial for a data governance structure to include a Technology Steward role, which is responsible for administering, maintaining, and warehousing the IT technological infrastructure requirements of data users, stewards, and the overall organization. Lastly, x) Continuing and Proactive Governance, where DGI emphasizes that different from the previous framework components, the continuous and proactive feature of governance relies on ensuring that organizations can ensure that their framework is not static and that metrics and indicators are continuously monitored and addressed [31].

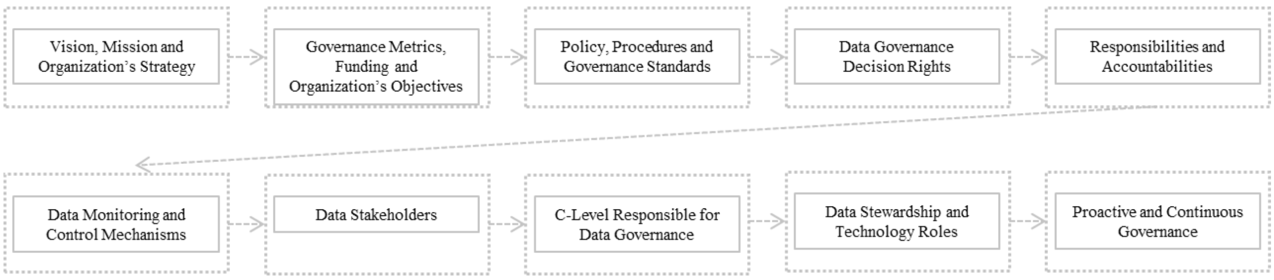


**Figure 3. The Data Governance Institute (DGI) framework detailed by Al-Ruithe et al. [31]**

### *D. Big Data Algorithms Systems (BDAS) Data Governance Framework*

Janssen et al. [23] proposed a data governance framework that focuses on the capability of organizations to manage data and data-related systems, especially their big data. These authors emphasize that this type of data is typically dependent on external data sources, thereby imposing a more significant risk and susceptible to manipulation or misuse. For these authors, this concept can be expressed in terms of trusted data-sharing frameworks, ensuring that data sharing is performed accurately, precisely, and securely in compliance and agreement with regulatory requirements, such as the GDPR (General Data Protection Regulation). A data sharing trusted process must encompass three different but related aspects: i) Identification, which includes the organization's capability to identify the person or entity that claims to have a specific identity; ii) Authentication, where organizations seek to ensure that a specific identified person or entity corresponds to what they are entitlement to be; iii) Authorization, after identifying and authenticating the person or entity, the organization must concede securely adequate accesses and permissions to specific applicable datasets, ensuring that data is shared on a needs to know basis, i.e., a minimum amount of data is shared to specific purposes [23].

Consequently, these authors propose a framework for data governance, named Big Data Algorithms Systems (BDAS), which includes a set of 13 design principles focusing on managing data, systems, data-sharing activities and its responsibilities, including the following: i) evaluation of data quality and data bias, aiming to provide data users with unswerving and free from biased data; ii) analysis and detection of patterns as to identify the root causes for shifts in the algorithms' outcomes as to validate these information and analyze its causes and data mechanisms; iii) organization and its data owners, users and partners, have only access to the minimum amount of data needed to a specific purpose, on a needs to know basis; iv) a bug bounty approach that is employed to foster users to spot and identify errors, inconsistencies and issues related to data and its algorithms, while information the organization; v) fostering transparency through the principle of information individuals, companies or government if applicable, of the data sharing activities that relates to them; vi) personal/sensitive data and non-personal/sensitive are categorized and separated; vii) empowerment of data owners to have control over their data activities, promoting data quality controls to ensure its accuracy, completeness and precision; viii) data gathering activities are based on data collected at the source, free from changes; ix) authorizations and accesses to data are adequately and regularly monitored to ensure the correctness of their attribution process; x) storage that is distributed, ensuring that data and its systems are less vulnerable and more controlled; xi) data stewardship through the definition and assignment of data roles and responsibilities for managing data and its systems; xii) segregation of duties and analysis of conflicts of duties to ensure that data operations are controlled and used sufficiently and adequately; xiii) establishing data as a valuable asset to support the organization's objectives and strategy, ensuring that organization-wide, the usefulness of data is recognized and precisely managed [23].

### *E. The Analytics Governance Framework*

The Analytics Governance Framework results from the enhancement of the Digital Government Development Agency (DGA) data governance framework performed by Kanying et al. [8]. This framework comprises level 0 and level 1 components, where level 0 is the primary domain of the framework and level 1 is its primary component description, as illustrated in Figure 4. Hence, the first level 0 component, the roles and responsibilities, is characterized by three level 1 components: i) the data governance office, which acts as the central body responsible for overseeing the data management and governance; ii) the data stewards' team, which is responsible for preserving the application of the data quality standards, including data policies, ensuring that the organizational data is aligned with internal and external standards; and iii) the structure of the analytics team, that is delineated as the group of individuals that ensure the harmonization of data, including roles such as data engineers and scientists [8].

In line, the second level 0 component is the policy and standards domain, which is composed of three level 0 features: 2.1) the analytics methodology and model documentation standard, 2.2) the analytics activities, and 2.3) the analytics references. Within the 2.1 feature, there are three level 1 components: i) the data definition (harmonization and standardization of the business understanding of data), the data rules (requirements that allow for data consistency and compliance), and data lineage (data point's journey life cycle, including business and technical metadata of a data point) [8]. To support this process, it is important to consider the data quality standards from the International Organization for Standardization (ISO), including standards ISO/IEC Standard 25012:2008 (Software engineering – Software product Quality Requirements and Evaluation), ISO/IEC 38505-1:2017 and TR 38505-2:2018 (Information technology — Governance of IT — Governance of data), ISO 8000-61:2016 (Data quality – part 61: data quality management: process). These standards provide a set of data quality and management guidelines and best practices that can assist an organization's data governance framework [61-65]; ii) the data preparation standards, which define the organizational procedures to ensure the data extraction, cleansing, transformation, and loading, resulting in data quality; and, iii) technology reliability, which ensures that systems are monitored in a manner that allows data to be used consistently, precisely, timely and is accessible.

Moreover, within 2.2. feature, there are three level 1 components: i) the business strategy, defining the set of organizational actions that support their primary objectives; ii) and iii) the analytics use case, vision, and priorities, highlighting the need for structured data scenarios that guide analytics engagements, meet business needs, and support the decision-making process. The last feature, 2.3, corresponds to the Analytics reference, which is composed of two level 1 components: i) the analytics roadmap, business glossary and definitions, outlining a path that supports a set of short- and long-term data analytics initiatives aligned with the organizational vision. Also, business glossary should be defined, considered a centralized and uniform data repository that includes business terminology data relationships and supports the overall business definitions [8].

Besides, the third level 0 is the analytics governance processes, which includes a set of five level 0 features: 3.1) design processes and system development, 3.2) operationalization and ability to explain, 3.3) testing and review processes, 3.4) approval and release, and 3.5) continual improvement. The 3.1 feature has one level 1 component, which defines the overall process that supports the business processes and systems design, ensuring that their development follows a step-by-step approach, including four levels: i) understanding and alignment, ii) design and structure, iii) standardization and control and iv) review and iteration. The 3.2 feature includes two level 1 components: i) the explainability assessment, where the organization should evaluate the rationale, robustness, and transparency of the outcomes retrieved from their analytics models, and ii) the model operationalization, which entails the overall process of integrating a model into the business operations. The 3.3 feature, testing and review processes, includes one level 1, check, measure, and report, that ensures that systems are constantly tested in development and quality environments prior to their production releases and that it is configured to be audited and reported on. The 3.4 and 3.5 features each contain one level 1, where the first relates to the approval and release processes, in which organizations must ensure that releases are approved in line with their internal responsibilities and authorities. The second refers to continual improvement, where a defined feedback channel should be established for improvements [8].

Last but not least, the fourth level 0 component is the analytics governance control, which includes three level 0, 4.1) data quality management, including one level 1 component which relates to the data control and quality assessment, where the organization should ensure the definition of metrics and monitor activities to measure the data quality and condition on factors such as precision, completeness, trust, accessibility, consistency and validity, 4.2) security and risk management, which encompasses two level 1 components, security and risk management and data risks assessment, which relate to the organizational process to identify and assess the security design, audit logs and data risks, and 4.3) people development and learning design, which includes one level 1 component skill gaps analysis and training roadmap, highlighting the need for organizations to ensure that their human resources have the adequate and necessary training that allows them to be able to keep up and play a vital role within the data governance framework [8].

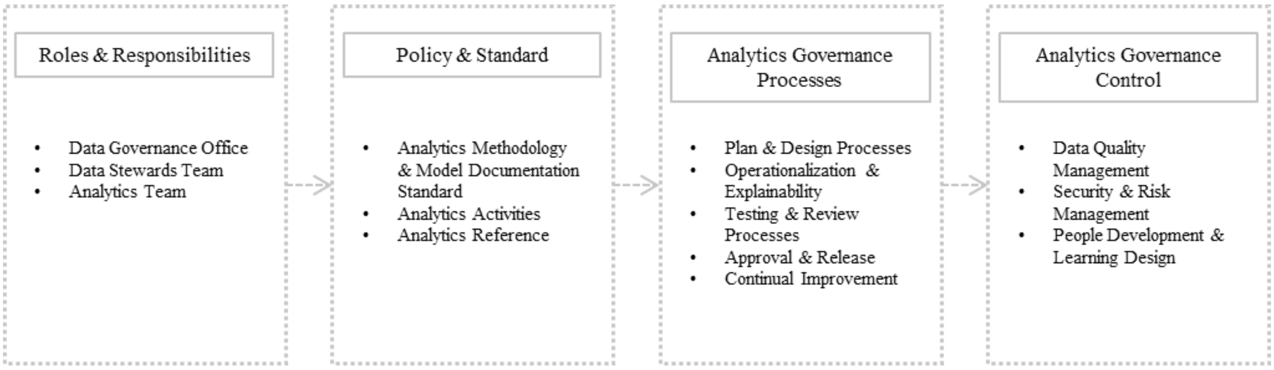


**Figure 3. Analytics Governance Framework adapted from Kanying et al. [8]**

### *F. The Conceptual Framework: Data Governance*

Abraham et al. [32] presented a data governance article that outlined the main concepts, providing a conceptual methodology for data governance. This methodology encompasses a total of six related dimensions, from 1) mechanisms for data governance, including the definition of its structure, procedures, and activities. In this dimension, it is expected that a company can establish and nominate data governance roles, responsibilities, policies, and operational procedures. This last feature can be defined through the design and implementation of formal documentation on data governance, including its primary strategy, procedures, policies, standards, business operations, and processes, performance and quality indicators, mechanisms for the continuous operations monitoring and its compliance with current legislation and rules in place. In line, at this stage, data quality metrics should be employed that allow for data to retain data quality assertions, from the completeness (all the data information, context and fields are all filled, preventing missing or incorrect data); timeliness (available data represent the most updated one and can be analyzed using time period filters); uniqueness (data prevents duplication and mistakes around data values and registries); accuracy (data are consistently correct and precise in their overall content and structure); consistent (data represent the business and user expectation, content, format and structure-wise); validity (where specific features, data rules and legislations or standards are ensured); and data lineage (knowledge on the data lifecycle journey through their data assets and roles) [66].

Besides, the second and third dimension 2) and 3), respectively, relates to the organizational and data scopes, where in 2) the objective is described the definition of the extent to which the data governance program will be stablished, and the 3) define and identify what data assets are currently being used and expected to be used, during the data governance program; The 4) and 5) dimension related to the domain scope and to the antecedents, where in the domain scope organizations are expected to be able to proceed with data decisions, including on its quality metrics, privacy and security, data lifecycle management and its storage mechanisms, and where antecedents are expected to encompass contingency plans, both internal and external that affect its overall ability to define a framework and to sustain its robustness. Last but not least, the sixth dimension is represented by the consequences, which reflect the potential impacts of having a data governance framework defined and that can be assessed through risk and control assessments.

