International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group | Volume 7 Issue 2 PP 46-49, 2026
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V7I2P106

Original Article

AI-Native and Agentic Data Governance: From Rule-Based Policies to Self-Healing Metadata Systems

Kuladeep Sandra
Independent Researcher, USA.

Received On: 24/02/2026 ***Revised On: 26/03/2026*** ***Accepted On: 04/04/2026*** ***Published On: 11/04/2026***

Abstract - *Data governance has progressed through three distinct generations. The first was manual: stewards, policy documents, spreadsheets, quarterly reviews. The second was rule-based automation: policy engines, schema contracts, automated access control derived from classification metadata. The third—emerging now—is AI-native: large language models and agentic systems that infer governance attributes, detect violations, and propose remediation autonomously. This paper surveys the progression and presents a position on what AI-native governance can and cannot do, grounded in operational experience running system-enforced governance in a 30-engineer data organization that achieved a 45 percent error reduction through deterministic automation alone. The case study evidence includes 500,000 files consolidated to 3,000 governed assets, $1.4 million in annual storage savings, 93 percent registration compliance and 87 percent classification coverage—establishing the baseline against which AI-native enhancements should be evaluated. This paper examines three research questions: how each generation of governance has worked and where each has failed; what AI-native techniques can productively add to the rule-based foundation; and what guardrails are non-negotiable when applying LLM-based agents to governance decisions with regulatory consequences. The central argument is that AI-native governance is an evolutionary step rather than a replacement, that the residual cases where deterministic automation falls short are exactly where LLMs offer the most value, and that the disciplined use of AI in governance requires accepting that some categories of decisions are properly human and should remain so.*

Keywords - *AI Governance, Automated Policy Enforcement, Intelligent Data Catalog, Active Metadata, LLM-Driven Governance, Policy Violation Detection, Metadata Enrichment.*

1. Introduction

Data governance has been a recognized discipline for at least two decades, but the operating practices have changed substantially over that period as the underlying data infrastructure has changed. The first generation of governance, call it manual governance, was built around the assumption that human stewards could keep up with the rate of data change in the platform. Stewards reviewed new datasets, classified columns, approved access requests, signed off on schema changes. The model worked at the scale of the data platforms it was designed for, which were considerably smaller than today's. As platforms grew, the model strained and eventually failed.

The second generation, rule-based automation, replaced human stewardship with deterministic systems. Classification was driven by pattern-matching tools like Microsoft Purview. Schema contracts were enforced at the producer side by Confluent Schema Registry or equivalent. Access control was derived automatically from the classification metadata through Apache Ranger or similar policy engines. Lineage was captured as a side effect of query execution rather than as a manual documentation exercise. Our team's experience, documented in detail in companion papers, shows that this generation produced dramatic improvements: 45 percent error reduction through governance-driven data quality improvements, 500,000 files consolidated to 3,000 governed assets, $1.4 million in annual storage savings, 93 percent registration compliance and 87 percent classification coverage. The numbers establish that rule-based governance is not just better than manual; it is enough better that organizations that have not made the shift are operating at a substantial disadvantage.

The third generation, AI-native governance, is emerging in 2024 and 2025 and is the topic of this paper. The premise is that LLM-based systems can handle the residual cases that rule-based automation cannot: ambiguous classifications where pattern matching is insufficient, ownership inference where the original creator has left and the successor is not obvious, policy violation detection where the rule is too complex to encode deterministically, and metadata enrichment where the value of the metadata depends on understanding context that no static rule can capture.

This paper addresses three research questions:

- RQ1: How have the three generations of data governance worked in practice, and what specific failure modes characterized each generation?
- RQ2: What AI-native techniques (LLM-based classification, agentic violation detection, intelligent metadata enrichment, ownership inference through lineage analysis) can productively add to the rule-based foundation?

- RQ3: What guardrails are non-negotiable when applying LLM-based agents to governance decisions, and which categories of governance work should never be fully automated?

The paper is organized as follows. Section 2 surveys the historical progression of governance generations. Section 3 documents the rule-based case study. Section 4 explores AI-native techniques. Section 5 addresses risks and failure modes. Section 6 offers forward-looking guidance. Section 7 concludes.

