Computer Fraud and Security
ISSN (online): 1873-7056

Governance Guardrails for Data Storage and Model Transmission: A Layered Security Framework

Dinesh Reddy Kommera
Independent Reseacher, USA

Abstract

Enterprise architectures increasingly rely on a series of connected data pipelines, analytical engines and predictive models, with complex interdependencies across various enterprise and national boundaries. As a result, governance limitations in storage and transmission components quickly spread downstream, affecting data reliability, regulatory compliance, and organizational trust across data stakeholders. Security solutions seldom combine the governance of storage and communication, leading to a gap between layers. This article presents a layered governance framework by organizing guardrails according to the data lifecycle, namely, data repositories, transformation pipelines, and model output channels, to address the cross-layer repercussions of governance decisions. The abstraction combines schema enforcement, access control, lineage tracking, and transmission validation. Systematic implementations encapsulate consistency, confidentiality, and semantic integrity within a structure leveraged across the entire lifecycle of data. Adopting this approach allows organizations to maintain governance continuity while supporting analytical innovation in distributed environments.

Keywords: Data Governance, Transmission Security, Model Validation, Access Control, Data Lineage, Schema Enforcement, Lifecycle Management, Distributed Systems

1. Introduction

Enterprise data architectures now consist of an abundance of inter-connected, data stores, streaming data pipelines, analytical data processing platforms and model-based machine learning inference serving systems. Data is in a constant state of flux as it arrives into and flows through a set of transformations, touches different service boundaries, and is consumed by other applications or human decision makers in a model-based environment. Studies of the data governance frameworks, have found that data governance strategies impact the economic development of an organization by 12 percent, indicating the planned importance of data governance frameworks [1].

In operational terms governance lapses such as these in such environments are more serious than the data breach itself. Data discrepancies introduced at ingestion can propagate down the line of analytics. Evaluated in such contexts, evidence shows that the incorporation of poor quality data and ambiguity in data dependencies relative to algorithms can easily introduce bias, distortion and risk to AI algorithms, causing systemic bias, an illegal decision, and/or large financial exposure to possible political conflict or crises [3]. Current data governance regulation calls for demonstrable control over the data life cycle from storage to processing, and even for dissemination.

Problem Statement: In spite of increasing awareness of governance requirements, storage and transmission protection mechanisms are often implemented in silos as independent controls. According to systematic literature reviews from 2017 and 2023, conceptual studies and structured literature reviews dominate the literature on data governance research. Empirical research has played a small role, comprising only six of the thirty-eight publications included in the review samples [14]. Also lacking is guidance on the governance of handoffs between the layers, from data repositories to processing engines, from analytic systems to model inference services, and from internal systems to external-facing APIs.

Research Gap: While existing literature in data governance is database-centric or network-centric, no thorough framework has yet been proposed to ensure data governance across heterogeneous layers of an architecture. Whereas the model proposed by Abraham et al. identifies six dimensions of data governance (governance mechanism, organizational scope, data scope, domain scope, antecedents, and consequences), it provides little guidance on how controls can be coordinated across dimensions. Likewise, the intersection of storage and transmission governance has received little attention in the literature.

Contribution: This article presents a layered governance framework to define collaborative guardrails for the data storage, transformation, and model transmission layers. This framework enables a unified implementation of structural validation, access control, lineage tracking, and semantic integrity controls across layer boundaries.

2. Background and Related Work

Enterprise data governance processes and controls have evolved alongside those architectural silos. The earliest enterprise repository management systems had database governance concepts such as integrity constraints, access controls, and audit controls. The customary definition of network security is encryption, authentication, and traffic monitoring of data on communication networks. Another description of governance is the process of defining, applying, and monitoring the patterns of rules and authorities for directing the proper functioning of, and ensuring the accountability for, the entire lifecycle of data and algorithms within and across organizations [3].

However, the emergence of distributed analytical architectures has blurred the boundary between governance regimes. Data lakes bring together data from multiple sources without any overall schema, applying validation only downstream. In a survey on data lake architectures, the concept of data lakes is introduced to reduce the challenges posed by big data (especially data variety): they provide storage and processing for any kind of data in its initial form [5]. Because most big data is unstructured, data lakes have been designed specifically to handle semi-structured and unstructured data which impose greater challenges than the data landscape described in the customary data warehouse [5].

