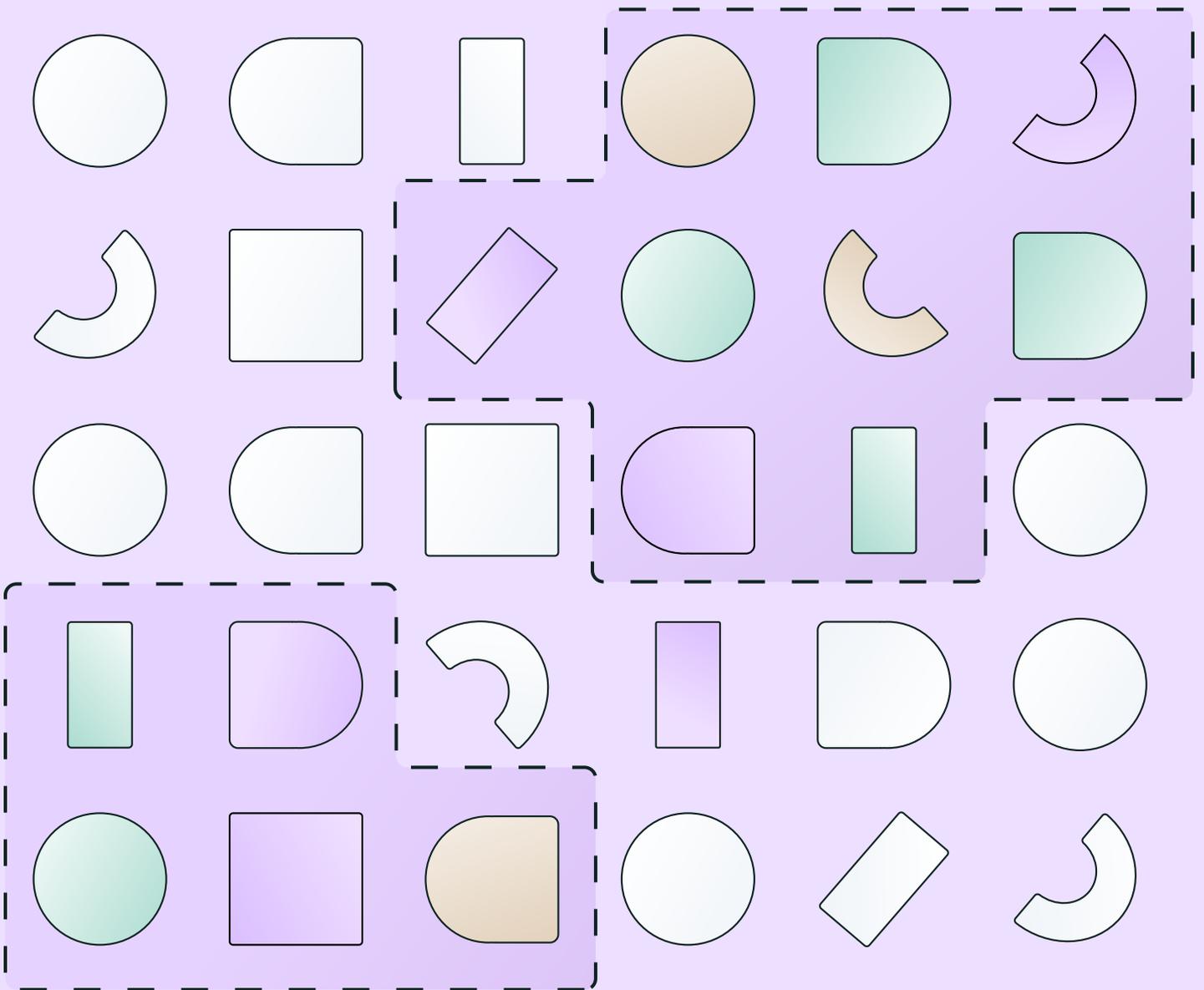


The Age Of Context

How AI is Reshaping Metadata Management



THE EVOLUTION OF METADATA CATALOGS

Table of Content

- 03 Introduction
- 04 The three generations of metadata catalogs
- 09 Metadata management platforms market landscape 2026
- 12 Conclusion

INTRODUCTION

Why metadata catalogs are transforming in 2026

Imagine an AI agent tasked with analyzing your Q4 revenue trends. It finds a dataset that seems relevant, but has no way to know it's actually outdated, contains test data, and that the key metric was redefined last month. The agent confidently delivers insights based on bad data.