### *G. The Analytical Framework for Data Governance in Digital Platforms*

The Analytical framework for data governance in digital platforms was proposed by Nokkala et al. [11], drawing it as an extension of the traditional single-entity frameworks, as shown in Figure 5. This framework encompasses a total of four domains: data ownership and access, original data quality, data use and value, and data stewardship, which leads into a fifth one, entitled the platform data quality to fit the purpose of digital platforms and data sharing [11]. The first domain relates not only to establishing ownership and adequate access rights to data stored within digital platforms, setting governance as an equation of what and who can perform those actions but also how this definition of ownership is regularly monitored and maintained. The second domain focuses on the original data quality, which corresponds to the data quality characteristics of the data gathered and input into digital platforms, including its metadata and the data lifecycle. The third domain focuses on data use and value, focusing on the data benefits, not only the internal data but also the external data, including competitors' one, as such this framework focuses on the need for a business model and data strategy that is clear and communicated to all relevant parties. The fourth domain includes data stewardship, a data role to support the definition of functions and responsibilities over an organization or platform data [11]. All of these domains are believed to support a platform data quality, which relates to the quality of the datasets that reside in digital platforms, including topics such as data consistency, accuracy, completeness, and availability, amongst others [11].

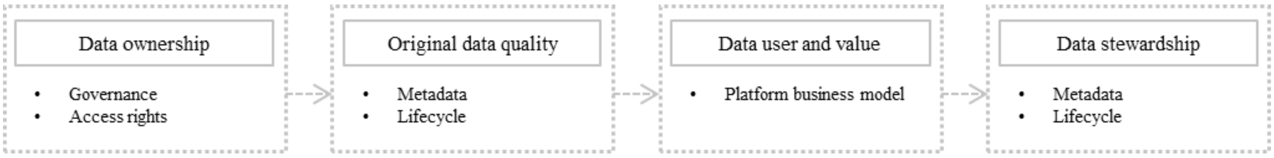


**Figure 4. Analytical Framework for data governance in digital platform adapted from Nokkala et al. [11]**

### *H. The Big Data Governance Framework*

The Big Data governance framework was drawn by Yang et al. [40], as illustrated in Figure 6, taking into consideration the extension of the traditional single entity frameworks to one that also encompasses the Big Data phenomenon and its characteristics, Velocity (the speed at which data is created and requires to be processed to meet the needs of business operations and their users), Variety (sourced from different domains with three typical types: unstructured, semi-structured and structured) and Volume (large amount of data). This framework encompasses four domains: enablers, governance, guiding principles, and goals. The first is responsible for the definition and implementation of the data governance procedures, guidelines, and rules, which must cover the overall key domains of data governance and management. This activity should typically be performed in accordance with the data roles and responsibilities, also known as data stewardship. The role of data steward is composed by a specific individual(s), although it can be set as a data stewardship council and represented by technical individuals who can solve the concerns of data users [40].

The second domain considers the five sub-domains of data governance and its distributed systems, namely i) the distributed data processing, where typical tools such as Hadoop, Spark, MapReduce, amongst others are used in other to keep up with the rapid data growth, ii) the distributed data storage, where no longer traditional dedicated servers are sufficient to maintain the data storage of an organizations, as such methods that aim at storing and integrating data via a distributed architecture are used, iii) integration and metadata management, the bottom landing zone of data, and supported by technology tools, ensures that data content are defined prior to the storage of data, iv) data quality management, composed by a set of strategies that focus on measuring, enhancing and certifying the quality and integrity of production, testing and legacy (archived) data and v) data security, privacy and ethics, which is focused on data ruling and legislation [28, 40].

The third domain encompasses four sub-domains of guiding principles, focused on i) accountability, defining the feature of data that ensures it is readily available, easily accessible and its credibility remains intact, and is usually established throughout the organizations' departments or business units through data ownership roles, ii) quality, corresponding to a feature that inter-connects both accountability, integrity and transparency, and that ensures that data quality standards and assertions are persevered in the data management processes, including, its accuracy, consistency, completeness, timeliness and veracity, iii) integrity, which corresponds to the feature of data where the veracity and principal characteristics of data remain consistent and adequate throughout the data lifecycle and where proper and robust internal controls are defined and periodically tested, and iv) transparency, where the focus is the protection of the organization from potential data breaches or misuses of sensitive data while safeguarding the appropriate documentation of data and that both internal and external third parties can audit and review data-related activities, processes and decisions [40]. Ultimately, the fourth domain relates to the goals, where the primary purpose of the data governance framework is to preserve and maximize the value derived from data, specifically in the decision-making process and data-handling processes, thereby fostering a culture of data-driven processes. The data governance framework goals have to be aligned with the organization's strategic aims, enabling them to leverage the key opportunities retrieved from big data with modern applications [40].

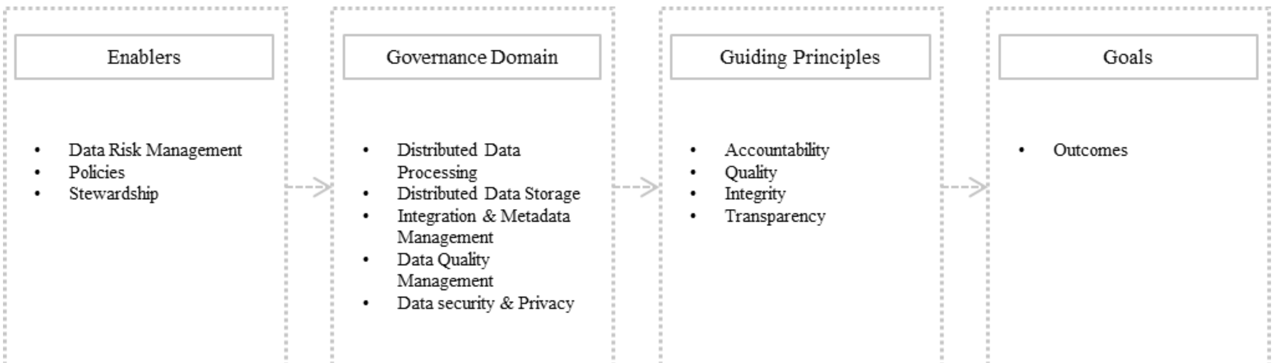


**Figure 5. Big Data governance framework adapted from Yang et al. [40]**

### I. The Data governance framework for Small and Medium-sized Enterprises (SMEs)

Okoro [38] recognized the specific and unique encounters imposed on SMEs. Thus, these authors have developed a framework, as presented in Figure 7, tailored to these organizations, prioritizing their ability to remain secure and compliant while maintaining ethical data management practices. This framework consists of five key pieces, namely i) data privacy and security, where SME organizations should ensure that their data collection stage is constrained to business requirements and related data, and proceed securely with their disposal as soon as the data are no longer necessary. Hence, organizations should define robust authentication instruments, data classification schemes and continuously monitor their data assets, including policies and procedures; ii) assessment of the impacts that results from data and its usage, including assessing the level of anonymization that the data contain as to reduce and mitigate privacy, concealment and personal data mismanagement risks, mainly through PIAs (Privacy Impact Assessments) and controls performed under the principles of the GDPR (General Data Protection Regulation); iii) definition and assignation of a Data Review Board, who should be composed by experts in the fields of data governance, compliance and privacy, and should address regularly the data lifecycle and governance topics, including the organizations data assets, data collection, analysis, storage, reporting and elimination; iv) fairness and prevention of bias, where organizations must ensure that their culture and data users strive to address the bias within the overall data lifecycle activities as previously described. Therefore, these authors invite organizations to review their data collection process and to ensure that their datasets contain a population that is representative of diverse characteristics and free from any biases that can impact the results of their data activities; and v) data governance, ethics and general compliance with data privacy rules, e.g. GDPR, where aligning the organization data governance with these standards, both ethical and legal is crucial to ensure that the data-related activities of organizations are transparent, trustworthy and adherent with all relevant regulations.


**Figure 6. The Data governance framework for Small and Medium-sized Enterprises (SMEs) presented by Okoro [38]**

### J. The Data Governance Center of Excellence framework

Sifter [26] asserted that reaching excellence in defining and implementing a data governance framework requires that it is backed by five aspects, namely i) foundational elements, ii) data portfolio management, iii) implementation management, iv) engineering and architecture, and v) operations and support. In this context, regarding the i) foundational elements, organizations should define and implement a strategy, charter, and vision for data governance that can be overseen, maintained, and enhanced, including transparent data accountability and ownership; ii) the data portfolio management features outline the operational side of the structure, namely through the and implementation of a regularly maintained data inventory, which should provide the data stewardship roles with the characterization of the organization's data, its data flow and integration processes. The data inventory tool can be considered as a data dictionary or data glossary, providing organizations with a catalogue of the existing data containing details on data ownership, formats, types, technical and business features, control metrics, metadata, amongst other relevant aspects [26, 67]. This tool must be regularly review and enhanced [67]; iii) the implementation management stage, where it is analyzed which are the best methods regarding organization's project, people, change and systems' access management, facilitating the fostering of a data governance culture and conceding the adequate training; and iv) the engineering and architecture stage, where organizations should ensure that their infrastructure is supported by the adequate standards, procedures and guidelines on software and change management, quality and data assurance, privacy and security. To support this, this author emphasizes the importance of having a robust processes, operations, and controls environment that allows the organization staff to communicate any incident or doubt whenever these situations occur promptly [26].

### K. The Quality of Data in Motion Framework

In a paper focused on data quality standards and their potential impacts on an organization's activities, Dutta [10] introduces six key areas where organizations can enhance their overall ability to govern, manage, and utilize their transactional data, which the author refers to as data in motion, illustrated in Figure 8. Within the first area, Discover, the author proposes that organizations initiate their data quality process by identifying, formalizing, and comprehending all critical data, as well as supporting operational architecture and systems. While doing so, firms should define metrics at relevant stages that result from the Discover activity, resulting in a repository of all key existing data identified through the organizations' processes, mapping these with internal and external systems infrastructure. The following key area is Define, where firms define their data-related risks that stem from their data quality requirements and domains, originating from transactional and non-transactional data (data at rest). In line, there should be a repository for data issues identified in data processes, through data observability, which can be put into practice as combined mechanisms that enable individuals to monitor, assure, and produce analysis and conclusions from data effectively [68]. In areas 3 and 4, Design and Monitor, organizations are expected to strive to enhance their data analytics techniques through automized processes that allow users to govern data expectations, inconsistencies, and incidents in a controlled and systematic manner. Historically, data analytics techniques involved selecting sample sizes rather than analyzing the whole population, however, with technological advancements, this approach is now more accessible. In line, the data governance process should not remain static over time and that there are mechanisms such as Key Performance Indicators (KPIs) and Data Quality Indicators (DQI), among other metrics, that allow for real-time control and oversight [10].


**Figure 7. Data Quality Framework presented by Dutta [10]**

### *L. The Data Governance Gamified framework*

The Data Governance Gamified framework proposition made by Hay [34], as shown in Figure 9, introduced an innovative methodology for data governance, specifically by leveraging gamification methods to enhance an organization's capacity to integrate a framework into its business. By applying gaming industry mechanics, the author believes that it leads to encouragement and fostering a robust data governance culture of participation and awareness. These mechanics are the following: i) game's fast feedback feature, where data users should receive immediate feedback on their data-related actions, while doing so the data users can earn points as a prize for their data-related actions involvement, including contributing to evident data stewardship, definitions, traceability and analysis or penalties for non-participation; ii) transparency and accessibility feature, where similar to games, the data governance processes are clearly accessible and open, allowing all data users to monitor the organizations' actions, enhancing the collaborative-driven culture; iii) Clear objectives, where users and data stewards work towards both short- and long-term goals, being those set at an individual and community level, promoting the data quality standards best practices; iv) data governance badges, where data users and stakeholders can earn badges and certificates that recognize their actions. Through this, organizations can track how their individuals level up in their data governance knowledge while promoting healthy competition to their collective workforce; and v) onboarding, which ensures that the practices can always be explained and passed to new users, independent of their experience or hierarchical position; and vi) collaboration and competition, where organizations promote team-building activities that encourage their individuals to participate in leaderboards and challenges, while driving collaboration between their workforce, including persons in organizational silos [34].