Additional legislative developments have introduced further governance requirements in some jurisdictions. For example, the General Data Protection Regulation imposes seven principles: lawfulness, fairness and transparency; purpose limitation; data minimization; accuracy; storage limitation; integrity and confidentiality; and accountability [1]. Financial services regulations stipulate documentation of audit trails of algorithmic decision making systems, while healthcare regulations govern their privacy and control during storage and transmission.

Zero trust architecture styles started to emerge as a response to these trends and are based on the principle of never trust and always verify, which means that the access of any subject, either internal or external to the network, has to be authenticated for every access request [7]. Zero trust networks have five assumptions, including that the network is always considered to be in a hostile environment (with threats internal and external and from the beginning to the end), network location is not sufficient to determine credibility, all devices, users, and network traffic should be authenticated and authorized, and all security policies should be dynamic and calculated based on as many data sources as possible [7].

3. Layered Governance Framework

In particular, we propose an architecture with three coordinated, but distinct and interoperable, layers of governance guardrails, namely storage, transformation, and transmission governance, that provide security and integrity for different aspects of the data life cycle, and are coordinated through metadata and policy propagation.

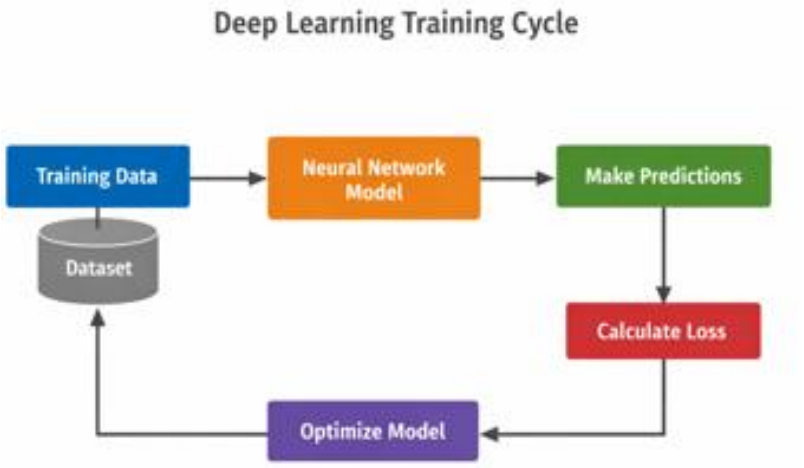

Figure 1: Deep learning training cycle diagram

3.1 Storage Layer Governance

Storage layer governance (governance of data at rest) is the first layer of governance controls. Transparency has been identified as a core governance principle (tri-commonality principle) across three frameworks: the Data Governance Institute framework, the General Data Protection Regulation, and the Open Data Charter [1]. The Data Governance Institute defines transparency as the degree to which all parties and auditors clearly understand how and when data-related decisions and controls were introduced into which processes [1].

Schema Enforcement and Logical Validation. Data repositories that aggregate data from multiple sources can become structurally inconsistent. In survey research on data quality tools, schema enforcement is recognized as the low-level function, where values of columns are checked against schema specifications and column types against type definitions [9]. Schema validation can be applied at value, row, column, and table levels. In conformance checks, the values are verified to match certain ranges, the strings to have a specified length, or the values to follow a regular expression [9].

Research in the domain of data quality management indicates a loss of approx 2% each month and 25% annually of customer data and hence the need for continuous validation. Validation logic built into the ingestion process prevents long-term degradation of the integrity of downstream analytics. There are four stages to a total data quality management process: audit, clean, prevent, and compliance. The process recognizes that data quality cannot be a reactive and corrective process, but rather a proactive and continuous process [8].

Access Management: Through role-based access controls, information sharing is limited to authorized users and services. The Data Governance Institute defines integrity as truthfulness and forthcoming communication regarding data-related drivers, constraints, options and impacts for data-related decisions and accountability as defining accountabilities for cross-functional data-related decisions, processes and controls [1]. Fine-grained permission models ease differentiated access based on user role, operational responsibility, and compliance requirements.