This scenario isn't hypothetical, it's why AI for analytics consistently fails to work at scale. [MIT's recent study](#) found that 95% of enterprise AI pilots fail to deliver measurable returns, not due to model limitations but because AI tools can't understand an organization's unique data logic, definitions, and workflows on their own.

The solution is [providing AI with context](#) through metadata, but not all metadata management platforms are created equal. The tools that worked for data teams won't necessarily cut it for AI agents, which can't ask clarifying questions or sense when something feels off.

In this guide, we'll explore how data catalogs are evolving from passive documentation repositories into active AI context layers, analyzing the metadata management space to understand what it takes to support AI agents that can actually be trusted with your data.

Whether you're a C-level executive, data leader, governance professional, or business leader, understanding this evolution is crucial for positioning your organization's AI strategy for success in 2025 and beyond.



95% of enterprise AI pilots fail to deliver measurable returns

The GenAI Divide, State of AI in Business 2025, MIT

The three generations of metadata catalogs

GENERATION 1

THE BUSINESS USER'S SOURCE OF TRUTH

The first generation of data catalogs emerged from a simple but powerful vision: create a single source of truth for organizational data. These early catalogs, essentially sophisticated business glossaries, focused on three core capabilities:

- ⊕ Providing a common glossary for business terminology
- ⊕ Managing access control
- ⊕ Serving as a discovery point where business users could find the data they needed

TL;DR

- ✓ Single source of truth vision
- ✓ Manual updates, perpetually stale
- ✓ Built for business, rarely adopted

However, this vision ran into fundamental problems. The reality is that the source of truth doesn't actually live in one place. Critical business logic is scattered across Git repositories or embedded in semantic layers. More problematically, these catalogs relied on **manual maintenance**: someone had to remember to update definitions, ownership, and lineage every time something changed. The result was predictable: catalogs were **always lagging, never fully deployed, and therefore never adopted**. Engineers built data products in their tools, while the catalog sat on the sidelines, perpetually out of sync. Trust eroded, and people stopped relying on something that felt outdated the moment it went live.

TL;DR

- ✓ Keeps pace with modern data stacks
- ✓ Engineers optimized, business left behind
- ✓ Tells what exists, not what to do

GENERATION 2

THE DATA PRACTITIONER'S METADATA PLATFORM

As the modern data stack evolved, so did catalogs. The second generation shifted focus from business users to data practitioners. These metadata management platforms addressed the complexity that came with distributed data architectures by providing automated crawling to discover assets, lineage tracking to understand dependencies, impact analysis to predict the effects of changes, and insights for optimization and cost reduction. They also introduced documentation managed in Git repositories and version-controlled alongside the code itself. This generation recognized that manual documentation can't keep pace with the velocity of change in modern data environments. Yet while these tools served data teams well, they still treated metadata as primarily descriptive, telling you what existed and how it connected, but not actively participating in how data was managed or consumed. And in doing so, they left the business behind. The interfaces and workflows were designed for engineers, not analysts or domain experts. Business users lost visibility into how data definitions evolved, and the shared understanding that Gen 1 had promised was never rebuilt.