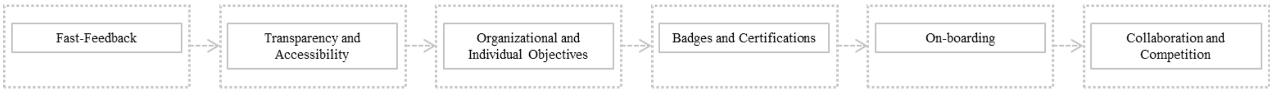


**Figure 8. The Data Governance Gamified framework presented by Hay [34]**

### *M. The Data Middle Platform Framework*

Mao et al. [37] proposed a framework operationalized in a way that fits the specifications of government institutions, including private and public, because they face decentralized data governance and management activities and cross-departmental silos. These authors believe that traditional frameworks fail to address these specifications, resulting in ineffective control over essential government processes. To address this, the authors designed the Data Middle Platform framework composed by six components, as demonstrated in Figure 10, namely i) objectives regarding service empowerment, where the focus of organization is to protect the sensitiveness features of data, namely personal ones, through enhancing data management capabilities, defining strategies and goals for their public services and promoting hierarchical cross-relationships; ii) establishing roles and assigning responsibilities across different levels, including the background, organization's layer that has the departments responsible for administrative tasks and the data governance decision-makers, councils and leaders, who define the plan, strategy and coordination of resources; the middle platform, organization's bridge that interacts, collaborates and connects with both the background and the foreground, being represented by the key data governance roles, including data users, custodians and stewards, individuals with C-Level Data function (CIO, Chief Information Officer, who governs and manages the IT infrastructure; CSO, Chief Service Officer, overseeing data-related activities and how they impact government services; and CDO, Chief Data Officer, who oversees the data governance framework and its overall strategy); iii) At the platform level, organizations should strive to develop a set of policies and procedures, assign data ownership and stewardship, and define mechanisms to monitor the governance performance, security protection and data quality and communicate those between the different layers; the foreground represents the government public service points that are accessible for citizens to interact and communicate with the government institutions, which requires regular monitorization and updates that in real-time meet evolving citizen needs; iv) communication and sharing openly across the different layers, which involves promoting stakeholders awareness for the importance and specifications of the organizations data governance framework.

Accordingly, at this stage, organizations should strive to integrate data sharing activities, where for collaboration, the Interagency Data Sharing (IDS) and for transparency and promoting public use of government data, the Open Government Data (OGD) actions; v) Compliance and monitoring, where government institution define i) their audits, regulation supervision and monitoring and data security and privacy protection mechanisms, and ii) data quality metrics and standards aligned with the ISO 2008; and vi) the operationalization of the framework, including policies and procedures, roles and responsibilities, data quality standards and mechanisms, ensuring that all previous components are stablished within the framework and the different data lifecycle stages, from data gathering (collection of data across the different sources, internally and externally, across different departments and public services), and processing (process of transforming, cleansing and structuring the data and its quality for an enhance data analysis), to analysis (capacity of data users to visualize and produce reports to support the decision-making process) and capitalization (management of metadata and monitorization over the data flows and its traceability across the organizations' IT infrastructure) [37].

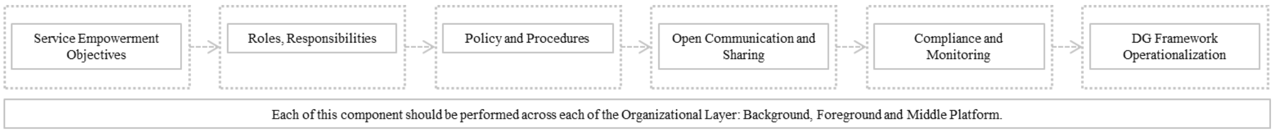


**Figure 9. The Data Middle Platform framework presented by Mao et al. [37]**

### *N. The Industry 4.0 Data Governance Framework*

In a paper presenting the Industry 4.0 specifications, Serrano & Zorrilla [12] and Zorrilla & Yebenes [39] presented the importance of defining a framework that suits the specific characteristics of Industry 4.0, namely its complex characteristics and inter-company interaction and collaboration, organizations and industry silos, legal compliance with the present-day regulation and challenging service level agreements (SLAs). Hence, their framework proposition, illustrated in Figure 11, considers data and governance as the core principles, where the framework is drawn around three principles, namely i) DaaS (Data-as-a-Service), focused on ensuring data is accessible and usable across the different organizations and their context; ii) MaaS (Monitoring-as-a-Service), enabling organizations to enforce their practices through real-time tracking over data governance related activities; and iii) PaaS (Platform-as-a-Service), specific to providing infrastructure that can be scalable for managing and monitoring data governance processes. To support these principles, a set of components should be interconnected comprising: maturity model and architecture development methodology, information base composed of standards, content metamodel, reference architecture, and Architecture Building Blocks supported by a reference model (ABBs). The ABBs can be of two types: i) the Architecture Principles, which define data governance guidelines and principles that are stored in a repository and govern according to organizational goals, compliance requirements, and roles and responsibilities, and ii) the Policy and Standards which includes the core governance and stewardship actions, namely definition of roles and responsibilities. These components enable organizations to define a standardized and structured data governance approach in Industry 4.0 [12, 39].

Besides, this framework incorporates international standards as the basis for the definition of the Architecture Building Blocks for the reference model, including: i) the TOGAF® Standard v9.2, a structured framework for enterprise architecture extended to fit a new governance entity recognized as "Policy" and aligned with ISO/IEC/IEEE 42010:2011 (Standard focused on providing guidelines for software architecture to ensure its consistency and comprehensiveness), ii) the standard regarding the Reusable Asset Specification, which has its leading utility of classifying each of the Architecture Building Blocks (AAB) into different asset categories, recognized as classes, namely Asset Name, Id and Description, Profile (Asset type and its lineage), Asset usage (set of activities that the asset allows), Asset relationships with other blocks, Requirements, Solution Building Blocks and description of the governance Architecture. Considering this, the proposed framework from the authors brings a new perspective to the data governance field, one that focuses on the specifications of Industry 4.0 while ensuring that organizations' data governance and management frameworks align with regulatory demands and are optimized for their data usage and operational efficiencies.


**Figure 11. The Industry 4.0 Data Governance framework by Serrano & Zorrilla [12] and Zorrilla & Yebenes [39]**

### *O. The Data Governance Framework for Big Data*

Another relevant framework is the one proposed by Panian and described by Kim & Cho [36], as shown in Figure 12. This framework is proposed to emphasize that a data governance program, to be successfully and effectively implemented in an organization, must involve and relate a set of different but interconnected elements, ranging from standards, policies, and procedures to the organization's context, structure, and data integration infrastructure.

Consequently, this framework is grouped into four different components: i) standards, ii) policies and procedures, iii) organization, and iv) data integration infrastructure, which are directly connected to six different data domains, namely data accessibility, availability, quality, consistency, security, and auditability.

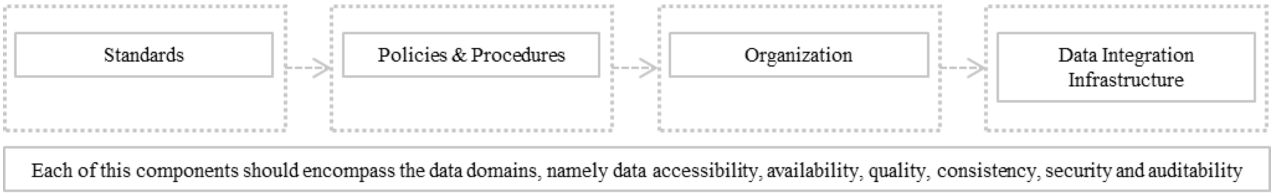


**Figure 10. The Data Governance Framework presented by Kim & Cho [36]**

### *P. The Responsible Data Governance Framework*

Fothergill et al. [41] highlighted the fact that data governance context is focused on its operational side, namely, on ensuring data quality, accessibility, accountability, and security, and fails to consider concerns on the ethical side of it, acknowledging the need for a responsible approach, focusing on integrating the RRI principles ("Responsible Research and Innovation") into a framework. These principles are integrated through the interrelationship of four features, named the AREA framework, illustrated in Figure 13: i) anticipation, ii) reflection, iii) engagement, and iv) responsiveness or Action. Accordingly, a framework must be facilitate the inclusion of these features, integrating previsions and considerations for future implications of its initiatives and being flexible to integrate possible modifications according to its outcomes. Additionally, these features are critical when cases of ethical issues arise in data life cycle management and governance. To incorporate a responsible framework, these authors propose a set of actions to address the features of the AREA framework, including strategies such as i) compliance and internal ethics check, the definition of a Data Protection Officer, ii) the definition of a data curation team focused on analyzing data ethical aspects, iii) periodic management meetings involving the data curation team, the DPO, advisory boards and compliance functions [41].

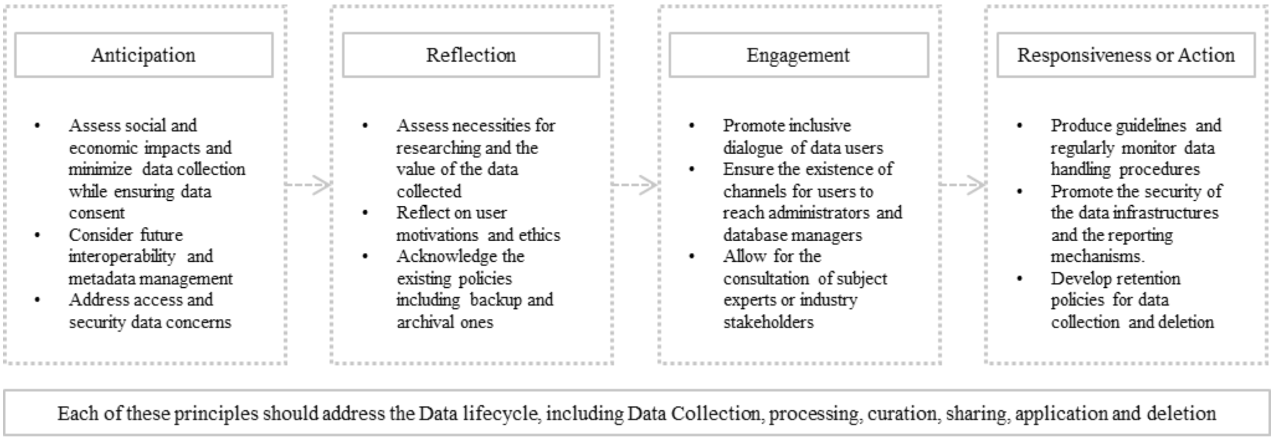


**Figure 11. The Responsible Data Governance Framework provided by Fothergill et al. [41]**

## 5- Unified Data Governance Framework

Within this section, we provide a discussion over each Research Question (RQ), from RQ1 through RQ3, while detailing the results obtained to answer them, as to design and define a novel and enhanced data governance framework that synthesizes and incorporates the best features of existing frameworks and integrates GenAI specific requirements.

### *5-1- Design of the Unified Data Governance Framework*

Considering **Research Question 1 (RQ1)**, we proceeded with the exploration and revision of multiple data governance frameworks, evaluating and comparing them through a systematic approach based on the critical attributes we consider vital for any framework which led to the results presented in Table 4, characterizing each of the frameworks analyzed namely, A - The Framework for Data Decision Domains, B - The IBM DG Framework, C - The DG Institute (DGI) Framework, D - The Big Data Algorithms Systems (BDAS) DG Framework, E - The Analytics Governance Framework, F - The Conceptual Framework for Data Governance, G - The Analytical Framework for DG in Digital Platforms, H - The Big Data Governance Framework, I - The DG Framework for SMEs, J - The Data Governance Center of Excellence, K - The Quality of Data in Motion Framework, L - The DG Gamified Framework, M - The Data Middle Platform Framework, N - The Industry 4.0 DG Framework, O - The Data Governance Framework for Big Data, P - The Responsible Data Governance Framework. Considering this, these frameworks were compared against a list of seventeen key attributes that we consider as the best practices from the merge of each framework's attributes and that all frameworks should encompass. Thus, our framework must include all attributes as illustrated in Figure 14.