Classification and Lifecycle Management. Data classification policies categorize data based on sensitivity and associated regulatory requirements. According to the GDPR principle of storage limitation, personal data may only be kept in a form which permits identification of data subjects for no longer than necessary [1]. Archiving and secure deletion procedures are examples of data lifecycle management, incorporating such transitions. For example, there is evidence that retention of obsolete data increases both storage costs and breach surface area [8].

3.2 Transformation Layer Governance

Transformation layer governance refers to controls over data processing that modifies, aggregates or otherwise derives new information from source records.

Processing Integrity. Transformation operations ensure data integrity during business logic processing. In the domain of survey research on the dimensions of data quality, consistency is the degree to which a data item has characteristics that are not contradictory to each other and that are in harmony with other items of data in the relevant context of use [9]. Idempotent processing designs yield invariant results regardless of repeated attempts after transient failures.

Lineage Tracking. Lineage tracking methods record the relations between the inputs, processing steps, and the outputs in a data processing workflow. In the context of data lake metadata management, inter-object metadata are metadata describing links between data objects. An example are parenthood links providing data lineage using processes when new objects are created from combinations of existing objects [5]. Lineage metadata enables a data analyst to trace the results of a derived calculation back to the source records, thereby enabling them to perform audits and root-cause analysis. Usage tracking consists of tracking create, read, and update operations by users in the data system [5].

Version Control. Transformation logic itself needs to be version-controlled for reproducibility. In metadata systems, data versioning is the ability to handle update operations while preserving old versions of data, in order to enable process reproducibility or to restore the consistency of data [5]..**3.3 Transmission Layer Governance**

Transmission layer governance refers to controls for data and model outputs flowing across system boundaries.

Confidentiality in Transit: Encryption protocols protect data that is in transit from eavesdropping or compromising. Zero trust architecture considers access control requirements to be that intra-network traffic must be authenticated and authorized and all resources must be protected, regardless of location. Mutual authentication methods are used to establish the identities of both the sending and receiving parties. The user authentication in the zero trust model is based

on policies regarding user location, device, and behavior, and device authentication is based on device identity and security policy application [7].

Structural Preservation: Controlled serialization formats ensure the receiving system can correctly read the data. Schema validation when receiving a message, which checks incoming data for its schema before passing it to the processing system, is needed for data quality checking. The schema defines syntax data types, which is a basic low-level functionality. Tools check them by testing parsing its value against its format [9].

Semantic Integrity: In model-driven contexts, moving predictive outputs without contextualization may lead to misinterpretation. The required model reporting information in a model card includes model information, model use cases, model performance factors, model performance metrics, model training data information and model ethical considerations [13]. In addition to that, the governance guardrails require the model card to contain the model version identifiers, model confidence scores, known limitations and intended use.

Traceability: Audit logs capture events that include sending timestamps, the identifiers for the source and destination systems, and their payload characteristics. In data governance for trustworthy artificial intelligence, traceability mechanisms are evidential trails for enforcing accountability in cases where the data to be transmitted enables decision-making with consequences [3]. Log retention policies strike a balance between storing costs and the need to document past events.

4. Implementation Architecture

The architecture that implements the coordinated governance guardrails across the storage, transformation and transmission layers consists of several components.

Policy Engine: Centralized policy engine that is responsible for governance rule definitions and for providing enforcement directives to individual layer control points. Literature identifies three kinds of governance approaches: planning and control approaches, based on annual cycles of objectives, budgets, and project evaluations; organizational (structure, responsibility, accountability, and reporting); and risk-based approaches, that identify risks and prescribe governance mechanisms [3]. Planning and control and risk-based approaches can be characterized as top-down approaches. Policy inheritance mechanisms propagate storage layer classifications further up the transformation and transmission layers.

Metadata Registry: This includes the schema definitions, classification labels, the lineage and version identifiers of all metadata. Data lake literature stresses the critical role of metadata as the primary means to avoid data swamps. Metadata can be categorized into intra-object (general description), inter-object (relationships between objects), and global metadata (context layer over any data set being processed and analyzed) [5]. Each layer has control points that consult the registry to verify and enforce compliance.

