



Agentic AI for Autonomous Metadata Management and Enterprise Data Catalogues: Automating Metadata Discovery, Classification, Semantic Enrichment, Data Lineage and Stewardship under DAMA and NDMO Frameworks

Uzair Momin^{1*} 

¹Deloitte (United States): New York, New York, US

*Corresponding Author

Uzair Momin

Deloitte (United States):
New York, New York, US

Article History

Received: 22.07.2026

Accepted: 15.09.2026

Published: 22.09.2026

Abstract: Agentic artificial intelligence is transforming enterprise metadata management from rule-based automation to goal-directed systems capable of discovering, interpreting, connecting, and governing metadata across heterogeneous data estates. This review evaluates how autonomous agents can facilitate metadata discovery, sensitive-data classification, semantic enrichment, lineage reconstruction, enterprise catalog maintenance, and stewardship, while maintaining accountability to established data-governance frameworks. A structured integrative review of DOI-verified peer-reviewed literature connects three typically distinct research streams: enterprise metadata and data catalogs, knowledge graphs and provenance, and large-language-model-based autonomous agents. The synthesis indicates that the most effective near-term approach is policy-constrained orchestration, where specialized agents propose metadata changes, ground decisions in catalogs and knowledge graphs, preserve evidence, and escalate significant actions to stewards. DAMA-oriented management domains provide the operational model for ownership, quality, architecture, metadata, and governance, while Saudi Arabia's NDMO context elevates requirements for classification, traceability, control evidence, and institutional accountability. The review introduces an auditable agentic metadata architecture and an autonomy-assurance loop. It concludes that enterprise value relies on provenance-preserving execution, confidence-aware escalation, semantic interoperability, human override, measurable metadata quality, and governance controls that recognize agent actions as auditable events rather than invisible automation.

Keywords: Agentic AI, Autonomous Metadata Management, Enterprise Data Catalogue, Data Governance, DAMA, NDMO, Data Lineage, Semantic Enrichment, Knowledge Graph, Data Stewardship.

Copyright © 2026 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY-NC 4.0) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

1. INTRODUCTION

Enterprise data estates are now distributed across warehouses, lakes, lakehouses, software-as-a-service applications, integration platforms, operational databases, analytical pipelines, and

machine-learning systems. Correspondingly, metadata is also dispersed: schemas and constraints are stored in databases; transformations are recorded in code and orchestration logs; business definitions are maintained in glossaries and

Citation: Momin, U. (2026). Agentic AI for Autonomous Metadata Management and Enterprise Data Catalogues: Automating Metadata Discovery, Classification, Semantic Enrichment, Data Lineage and Stewardship under DAMA and NDMO Frameworks; *Glob Acad J Econ Buss*, 8(5), 1289-1299. <https://doi.org/10.36348/gajeb.2026.v08i05.021>

documents; ownership is reflected in organizational practices; and sensitivity labels often vary across platforms. As a result, metadata management serves as an operational control plane for locating, interpreting, trusting, and governing data, rather than merely providing documentation. Research on data governance consistently identifies value creation and risk reduction as joint objectives. Recent reviews further highlight that governance effectiveness is determined by coherent decision rights, policies, data quality, and organizational mechanisms, rather than by technology alone (Abraham *et al.*, 2019; Walsh *et al.*, 2022; Blišnák *et al.*, 2024).

Enterprise data catalogs aim to operationalize this control plane by inventorying assets and integrating technical, operational, and business metadata. However, catalog research identifies ongoing challenges with integration, federation, curation, and adoption. While data catalogs can bridge data supply and demand, their effectiveness diminishes when metadata harvesting is incomplete, terminology is inconsistent, or ownership and access workflows are not incorporated within the catalog (Ehrlinger *et al.*, 2021; Jahnke & Otto, 2023; Petrik *et al.*, 2024). A recent multivocal review demonstrates that automation and standardization are essential for scalable metadata life-cycle management, yet many catalog tools continue to offer inconsistent support for governance, collaboration, and data-product practices (Tonnarelli *et al.*, 2025). These limitations underscore that simply introducing another catalog interface does not establish a self-sustaining governance environment.

Agentic AI introduces an architectural proposition. Rather than executing one model inference or a fixed workflow, an agent observes context, reasons about a goal, selects tools, performs actions, evaluates outcomes, and may coordinate with other agents. Surveys identify planning, memory, perception, tool use, interaction, and adaptation as components (Wang *et al.*, 2024; Guo *et al.*, 2024; Li *et al.*, 2024; Acharya *et al.*, 2025). Applied to metadata, agents could inspect assets, infer classifications, reconcile terms, reconstruct lineage, generate stewardship tasks, and verify compliance. The challenge is to obtain autonomy without converting probabilistic reasoning into ungoverned authority. This review evaluates agentic metadata through DAMA- and NDMO-oriented governance lenses (Goel *et al.*, 2024; Aljarallah & Alfadhel, 2025).

2. Aim of the Study

The study aims to develop a critical, governance-centered understanding of how agentic AI can automate enterprise metadata management

and data catalogs without weakening human accountability, semantic reliability, or compliance evidence. It evaluates where autonomous agents can add value across metadata discovery, classification, enrichment, lineage, and stewardship; identifies the technical and organizational controls required for safe deployment; and derives a reference architecture aligned with DAMA-oriented data-management responsibilities and NDMO-oriented expectations for controlled, traceable, and auditable data governance.

3. Research Objectives

Objective 1: Synthesize recent evidence on enterprise metadata management, data catalogs, lineage, provenance, semantic metadata, and agentic AI to identify capabilities that can be combined into autonomous metadata operations.

Objective 2: Critically assess agent roles for metadata discovery, data classification, semantic enrichment, lineage generation, and stewardship, distinguishing high-confidence automation from actions requiring human approval.