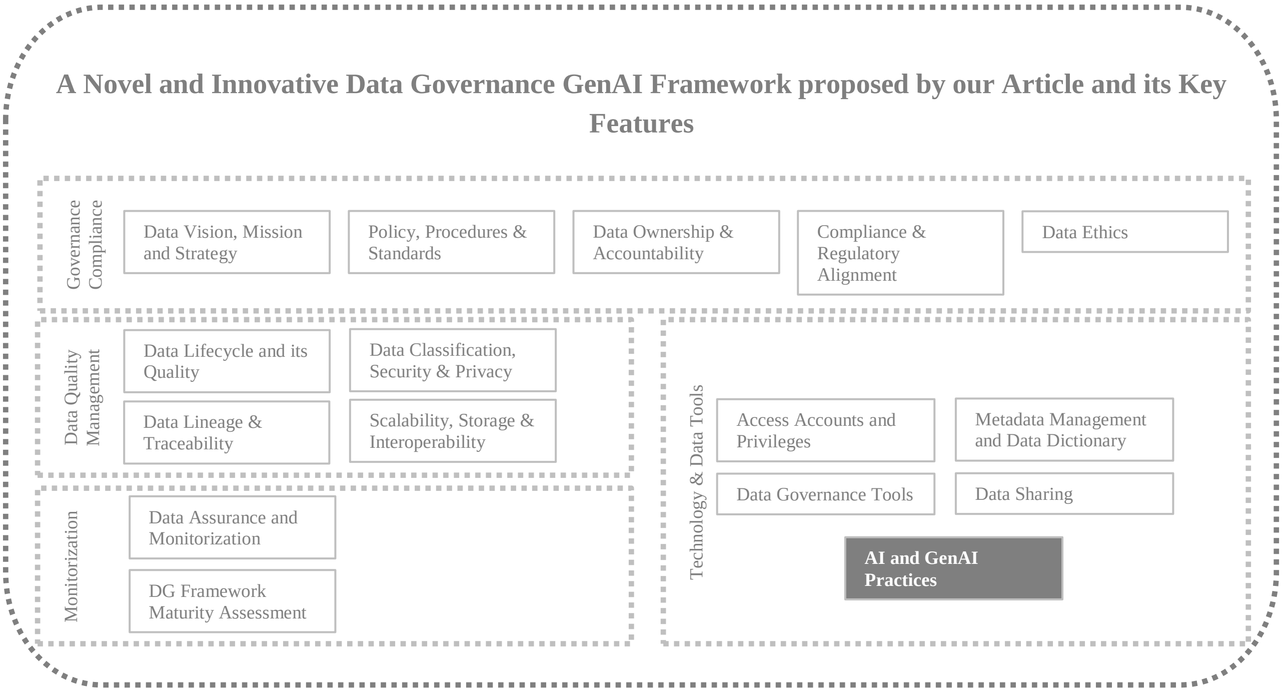


**Figure 12.** **Governance GenAI Framework proposed by the authors and its Key Features**

These components and features can be described as:

- **i) Data Vision, Mission, and Strategy**, the starting point for any framework definition, which reflects and encourages the definition of a clear mission, vision, and strategy regarding a given organization's data-related assets, activities and users, while defining a short- and long-term roadmap of initiatives aligned with business objectives to ensure the definition of a robust data governance framework supported by an equally strong data culture;
- **ii) Policy, Procedures, & Standards**, which are the basic rules that support a framework and the overall practices that relate to data and their users, including the definition of policies and procedures that guide all data-related activities, from governance and structure, quality management to data retention, security, storage, and destruction;
- **iii) Data Ownership & Accountability**, which relates to the attribute that deep-dives into how a given framework should be operationalized in terms of data roles and ownership across the organization's data assets, including the definition of these roles, from data stewards, custodians, to data owners, users, and executive sponsors;
- **iv) Data Lifecycle and its Quality**, which encompasses the attributes that organizations must establish that ensure data quality attributes, from integrity, completeness, consistency, availability, and accessibility, among others, remain preserved throughout the data flows and data lifecycle;
- **v) Compliance & Regulatory Alignment**, yielding the framework with tools to ensure that it remains adherent throughout time to legal regulations and directives. It details how governing bodies should develop policies and procedures for data use, ensuring alignment with the organization's overall vision, mission, and objectives;
- **vi) Metadata and Data Dictionary**, relates to the process of identifying, indexing, and cataloguing data assets and their attributes, to make them easily understandable, obtained, and regularly updated, including the business metadata and notions, technical and operational metadata;
- **vii) Data Classification & Security**, which regards organizations' approaches to ensure that access, breaches, and threats regarding cybersecurity are prevented, monitored, and controlled, including techniques such as data redundancy, encryption, anonymization, multi-factor authentication, and data masking, among others;
- **viii) Data Assurance and monitoring**, that relates to the activities that support organizations in preventing, detecting, and mitigating their data-related risks through regular audits and tests over their internal control;
- **ix) Scalability, Storage, & Interoperability**, through which data governance frameworks can address the growing needs of organizations, including the accommodation of a rise in data complexity, volumes, sources, and emerging systems while enabling the usage of different integrated systems;
- **x) Data Ethics**, which relates to the ethical aspects and considerations that organizations must address to ensure that their data-related activities are performed in a responsible, fair, transparent, and unbiased manner, especially in the data-gathering, analysis, and decision-making phases. Our framework introduces a dedicated Data Ethics domain designed to ensure that organizations address critical dimensions of ethical AI governance, including

fairness, model accountability, and explainability, while supported by GenAI initiatives that aim to define and measure these concepts through well-defined indicators, enabling continuous monitoring and ethical assurance across all stages of the data and AI lifecycle;

- **xi) Access Accounts and Privileges**, which defines the organization's approaches to managing the access creation, change, and removal processes to specific data within the organization's data storage and systems infrastructure, preventing inadequate roles assignation and unauthorized access to programs and their data;
- **xii) Data Lineage & Traceability**, relating to the ability that ensures organizations are rapidly able to demonstrate and audit the origin, data flows, and transformations that their data points have over time and organizations' systems;
- **xiii) Data Sharing**, where organizations' data governance framework focuses on securing and setting principles for data sharing across different business units, organizations, third-party suppliers, and competitors, among others, allowing data sharing to be performed under complete and robust policies that secure this collaboration;
- **xiv) Data Governance Tools**, relates to the software enablers supported in a technological infrastructure that aid the execution of robust and more precise data governance and quality initiatives, including management and monitoring of data quality, metadata catalogue, and business data dictionaries, data traceability and their lineage;
- **xv) DG Framework Maturity Assessment,** where frameworks provide a set of principles that allow organizations to perform and apply structured assessment models to evaluate their data governance maturity and define at which level organizations stay – from initial or ad-hoc (usually the lowest level), managed, defined to measured and optimized (typically the highest level). These practices help organizations acknowledge where they are in terms of data governance maturity, enabling the definition of a roadmap of activities to increase that given status;
- **xvi) AI and GenAI Practices**, where with the recent exponential evolution of Gen AI tools – ChatGPT, Copilot, Bard, among others – data governance frameworks must encompass a set of initiatives that help organizations in managing, regulating, and controlling their AI-driven systems and models, ensuring that their usage of these systems is performed in a governed manner, including initiatives that cover their training and testing data sources, data sources, model transparency and bias levels, model outputs, among other critical activities.

### *5-2- Phased Implementation of the Proposed Framework*

Considering the number of dimensions and their overall specifications, we consider that our framework should be employed based on a phased and systematic methodology, which is represented by five layers each encompassing a set of dimensions. Each layer is presented as a five-layered pyramid metaphor, considering each component selected based on their strategic priority, time to implement and organizational cultural readiness level. Hence, the phased process corresponds to the following, namely:

- **1st Layer – Governance & Compliance**, where organizations should dive into define and implement the dimensions of i) Data Vision, Mission, and Strategy, ii) Policy, Procedures, & Standards, iii) Data Ownership & Accountability, v) Compliance & Regulatory Alignment and x) Data Ethics;
- **2nd Layer – Data Quality Management**, representing the stage where organizations strive to implement the data governance more operational and technological dimensions, such as iv) Data Lifecycle and its Quality, vii) Data Classification & Security, ix) Scalability, Storage, & Interoperability and xii) Data Lineage & Traceability;
- **3rd Layer and 4th Layer – Technology and Data Tools** phase 1 and 2, where dimensions related to specific data governance operational processes should be established, including dimensions such as at phase 1, vi) Metadata and Data Dictionary, xiii) Data Sharing and xi) Access Accounts and Privileges and phase 2, xvi) AI and GenAI Practices and xiv) Data Governance Tools;
- **5th Layer – Data Assurance & Monitoring**, where at this dimension organizations look at periodically assess, assure and monitor their data governance framework maturity and define a roadmap of initiatives to increase their level of maturity and enhance the overall practices, including the implementation of dimensions such as xv) DG Framework Maturity Assessment and viii) Data Assurance and monitoring.

The application of this phased cycle, as illustrated in Figure 15, is only possible to be effectively implemented if organizations can define and establish the scope of their data governance framework, including processes, systems and people, enabling them to answer what their data governance intend to focus on, including granular activities such as reports, systems, departments, amongst other. Our data governance framework is designed as a modular and layered model to accommodate organizations with varying levels of data maturity and to allow for progressive definition and implementation of different data governance instruments. The framework comprises five interdependent layers that can be implemented progressively and adapted to each organization's context and readiness: 1st Governance & Compliance, 2nd Data Quality Management, 3rd and 4th Technology and Data Tools and 5th Data Assurance & Monitoring. This layered architecture enables a phased and flexible implementation pathway, allowing organizations to build and strengthen their data governance practices over time, according to their maturity level and strategic priorities.

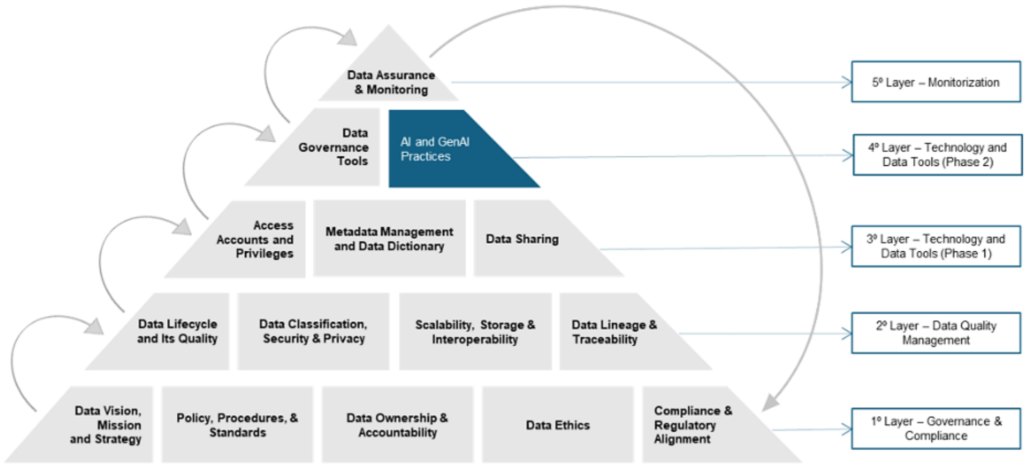


**Figure 13.** **Phase and Cyclic Implementation of the Novel and Innovative Data Governance GenAI Framework.**

Table 4 provided a complete list of the frameworks compare denoting with a tick mark (✓) per attribute the frameworks that have practices that cover that specific attribute in their detailed approaches and applications.