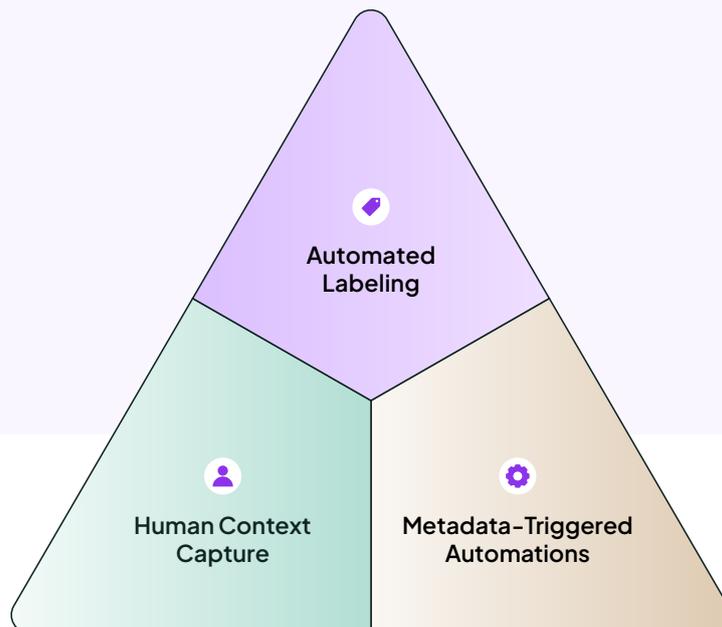
GENERATION 3

THE AI AGENT'S CONTEXT LAYER

A new interface has entered the data stack: AI agents. They're becoming the way business and technical users interact with data, through natural language. But AI has one critical need: context. It doesn't know what "revenue" means in your business, which dashboards are trustworthy, or how metrics are calculated. Unlike a human, it can't tap a veteran analyst on the shoulder to ask what a field really means. To operate reliably, AI needs a dynamic context layer where meaning, trust, and business logic are provided through metadata, and this transformation requires three critical capabilities:

TL;DR

- ✓ AI agents get dynamic context
- ✓ Human knowledge becomes machine-executable
- ✓ Orchestrates actions, not just describes





Automated labeling

We need dynamically applied labels based on live conditions (tags like “certified,” “well-documented,” “contains PII,” or “high query cost” that update automatically as underlying conditions change). Assets are constantly re-evaluated against rules, ensuring AI agents always work with current state information rather than stale documentation. These labels are called **Active Metadata Tags**.



Human context capture. The vast amount of context that lives in human heads must be captured, but in a fundamentally different way than the traditional glossaries. This context needs to be coded, version-controlled, and easily editable by business users without technical skills. It must map directly to the data model resources, creating a clear relationship between business context and the data.



Metadata triggered automations

Metadata must trigger action, not just be collected and stored. This is essential if we want to fully leverage AI capabilities. You should be able to create workflows (potentially managed by AI agents) to automatically archive unused data assets, alert when downstream dashboards might break due to schema changes, flag when new models don’t follow modeling standards, identify and potentially pause costly pipelines that are going unused, or escalate when highly-used dashboards lose their certification status. This shift from information to action is what transforms a catalog from a reference tool into an **active metadata management platform**.

THE THREE GENERATIONS OF METADATA CATALOGS: CAPABILITY COMPARISON

Capability	Gen 1 Business Glossary	Gen 2 Metadata Platform	Gen 3 AI Context Layer
Automated Discovery	✗	✓	✓
Lineage Tracking	✗	✓	✓
Dynamic Trust Signals	✗	⊗ Static tags only	✓
Business Context	⊗ Manual glossary	⊗ Git docs (engineer-focused)	✓
Workflows	✗	✗	✓
AI Agent Ready	✗	✗	✓