Objective 3: Map agentic metadata functions to governance responsibilities associated with DAMA-oriented practices and NDMO-oriented controls, emphasizing accountability, classification, quality, traceability, and policy enforcement.

Objective 4: Identify architectural safeguards for trustworthy autonomy, including grounding, provenance capture, confidence thresholds, segregation of duties, human override, observability, and explainable decision records.

Objective 5: Derive research gaps, implementation priorities, and measurable evaluation criteria for enterprise-scale deployment of autonomous metadata and catalog capabilities.

4. Research Questions

RQ1: Which metadata-management activities are technically suitable for agentic automation, and which should remain approval-gated?

RQ2: How can knowledge graphs, provenance, and catalog context reduce semantic ambiguity and unsupported agent decisions?

RQ3: How should DAMA and NDMO governance responsibilities constrain agent permissions, evidence requirements, and stewardship escalation?

RQ4: Which evaluation measures can demonstrate that agentic metadata management improves coverage and timeliness without degrading correctness, explainability, or accountability?

5. REVIEW METHODOLOGY

A structured integrative review was selected because the research question spans information systems, database and metadata engineering, knowledge representation, AI agents, and governance. A narrowly defined systematic review

within one discipline would omit relevant mechanisms, whereas an unstructured narrative review would make source selection difficult to audit. The protocol therefore combined concept-driven searching, citation chaining, publisher verification, and thematic synthesis. Source discovery concentrated on peer-reviewed journal and conference literature available through IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, Wiley Online Library, Taylor & Francis, ACL Anthology, IJCAI proceedings, and the Dublin Core proceedings. Scopus and Web of Science were treated as quality and discoverability targets where publisher or institutional records exposed indexing links; this manuscript does not claim a subscription-level exhaustive export from either database.

Search concepts were combined iteratively. Core combinations included “metadata management” AND “data catalog”, “enterprise metadata” AND governance, “data lineage” AND provenance, “semantic metadata” AND “knowledge graph”, “metadata enrichment” AND AI, “autonomous agents” OR “agentic AI”, “LLM multi-agent”, “LLM knowledge graph construction”, “DAMA” AND data governance, and “NDMO” AND data governance. Searches prioritized 2020-2025 publications, retaining earlier works only for indispensable governance or explainability foundations. The evidence set was limited to thirty DOI-verified peer-reviewed works so each source could be checked for identity and relevance. No initial-hit or exclusion counts are reported because no reproducible database-export log was preserved; inventing such numbers would falsely imply a registered systematic-review flow.

Inclusion required a peer-reviewed contribution with a resolvable DOI and a clear connection to at least one analytical domain: enterprise data governance; metadata architecture or catalogs; automated metadata discovery or profiling; lineage or provenance; semantic enrichment or knowledge graphs; trustworthy AI; or autonomous/multi-agent systems. Studies were excluded when they were vendor marketing, unreviewed preprints without a peer-reviewed version, non-technical opinion pieces, duplicated conference-to-journal versions where the later version subsumed the earlier work, or sources whose bibliographic details could not be independently verified. Saudi-specific NDMO evidence was retained when it directly examined implementation of NDMO standards. The supplied background paper informed theme scoping but was excluded from the evidentiary corpus, preventing circular reliance on its references.

Quality appraisal considered venue and publisher credibility, clarity of research method, evidence-to-claim fit, replicability of technical

descriptions, relevance to enterprise metadata operations, and whether limitations were acknowledged. Greater interpretive weight was given to systematic reviews, established journal articles, well-described empirical or design studies, and major peer-reviewed conference proceedings. Synthesis proceeded through a concept matrix linking each paper to five functions: discovery, classification, semantic enrichment, lineage, and stewardship. A second coding layer captured agent mechanisms such as planning, tool use, memory, graph reasoning, verification, escalation, and human oversight. The final synthesis was comparative rather than vote-counting: conflicting findings were examined as boundary conditions, especially where LLM flexibility trades off against precision or auditability.

Methodological limitations were handled explicitly. Terminology is unstable, and “agentic AI” is often applied commercially to conventional automation; this review therefore defines it functionally as goal-directed perception, planning, tool-mediated action, and feedback. Direct evidence on production autonomous metadata systems remains limited, so architectural recommendations are inferential syntheses of validated component capabilities rather than claims of proven enterprise-wide autonomy.

6. Thematic Literature Review and Critical Analysis

6.1 From Passive Metadata Repositories to Active, Agent-Operated Catalogs

Metadata platforms historically emphasize capture, storage, search, and curation. Data-lake research demonstrates why this is necessary: without flexible metadata models, heterogeneous assets become difficult to interpret, relate, and govern. Sawadogo and Darmont (2021) position metadata management as a central design issue for data lakes, while Eichler *et al.*, (2021b) propose a generic model capable of representing metadata at different granularities and for use cases such as versioning and lineage. Their enterprise case study separately shows that metadata transparency and sharing are organizational as well as technical challenges (Eichler *et al.*, 2021a). These findings imply that an agent cannot be effective if the underlying catalog model is narrow, stale, or disconnected from business context.

Catalog research sharpens the requirements. Ehrlinger *et al.*, (2021) identify semantic classification and organization as core catalog functions, Jahnke and Otto (2023) show that enterprise value depends on integration across data sources and metadata tools, and Petrik *et al.*, (2024) extract functional requirements for enterprise catalog design. Tonnarelli *et al.*, (2025) extend this

picture by mapping catalog features to metadata life-cycle phases and modern data architectures. Collectively, the literature supports an “active metadata” direction, but agentic AI changes the unit of automation. Instead of only harvesting metadata on schedule, an agent can detect an event, infer what it means, select a catalog or lineage tool, propose updates, seek supporting evidence, and initiate a stewardship workflow.