2. The Three Generations of Governance

2.1. Manual Governance

The first generation was built around human stewards. Each business unit had assigned stewards who were responsible for the governance of their data. The stewards reviewed new datasets, applied classifications, approved access requests, and tracked their work in governance tools (Collibra was a common one). The model worked when the rate of data creation was small enough that human review could keep up.

The failure mode was capacity. As the rate of data creation grew, the stewards fell behind. The backlog of unreviewed datasets grew. Schema changes shipped without review. Access requests piled up. We experienced this directly: steward roles were on paper for 18 months before we restructured the program. The stewards were competent and motivated. The model asked them to do work that exceeded human capacity, and no amount of effort or process improvement changed that.

2.2. Rule-Based Automation

The second generation replaced human review with deterministic automation. Microsoft Purview classified columns by matching their values and names against patterns and trained classifiers. Confluent Schema Registry enforced schema compatibility at the producer side, refusing to accept producer updates that broke the contract. Apache Ranger enforced access control policies at query time, denying queries that would have returned data the requester was not authorized to see. Lineage was captured automatically through query instrumentation.

The metrics improved dramatically. Classification coverage went from the low double digits (under manual stewardship) to 90 percent. Registration compliance went from inconsistent to 95 percent. Access violations went from a recurring incident category to zero unresolved. Data quality errors fell by approximately 45 percent as the governance metadata enabled targeted quality improvements. The number of governed assets the team could maintain went from a few hundred under manual stewardship to several thousand under automation.

The remaining failure modes are the cases where deterministic rules are not enough. A column whose name is ambiguous and whose values do not match any known pattern. A schema change whose backwards compatibility depends on knowing what consumers do. An ownership question for a table whose creator has left and whose successor is unclear. The 5 to 10 percent of cases that the deterministic system cannot handle is where the third generation lives.

2.3. AI-Native Governance

The third generation adds LLM-based components that handle the residual cases. The goal is not to replace the rule-based foundation but to extend it. LLMs are well suited to the residual cases precisely because the residual cases are the ones that require interpreting natural language context, making judgment calls under uncertainty, and synthesizing information from multiple sources exactly the tasks LLMs are good at. The risk is that LLMs are also capable of producing confident outputs that are wrong, which is a more dangerous failure mode in governance than in many other applications.

3. The Rule-Based Case Study

Our team's experience with rule-based governance is the empirical anchor for this paper. The platform processes more than 500,000 individual files (post-compaction, 3,000 well-sized governed assets), serves 6 business units across three time zones, and runs on a hybrid on-premise and multi-cloud architecture. The technology stack is Apache Iceberg for the lakehouse, Spark for batch processing, Kafka for streaming, Trino for federated queries, and Microsoft Purview plus Apache Ranger plus Confluent Schema Registry as the governance components.

The architectural pattern is that classification metadata flows from Purview into Ranger automatically, so a column flagged as PII by Purview inherits the appropriate access policy without manual intervention. Schema contracts are enforced at the producer side, with the schema registry refusing to accept producer updates that break compatibility. Lineage is captured by query instrumentation in Trino and Spark. The result is that governance is the default behavior of the platform rather than an exception that requires manual intervention.

The outcomes (45 percent error reduction, 500,000-to-3,000 consolidation, $1.4 million in annual savings, 95/90/zero compliance metrics) are documented in our companion papers [1], [2], and [3]. The relevant point for this paper is that rule-based governance is good enough to ship, and the question is what the next generation should add.

4. AI-Native Techniques

4.1. Llm-Based Classification

The first technique is LLM-based classification for the residual cases that pattern-matching tools cannot handle. A column called score whose values do not match any standard PII pattern is ambiguous to a deterministic classifier. An LLM that has access to the column name, the surrounding columns, sample values, and the table description can often infer the correct classification. The classification still has to be confidence-calibrated and escalated to a human if the

confidence is low, but the LLM extends the reach of the classifier into cases that pattern matching cannot solve.

4.2. Ownership Inference

The second technique is ownership inference through lineage analysis. When a table's assigned owner has left the organization and there is no obvious successor, an LLM can examine the lineage graph to identify which teams are using the table, which engineers have committed to its production code, and which downstream consumers depend on it. The proposed owner can be surfaced to a human for confirmation rather than assigned autonomously, but the work of identifying the candidate is something the deterministic system was not doing well.