Audit Aggregator: Audit events originating at storage, transformation, and transmission control points are aggregated to a single audit log by an aggregator. The correlation capabilities enable the reconstruction of the entire data flow across layers. Thus, the design of system-level accountability should cover all aspects ranging from the internal operations of algorithms to the organization of usage, the feeding of algorithms with data, the control of data, the verification of outcomes, and the auditing of entire systems [3].

Monitoring Dashboard: Real-time monitoring surfaces governance metrics such as validation failure rates, access anomalies and transmission integrity statistics. Data quality assessment (DQA) literature often identifies six features metadata systems should support: semantic enrichment, data indexing, generating links, data polymorphism, data versioning, and usage monitoring [5]. When threshold values are exceeded, users can be notified if observed values diverge from expectations.

5. Governance Control Mechanisms

Objective	Storage Layer	Transformation Layer	Transmission Layer
Structural Integrity	Schema enforcement, field validation	Idempotent processing, transaction boundaries	Schema validation on receipt, contract testing

Confidentiality	Role-based access, encryption at rest	Processing isolation, temporary data protection	Encryption in transit, mutual authentication
Traceability	Access logs, modification history	Lineage capture, version control	Transmission logs, correlation identifiers
Lifecycle Control	Retention policies, archival automation	Intermediate data cleanup, cache expiration	Message time-to-live, delivery confirmation
Semantic Clarity	Classification metadata, documentation	Transformation documentation, output labeling	Context metadata, usage scope indicators

Table 1: Governance Control Mechanisms by Layer and Objective [3, 9]

A survey of seven open-source data quality tools identifies 25 low-level functionalities of the tools and maps them to the six dimensions of data quality given in ISO/IEC 25012: accuracy, completeness, consistency, currentness, accessibility, and compliance [9]. The relationship between the dimensions of data quality and the low-level functionalities is mostly many-to-many, since a single low-level functionality usually maps to multiple data quality dimensions [9].

Dimension	Direct Assessment Methods	Indirect Assessment Methods
Accuracy	Range/set validation, regex compliance, data type conformance	Statistics checks, distribution analysis, anomaly detection
Completeness	Volume quantification, missing value identification	Distinct value counts, unique element checks
Consistency	Referential integrity, row-level comparisons, ordering validation	Descriptive statistics, cardinality profiling
Currentness	Timestamp recency checks, freshness validation	Table matching for latency assessment
Accessibility	Data type conformance, schema presence verification	Volume and cardinality quantification
Compliance	Format adherence, pattern matching, business rule validation	Distribution checks, uniqueness verification

Table 2: Data Quality Dimensions and Assessment Approaches [8, 9]

Priority of implementation should be linked to risk profiles and regulatory obligations, although studies show organizations often overestimate the quality of their data and underestimate the cost of errors. Less than 50 percent of organizations have a high degree of confidence in the quality of their data [8].

6. Framework Validation and Discussion

To check the theoretical concepts and practical relevance, the proposed layered governance framework is validated against known data quality dimensions, data governance principles, and requirements of a zero trust architecture.

6.1 Compliance with ISO/IEC 25012 Data Quality Dimensions

The ISO/IEC 25012 standard identifies fifteen data quality attributes. They can be classified into two groups: intrinsic attributes and system dependent attributes. A survey of data quality tools identified six dimensions as the most commonly implemented: accuracy, completeness, consistency, currentness, accessibility and compliance [9]. The proposed framework is able to handle all of these dimensions using different governance mechanisms in the storage, transformation and transmission layers.

Finally, accuracy is checked through schema enforcement, range checks, and pattern-based validation at the storage layer, and through semantic checks (e.g., model documentation requirements) at the transmission layer. The literature

uses the term accuracy to refer to the degree to which data has attributes that correctly represent the true value of the intended attribute of a concept or event in a specific context of use; both syntactic and semantic aspects relate to the same definition [9]. Syntactic accuracy relates to data type conformance and format validation. Semantic accuracy relates to business rules and referential integrity checks.

Completeness governance is achieved by measuring volume, discovering missing values, and verifying schema in all the layers. The framework complies with the completness governance guideline, which consists in direct measurement (e.g. missing value detection) and indirect measurement (e.g. cardinality and number of distinct observations). Empirical evidence shows that completeness indicates how much the subject data associated with an entity have values for all its attributes or instances of a related entity [9].