The most defensible architecture is specialized rather than monolithic. A discovery agent can enumerate assets and collect schemas, tags, owners, query statistics, transformation code, and operational signals. A classification agent can combine deterministic rules, pattern recognition, and contextual reasoning to propose sensitivity, domain, or criticality labels. A semantic agent can reconcile

column names with glossary terms and ontology concepts. A lineage agent can correlate orchestration metadata, SQL, Spark plans, and runtime events. A stewardship agent can prioritize unresolved conflicts, assign tasks, and track decisions. This division mirrors multi-agent research, where role specialization and communication help decompose complex goals (Guo *et al.*, 2024; Li *et al.*, 2024), while also enabling enterprise segregation of duties.

Table 1 summarizes this decomposition. Agents should record proposed assertions with evidence, confidence, source, timestamp, and policy basis before authoritative metadata is changed. This “propose-verify-promote” pattern distinguishes autonomous assistance from uncontrolled mutation and preserves evidence about how AI-produced metadata was created.

Table 1: Agentic metadata functions, evidence, outputs, and governance gates

Metadata function	Primary agent role	Evidence inputs	Proposed output	Governance gate
Discovery	Discovery / scanner agent	Schemas, APIs, logs, registries, pipeline metadata	Observed asset and structural metadata	Auto-promote directly observed facts after validation
Classification	Classification agent	Patterns, sampled values, glossary, lineage, policies	Sensitivity, domain, or criticality assertion	Confidence and risk threshold; steward review for ambiguity
Semantic enrichment	Semantic agent	Glossary, ontology, knowledge graph, documentation	Canonical terms, definitions, and relationships	Constraint checking; steward review for conflicts or new concepts
Lineage	Lineage agent	SQL, DAGs, Spark plans, code, runtime events	Observed or inferred lineage edges	Observed edges may promote; inferred edges remain provisional
Stewardship	Workflow / policy agent	Ownership graph, issues, policies, prior decisions	Task routing, decision packet, recommended action	Human owner or steward approves consequential changes

6.2 Automated Discovery and Classification: Precision before Autonomy

Automated discovery is the lowest-risk entry point because much technical metadata is directly observable. Connectors can inspect schemas, file formats, APIs, query logs, model registries, and pipeline definitions. Aikoh *et al.*, (2020) demonstrate that profiling can extract useful metadata characteristics, while catalog studies consistently treat automated ingestion as necessary for scale. An agent adds value by deciding which extractor to run, detecting deltas, interpreting previously unseen structures, and requesting missing context. However, discovery completeness and correctness are different metrics: finding an asset does not establish its business meaning, legal status, or owner. Agent design should therefore maintain explicit states such as observed, inferred, steward-confirmed, and policy-certified rather than collapsing them into one catalog record.

Classification is more consequential because labels may drive access, retention, masking, or sharing decisions. Agentic classification can combine lexical evidence, schema patterns, sampled values, existing glossary mappings, lineage neighbors, and organizational context. Yet confidence must be calibrated by class and consequence. A high-confidence inference that a field contains an email pattern may be safely auto-proposed, while classifying a derived analytical dataset as non-sensitive should require stronger evidence because false negatives can propagate downstream. NDMO-oriented governance increases this requirement: classification is not merely a search facet but an auditable governance decision. The King Saud University study illustrates that NDMO implementation involves strategic and executive mechanisms and identifies data inconsistency and organizational constraints as material challenges (Aljarallah & Alfadhel, 2025). Agentic automation

should therefore surface those inconsistencies rather than silently normalize them.

A useful control model separates recognition, decision, and enforcement. Recognition agents identify candidate sensitive entities or domain labels. Decision policies combine confidence thresholds, business rules, and risk tiers. Enforcement remains with deterministic access-control, retention, or masking systems after required approval. This architecture prevents a generative model from directly changing security posture. It also aligns with broader trustworthy-AI governance, where data, processes, algorithms, and scrutiny mechanisms must be organized jointly rather than treating the model as an isolated component (Janssen *et al.*, 2020). The practical target is not maximum autonomous action rate; it is maximum trustworthy throughput, measured as the proportion of metadata decisions completed correctly with acceptable evidence and escalation cost.

6.3 Semantic Enrichment, Knowledge Graphs, and Business Context

Semantic enrichment is where language models can offer the greatest improvement over conventional harvesting because enterprise meaning is often expressed in names, descriptions, policies, tickets, dashboards, and informal documentation. Knowledge graphs provide a structured representation for linking assets to concepts, owners, processes, policies, metrics, applications, and other assets. Hogan *et al.*, (2021) describe knowledge graphs as a means of representing and exploiting heterogeneous, dynamic knowledge, while Honick *et al.*, (2022) show how entity reconciliation can enrich metadata and create graph relationships. Reznik *et al.*, (2022) likewise emphasize structured and linked metadata as a route to improved findability and contextualization.

LLMs can accelerate graph construction but cannot be assumed to supply ground truth. Pan *et al.*, (2024) characterize a complementary relationship between knowledge graphs and LLMs: graphs can ground inference and improve interpretability, while LLMs can assist graph construction and completion. Zhu *et al.*, (2024) provide a particularly relevant caution: their evaluation finds strong reasoning potential but shows that LLMs are not uniformly superior as few-shot information extractors. Zhang and Soh (2024) address schema scale by separating extraction, definition, and canonicalization, demonstrating why semantic normalization should be an explicit phase rather than an implicit by-product of prompting.

For enterprise metadata, the implication is a bidirectional semantic loop. The catalog and

knowledge graph should constrain the agent with approved terminology, known relationships, ownership domains, and policy concepts. The agent can then propose new mappings or relationships when it encounters unknown assets. Those proposals should be canonicalized against the graph, checked for contradictions, and either accepted automatically under low-risk thresholds or routed to a domain steward. This reduces two common catalog defects: synonym fragmentation, in which equivalent concepts proliferate under different labels, and false semantic convergence, in which superficially similar terms are merged despite different business meanings.