The enabling infrastructure

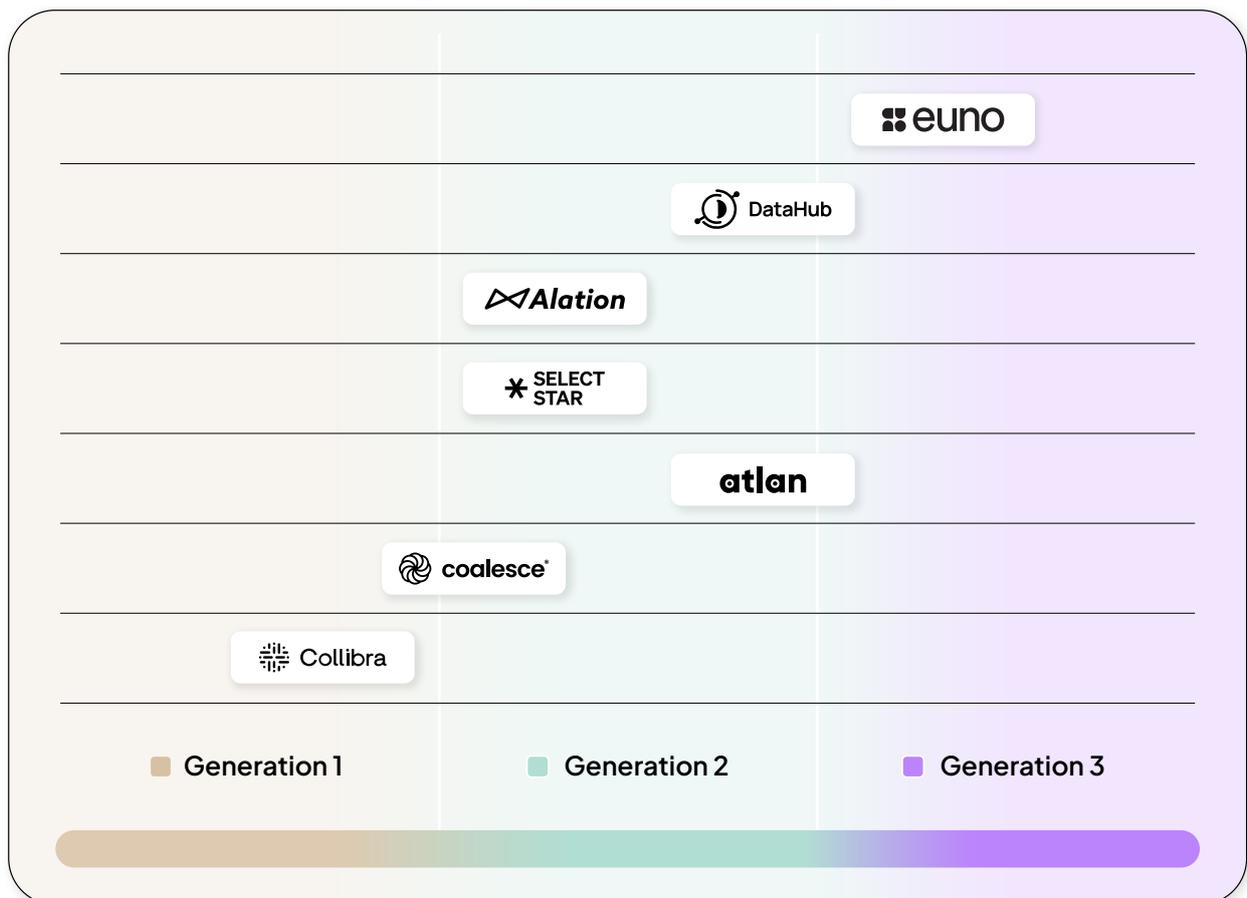
Generation 3 cannot exist without the foundation that Generation 2 provides: the automated discovery, lineage tracking, and metadata collection remain essential. However, each generation requires different underlying infrastructure to serve its primary use case. Some tools and platforms are adapting existing infrastructure, while others are building AI-first from the ground up.

The adaptation approach creates complexity: platforms must support existing customer features while building new AI capabilities. This dual requirement inevitably slows innovation cycles, and in a rapidly moving AI landscape, this becomes a strategic consideration.

Metadata Management Platforms

MARKET LANDSCAPE 2026

The AI era is reshaping the metadata management platforms market. Some platforms remain rooted in their glossary origins, others have matured into data practitioner-focused, and a few are beginning to experiment with the context capabilities required for AI. Here's how they stack up in 2026.