The various consistency controls (transformation layer lineage tracking, version control at the transmission layer, and cross-table validation at the schema layer) ensure that the data is not contradictory or inconsistent with other data at any point in its lifetime. Currentness is addressed with timestamp validation and freshness checks. Accessibility and compliance are addressed with schema conformance, pre-defined format standards, and information policies.

6.2 Mapping to Data Governance Principles

The academic literature on data governance frameworks for trustworthy artificial intelligence proposes twelve principles that such frameworks should fulfill [3]. The three-layer architecture of the framework can be aligned with these twelve principles.

Stewardship refers to data gathering, ensuring data quality and data security. Stewardship is governed by storage layer access controls, classification policies and lifecycle management procedures. This is supported by data stewardship literature, which found that data stewardship roles are needed to formalize the accountability of data and information resource management on behalf of and in the best interest of others (including data quality, validity, and security) [3]. The framework operationalizes this through RBAC and ownership.

The controlled opening principle allows for controlled transparency and public accountability, while preserving privacy. This principle is enforced by fine-grained permission models and classification-based access control policies. Transmission layer semantic integrity controls ensure that model outputs containing shared context metadata do not expose proprietary implementation details.

Risk-based governance is seen throughout, from the monitoring dashboard to the alerts. Such an approach is particularly relevant for the case of Big Data Algorithmic Systems, due to the risks related to privacy violations, data misuse, discrimination, bias and wrong decisions [3]. By monitoring these three layers, the framework identifies risks and creates appropriate responses.

System-level controls are one of the building blocks of the integrated architecture. The integrated architecture research that supports the system-level accountability design requires that the following mechanisms be in place: an algorithm usage organization, a data feeding process, data control, results checking, and general auditing capabilities [3]. The audit aggregator and metadata registry directly support these requirements.

6.3 Industry Action on Zero Trust Architecture

Zero trust architecture is based on a set of assumptions, which the proposed framework responds to substantively, through addressing the basic five assumptions of zero trust identity as described in the literature: perpetual vulnerability of the network, constant presence of internal and external threats, lack of a trusted location, necessity of universal authentication and authorization, and contextual policy calculation [7].

For authentication, the framework employs transmission layer and storage layer authentication mechanisms. For transmission layer, it adopts zero-trust principles: location, device and behavior assessment based authentication mechanisms whereas the storage layer employs access control policies based on device identity and security [7]. Dynamic authentication requirements can be defined through rules in the policy engine of the framework.

The principle of least privilege and need to know fits within the fine-grain permission models, and classification-based models for access control in the framework. Zero trust requires the use of role-based access control models for resources and the use of minimum required permissions to complete the work at any given point in time [7]. This principle is implemented through access controls at the storage layer and authorizations at the transmission layer.

The framework's continuous monitoring and real-time adaptive policies support adaptive governance. A related concept states that zero trust systems should use multiple information sources to dynamically adapt context-aware access policies [7]. Monitoring dashboards and alerts provide the feedback loop necessary for zero trust principles.

6.4 Comparison With Existing Approaches To Governance

A systematic literature review found three models in the field of data governance: planning and control, organizational, and risk-based [3]. The proposed framework uses elements of all three models instead of focusing on a single one.

Planning and control elements of the framework appear in the policy engine and compliance monitoring capabilities. These guide the annual processes of setting objectives and budgets, and reviewing projects. The elements of organization are framed as responsibility structures and cross-functional coordination mechanisms. The elements of risk are framed as continuous monitoring, anomaly detection, and alert-based response capabilities.

As shown in a review of data governance frameworks, the frameworks should be contextualized to the circumstances of economic sectors or business models, and the governance principles leading to the framework choice are equally important to the framework [1]. This framework takes into account factors specific to economic sectors in the configurable policy definitions, while sharing components across implementations.

The principles of transparency, integrity, and accountability appear in the Data Governance Institute Data Governance Framework, the General Data Protection Regulation (GDPR), and the Open Data Charter [1]. Our Data Governance Framework supports these principles with the tools of audit logging and data lineage for transparency, data validation and access control for integrity, and data ownership and compliance checks for accountability.