6.4 Autonomous Lineage and Provenance as the Trust Backbone

Lineage records how data moves and transforms; provenance extends the concern to the entities, activities, and agents involved in producing an information object. Automated lineage is essential because manually maintained diagrams decay quickly in changing pipelines. Schoenenwald *et al.*, (2021) demonstrate extraction and visualization of Spark lineage from execution plans, showing how operational artifacts can generate more reliable metadata than manual documentation. Pan *et al.*, (2023) show that provenance is also security-relevant because trustworthy histories support forensic analysis and accountability, while Chapman *et al.*, (2024) demonstrate the analytical value of fine-grained provenance for understanding data-science pipelines.

Agentic lineage should therefore prioritize evidence correlation over narrative generation. An agent can parse SQL, orchestration DAGs, Spark plans, API mappings, transformation code, and query history; reconcile identifiers; and create a candidate lineage graph. But when direct evidence conflicts, the system should preserve competing hypotheses instead of inventing a clean path. Generative reasoning is useful for interpreting opaque transformation logic or proposing missing links, yet inferred edges must remain distinguishable from observed edges. This distinction is crucial for impact analysis, regulatory explanation, model reproducibility, and incident response.

Lineage also supplies the context needed to make other agents safer. A classification agent can propagate sensitivity cautiously along verified transformations. A semantic agent can infer likely business domains from upstream and downstream neighbors. A stewardship agent can identify the owner most able to resolve a disputed term. A policy agent can assess whether data crossed a boundary that requires review. In this sense, provenance is not one feature among many; it is the trust backbone that

connects autonomous actions to evidence. Figure 1 presents the resulting architecture: operational sources feed specialized agents, assertions are stored in a provenance-aware semantic layer, the enterprise

catalog exposes governed metadata, and policy and stewardship controls mediate promotion to authoritative status.

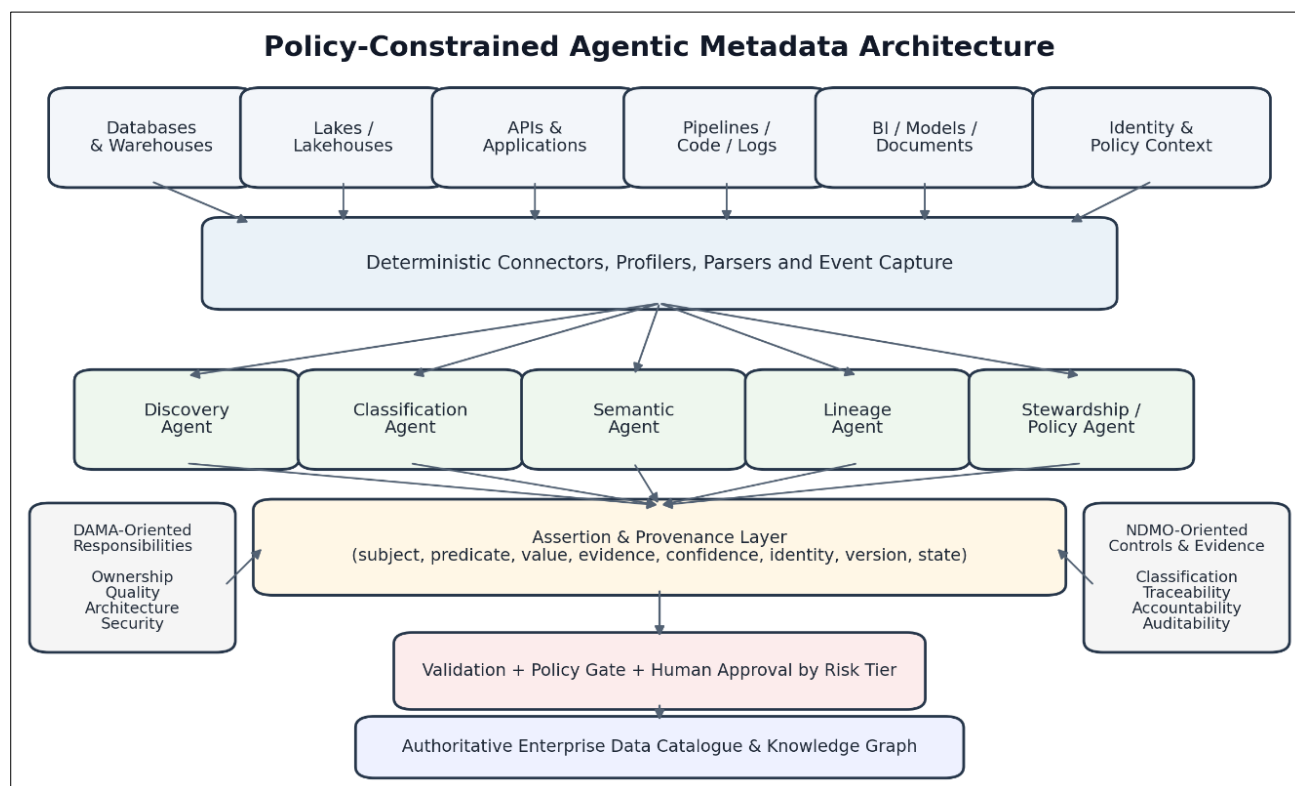


Figure 1: Policy-constrained agentic metadata architecture

6.5 Stewardship, DAMA and NDMO: Redesigning Human Accountability

Autonomous metadata management does not eliminate stewardship; it changes stewardship from routine entry to exception handling, adjudication, and control design. Data-governance research warns against technology-centric programs that neglect decision rights and organizational motivations (Abraham *et al.*, 2019; Walsh *et al.*, 2022). Goel *et al.*, (2024) operationalize DAMA dimensions in a data-governance framework for process mining, reinforcing the relevance of domains such as metadata, quality, architecture, security, integration, and governance when data is used in analytical processes. Bernardo *et al.*, (2024) similarly emphasize the interdependence of governance and data quality. Agentic systems should be embedded within these responsibilities rather than installed as a parallel automation layer.