4.3. Active Metadata and Drift Detection

The third technique is active metadata: governance decisions driven by the real-time state of the data rather than by static policies. An LLM-based agent watching the data quality metrics on a production table can detect anomalies that suggest the underlying data has drifted, identify the most likely root cause through the lineage graph, and surface the diagnosis to the responsible team. The agent does not fix the problem; it identifies the problem and accelerates the human response.

4.4. Policy Violation Detection

The fourth technique is policy violation detection at the level of natural language policies. Some governance policies are not easily expressed as deterministic rules; a policy that says personal data should not be shared with third parties without explicit consent requires interpreting what shared and third party mean in any specific context. An LLM-based agent can read the policy, examine the data flows, and flag flows that may violate the policy for human review.

5. Risks and Failure Modes

Hallucination in sensitive classifications: The most serious risk is that an LLM confidently classifies sensitive data as non-sensitive because of a pattern in its training that does not match reality. The mitigation is confidence-based tiering: low-confidence classifications escalate to humans, high-confidence classifications proceed with logging and periodic audit, and the audit catches systematic errors before they accumulate.

Regulatory accountability: The second risk is that a regulator asks who is responsible when an AI system makes a governance decision. The answer in current regulatory frameworks is that the human supervisor is responsible, which means the supervisory architecture has to be real and not nominal. Logging the agent's decisions, the inputs it had, and the reasoning it used is the minimum standard.

The steward paradox: The third risk is the one we already lived through: tools and policies are necessary but not sufficient. Steward roles failed even with perfect tools and policies because the work the model asked of stewards was beyond human capacity. AI agents do not face the capacity constraint, but they face their own failure modes. The lesson is not that AI replaces human judgment; it is that AI handles the cases humans could not handle anyway, leaving humans to do the work that is properly theirs.

6. Forward-Looking Guidance

AI techniques that are ready for governance today: Classification of ambiguous columns with confidence-based escalation, ownership inference with human confirmation, lineage-based root cause diagnosis with human action. These work because the LLM is providing information that a human acts on, not making autonomous decisions. AI techniques that need more research before deployment: Autonomous remediation of policy violations, agent-driven access grants, fully automated schema change approvals. These are areas where the consequence of error is high enough that the human-in-the-loop should remain present.

Non-negotiable guardrails: Every agent action must be logged with reasoning and inputs; every high-stakes decision must escalate to a human; every classification of personal data must be human-verifiable; every action must be reversible. Measuring success beyond adoption metrics: The metrics that matter are accuracy on ground truth where ground truth is available, false-positive rates on sensitive classifications, escalation rates to humans (high escalation rates suggest the agent is operating beyond its competence), and the quality of the agent's reasoning explanations as judged by human reviewers.

7. Conclusion

Returning to the three research questions:

- RQ1: Manual governance failed because it asked humans to do work beyond human capacity. Rule-based automation fixed the bulk of the problem but left a residual of cases that deterministic rules cannot handle. AI-native governance is positioned to handle the residual.
- RQ2: The AI-native techniques most ready for productive use are LLM-based classification of ambiguous columns, ownership inference through lineage analysis, active metadata for drift detection, and policy violation detection on natural-language policies. All of them work best when the LLM provides information that humans act on rather than making autonomous decisions.
- RQ3: The non-negotiable guardrails are auditability of every agent action, confidence-based escalation to humans, human verifiability for sensitive classifications, and reversibility of agent actions. Governance functions that should never be fully automated are those involving regulatory accountability, those requiring business context the agent cannot infer, and those whose errors have asymmetric high-cost consequences.

The closing observation is that the progression from manual to rule-based to AI-native governance is not a replacement story; it is an accumulation story. Each generation builds on the previous one. The organizations that have not yet shifted from manual to rule-based should make

that shift first. The organizations that have made the shift to rule-based should consider AI-native enhancements for the residual cases. The organizations that try to skip directly from manual to AI-native will discover that the AI-native components depend on the metadata infrastructure that rule-based governance produces, and they will be building on a missing foundation.