Figure 16 demonstrates each attribute analyzed per framework, denoting the afore conclusions:

- i) most of the frameworks analyzed contain a set of practices that cover the areas of Data Vision, Mission and Strategy (14/16, 88%), Data Ownership & Accountability, and Data Lifecycle and its Quality (13/16, 81%), Policy, Procedures & Standards (12/16, 75%) and Data Assurance and Monitorization (10/16, 63%). This result demonstrates that a framework must encompass the definition of a vision and strategy, supported by robust and formalized policies and procedures and roles and responsibilities; Besides, about 31% to half of the frameworks relate to techniques that regard Compliance & Regulatory Alignment (5/16, 31%); Data Quality Management process: Scalability, Storage & Interoperability (8/16, 50%) and Data Lineage & Traceability (5/16, 31%); and Technology and data tools, namely Metadata Management and Data Dictionary (8/16, 50%) and Data Sharing (5/16, 31%). This result aligns with the literature coverage over recent times, where these topics, although newer, are becoming popular among studies and thus within frameworks, thanks to tools like metadata catalogues and data dictionaries. Also, newer frameworks tend to provide an extended version of a governance approach that focuses on specific purposes or sectors rather than general ones;
- ii) about 19% to a quarter of the frameworks related to topics regarding Data Classification, Security & Privacy (4/16, 25%), Access Accounts and Privileges (4/16, 25%), Data Governance Tools (4/16, 25%) and DG Framework Maturity Assessment (3/16, 19%). This demonstrate that data security, privacy, and access management remain outside the scope of most frameworks due to their specific complexities. However, their increased importance is undeniable, and we firmly believe that it must not be disregarded. In line, frameworks disregard the connection of their initiatives with data governance software tools, including such as e.g. Atlan, Collibra Data Governance, Atalion Data Catalog, Informatica, Azure Data Catalog, Google Cloud Data Catalog, SAP Master Data Governance, OvalEdge, PowerBI, Transcend, Erwin, Apache Kafka, Talend Data Fabric, Salesforce Security and Privacy, CastorDoc, Snowflake Data Governance, Satori Security Data Platform, Teradata Vantage, Lyftrondata, Securiti, Microsoft Purview, Secoda, Netwrix Auditor, Avo, Immuta, Informatica Cloud Data Quality, DataPrivilege, Apache Atlas, Ataccama One, among other options. In line, regarding the last attribute, Data Governance Maturity Assessment, our results show a relatively limited number of frameworks consider practices regarding regular evaluations of a framework and its strategy to increase this maturity. What we have learned is that these assessment techniques take their form as sole frameworks themselves and not as an attribute within a given framework, thus, most maturity assessment are specific frameworks, i.e. two five-level maturity assessment frameworks presented by Zorrilla & Yebenes [12] and the Data Governance Maturity Model (DGMM) presented by IBM and analyzed by Belghith et al. [69].

iii) one isolated framework focused on data ethics and transparency, which suggests that, in terms of data integrity ethics, there is considerable room for further studies and improvement in its overall maturity. Considering the recent exponential increase in usage of AI models, including Gen AI ones, data ethics, bias control, and transparency are becoming more and more important to ensure that data-related activities are performed in a responsible, fair, and unbiased way. Also, none of the frameworks analyzed demonstrate any activity or practice that a framework could have to support AI models and systems, including GenAI.

**Table 4.** **Systematic Comparison of Data Governance (DG) Frameworks Based on Key Main Domains**

| # | DG Framework Tittle | i) Data Vision, Mission and Strategy | ii) Policy, Procedures & Standards | iii) Data Ownership & Accountability | iv) Data Lifecycle and its Quality | v) Compliance & Regulatory Alignment | vi) Metadata and Data Dictionary | vii) Data Classification & Security | viii) Data Assurance and Monitor | ix) Scalability, Storage & Interoperability | x) Data Ethics | xi) Access Accounts and Privileges | xii) Data Lineage & Traceability | xiii) Data Sharing | xiv) Data Governance Tools | xv) DG Framework Maturity Assessment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | The Framework for Data Decision Domains | ✔ | | | ✔ | | ✔ | | | | | ✔ | | | | |
| B | The IBM DG Framework | ✔ | ✔ | ✔ | | | ✔ | ✔ | ✔ | ✔ | | ✔ | ✔ | | ✔ | ✔ |
| C | The DG Institute (DGI) Framework | ✔ | ✔ | ✔ | | | | | ✔ | | | | ✔ | | ✔ | ✔ |
| D | The Big Data Algorithms Systems (BDAS) DG Framework | | | | ✔ | | | ✔ | | ✔ | | ✔ | | | | |
| E | The Analytics Governance Framework | ✔ | ✔ | ✔ | | | ✔ | | ✔ | | | | | | | |
| F | The Conceptual Framework | ✔ | ✔ | ✔ | ✔ | ✔ | | | | | | | | | | ✔ |
| G | The Analytical Framework for DG in Digital Platforms | | | ✔ | ✔ | | | | | | | ✔ | | | | |
| H | The Big Data Governance Framework | ✔ | ✔ | ✔ | ✔ | | | | | | | | | | ✔ | |
| I | The DG Framework for SMEs | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ | | | ✔ | ✔ | | |
| J | The Data Governance Center of Excellence | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ | | | | ✔ | | |
| K | The Quality of Data in Motion Framework | ✔ | ✔ | ✔ | ✔ | | ✔ | | ✔ | ✔ | | | | | | |
| L | The DG Gamified Framework | ✔ | | ✔ | ✔ | | | | ✔ | ✔ | | | | ✔ | ✔ | |
| M | The Data Middle Platform Framework | ✔ | ✔ | ✔ | ✔ | ✔ | | | ✔ | ✔ | | | ✔ | ✔ | | |
| N | The Industry 4.0 DG Framework | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ | | ✔ | ✔ | | | ✔ | ✔ | | |
| O | The Data Governance Framework | ✔ | ✔ | ✔ | ✔ | | ✔ | | ✔ | | | | | | | |
| P | The Responsible Data Governance Framework | ✔ | ✔ | | ✔ | | | | | | ✔ | | | | | |

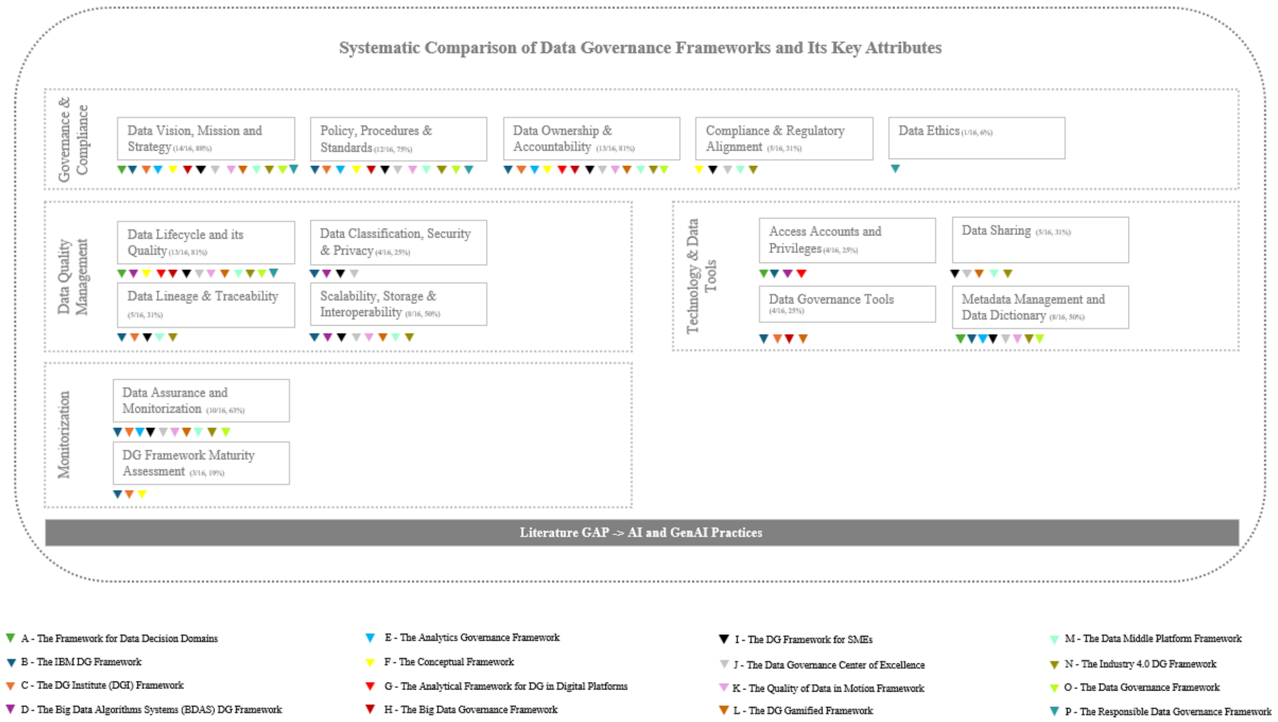


**Figure 14.** Graphical Results of Systematic Comparison of Data Governance Frameworks based on Key Main Domains

### *5-3- Relationship Between Data Governance and GenAI*

Considering **Research Question 2 (RQ2)** and as the practices of AI and their need to be integrated into a data governance framework continue to grow and evolve, legislation has become a vital source of information that supports organizations in framing the limits of their AI and data governance as ethical, trustworthy and responsible practices. Hence, government institutions from several locations have begun formalizing, implementing, or proposing a regulation that guides their organizations [42, 70]. From the legislation point of view, the literature highlights their importance and the challenges that organizations can face if they are not compliant, including legislation such as the European Union's Data Governance AI Act (DGA), which was introduced as one of the first legislation around the AI, and considered as key in shaping the global legislation, specifically by introducing key legislative domains, including an AI systems' risk-based classification, trustworthy and transparent requirements and strict governance over biometric surveillance and high-risk AI systems [71]. This Act underwent significant revisions and adaptations following the rapid emergence of the ChatGPT system, among others [43]. This AI system is recognized as one of the pivotal milestones in the history of GenAI [42, 70]. Similarly, in Canada, an Artificial Intelligence and Data Act (AIDA) was proposed as a standard to support organizations in conceptualizing their framework standards addressing the data risks and challenges associated with the AI models usage. This Act focuses on establishing a clear relationship between a risk-based approach for AI and complementary disciplines, such as data privacy and security, with a special focus on the intersection of AI and the protection of personal sensitive Data (PIPEDEA, Canadian Personal Information Protection, and Electronic Documents Act). To do so, this Act defines that organizations' AI systems must identify, monitor, mitigate, and be transparent on the potential risks that AI can pose to their internal and external environment before their deployment, setting oversight rules and penalties for non-compliance with the Act [70]. Besides, the AU (Africa Union) Data policy was introduced in 2022 to denote the importance of cross-sector collaboration and interaction between African regions to manage and govern data, which can be vital in ensuring that infrastructures can take the most out of their data usage, generated by different and several means, and that does not harm the overall public, including people's sensitive data [42].

Moreover, GenAI must not be isolated when it comes to available data governance frameworks, as similar to any data-related activity, it involves the governance of data, business processes, systems, and people. Hence, a given framework must ensure that authorities regarding GenAI systems are defined and controlled, establishing and communicating the necessary organizational changes to incorporate the specific requirements of this type of AI [12]. Similar to any traditional program, one that does not yet consider the specifications of the discipline of GenAI, organizations must consider that to advance with a data governance framework that includes this topic, it requires a significant investment of tools and resources – including people, time, funds and the systems – for an organization to address this ever emergent science fully [26, 72, 73]. Nonetheless, organizations already acknowledge GenAI, however their boards tend to focus on investing and enhancing their cybersecurity, usually neglecting financial and human resources investments in establishing a more robust framework. Therefore, we agree that cybersecurity is significant in data governance and GenAI. However, these two disciplines involve additional components that should also be considered as priorities, for which so far, studies have shown that are often neglected [72, 74]. Academic surveys have denoted that the majority of businesses are still lacking a robust framework for their traditional data usage, symbolizing that there is still a long path in maturity terms [72, 75, 76]. In light of this, while data governance has enormous potential, guidelines and academic surveys on it are rare and usually presented in theoretical terms [77]. Besides, organizations are often more concerned with proceeding with the application of AI tools rather than focusing on their training dataset quality and each of the data lifecycle stages [4, 72]. As GenAI evolves and grows more complex, the challenges associated with it and data governance are intensifying daily. Nonetheless, existing literature continues to highlight a critical gap – there is currently no standardized or unified framework specifically designed to address the unique specifications of GenAI within organizations through a formalized and methodologically consistent approach [4, 19, 72].