Under a DAMA-oriented model, agents can perform operational work while decision accountability remains assigned to human roles. Data owners define acceptable use and risk appetite; stewards approve definitions, resolve semantic conflicts, and review critical classifications; architects govern metadata models and integration

patterns; security and privacy functions define enforcement rules; and the data office monitors quality, lineage, and policy evidence. Agents operate as delegated processors with explicit scopes. Their identity, permissions, tool calls, input evidence, outputs, and approvals should be recorded like those of service accounts, but with richer reasoning traces and confidence metadata.

NDMO alignment adds a national-control perspective. In Saudi organizations, data governance is expected to be demonstrable through institutional mechanisms, not merely inferred from technical capability. Aljarallah and Alfadhel (2025) show that implementation involves both strategic and execution dimensions; this supports a design in which autonomous metadata actions are tied to documented governance roles and standardized evidence. An agent should therefore be able to answer: which control or policy motivated this action, which data asset was affected, what evidence was examined, who owns the decision, what changed, and whether human approval was required. These questions make autonomous operation inspectable to internal assurance and external oversight.

Table 2 translates common agent failure modes into governance controls. Risk-tiered autonomy resolves the speed-versus-authority tension: reversible, evidence-backed, low-impact updates can be promoted automatically; ambiguous semantic or classification decisions can require

steward confirmation; and actions affecting access, retention, legal basis, or external sharing should remain deterministic and approval-gated. This creates a graduated operating model rather than a binary choice between manual and autonomous governance.

Table 2: Representative agent failure modes and governance controls

Failure mode	Governance consequence	Preventive control	Detective / corrective control
Hallucinated semantic mapping	Incorrect glossary alignment or false concept merge	Retrieval grounding, ontology constraints, approved vocabulary	Contradiction checks; steward adjudication; rollback
False-negative classification	Sensitive data may be under-protected	Risk-weighted thresholds; multi-signal evidence; no generative enforcement	Sampling, drift monitoring, downstream lineage review, mandatory escalation
Fabricated or missing lineage	Unreliable impact analysis and audit evidence	Prefer runtime artifacts; label inferred edges explicitly	Evidence reconciliation; lineage completeness checks; preserve competing hypotheses
Privilege overreach or tool misuse	Unauthorized metadata or policy changes	Least privilege, allow-listed tools, scoped transactions, separation of duties	Complete tool-call logs, policy alerts, transaction rollback, kill switch
Stale or drifted metadata	Catalogue loses operational trust	Event-driven refresh, freshness SLOs, bounded memory	Delta detection, stale-asset alerts, periodic steward sampling
Unsupported claim propagation between agents	Speculation becomes falsely authoritative	Assertion provenance and authorization context on every message	Trace claim ancestry; quarantine unsupported assertions; recalibrate agents

6.6 Technical Architecture for Policy-Constrained Agentic Metadata

A production architecture should separate reasoning from authority. The perception layer connects to databases, warehouses, lakehouses, APIs, pipelines, code repositories, BI tools, model registries, and document stores. Extraction tools produce deterministic observations wherever possible. Above them, specialized agents reason over a shared context composed of catalog metadata, business glossary, policy library, lineage graph, knowledge graph, prior stewardship decisions, and identity context. A planner decomposes goals but is not itself authorized to commit high-impact changes. Tool gateways enforce allow-lists, parameter validation, row or object scopes, rate limits, and transaction boundaries.

The assertion layer is critical. Every proposed metadata statement should be stored as an assertion object containing subject, predicate, value, evidence references, generator identity, model or rule version, confidence, timestamp, and validation state. Deterministic validators check schema compatibility, ontology constraints, policy rules, duplicate conflicts, and lineage consistency. Only validated assertions are promoted into authoritative catalog fields. This approach converts probabilistic output into governed state transitions. It also

supports rollback because the system can distinguish the underlying observed metadata from later agent interpretations.

Agent memory should be bounded and governable. Wang *et al.*, (2024) identify memory, planning, tool use, and evaluation as core agent components, but enterprise memory should consist of approved definitions, prior resolutions, policies, catalog assertions, and performance feedback subject to retention and access rules. Multi-agent communication should also carry provenance and authorization context so that unsupported claims cannot be laundered through delegation (Guo *et al.*, 2024; Li *et al.*, 2024).

Explainability must focus on actionable evidence rather than verbose model rationales. Barredo Arrieta *et al.*, (2020) distinguish multiple forms of explainability and emphasize the need for responsible, interpretable AI. For metadata decisions, the useful explanation is compact: source evidence, applied rules, graph paths, confidence, alternatives considered, and reason for escalation. Qin and Yu (2024) argue that metadata itself is central to trustworthy AI across datasets, models, pipelines, and artifacts; agentic metadata systems extend this logic by requiring metadata about the agents that create metadata. Figure 2 depicts the assurance cycle:

sense, plan, act in a constrained workspace, verify against evidence and policy, then either promote,

escalate, or roll back, with feedback captured for future calibration.

