

CASE STUDY

From Self-Hosted OpenMetadata to a Managed **Data-Product** Platform

How a global premium automaker consolidates its data landscape and lays the foundation for agentic AI with Collate



Migrated

From self-hosted OpenMetadata OSS into a managed AI context layer with Collate

Reduced

Operational burden of running a context layer at enterprise scale

Agent-ready context

Knowledge graph and semantic understanding for agentic AI

A global premium automaker, selling hundreds of thousands of vehicles a year with tens of thousands of employees worldwide, runs on data across manufacturing, supply chain, and connected-vehicle operations. Departmental autonomy had left the company with a fragmented data estate and overlapping tools. To drive better trust in data and AI, the data platform team standardized on the OpenMetadata open-source project, but then hit the operational frictions of self-hosting it at enterprise scale. The team migrated to Collate's managed OpenMetadata service in its own cloud environment, consolidating onto a single managed context layer, enabling new Collate capabilities, and building toward an AI-driven, data-product approach to governance.

INDUSTRY

Automotive

DEPLOYMENT

BYOC (bring-your-own-cloud) — enterprise cloud tenant

TECHNOLOGIES

Cloud data warehouse, cloud platform, BI tools, dbt, OpenMetadata, open-source data-quality tooling

"Collate gives us that same standard as a managed context layer inside our own cloud environment: a single lens into our data landscape that lets us build governed data products instead of maintaining infrastructure."

Senior Product Manager, Data Platform, a global premium automaker

A Fragmented Data Landscape That Siloed Data Understanding

The automaker built its data capabilities through years of departmental autonomy, with individual teams choosing their own tools and setting their own standards. When the company centralized around a single data platform, that history surfaced as a fragmented estate: overlapping catalog tools, inconsistent metadata, and no shared definition of what "good" data looked like. The data platform team had to unify the estate without slowing the domains that depended on it.

LEGACY CATALOG TOOLING

The company previously ran a legacy commercial data catalog but found it offered limited new capability for what it cost, with weak account support. This led the team to explore OpenMetadata as an alternative, though it meant maintaining parallel tools with divergent metadata.

THE COST OF SELF-HOSTING AT SCALE

Scaling OpenMetadata into production across the enterprise was non-trivial. For the number of departments and data assets the company had, the work of patching, upgrading, scaling, and supporting a self-hosted context layer grew heavy. It pulled a specialized platform team toward infrastructure maintenance and away from governance work.

A STRICT ENTERPRISE SECURITY BAR

The company handles sensitive manufacturing and supply-chain data and enforces a rigorous security policy, including a prohibition on broad query access to its cloud data warehouse. Any tooling had to run inside the company's own cloud environment and clear its minimum cybersecurity requirements before touching production data.

AMBIGUITY BLOCKING AI ADOPTION

When the company started building AI agents on its data warehouse's built-in AI service, semantic ambiguity became a blocker. The same term, like "order," could mean a production order or a sales order depending on context. Without a shared semantic layer, AI agents couldn't reason reliably over the company's data.

A Managed, AI-Driven Context Layer with Collate

The automaker migrated from self-hosted OpenMetadata to Collate's managed service, deployed to a BYOC (bring-your-own-cloud) instance inside its own cloud environment. The conversion let the team keep the open standard it had chosen, hand off the operational work, and consolidate away from overlapping tools.

SAME STANDARD, MANAGED OPERATIONS

Collate runs on the OpenMetadata standard the company had already adopted, and the migration process was quick, automated, and preserved the work the team had invested into the open ecosystem. As a managed service inside the company's cloud, Collate took patching and scaling work off the platform team.

SECURITY-CLEARED, BYOC DEPLOYMENT

Collate is deployed entirely within the company's own cloud environment, with credentials held in the company's own key vault. Instead of broad warehouse access, the deployment maps warehouse access to the company's existing identity provider. This meets security requirements while still capturing metadata context and semantic understanding across the data landscape automatically.

A SINGLE CENTRAL CONTEXT LAYER

The legacy commercial catalog was deprecated, and the Collate service now serves as the lens into the company's data landscape. Collate connects the cloud data warehouse, BI tools, data-quality tooling, and dbt, and gives business domains a single place to discover, understand, and trust data.

A SEMANTIC LAYER FOR AI UNDERSTANDING

To resolve the ambiguity blocking AI adoption, the company is adopting Collate's auto-generated knowledge graph and ontology as a semantic layer, exposed through API and MCP (Model Context Protocol, a standard for connecting AI agents to data). External agents reason over governed definitions instead of a custom-built graph database, distinguishing between a "production order" and a "sales order" to better understand the data.

Building the Foundation for Governed Data Products and Agentic AI

The automaker replaced the operational drag of a siloed toolset and self-hosting with a single managed context platform, setting up a roadmap for increased data trust and agentic AI initiatives across its enterprise.

CENTRALIZING DATA QUALITY AND CONTRACTS

The data team leverages Collate's data quality test cases, rules, and dashboarding to understand data quality across its landscape. Data contracts provide clear expectations between data producers and consumers across manufacturing, supply chain, and other business domains.

TRUSTED DATA PRODUCTS

The company runs Collate to power a data product strategy built on standardized templates that bundle quality, contracts, and metadata into reusable, governed packages for data team self-service. Collate is the central place where the company defines, publishes, and discovers these trusted data products across its enterprise.

A CO-DEVELOPMENT PARTNERSHIP

The company has become a roadmap-shaping design partner. Working with Collate, the team has helped design customizable intake forms for data products, domains, and glossary terms, with required fields and custom properties to ensure consistent and standardized processing for these data assets.

READY FOR AGENTIC AI

The team is moving toward computational governance that cuts manual compliance work and supports trustworthy agentic AI over its data estate. The company can ground AI agents in its warehouse's AI service in governed, unambiguous definitions built on Collate's knowledge graph and semantic layer.



"We chose OpenMetadata and ran it ourselves, but at our scale self-hosting kept pulling our team toward infrastructure work. Collate gives us that same standard as a managed context layer inside our own cloud environment: a single lens into our data landscape that lets us build governed data products instead of maintaining infrastructure."

Senior Product Manager, Data Platform

A global premium automaker



getcollate.io

POWERED BY OPENMETADATA