Lights-off Data Factory: A Governance-First Architecture for Level-5 Autonomous Data Systems

Author Details: Sukant Pandey, datamgmt2015@gmail.com

Abstract

Enterprise data governance systems increasingly fail to scale under the demands of autonomous analytics and AI-driven data consumption. Existing data management practices rely heavily on human-in-the-loop decision authority, introducing latency, inconsistency, and systemic risk. This paper proposes a governance-first framework for autonomous data systems composed of three orthogonal components: the Autonomous Data Governance Maturity Model (ADGMM), the Autonomous Epistemic Governance & Integrity System (AEGIS), and Canonical Entity Reasoning & Epistemic Stewardship (CERES). The framework repositions operational domains such as Master Data Management, Data Quality, and Data Integration as governed execution layers rather than sources of authority. Autonomy is defined as a system's ability to estimate uncertainty, reason semantically, and self-correct without continuous human arbitration.

Keywords: Data Governance, Autonomous Systems, Knowledge Graphs, Epistemic Uncertainty, Data Architecture, Master Data Management, Data Quality, Data Stewardship.

I. Introduction

Enterprise data systems increasingly operate at machine scale while governance mechanisms remain human-centric. Governance workflows such as stewardship review, exception queues, and manual approvals introduce structural bottlenecks that limit scalability and delay decision-making. While automation has improved operational efficiency, authority over ambiguous or high-impact decisions remains embedded in human processes.

Recent advances in machine learning and large language models have not resolved this limitation. Instead, they have exposed the fragility of governance architectures that lack explicit mechanisms for uncertainty estimation, internal verification, and autonomous correction. This paper argues that these failures stem from architectural assumptions rather than insufficient automation.

II. Background and Related Work

Traditional data governance frameworks emphasize organizational roles, policies, and stewardship models [1][2]. While effective at moderate scale, these approaches do not address autonomy as a system-level property. Human-in-the-loop models, commonly adopted to manage uncertainty, scale linearly with human effort and introduce variability and delay [3][4].

Recent work in knowledge graphs, semantic reasoning, and epistemic uncertainty highlights the importance of reflective architectures capable of self-assessment and correction [5][6][7]. However, these insights are rarely integrated into governance system design.

III. Problem Statement

Three structural constraints limit existing governance systems:

Decision authority is embedded in human workflows.

Autonomy is not explicitly measured.

Semantic reasoning is treated as an operational concern rather than an epistemic one.

These constraints apply uniformly across MDM, DQ, DI, and metadata management.

IV. Governance-First Framework

ADGMM (Autonomous Data Governance Maturity Model)

ADGMM defines five levels of governance autonomy, ranging from manual governance to epistemically autonomous systems capable of estimating uncertainty and self-correcting decisions. Maturity is measured at the governance layer rather than within operational domains.

AEGIS (Autonomous Epistemic Governance & Integrity System)

AEGIS specifies an architectural pattern for enforcing governance as an executable system. Key characteristics include separation of generation and verification, internal critic mechanisms, and uncertainty-driven escalation.

CERES (Canonical Entity Reasoning & Epistemic Stewardship)

CERES provides a semantic and epistemic reasoning core that treats entities, schemas, and policies as revisable knowledge claims. This enables consistent reasoning across MDM, DQ, DI, and metadata domains.

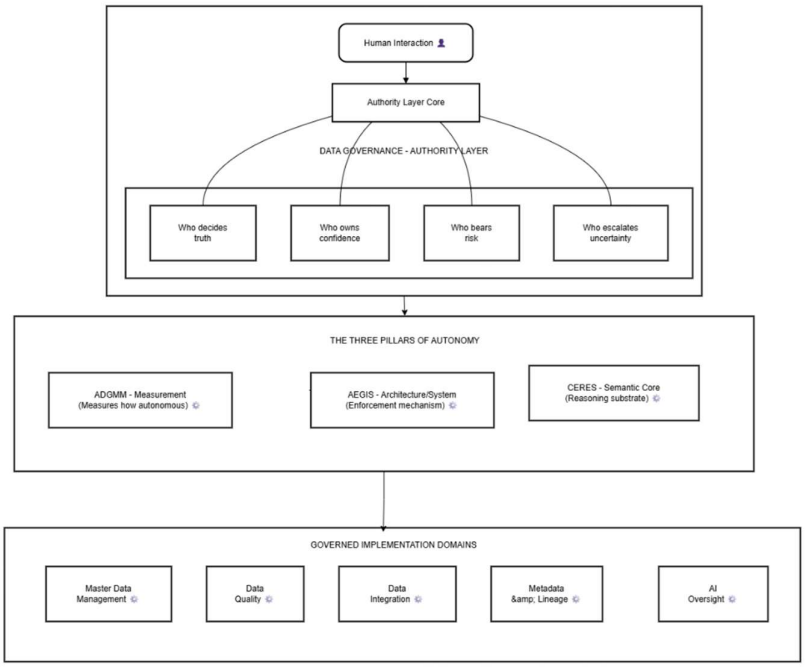

Figure1: The relationship between the three layers of Governance-First Architecture for Level-5 Autonomous Data Systems

V. Implications

By elevating governance to a first-class architectural concern, operational domains become execution layers governed by autonomous decision systems. This approach improves scalability, auditability, and consistency while reducing human arbitration.

VI. Conclusion

Human-centric governance architectures impose a hard ceiling on autonomy. A governance-first approach—measuring autonomy, enforcing it architecturally, and grounding it semantically—provides a foundation for Level-5 autonomous data systems.

References

1. DAMA International, DAMA-DMBOK, 2017.
2. Gartner, Data Governance Foundations, 2023.
3. Amershi et al., "Guidelines for Human-AI Interaction," CHI, 2019.
4. Christen, Data Matching, Springer, 2012.
5. Hogan et al., "Knowledge Graphs," ACM CSUR, 2021.
6. Gal & Ghahramani, "Bayesian Approximation," ICML, 2016.
7. Shinn et al., "Reflexion," arXiv, 2023.