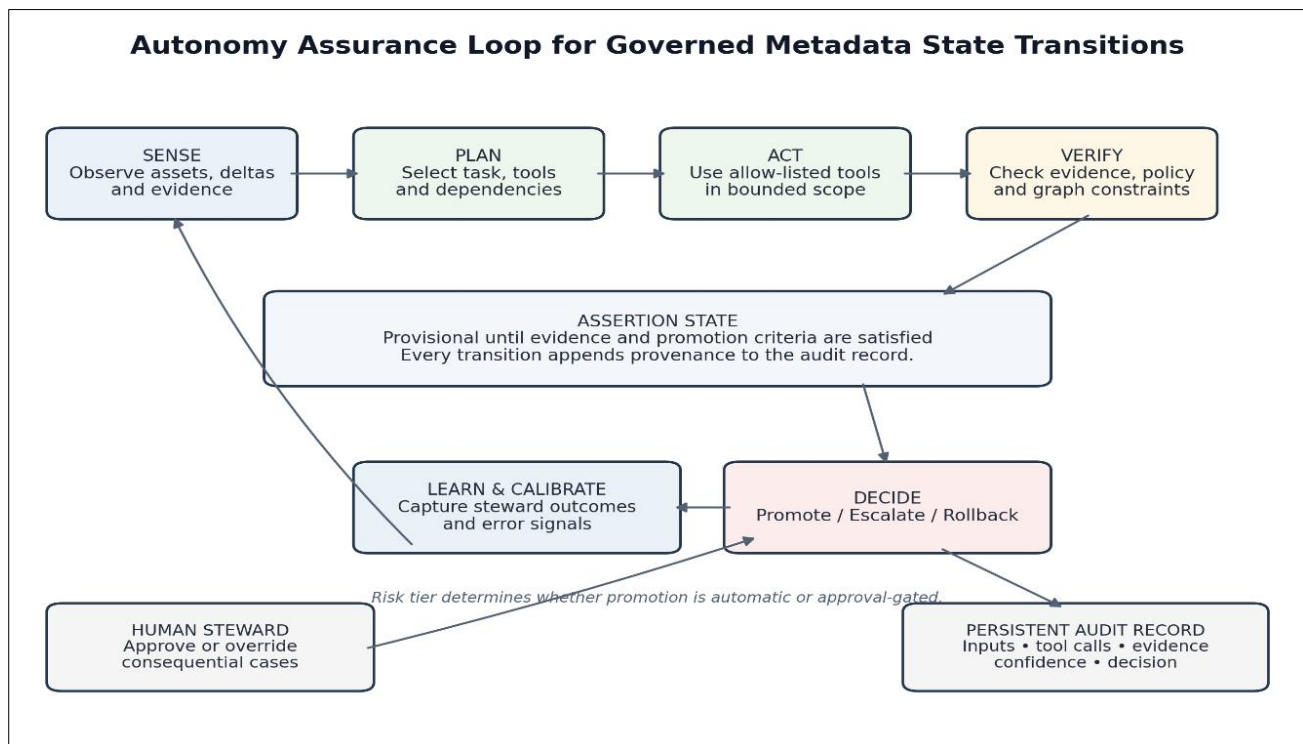


Figure 2: Autonomy assurance loop for governed metadata state transitions

7. DISCUSSION

The synthesis indicates that agentic AI is most credible as a governance-aware orchestration layer over existing metadata infrastructure, not as a replacement for catalogs, lineage systems, or stewardship. Enterprise catalog research repeatedly identifies integration and lifecycle gaps, while agent research contributes planning and tool coordination. Combining them resolves a practical mismatch: metadata platforms possess state but limited adaptive reasoning; LLM agents possess flexible reasoning but limited authority and factual reliability. A safe architecture lets the catalog and provenance graph provide state and constraints while agents perform interpretation and workflow coordination.

This reframes autonomy as a spectrum of governed state transitions. Discovery can be highly autonomous because it observes technical facts. Semantic mapping can be semi-autonomous because ambiguity is common, but reversibility is high. Classification can be autonomous only for well-calibrated, evidence-rich cases because false negatives may have security consequences. Lineage can automatically promote directly observed edges while retaining inferred edges as provisional. Stewardship routing can be automated, but ownership changes and policy exceptions require accountable approval. This decomposition is more

useful than assigning one enterprise-wide “autonomy level.”

The review also reveals a difference between semantic accuracy and governance validity. A model may correctly infer that two fields refer to a customer identifier, yet governance validity additionally asks whether the approved business term, data owner, classification, and permitted use were applied. DAMA-oriented management supplies the cross-functional responsibility model needed to answer those questions. NDMO-oriented governance raises the importance of demonstrable evidence and control alignment in Saudi settings. Consequently, benchmark accuracy alone cannot justify production deployment; evaluation must include auditability, policy compliance, escalation quality, and the stability of metadata over time. Operational evaluation should therefore compare autonomous decisions against steward-confirmed baselines over time, separating discovery coverage from semantic correctness and tracking whether confidence thresholds remain calibrated when schemas, policies, data domains, and organizational responsibilities change materially.

Knowledge graphs provide an intermediate representation between free text and rigid catalog schemas. They can ground agents in approved concepts while agents propose graph expansion. Yet

LLM-KG research cautions that stronger reasoning does not guarantee stronger extraction (Pan *et al.*, 2024; Zhu *et al.*, 2024). The appropriate response is asymmetric trust: observed metadata and approved governance artifacts carry higher evidential status than agent-generated assertions, and contradictions should be surfaced rather than smoothed away.

8. Research Gaps and Future Research

First, there is little peer-reviewed evidence from longitudinal production deployments of agentic metadata systems. Future research should compare rule-based, single-agent, and multi-agent approaches using shared enterprise-like benchmarks that include heterogeneous schemas, policy constraints, evolving pipelines, and ambiguous business terms. Evaluations should report precision, recall, calibration, latency, compute cost, human review time, rollback rate, and downstream governance effects rather than only language-model task scores.