### *5-4- GenAI Initiatives to be Included in Our New and Enhanced Data Governance Framework*

Concerning **Research Question 3 (RQ3)**, we present the initiatives our framework contains regarding AI and GenAI. Accordingly, the literature denotes a clear consensus on the importance that AI has within the context of a data governance framework and its urgent need to be capable of integrating the specifications of AI [48]. AI is currently being discussed broadly within literature, namely in defining the challenges (legal, regulatory, ethical) that organizations can face when applying AI, where its integration with the data governance field is still scarcely studied and analyzed, being the frameworks presented in the literature not considering this integration [48]. When considering this integration, we propose organizations to perform the following initiatives:

- **Intensive and comprehensive training and continuous support**, where organizations' data governance and responsibilities are defined in a manner that ensures the admission of AI specialists, who can be vital guidance and assistance for data owners and users transversal to the organization and that promote the occurrence of training sessions where both fundamental and in-depth Generative AI concepts are approached [48].
- **Policies and Procedures** that involve the Data Modelling and Business Glossary (data catalogue), Data Virtualization, and Data engineering to ensure that all of the five layers of information are formalized and periodically inspected and that GenAI models are explicitly presented and controlled by their stakeholders, including the following layers i) training datasets for the initial models, in file text, audio, image or video format, ii) domain-specific fine-tuning training datasets, labelled to allow for the improvement of the initial training set and challenge of the AI model, iii) the neural network architectures, which corresponds to a config file, iv) AI trained models with its overall parameters, presented within a weights file and v) the scripts for training and outcome generation, which are included within software code [78];
- **Access management and monitoring of AI systems**, including control of system and data permissions to AI models, from training data and model development to documentation and output, as well as follow-up adaptations. Hence, organizations should establish robust authentication mechanisms, authorization and authentication standards and mechanisms as well as layers of responsibilities and ownerships that can ensure adequate access to data and applications [51]. In line, the EU Data Governance AI Act defines that data monitoring mechanisms must be established to prevent bias, discrimination, and unethical data use, which can be achieved through i) regular internal and external audits and assessment of AI models and their training datasets, mandating that the AI developers and users formalize their data processing and analysis approaches; and, ii) definition and implementation of an independent board responsible for data oversight which a focus on any data governance practice and policies that relates to AI systems [71];
- **Continuous monitorization over the training data**, including implementing mechanisms such as version control and log monitoring, which allows users not only to acknowledge and identify data consultations, insertions, changes, and updates or deletions but also, to visualize all of the key user interactions with that data, allowing the management of multiple versions (audit trail) including quick data restores [51];
- **Promotion of mechanisms for the real-time monitoring compliance standards**, namely to keep up with the AI's evolving requirements, through i) AI auditing techniques to automate and continuously monitor their policies and procedures that support their AI models and their outcomes, including risks classification systems and impact assessments and ii) create a set of AI governance sandboxes, where both AI developers and business users can test AI models in a controlled environment before its official production deployment [43];
- **Definition and implementation of measurable indicators to evaluate the data governance capability to address GenAI risks**, including i) transparency and explainability (ensuring AI processes and decisions are clearly trackable, understandable and that these decisions can be inferred, analyzed, interpreted and justified), through key metrics and methods such as the Shapley values (quantification of each feature contribution to a specific prediction, ensuring fairly distribution of features' importance), Model-agnostic metrics (LIME, generation of simple and interpretable explanations for individuals predictions by analyzing the effects of the features on model outputs,

aggregating these scores to provide a holistic view of feature contributions) and visual explanations (through the usage of methods such as the Grad-CAM, a model that class-discriminates localization producing visual explanations from any CNN-based network and Guided Backpropagation, technique that provides a pixel-space gradient visualization of pixels that had a positive influence on the prediction); and ii) bias (evaluating how the framework reduces unfair, discriminatory and premediated outcomes), through the implementation of metrics and methods such as the statistical parity (fairness metric focus on ensuring equal proportionality within positive outcomes across different groups, eliminating disparities), the Gini coefficient measures (assessing inequality in the distribution of the outcomes predicted, ranging from 0 to 1, where 1 sets the maximum inequality and 0 the lowest), and accuracy measures (raging from pre-processing, in-processing and post-processing methods to analyze, modify and control datasets of data before, during and post training to reduce and mitigate bias) [79];

- **GenAI systems,** aim at producing data that was not yet seen and explored, as such, frameworks must define standards that i) rule, identify, and catalogue new data introduced within the organization databases, including perming testing procedures over its data quality attributes (accuracy, completeness, timeliness, among others), ii) define clear hierarchical responsibilities and accesses attributions, iii) set protocols over the monitorization of the changes and accesses to these data made over time [51];
- **Data privacy and security standards** that focus on safeguarding the lifecycle of GenAI, enforcing rules that keep data assets – training data, AI model, its documentation and log, outputs, and others – safe independently on their source and storage employing tools such as firewalls, antivirus, encryption on the network [51];
- **Alignment of AI systems with Ethical AI guidelines**, where the framework must encompass the ethical principles of fairness, accountability and transparency through a set of data principles that establish standards that ensure that AI systems, their training data sets, and models are continuously tested for fairness, accountability, and transparency before deployment and its explainability is transparent and understood [71];
- **The Data governance over the AI algorithms transparency and requirements**, where organizations must ensure that their AI developers and users document and define how algorithms' predictions and analysis are being generated and what the controls performed over their training datasets to inform those outputs [15].
- **Adoption of AI data governance initiatives** that are presented by the regulation guidelines and standards, including i) data reuse following legal and ethical guidelines, where organizations must implement mechanisms, data-sharing agreements, anonymization methods, and cross-sectoral collaboration guidelines; ii) data altruism and intermediation services, where organizations set their grounds in enabling their stakeholders to voluntarily share and provide their data for research and AI innovation purposes, as such organizations must identify data that can be shared and define its security, anonymization, and trust methods [71];
- **Definition and designation of a joint accountability model,** where structures allow organizations' stakeholders to share responsibility structures, in which AI developers and users collaborate directly and indirectly with policymakers and data users to govern and monitor GenAI outputs while preventing AI developers to be dominant and single entities with full responsibility over the GenAI outcomes [52];
- **Attribution and implementation of an AI governance council**, which should be composed of stakeholders from different expertise, including legal professionals, AI developers responsible, ethicists, and business directors to oversee the implications and value of the organization's GenAI models [43];
- **Internalize and develop owned GenAI models and chatbots**, ensuring that the organization is responsible for the development of its models and chatbots and that specific guidelines are formalized, from obtaining training data to generating and analyzing AI-generated outcomes. The chatbot interfaces should promote educational moments for employees on the data governance policies, including ethical implications of AI usage, and contain periodic reminders about best practices to create awareness and generate a mindful culture [80];
- **Incorporation of data analytics maturity frameworks,** to monitor the maturity of GenAI models and their training and generated data. A framework such as the Consensual Big Data Assessment System (CBDAS) can be applied to evaluate the organization's readiness and maturity for adopting AI models, assessing the maturity concerning organization's AI data, encompassing its data strategy, data architecture, IT systems, infrastructure, and human capital skills. This assessment provides a structured set of 59 questions, where 44 focus on the maturity level of the critical success factors (CSFs) and the remaining address the business needs [81];
- **Adoption of the data principles presented by the FAIR framework**, where each principle represents the following: i) Findable, ensures that data are always easily identifiable for any user who needs them, where data should have a unique and permanent identifier and be accompanied by metadata; ii) Accessible, focuses on data availability and accessibility, ensuring that data can be accessed immediately when required and ensuring that storage capacity is sufficient to accommodate all data; iii) Interoperable, data is seamlessly integrated into tools and systems and utilized with other data, where organizations store data in formats aligned with the business standards; and iv) Reusable, ensuring that data are available for reuse and can be analyzed multiple times as needed, while safeguarding the organizations resources in an adequate and readily accessible manner [82].

# 6- Discussion

## *6-1- Research Focus Group Activity – Demonstration and Evaluation*

During this stage, our proposed framework is demonstrated and assessed in terms of its utility, contribution and validity to the data governance and GenAI fields [83]. The main goal of this stage is to obtain insights from experts in these fields by promoting their interaction with our value proposition, namely the unified data governance framework, the relationship between data governance and GenAI, and the proposed GenAI initiatives. To achieve this, we organized a focus group that included various discussion topics, namely:

- Topic 1 - Challenges of data governance and the issues that organizations and practitioners' encounter;
- Topic 2 - The overall contribution of the proposed framework to Data Governance and Generative AI;
- Topic 3 - Reflections on what has been studied, including the relevance and validity of our proposal;
- Topic 4 - Opinions, feedback for improvement, and views on future research directions and next steps.

Considering this, the focus group activity was divided into two parts, where the first three topics, were closed-ended questions, enabling us to gather and comprehend the interviewees' opinions on the framework, the relationship between the two fields, and the proposed initiatives. This aided in measuring their overall usefulness and significance both to the academic and business communities. The subsequent and final topic centered on obtaining positive and negative feedback on what was proposed, which we regard as a vital source of knowledge drawn from experts' experiences to enhance our proposition and define next steps. Thus, it was essential to identify areas of strength and improvement, and others that set the basis for future research. Moreover, and in order to obtain key and robust contributions, we have defined a selection criteria for participants, ensuring that interviewees represent individuals who are experienced i) in scientific research in the fields of Data Governance and GenAI, ii) in leading and managing the definition, implementation, and monitoring of data governance frameworks in various organizations, and iii) in conducting data governance framework maturity assessments using rigorous methods contributed to feedback that is objective, essential, and less biased. The list of participants and their backgrounds are summarized in Table 5, specifically:

**Table 5. Identification and Characterization of the Focus Group Participants**

| Participants | Background/ Expertise |
|---|---|
| Interviewee (I1) | A person with almost 16 years of experience leading and managing data governance frameworks in three different multinationals' organizations from the Industry Segment. Also, part of one of the Major and most important Associations in the Data Governance Field, the Data Management Association (DAMA). |
| Interviewee (I2) | A person with 6 years of experience in the Financial Services (Banking and Insurance) segment, namely in data governance assurance, performs periodic assessments of the maturity of these organizations' data governance frameworks, ranging from more mature to less mature ones. |
| Interviewee 3 (I3) | A person with 15 years of experience working with databases and warehouses, and 5 years in Lecturing Data Governance courses at a prestigious University. This person's field of research focuses on databases, Information Technology, systems and Data Governance. |

Consequently, the participants were asked to evaluate the proposed framework and its key contributions based on specific evaluation criteria, including relevance, clarity, value, and applicability. The insights provided by the participants addressed the four key topics outlined above, contributing directly to the overall study findings, namely:

- Topic 1: The interviewees identified various ongoing challenges in data governance and GenAI, both at an organizational and individual level. From the challenges shared, the most discussed regarded the ambiguity within the goals of governance initiatives and the daunting complexity of current frameworks, where Interviewees agreed that many organizations have difficulty in defining the scope of what requires governance, often attempting to consider all data and assets at once, which from they have experience in the past and consider unsustainable. Besides, this lack of focus is exacerbated by uneven adoption across the different organizations' business units, resulting in siloed practices and fragmented cultures. Technical challenges were also a significant concern that has yet to be solved, particularly emphasized by the inability of existing tools to automate data lineage across systems and the significant costs;

- Topic 2 and 3: The interviewees expressed a consensus that the unified framework, extended to feature GenAI, effectively addresses numerous challenges by providing a more innovative, organized and comprehensive approach, being one of the most appreciated features, the integration of GenAI specifications, an aspect that the sixteen frameworks evaluated overlooked. In fact, the interviewees stressed that this is not only overlooked in the literature, as organization's practices are also still little to non-existent. Besides, the interviewees emphasized the framework's adaptability to various organizational contexts and potential compliance with regulatory requirements, especially in industry, banking, and insurance. Nevertheless, some interviewees emphasized the need for more actionable and step-by-step guidance and operational clarity. They suggested that future research contributions should focus on clearly identifying strategies for implementing data governance. These should be segmented and grouped by phases, extended over time, and connected to clearly defined data products, such as meaningful reports or datasets spanning multiple departments;

- Topic 4: The interviewees pointed areas which must be lines of research for these two fields, namely through i) evaluate the potential of this research through practical case studies in an organization, analyzing what dimensions are easier and adopted in a shorter period from those where the resistance of employees and the difficulties in implementing the dimensions is more complex; ii) the analysis of the advantages that companies can obtain from having a dedicated governance committee, established at a high hierarchical level; and, iii) study on the need for more advanced and complete tools to support the automation of the application of a given framework, specially to accommodate the specifications of GenAI and data lineage. Lastly, there was a widespread discussion on the success of a framework, where the interviewees reinforced that only if a framework is backed by a well-stablish culture, clear functions and roles and stakeholder participation, can such framework have success not only in terms of using data as a strategic asset.