6.5 Limitations and Boundary Conditions

The framework does also have certain underlying assumptions to be borne in mind whilst applying it. The framework assumes that there is a commitment to the theme of governance by the organization as a business priority and that, from the top of the business down, data quality should be championed as a planned, long term initiative to establish and maintain quality standards [8].

Technical challenges include the need to write precise metadata to enforce the policies and the potential performance cost of validations. The data quality tools literature has also explored techniques for functional approximations of computationally intensive operations that yield accuracy-performance tradeoffs in high-throughput systems [9].

While the framework largely covers internal data quality dimensions, it does not cover the system-dependent dimensions of efficiency, portability, and recoverability as defined in ISO/IEC 25012 [9]. Organizations needing these additional dimensions should use additional controls at the infrastructure level.

From 2017 to 2023, the empirical validation of data governance frameworks was limited, with most of the research on data governance being conducted as conceptual research or systematic literature reviews [14]. This framework could benefit from empirical validation through case studies in a range of organizational contexts.

7. Organizational Coordination

Technical guardrails, meanwhile, require alignment across the organization, and operationalizing them generally requires multiple teams, such as IT, compliance, data governance, and operations.

Accountability Structures. Policies should define ownership of data assets, the data transformation process and interfaces to transmission services. Data Governance Institute calls for defining detailed accountabilities for cross-functional data-related decision-making, processes and controls. GDPR further requires the organization to take responsibility for the data under its control [1]. Data stewards have decision rights to make classification and access authorization decisions. Literature on data stewardship suggests data stewards should only share information in responsible ways, define accountabilities for managing information resources on behalf of and to the advantage of other stakeholder entities, and make sure data quality, validity, and security are guaranteed [3].

Governance Councils. Cross-functional governance councils act as a forum to review new data initiatives, proposed model deployments, and adjudicate competing requirements. The implementation studies on zero trust identify the need for ethics committees, as well as other oversight committees, to make Big Data Algorithmic Systems decisions, to

improve employees' understanding of data quality and its sharing, and to integrate governance into architectural considerations through council oversight [3].

Continuous Improvement. Governance frameworks must evolve as regulatory environments, volumes of data and architectural patterns change. The intersection of data governance and AI governance has been identified as a key application area in systematic literature reviews and recent studies further note the demand for improved AI governance frameworks with the advent of AI [14]. Furthermore, business processes, customer expectations, source systems and rules are constantly changing, which means that data quality management systems should also be dynamic [8].

8. Practical Considerations

Organizations seeking to implement layered governance frameworks face practical challenges.

Legacy Integration. Legacy systems may not support the governance mechanisms the framework assumes are in place. Data lake architecture research typically maps to either pond or zone architectures, where data assigned to zones instead of repositories based on degree of refinement [5]. Adapter components may be introduced between legacy interface points and governance control points, using intercept patterns to capture audit events, and enforce governance policy.

Performance Impact: The use of governance controls has the potential to add latency overhead. According to a survey of data quality tools, there are several approximate computation methods that can be used on resource-intensive computation requests, such as HyperLogLog++ sketching for cardinality estimation and KLL sketching algorithms for quantile estimation [9]. The location of control points was carefully chosen to minimize redundancy while ensuring coverage on both sides.

Scalability. Governance mechanisms must be able to scale to increased data volumes. Data lake research identifies that storage and processing scalability are key to successful data lake implementations [5]. Distributed policy enforcement focuses on deploying validation logic to edge locations as opposed to deploying components in centralized bottleneck locations.

Multi-Cloud Complexity. Hybrid and multi-cloud implementations introduce heterogeneous infrastructure requiring governance coordination across provider boundaries. According to zero trust architecture research, zero trust methods need to address the fact that since cloud computing and Internet of Things technologies have been developed, it is no longer possible to define whether objects are internal based on their location in space [7]. Abstraction layers provide uniformity in control interfaces across platforms.