Second, autonomous lineage and semantic enrichment need uncertainty-aware representations. Research should develop provenance models that distinguish observed, inferred, approved, superseded, and disputed assertions, and should test how uncertainty propagates through classification and impact analysis. Fine-grained provenance research provides a foundation (Chapman *et al.*, 2024), but agent-generated evidence introduces additional questions about model version, prompt or policy context, tool outputs, and multi-agent delegation.

Third, DAMA and NDMO alignment requires formal policy compilation. Future work should translate governance requirements into machine-testable constraints while preserving the role of human interpretation. This includes mapping policy clauses to catalog attributes, lineage predicates, approval thresholds, and evidence schemas. Saudi studies should examine how NDMO controls can be operationalized as continuous metadata checks without reducing compliance to superficial dashboard completion.

Fourth, semantic governance needs benchmarks containing enterprise synonymy, homonymy, multilingual terminology, and domain-specific definitions. Studies should test whether retrieval-grounded agents reduce false merges across Arabic-English contexts. Fifth, organizational research should measure how agentic systems alter stewardship roles, skills, accountability, and trust, including whether continuous automation strengthens governance discipline or encourages inappropriate deference to machine-generated metadata.

9. Practical, Managerial and Policy Implications

Organizations should begin with bounded use cases that have observable evidence and reversible outputs. Automated asset discovery, stale-metadata detection, duplicate-term identification, lineage extraction, and stewardship routing are appropriate early targets. Sensitive-data downgrades, owner reassignment, access changes, retention decisions, and policy exceptions should remain approval-gated. A phased program should establish the metadata model, connectors, identity controls, provenance store, and evaluation baseline before introducing higher autonomy.

Managers should define an agent RACI alongside the human data-governance RACI. Each agent needs an accountable business owner, authorized tools, permitted data domains, decision thresholds, escalation paths, and a kill-switch procedure. Catalog KPIs should shift from record counts toward coverage, freshness, confirmed semantic mappings, lineage completeness, unresolved exceptions, steward cycle time, and verified error rates. Incentives should reward closed governance outcomes rather than raw automation volume.

For Saudi entities, NDMO alignment should become machine-observable where feasible. Catalog attributes, classifications, lineage, stewardship approvals, and policy checks should produce durable evidence linked to assets and responsible roles. Procurement should require exportable metadata, open interfaces, provenance access, role-based controls, and clear separation between agent-inferred and human-approved information. Claims of “autonomous governance” should be evaluated against measurable evidence and override mechanisms.

10. Limitations

This review integrates adjacent literatures because direct empirical studies of fully autonomous enterprise metadata governance remain scarce. The thirty-source corpus is deliberately curated and DOI-verified rather than exhaustive, and no fabricated database-screening flow is presented. Some recommendations therefore synthesize validated component capabilities into an architecture that still requires production testing. NDMO-specific peer-reviewed evidence is also limited compared with general data-governance research, so the Saudi interpretation emphasizes control principles and institutional accountability rather than claiming comprehensive validation of every NDMO specification. Rapid changes in agent models and catalog products may also alter implementation choices, although the proposed evidence, authorization, provenance, and human-override

principles are comparatively technology-independent.

11. CONCLUSION

Agentic AI can materially improve enterprise metadata management when autonomy is treated as governed delegation rather than unrestricted generation. The literature supports automation of discovery, profiling, catalog integration, lineage extraction, semantic reconciliation, graph enrichment, and workflow coordination, but it also shows why flexible reasoning cannot substitute for provenance, policy, and accountable roles. The most robust design uses specialized agents over a shared catalog and knowledge graph, deterministic extractors for observable facts, assertion-level provenance for inferred metadata, confidence-aware verification, and risk-tiered human approval.

DAMA and NDMO provide complementary governance lenses. DAMA-oriented practices clarify the enterprise responsibilities that agents must support across metadata, quality, architecture, security, integration, and stewardship. NDMO-oriented governance emphasizes demonstrable institutional control, classification, traceability, and compliance evidence in the Saudi context. Neither framework is strengthened by invisible automation. Agent actions must therefore become auditable metadata events with identifiable sources, permissions, decisions, and approvals.

The central principle is that agents may accelerate metadata production, but authoritative metadata should emerge through evidence-weighted state transitions. Discovery can be observed, meaning inferred, policy evaluated, and actions proposed; promotion to enterprise truth requires verification proportional to risk. This preserves agentic adaptability while limiting hallucination, semantic drift, and uncontrolled policy effects.