Euno represents a Gen 3 approach, focusing specifically on AI agent needs. Euno emphasizes three core capabilities: [active metadata tags](#) that stay up to date, **automated workflows** so action happens automatically, and **structured business context capture** so business knowledge is never lost. Under the hood, Euno’s architecture is designed to answer the kind of tough, multi-step questions AI will constantly ask, not just “show me dashboard related to MRR?” but “which dashboards rely on this finance model, and are they certified?” To do this reliably at scale, Euno runs on a graph database with a query language purpose-built for flexibility and speed, and its **MCP server** allows AI agents to connect directly and retrieve the precise context they need to deliver reliable results.



[Atlas](#) is a strong Gen 2 player, with deep roots in automated metadata collection, lineage, and data practitioner use cases. At [Activate 2025](#), they showcased their ambition to move toward Gen 3 with innovations like the MCP server, designed to feed enterprise context directly into AI agents. It’s a bold and forward-looking roadmap, but their underlying infrastructure was originally built for descriptive metadata, not for the fully dynamic context layer AI demands. Their roadmap shows strong commitment to AI-first metadata management, though the transition from their Gen 2 foundation will be worth monitoring.



[DataHub](#), as an open-source project, has established itself as a strong Gen 2 player with robust automated discovery, lineage, and practitioner-focused capabilities. In their recent [Series B announcement](#), they made it clear they’re leaning into the AI era by accelerating R&D with a focus on **AI governance and context management capabilities** — signaling a willingness to evolve toward Gen 3. They are saying the right things, and we believe their upcoming summit at the end of October will likely feature important roadmap announcements in this direction. Definitely one to watch.



[Alation](#) was originally built as a business glossary platform, focused on giving organizations a reliable “single source of truth” for definitions, access, and discovery. Over time it has made steady progress toward Gen 2 metadata capabilities with automated discovery, lineage, governance, etc. With its recent [acquisition of Numbers Station](#), Alation has made it clear that it’s moving more aggressively into AI use cases. The deal adds agentic workflows that combine trusted metadata with AI agents acting over structured data, which helps Alation provide a more end-to-end solution. The question is whether Alation’s metadata architecture can scale for the demands of highly responsive agents, and how well it will interoperate with other agents (e.g., Snowflake Cortex, Databricks Genie etc.) that organizations may already be adopting.



[Coalesce Catalog](#) (formerly CastorDocs) automates discovery, provides lineage, supports governance, and adds collaboration features that help data teams manage complexity. Its AI vision, however, is not about building a context layer for autonomous agents. Instead, it focuses on empowering humans with AI — things like natural-language discovery, SQL coding assistants, and AI-driven documentation. This keeps Coalesce catalog firmly oriented toward improving the human data experience, rather than enabling AI agents to act as first-class citizens in the data ecosystem.



[Select Star](#) is branding itself as a Metadata Context Platform for Data & AI, even introducing an [MCP server](#) to connect its catalog with LLMs. In practice, though, the context it provides largely mirrors what’s already in the catalog, enabling users to query lineage, glossary entries, and relationships through natural language. What it doesn’t yet deliver are dynamic trust signals (like quality, certification, or usage) that AI agents would need to act autonomously. For now, Select Star remains more a tool to help humans, using LLMs to navigate metadata, rather than a true context layer powering AI-driven actions.



[Collibra](#) has built its reputation as a metadata catalog (focused on business glossaries, definitions and some governance and lineage) helping organizations build trust and clarity around their data. Their vision for the AI era leans heavily toward enhanced discovery: they've introduced [Collibra AI Copilot](#), which uses natural-language questions to help all users find data assets, glossary terms, or documentation. Recently, they've announced the integration with [Deasy Labs](#) which brings automated discovery and enrichment of unstructured data. These aren't yet full context layers for AI agents, but they make Collibra's catalog more AI-aware.

CONCLUSION

The evolution from business glossary to AI context layer represents more than technological advancement, it's a fundamental reimagining of what a metadata platform should be. We're moving from a passive repository of information to an active participant in data governance, from human-readable documentation to machine-executable context, and from describing what exists to orchestrating what happens next.

How can your organization prepare for the AI era? See how a true Gen 3 context layer works in practice.