Conflicts of Interest

The author declares that there is no conflict of interest concerning the publishing of this paper.

Acknowledgements

The author thanks the data governance, platform engineering, and risk colleagues whose deployments and policy frameworks shaped the position presented here.

References

[1] K. Sandra, "Data modernization and governance case study," companion paper KD-09, Unpublished manuscript, 2026. Google Scholar

[2] K. Sandra, "Storage optimization through deterministic governance," companion paper KD-14, Unpublished manuscript, 2026. Google Scholar

[3] K. Sandra, "Data quality metrics and governance impact," companion paper KD-20, Unpublished manuscript, 2026. Google Scholar

[4] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao, "ReAct: Synergizing reasoning and acting in language models," in Proc. Int. Conf. Learning Representations (ICLR), 2023. Google Scholar | Publisher Site

[5] T. Schick, J. Dwivedi-Yu, R. Dessì, R. Raileanu, M. Lomeli, L. Zettlemoyer, N. Cancedda, and T. Scialom, "Toolformer: Language models can teach themselves to use tools," in Advances in Neural Information Processing Systems (NeurIPS), 2023. Google Scholar | Publisher Site

[6] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou, "Chain-of-thought prompting elicits reasoning in large language models," in Advances in Neural Information Processing Systems (NeurIPS), 2022. Google Scholar | Publisher Site

[7] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, et al., "Language models are few-shot learners," in Advances in Neural Information Processing Systems (NeurIPS), 2020. Google Scholar | Publisher Site

[8] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, et al., "On the opportunities and risks of foundation models," arXiv:2108.07258, 2021. Google Scholar | Publisher Site

[9] A. Halevy, F. Korn, N. F. Noy, C. Olston, N. Polyzotis, S. Roy, and S. E. Whang, "Goods: Organizing Google's datasets," in Proc. ACM SIGMOD Int. Conf. Management of Data, 2016, pp. 795–806. Google Scholar | Publisher Site

[10] J. M. Hellerstein, V. Sreekanti, J. E. Gonzalez, J. Dalton, A. Dey, S. Nag, K. Ramachandran, S. Arora, A. Bhattacharyya, S. Das, M. Donsky, G. Fierro, C. She, C. Steinbach, V. Subramanian, and E. Sun, "Ground: A data context service," in Proc. Conf. Innovative Data Systems Research (CIDR), 2017. Google Scholar | Publisher Site

[11] DAMA International, DAMA-DMBoK: Data Management Body of Knowledge, 2nd ed. Basking Ridge, NJ: Technics Publications, 2017. Google Scholar | Publisher Site

[12] Microsoft, "Microsoft Purview documentation," 2024. [Online]. Available: https://learn.microsoft.com/en-us/purview/ Google Scholar | Publisher Site

[13] Apache Software Foundation, "Apache Ranger documentation," 2024. [Online]. Available: https://ranger.apache.org Google Scholar | Publisher Site

[14] Apache Software Foundation, "Apache Iceberg documentation," 2024. [Online]. Available: https://iceberg.apache.org Google Scholar | Publisher Site

[15] Confluent, "Confluent Schema Registry documentation," 2024. [Online]. Available: https://docs.confluent.io/platform/current/schema-registry/index.html Google Scholar | Publisher Site

[16] Trino Software Foundation, "Trino documentation," 2024. [Online]. Available: https://trino.io Google Scholar | Publisher Site

[17] European Parliament and Council, "Regulation (EU) on artificial intelligence (EU AI Act)," 2024. [Online]. Available: https://eur-lex.europa.eu Google Scholar | Publisher Site

[18] National Institute of Standards and Technology, AI Risk Management Framework (AI RMF 1.0), 2023. [Online]. Available: https://www.nist.gov/itl/ai-risk-management-framework Google Scholar | Publisher Site

[19] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, and T. Gebru, "Model cards for model reporting," in Proc. ACM Conf. Fairness, Accountability, and Transparency (FAT*), 2019, pp. 220–229. Google Scholar | Publisher Site

[20] M. Kleppmann, Designing Data-Intensive Applications. Sebastopol, CA: O'Reilly Media, 2017. Google Scholar | Publisher Site