Establishing effective governance guard rails in the data storage and model transmission layers meets the security and compliance requirements of contemporary enterprise architecture. The layered model eases maintaining governance as data is transformed and transmitted from one layer to the next. Bundling schema enforcement, access control, lineage capture, and semantic integrity validation in a single architecture avoids governance gaps at the boundaries between architectural layers.

Preparing for this in practice will require not just the technical design and implementation, but organizational orchestration, too. Empirical evidence suggests data governance is a way to deal with questions of transparency, accountability, fairness, discrimination and trust, and there are enabling mechanisms, like governance structures, responsibilities and accountabilities, planning and control cycles, and risk management [3]. Explicit accountability structures, cross-functional governance councils and continuous improvement processes help sustain the effectiveness of guardrails over time.

Changes in regulations as well as architectural patterns will force frameworks to adapt over time. From 2017 to 2023, an increasing portion of scholarship has focused on the challenge of governing AI as a data governance use case [14]. Thanks to the modularity principles the proposed approach can support organizations to develop basic data architectures and governance in the storage, transformation, and transmission layers by ensuring the confidentiality, integrity, transparency, and adaptability of the data despite the increasing complexity of the digital context.

References

[1] Tosin Ekundayo et al., "Identifying The Core Data Governance Framework Principle: A Framework Comparative Analysis," Organization Leadership and Development Quarterly, 2023. https://www.researchgate.net/profile/Justine-Chinoperekweyi-2/publication/366759618

[2] Jiban K. Pal, "Metadata initiatives and emerging technologies to improve resource discovery," Annals of Library and Information Studies, 2010. https://www.researchgate.net/profile/Jiban-Pal/publication/256484108

[3] Marijn Janssen et al., "Data governance: Organizing data for trustworthy Artificial Intelligence," Government Information Quarterly, 2020. https://repositorio.inesctec.pt/server/api/core/bitstreams/8e3184ec-8dc3-4dfe-ac5f-61969a66ad74/content

[4] KRISTIN WEBER et al., "One size does not fit all: A contingency approach to data governance," Journal of Data and Information Quality, 2009. DOI: http://doi.acm.org/10.1145/1515693.1515696

[5] Pegdwend´e Sawadogo and J´erˆome Darmon, "On Data Lake Architectures and Metadata Management," arXiv, 2021. https://arxiv.org/pdf/2107.11152

[6] Jan Philipp Albrecht, "How the GDPR Will Change the World," EDPL, 2016. https://web.archive.org/web/20200321130451id_/https://edpl.lexxion.eu/data/article/10073/pdf/edpl_2016_03-005.pdf

[7] Yuanhang He et al., "A Survey on Zero Trust Architecture: Challenges and Future Trends," Wiley, 2022. https://onlinelibrary.wiley.com/doi/pdf/10.1155/2022/6476274

[8] Richard Marsh, "Drowning in dirty data? It's time to sink or swim: A four-stage methodology for total data quality management," Database Marketing & Customer Strategy Management, 2004. https://link.springer.com/content/pdf/10.1057/palgrave.dbm.3240247.pdf

[9] VASILEIOS PAPASTERGIOS et al., "Unfolding Data Quality Dimensions in Practice: A Survey," ACM Journal of Data and Information Quality, 2026. https://dl.acm.org/doi/pdf/10.1145/3786328

[10] Vijay Khatri and Carol V. Brown, "Designing Data Governance," Communications of the ACM, 2010. https://dl.acm.org/doi/pdf/10.1145/1629175.1629210

[11] Boris Glavic, "Data Provenance: Origins, Applications, Algorithms, and Models," Emerald Insight, 2021. DOI: https://doi.org/10.1561/1900000068

[12] Muhammad Zubair et al., "Network Security and Cryptography Challenges and Trends on Recent Technologies," Journal of Applied and Emerging Sciences. [Online]. Available: https://www.researchgate.net/profile/Aqib-Ali-6/publication/372244622.pdf

[13] Margaret Mitchell et al., "Model cards for model reporting," Proceedings of the Conference on Fairness, Accountability, and Transparency, 2019. DOI: https://doi.org/10.1145/3287560.3287596

[14] KAROL BLIŽNÁK et al., "A Systematic Review of Recent Literature on Data Governance (2017–2023)," IEEE Access, 2024. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10707270