### *6-2- Research Focus Group Activity – Discussion of the Demonstration and Evaluation*

Considering the above-described results, we consider that this activity supports us in reaching a conclusive overall discussion on several aspects. In fact, the real organizational examples obtained from the participants' tenure were a vital source of accurate information to support not only this work but any research on these fields. Regarding the first topic, it was explained by the participants that the current challenges that these fields are facing, all reside in three major components, people, process and systems, where the first, people, can be considered as one of the major that heavily impact the success and effectiveness of any framework. Equally, participants agreed that organizations have tremendous difficulties in answering the question of what the scope of their framework is. From the participants' experience, it was noted that organizations are trying to govern all the data, their data assets and activities, which is considered a nearly impossible task, due to the constraints of not only human capital, financial and time resources. Therefore, the participants denoted that organizations must define clearly what is the data, data assets, activities that they pretend to govern and manage in a more detailed manner, distinguishing from those that they know data exists but it is not significantly relevant e.g. non-relevant data, such as quality and development environment data, sandbox data, public data, amongst other. In line with our background research, participants acknowledge that cultural resistance to change is still a strong pain point, where frameworks demands are not yet comprehended and valued throughout the organizations' different business units, leading to silos and fragmented practices. Similarly, from the participant's experience with several data governance tools, it was noted that although they provide benefits and foster a framework, their functionalities and features are still at a very early stage of maturity, where for example participants denoted that these tools were very capable of mapping and identifying the interfaces and data integrations in a given accounting system, but have a limited capacity to map and describe all the integrations between different systems, usually performing this manually. Also, these tools are typically commercialized through expensive licenses, where the participants argue that such tools are not yet perceived as a value-added investment, where the End-User Computing (EUCs) mechanisms, such as Microsoft Office abilities (e.g. Excel, Access, Word, amongst others) prevail over these more evolved but yet limited tools.

Regarding the overall contribution of our work, the interviewees agreed positively on the criticality and significance of such work. To support this consensus, the participants highlighted that this research by digging into the existing different frameworks, including both general frameworks, but also context-specific ones, has a great potential in providing a unique, up-to-date and robust source of knowledge for both companies and academia, including roles such as data analysts, engineers and architects. In line, the participants highlighted that specific context frameworks tend to present more technical and complex terms, focusing on specialized researchers and users. Besides, interviewees highlighted the benefits of exploring the best features of each framework and framing those into segregated but interrelated dimensions through a robust methodologically approach, which strived to look for the best quality criteria and provide a highly valuable research, that brought both a systematic identification and comparison of both general and context specific frameworks, an in-depth analysis over the GenAI integration with data governance and lastly, the integration of a specific GenAI dimension into our framework. Regarding the interviewees' suggestions on the overall validity of the proposed framework, it was proposed that our framework presented should resemble a pyramid scheme layered down by the different importance and priority levels of implementation of each given dimension of the framework. Accordingly, the participants consider that one the challenges is that organizations pretend to implement a robust framework, "as soon as possible" and encompass all the data, which corresponds to a "nearly impossible" activity. Thus, by providing a framework that highlights explicitly the different layers by order of importance, organizations can prioritize and start implementing their framework in a phased and spread manner. This is relevant as although the framework's dimension are related, they should be employed in an ordered cycle that in the early stages prioritizes the definition and formalization of the data governance, mission, strategy, scope, structure, policies, procedures, roles and responsibilities, and then after organizations should strive to optimize operational activities, such as technologies, data quality standards, data risks and controls, monitorization, amongst others activities. Equally, one of the arguments presented by the participants is that frameworks provided by the literature present their dimensions in a static visual figure, which in turn, can be misleading to organizations where they will strive to approach all the dimensions at once.

Additionally, regarding future research directions, the participants' inputs were highly relevant, providing research topics such as an in-depth research on i) the existing tools and applications that can be used to automate, support and robust the data governance framework implementation process; ii) the clarification and standardization of the majority of the notions presented within the data governance available literature, where participants emphasized that each literature seems to have their unique perspective on the definition of data governance terms, for example, a data dictionary is seen throughout the literature in different notions, which makes it harder for organizations and data governance practitioner to distinguish and decide what is the better suited and most factual and reliable definitions; iii) application of this data governance framework in a real organizational environment and analysis of the overall results, included metrics on the time that each phase took to be effectively implemented, changes on the people (cultural and awareness) and on processes and systems (data quality indicators and key risks indicators); iv) the application of inter-rater reliability or triangulation of our findings to validate and corroborate qualitative insights.

### *6-3- Research Contributions*

Throughout our research, we explored through a rigorous systematic approach the literature and experience available on these fields, as to acknowledge the objectives defined and thus address the topics within our research questions. Based on these activities, we strongly believe that our work substantially supports these fields and brings added value to any organization or researcher, aiding them in becoming more aware, and having access to a robust, unified and extensive basis of knowledge, as such we concluded that our contribution provides robust insights on:

- A solid, unified and comprehensive data governance and management unified framework which results from the integration of the best features of several general and context-specific data governance frameworks, extended to encompass a specific and inter-related dimension that focus on providing initiates for organizations to address the GenAI features and specifications gaps identified by several different authors including Gardner et al. [20], Hikmawati et al. [21], and Kanying et al. [8].
- ii) A unique and systematic comparison analysis on data governance frameworks and the in-depth analysis of those frameworks both at an individual-level but also at a comparable set, based on key data governance aspects and criteria, identifying how they converge and how they differ from each other. This allowed not only to provide a sufficient basis of knowledge and maturity in data governance frameworks and practices, gaps identified by MacFeely et al. [2], Nokkala et al. [11], and Scholz et al. [3];
- iii) An extensive and comprehensive research on the role of leadership in a data governance framework, not only through the set of a C-level role, such as the Chief Data Officer and the definition of both data governance and AI committees fostering the research of how can organizations look into cultural and motivational changes to address their employee's resistance and business units' silos and engage the stakeholders who are vital for the coordination, funding, investment and engagement of any data governance framework. These gaps were identified by various sources, including Cheong & Chang [24], Janssen et al. [23], Lis & Otto [6];
- iv) Our proposed framework and overall literature analysis, includes specific dimensions and understandings on how can a organizations and its framework address the rapid technological advancements (ix – Scalability, Storage & Interoperability dimension), the constantly involving data protection measures and legislations (v – Compliance & Regulatory Alignment dimension), the increased usage of IoT ecosystem and data governance tools within the business units context (xii – Data Lineage & Traceability, xiii – Data Sharing and xiv – Data Governance Tools dimensions), definition of performance metrics and data quality indicators (iv – Data Lifecycle and its Quality and vii – Data Assurance and Monitor dimension) and the establishment of an auditing culture, policies and maturity assessment mechanisms (xv – Data Governance Framework Maturity Assessment dimension), gaps denoted by Coche et al. [28], Liu et al. [27], and Milne & Brayne [29].

## 7- Conclusion

Our research methodology, structured into three crucial phases—exploration, conceptual, and conclusive—allowed us to identify and analyze the different fundamentals and specifications of data governance and GenAI. It also enabled us to verify that none of the sixteen frameworks have been extensively studied while considering the current complexities of the challenges faced in these fields. However, through our work and by comparing and consolidating both generic and context-specific frameworks, we were able to present an innovative and enhanced approach to these areas. Therefore, we believe that data governance practitioners will be in a better position to address the major challenges currently affecting these fields and will be supported throughout their journey with a robust and unique knowledge base. Furthermore, by gathering and presenting insights from various experts on the complex components of the data governance field, we offer a framework that is both validated and contextualized to current and future organizational environments and requirements, addressing significant and previously unaddressed gaps. In addition, the feedback received from the focus group participants was strongly positive, confirming the added value, timeliness, and relevance of this work in filling a gap they have experienced in the past and are currently facing within their organizations and studies, while also identifying future avenues for further research.

In summary, our study enables organizations and data governance practitioners to reconsider their data governance approaches and frameworks and to adopt this holistic, phased approach while clearly defining their data governance objectives. This includes integrating not only structural, cultural, and governance changes but also technological and operational ones. Through our roadmap of initiatives and data governance dimensions, organizations will be better equipped to adapt to the current transition and move toward a more mature, resilient, controlled, and ethically sound data governance framework and its overall practices. We strongly emphasize the need for continuous research in the field of data governance to ensure that organizations can keep pace with the constantly evolving conditions of the fast-paced environments in which they operate.

## 8- Declarations

### 8-1- Author Contributions

Conceptualization, B.B., H.M., M.N-Z., J.B., and V.S.; methodology, B.B., H.M., M.N-Z., J.B., and V.S.; software, B.B., H.M., M.N-Z., J.B., and V.S.; validation, B.B., H.M., M. N-Z., J.B., and V.S.; formal analysis, B.B. and H.M.; investigation, B.B., H.M., M.N-Z., J.B., and V.S.; resources, H.M., M.N-Z., J.B., and V.S.; data curation, J.B. and V.S.; writing—original draft preparation, B.B.; writing—review and editing, B.B., H.M., M.N-Z., J.B., and V.S.; visualization, B.B., H.M., M.N-Z., J.B., and V.S.; supervision, H.M., M.N-Z., J.B., and V.S.; project administration, H.M., J.B., and V.S.; funding acquisition, H.M., M.N-Z., J.B. and V.S. All authors have read and agreed to the published version of the manuscript.

### 8-2- Data Availability Statement

The data presented in this study are available in the article.

### 8-3- Funding

This work was supported by national funds through FCT (Fundação para a Ciência e a Tecnologia), under the project - UID/04152/2025 - Centro de Investigação em Gestão de Informação (MagIC)/NOVA IMS - https://doi.org/10.54499/UID/04152/2025 (2025-01-01/2028-12-31) and UID/PRR/04152/2025 https://doi.org/10.54499/UID/PRR/04152/2025 (2025-01-01/ 2026-06-30).

### 8-4- Acknowledgments

The authors acknowledge all the support given to this research by INESC TEC - Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência.

### 8-5- Institutional Review Board Statement

Not applicable.

### 8-6- Informed Consent Statement

Not applicable.

### 8-7- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

## 9- References

[1] Bennett, S. (2015). Why information governance needs top-down leadership. Governance Directions, 67(May), 207–213.

[2] Macfeely, S., Me, A., Fu, H., Veerappan, M., Hereward, M., Passarelli, D., & Schüür, F. (2022). Towards an international data governance framework. Statistical Journal of the IAOS, 38(3), 703–710. doi:10.3233/SJI-220038.

[3] Scholz, N., Wieland, J., & Schäffer, T. (2022). Towards a Framework for Enterprise & Platform Ecosystem Data Governance. 28th Americas Conference on Information Systems, AMCIS 2022.

[4] Ragan, R., & Strasser, T. (2020). Understanding the how and why of CU data governance. Credit Union TIMES, 31, 10–10.

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

[6] Lis, D., & Otto, B. (2020). Data governance in data ecosystems - Insights from organizations. In 26th Americas Conference on Information Systems, AMCIS 2020.

[7] Joshi, D., Pratik, S., & Rao, M. P. (2021). Data governance in data mesh infrastructures: The Saxo bank case study. Proceedings of the International Conference on Electronic Business (ICEB), 21, 599–604.

[8] Kanying, T., Thammaboosadee, S., & Chuckpaiwong, R. (2023). Formulating Analytical Governance Frameworks: An Integration of Data and AI Governance Approaches. ACM International Conference Proceeding Series, 1–9. doi:10.1145/3628454.3628461.

[9] Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. AI and Ethics, 2(4), 603–609. doi:10.1007/s43681-022-00143-x.

[10] Dutta, A. (2016). Ensuring the Quality of Data in Motion: The Missing Link in Data Governance. Computer Weekly, 1–4.

[11] Nokkala, T., Salmela, H., & Toivonen, J. (2019). Data governance in digital platforms. In Proceedings of the Twenty‐ Eighth Americas Conference on Information Systems (AMCIS), Minneapolis, United States.

[12] Zorrilla, M., & Yebenes, J. (2022). A reference framework for the implementation of data governance systems for industry 4.0. Computer Standards and Interfaces, 81, 103595. doi:10.1016/j.csi.2021.103595.

[13] Ligot, D. V. (2024). AI Governance: A Framework for Responsible AI Development. In SSRN Electronic Journal. doi:10.2139/ssrn.4817726.