REFERENCES

- Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, 49, 424-438. <https://doi.org/10.1016/j.ijinfomgt.2019.07.008>
- Acharya, D. B., Kuppan, K., & Divya, B. (2025). Agentic AI: Autonomous intelligence for complex goals-A comprehensive survey. *IEEE Access*, 13, 18912-18936. <https://doi.org/10.1109/ACCESS.2025.3532853>
- Aikoh, K., Isoda, Y., & Sugimoto, K. (2020). Data profiling method for metadata management. In 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA) (pp. 779-780). IEEE. <https://doi.org/10.1109/DSAA49011.2020.00113>
- Aljarallah, Y., & Alfadhel, M. (2025). Data governance evaluation of the Data Management Office at King Saud University based on the National Data Management Office standards. *Journal of Information Systems Engineering and Management*, 10(10s), 794-805. <https://doi.org/10.52783/jisem.v10i10s.1530>
- Barredo Arrieta, A., Diaz-Rodriguez, N., Del Ser, J., Bannetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Bernardo, B. M. V., São Mamede, H., Barroso, J. M. P., & Santos, V. M. P. D. (2024). Data governance & quality management: Innovation and breakthroughs across different fields. *Journal of Innovation & Knowledge*, 9(4), 100598. <https://doi.org/10.1016/j.jik.2024.100598>
- Blišnák, K., Munk, M., & Pilková, A. (2024). A systematic review of recent literature on data governance (2017-2023). *IEEE Access*, 12, 149875-149888. <https://doi.org/10.1109/ACCESS.2024.3476373>
- Chapman, A., Lauro, L., Missier, P., & Torlone, R. (2024). Supporting better insights of data science pipelines with fine-grained provenance. *ACM Transactions on Database Systems*, 49(2), Article 6, 1-42. <https://doi.org/10.1145/3644385>
- Ehrlinger, L., Schrott, J., Melichar, M., Kirchmayr, N., & Wöß, W. (2021). Data catalogs: A systematic literature review and guidelines to implementation. In *Database and Expert Systems Applications - DEXA 2021 Workshops (CCIS 1479)*, pp. 148-158. Springer. https://doi.org/10.1007/978-3-030-87101-7_15
- Eichler, R., Giebler, C., Gröger, C., Hoos, E., Schwarz, H., & Mitschang, B. (2021a). Enterprise-wide metadata management: An industry case on the current state and challenges. *Business Information Systems*, 1, 269-279. <https://doi.org/10.52825/bis.v1i.47>
- Eichler, R., Giebler, C., Gröger, C., Schwarz, H., & Mitschang, B. (2021b). Modeling metadata in data lakes: A generic model. *Data & Knowledge Engineering*, 136, 101931. <https://doi.org/10.1016/j.datak.2021.101931>
- Goel, K., Martin, N., & ter Hofstede, A. (2024). Demystifying data governance for process mining: Insights from a Delphi study. *Information & Management*, 61(5), 103973. <https://doi.org/10.1016/j.im.2024.103973>

- Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V., Wiest, O., & Zhang, X. (2024). Large language model based multi-agents: A survey of progress and challenges. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence* (pp. 8048-8057). <https://doi.org/10.24963/ijcai.2024/890>
- Hogan, A., Blomqvist, E., Cochez, M., D'Amato, C., de Melo, G., Gutierrez, C., Kirrane, S., Labra Gayo, J. E., Navigli, R., Neumaier, S., Ngonga Ngomo, A.-C., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys*, 54(4), Article 71, 1-37. <https://doi.org/10.1145/3447772>
- Honick, B. J., Polley, K. L., & Qin, J. (2022). Entification of metadata for knowledge-graph-guided discovery. *Proceedings of the Association for Information Science and Technology*, 59(1), 111-120. <https://doi.org/10.1002/pra2.609>
- Jahnke, N., & Otto, B. (2023). Data catalogs in the enterprise: Applications and integration. *Datenbank-Spektrum*, 23, 89-96. <https://doi.org/10.1007/s13222-023-00445-2>
- 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), 101493. <https://doi.org/10.1016/j.giq.2020.101493>
- Li, X., Wang, S., Zeng, S., Wu, Y., & Yang, Y. (2024). A survey on LLM-based multi-agent systems: Workflow, infrastructure, and challenges. *Vicinagearth*, 1, Article 9. <https://doi.org/10.1007/s44336-024-00009-2>
- Pan, B., Stakhanova, N., & Ray, S. (2023). Data provenance in security and privacy. *ACM Computing Surveys*, 55(14s), Article 323, 1-35. <https://doi.org/10.1145/3593294>
- Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*, 36(7), 3580-3599. <https://doi.org/10.1109/TKDE.2024.3352100>
- Petrik, D., Untermann, A., & Baars, H. (2024). Functional requirements for enterprise data catalogs: A systematic literature review. In *Software Business. ICSOB 2023 (LNBIP 500, pp. 3-18)*. Springer. https://doi.org/10.1007/978-3-031-53227-6_1
- Qin, J., & Yu, B. (2024). Metadata in trustworthy AI: From data quality to ML modeling. *Proceedings of the International Conference on Dublin Core and Metadata Applications*, 2024. <https://doi.org/10.23106/dcmi.953354037>
- Reznik, T., Raes, L., Stott, A., De Lathouwer, B., Perego, A., Charvat, K., & Kafka, S. (2022). Improving the documentation and findability of data services and repositories: A review of (meta)data management approaches. *Computers & Geosciences*, 169, 105194. <https://doi.org/10.1016/j.cageo.2022.105194>
- Sawadogo, P. N., & Darmont, J. (2021). On data lake architectures and metadata management. *Journal of Intelligent Information Systems*, 56(1), 97-120. <https://doi.org/10.1007/s10844-020-00608-7>
- Schoenenwald, A., Kern, S., Viehhauser, J., & Schildgen, J. (2021). Collecting and visualizing data lineage of Spark jobs. *Datenbank-Spektrum*, 21, 179-189. <https://doi.org/10.1007/s13222-021-00387-7>
- Tonnarelli, M., Kumara, I., Driessen, S., Tamburri, D. A., van den Heuvel, W.-J., & Oor, P. (2025). Data catalog tools: A systematic multivocal literature review. *Journal of Systems and Software*, 230, 112584. <https://doi.org/10.1016/j.jss.2025.112584>
- Walsh, M. J., McAvoy, J., & Sammon, D. (2022). Grounding data governance motivations: A review of the literature. *Journal of Decision Systems*, 31(sup1), 282-298. <https://doi.org/10.1080/12460125.2022.2073637>
- Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6), 186345. <https://doi.org/10.1007/s11704-024-40231-1>
- Zhang, B., & Soh, H. (2024). Extract, define, canonicalize: An LLM-based framework for knowledge graph construction. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 9820-9836). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-main.548>
- Zhu, Y., Wang, X., Chen, J., Qiao, S., Ou, Y., Yao, Y., Deng, S., Chen, H., & Zhang, N. (2024). LLMs for knowledge graph construction and reasoning: Recent capabilities and future opportunities. *World Wide Web*, 27(5), Article 58. <https://doi.org/10.1007/s11280-024-01297-w>