[14] Baum, K., Bryson, J., Dignum, F., Dignum, V., Grobelnik, M., Hoos, H., Irgens, M., Lukowicz, P., Muller, C., Rossi, F., Shawe-Taylor, J., Theodorou, A., & Vinuesa, R. (2023). From fear to action: AI governance and opportunities for all. Frontiers in Computer Science, 5. doi:10.3389/fcomp.2023.1210421.

[15] Cugurullo, F., & Xu, Y. (2025). When AIs become oracles: Generative artificial intelligence, anticipatory urban governance, and the future of cities. Policy and Society, 44(1), 98–115. doi:10.1093/polsoc/puae025.

[16] Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data governance as a collective action problem. Information Systems Frontiers, 22(2), 299-313. doi:10.1007/s10796-019-09923-z.

[17] Boppiniti, S. (2018). Unraveling the Complexities of Healthcare Data Governance: Strategies, Challenges, and Future Directions. Transactions on Latest Trends in IoT, 1(1), 73–89.

[18] Rascao, J. P. (2021). Data Governance in the Digital Age. Handbook of Research on Digital Transformation and Challenges to Data Security and Privacy, IGI Global, 34–62. doi:10.4018/978-1-7998-4201-9.ch003.

[19] Paredes, D. (2016). The six steps to become a successful CDO. CIO, (13284045), 4.

[20] 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 and Systems, 16(1), 125. doi:10.1186/s12961-018-0401-2.

[21] Hikmawati, S., Santosa, P. I., & Hidayah, I. (2021). Improving Data Quality and Data Governance Using Master Data Management: A Review. International Journal of Information Technology and Electrical Engineering, 5(3), 90. doi:10.22146/ijitee.66307.

[22] Lam, K., Iqbal, F. M., Purkayastha, S., & Kinross, J. M. (2021). Investigating the ethical and data governance issues of artificial intelligence in surgery: Protocol for a Delphi study. JMIR Research Protocols, 10(2), 26552. doi:10.2196/26552.

[23] 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.

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

[25] Chakravorty, R. (2020). Common challenges of data governance. Journal of Securities Operations & Custody, 13(1), 23. doi:10.69554/gaku4007.

[26] Sifter, C. J. (2017). Establishing a data governance center of excellence within your bank. New Jersey Banker, 24–25.

[27] Liu, K., Deng, Z., & Zhang, M. (2023). Research on Capability Maturity Evaluation Model of Power Grid Data Management. Procedia Computer Science, 228, 1030–1037. doi:10.1016/j.procs.2023.11.135.

[28] 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.

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

[30] Sanchez, S. C., Hakim, G. J., & Saenger, C. P. (2021). Climate model teleconnection patterns govern the niño-3.4 response to early nineteenth-century volcanism in coral-based data assimilation reconstructions. Journal of Climate, 34(5), 1863–1880. doi:10.1175/JCLI-D-20-0549.1.

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

[32] 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.

[33] 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.

[34] Hay, J. (2014). Data governance gamification. Business Intelligence Journal, 19(1), 30-35.

[35] Khatri, V., & Brown, C. V. (2010). Designing data governance. Communications of the ACM, 53(1), 148–152. doi:10.1145/1629175.1629210.

[36] Kim, H. Y., & Cho, J. S. (2018). Data governance framework for big data implementation with NPS Case Analysis in Korea. Journal of Business and Retail Management Research, 12(3), 36–46. doi:10.24052/jbrmr/v12is03/art-04.

[37] Mao, Z., Wu, J., Qiao, Y., & Yao, H. (2022). 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.

[38] Okoro, R. (2021). Proposed Data Governance Framework for Small and Medium Scale Enterprises (SMES). All Graduate Theses, Dissertations, and Other Capstone Projects, 1–48.

[39] Yebenes Serrano, J., & Zorrilla, M. (2021). A Data Governance Framework for Industry 4.0. IEEE Latin America Transactions, 19(12), 2130–2138. doi:10.1109/TLA.2021.9480156.

[40] Yang, L., Li, J., Elisa, N., Prickett, T., & Chao, F. (2019). Towards Big data Governance in Cybersecurity. Data-Enabled Discovery and Applications, 3(1), 10. doi:10.1007/s41688-019-0034-9.

[41] Fothergill, B. T., Knight, W., Stahl, B. C., & Ulnicane, I. (2019). Responsible data governance of neuroscience big data. Frontiers in Neuroinformatics, 13. doi:10.3389/fninf.2019.00028.

[42] Effoduh, J. O., Akpudo, U. E., & Kong, J. D. (2024). Toward a trustworthy and inclusive data governance policy for the use of artificial intelligence in Africa. Data and Policy, 6, 34. doi:10.1017/dap.2024.26.

[43] Chen, S. Y. (2023). Generative AI, learning and new literacies. Journal of Educational Technology Development and Exchange, 16(2), 1–19. doi:10.18785/jetde.1602.01.

[44] Meyers, C. (2014). How Data Management and Governance Can Enable Successful Self-Service BI. Business Intelligence Journal, 19(4), 23–27.

[45] Bordey, G. (2018). Agile in Data Governance Design. Business Intelligence Journal, 23(2), 23–32.

[46] Burniston, T. R. (2015). Data Governance: A regulatory and Business Imperative. American Bankers Association. ABA Banking Journal, 107(4), 56.

[47] 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. doi:10.1016/j.ijinfomgt.2018.12.017.

[48] Minkkinen, M., & Mantymaki, M. (2023). Discerning Between the “Easy” and “Hard” Problems of AI Governance. IEEE Transactions on Technology and Society, 4(2), 188–194. doi:10.1109/TTS.2023.3267382.

[49] Liza, F. F. (2022). Challenges of Enforcing Regulations in Artificial Intelligence Act: Analyzing Quantity Requirement in Data and Data Governance. Proceedings of the 2022 1st International Workshop on Imagining the AI Landscape after the AI Act: In conjunction with the first International Conference on Hybrid Human-Artificial Intelligence. CEUR-WS, 3221, 9.

[50] Narne, H. (2024). Generative Artificial Intelligence Device to analysis the data for SAP based Data Management. Intelligent Computing and Emerging Communication Technologies, ICEC 2024, 1–6. doi:10.1109/ICEC59683.2024.10837215.

[51] Chinta, S. (2019). The role of generative AI in oracle database automation: Revolutionizing data management and analytics. World Journal of Advanced Research and Reviews, 4(1), 054–063. doi:10.30574/wjarr.2019.4.1.0075.

[52] Janssen, M. (2025). Responsible governance of generative AI: Conceptualizing GenAI as complex adaptive systems. Policy and Society, 44(1), 38–51. doi:10.1093/polsoc/puae040.

[53] Ciriello, R. F., Chen, A. Y., & Rubinsztein, Z. A. (2025). Compassionate AI Design, Governance, and Use. IEEE Transactions on Technology and Society, 6(3), 270–275. doi:10.1109/TTS.2025.3538125.

[54] Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. Journal of Information Technology Case and Application Research, 25(3), 277–304. doi:10.1080/15228053.2023.2233814.

[55] Shi, Y. (2023). Study on security risks and legal regulations of generative artificial intelligence. Science of Law Journal, 2(11), 17–23. doi:10.23977/law.2023.021104.

[56] Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly: Management Information Systems, 28(1), 75–105. doi:10.2307/25148625.

[57] Ostrowski, Ł., Helfert, M., & Xie, S. (2012). A conceptual framework to construct an artefact for meta-abstract design knowledge in design science research. Proceedings of the Annual Hawaii International Conference on System Sciences, 4074–4081. doi:10.1109/HICSS.2012.51.

[58] Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., & Rossi, M. (2018). Design science research contributions: Finding a balance between artifact and theory. Journal of the Association for Information Systems, 19(5), 358–376. doi:10.17705/1jais.00495.

[59] Kurniawan, D. H., Ruldeviyani, Y., Adrian, M. R., Handayani, S., Pohan, M. R., & Rani Khairunnisa, T. (2019). Data Governance Maturity Assessment: A Case Study in IT Bureau of Audit Board. Proceedings of 2019 International Conference on Information Management and Technology, ICIMTech 2019, 629–634. doi:10.1109/ICIMTech.2019.8843742.

[60] 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.

[61] ISO. (2005). International Organization for Standardization ISO/IEC 25000:2005, Software Engineering - Software Product Quality Requirements and Evaluation (SQuaRE). ISO, Geneva, Switzerland.

[62] ISO. (2016). ISO 8000‑61:2016: Data quality — Part 61: Data quality management: Process reference model. ISO, Geneva, Switzerland. Available online: https://www.iso.org/standard/63086.html (accessed on January 2026).

[63] ISO. (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. ISO, Geneva, Switzerland.

[64] ISO. (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. ISO, Geneva, Switzerland.

[65] Rafique, I., Lew, P., Abbasi, M. Q., & Li, Z. (2012). Information Quality Evaluation Framework: Extending ISO 25012 Data Quality Model. International Journal of Computer, Electrical, Automation, Control and Information Engineering, 6(5), 568–573. doi:10.5281/zenodo.1072956.

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

[67] Clarke, N. (2019). How to ensure provision of accurate data to enhance decision-making. Journal of Securities Operations & Custody, 11(2), 112. doi:10.69554/tkla5594.

[68] 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.

[69] Belghith, O., Skhiri, S., Zitoun, S., & Ferjaoui, S. (2021). A Survey of Maturity Models in Data Management. Proceedings of 2021 IEEE 12th International Conference on Mechanical and Intelligent Manufacturing Technologies, ICMIMT 2021, 298–309. doi:10.1109/ICMIMT52186.2021.9476197.

[70] Wakunuma, K., & Eke, D. (2024). Africa, ChatGPT, and Generative AI Systems: Ethical Benefits, Concerns, and the Need for Governance. Philosophies, 9(3), 80. doi:10.3390/philosophies9030080.

[71] Ruohonen, J., & Mickelsson, S. (2023). Reflections on the Data Governance Act. Digital Society, 2(1), 10. doi:10.1007/s44206-023-00041-7.

[72] Bernardo, B. (2022). Artificial Intelligence and Digital Forensics on Data Governance Breaking Through its Importance to Organizations and its Operations. In Forensic Science & Addiction Research, 5(4). doi:10.31031/fsar.2022.05.000625.

[73] Lancaster, J., Ledford, L., & Stephens, J. (2019). Structure your data governance. Business Officer, 52, 19.

[74] Lee, A. P. (2019). Why data governance should be part of your boardroom conversations. NACD Directorship Magazine, 45, 68–68.

[75] Johnston, M. (2016). Are you prepared for your next data disaster. NetworkWorld Asia, 13(2), 52-52.

[76] 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.

[77] Cerrillo-Martínez, A., & Casadesús-De-mingo, A. (2021). Data governance for public transparency. Profesional de La Informacion, 30(4), 1–13. doi:10.3145/EPI.2021.JUL.02.

[78] Ferrari, F., van Dijck, J., & van den Bosch, A. (2025). Observe, inspect, modify: Three conditions for generative AI governance. New Media and Society, 27(5), 2788–2806. doi:10.1177/14614448231214811.

[79] Nastoska, A., Jancheska, B., Rizinski, M., & Trajanov, D. (2025). Evaluating Trustworthiness in AI: Risks, Metrics, and Applications Across Industries. Electronics (Switzerland), 14(13), 2717. doi:10.3390/electronics14132717.

[80] Amendi, R., Halim, E., & Hartono, H. (2024). Exploring Ethical Implications: Unraveling Factors Influencing Data Governance Awareness Behavior in Generative AI Chatbot. Proceedings - 2024 2nd International Conference on Technology Innovation and Its Applications, ICTIIA 2024, 1–6. doi:10.1109/ICTIIA61827.2024.10761290.

[81] Malacaria, S., Grimaldi, M., Greco, M., & Mauro, A. De. (2023). Business Talk: Harnessing Generative AI with Data Analytics Maturity. International Journal on Cybernetics & Informatics, 12(7), 01–10. doi:10.5121/ijci.2023.120701.

[82] Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. 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). Comment: The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. doi:10.1038/sdata.2016.18.

[83] Cronholm, S., & Göbel, H. (2016). Evaluation of the Information Systems Research Framework: Empirical Evidence from a Design Science Research Project. The Electronic Journal Information Systems Evaluation, 19(3), 158.