

CASE STUDY

Unifying a fractured data landscape across a global ecosystem

How Rakuten consolidated siloed catalogs into a single group-wide OpenMetadata platform, and laid the foundation for AI agents



2-3 mo → <1 hr

Projected time to discover whether a dataset is useful for a use case, once OpenMetadata is the group-wide platform.

30-60% → 80%

Projected lift in data utilization across onboarded business platforms.

Thousands of tables

Scale of the data estate for automatic description generation.

Rakuten runs one of the world's largest digital ecosystems. Founded in Japan in 1997, the company now spans more than 30 countries and regions, over 70 services, and more than 2.1 billion users across e-commerce, mobile, fintech, and media, with global gross transaction value exceeding ¥48 trillion. Beneath that ecosystem sit many independent data platforms, each maintained by its own engineering team, with catalogs fractured across Excel files, Confluence pages, and bespoke web apps. To consolidate discovery, governance, and utilization across the group, and to build a foundation ready for AI use cases, Rakuten's Data Platform team adopted **OpenMetadata** as a single, group-wide platform.

INDUSTRY

Internet Services & E-commerce

HEADQUARTERS

Tokyo, Japan

FOUNDATION

OpenMetadata (open-source, self-managed on Kubernetes)

TECHNOLOGIES

MySQL, PostgreSQL, Cassandra, MongoDB, Kafka, Spark, Airflow, Rakuten-wide data federation platform

"We see that OpenMetadata is a complete package solution for Rakuten Group."

Muqtafi Akhmad, Data Platform Department, Rakuten Group

A Fractured Catalog Landscape Across a Sprawling Ecosystem

The same broad ecosystem of complementary services that gives Rakuten its strength also made its data hard to govern. Many data platforms, each owned by a different engineering group, meant many catalogs, many governance models, and no shared view of what data existed or how businesses related to one another. Data consumers could not find relevant data or judge whether it was useful, and pipelines ran up costs even when no one used their output.

FRACTURED, SILOED CATALOGS

Each team maintained its own catalog in whatever form it chose: a spreadsheet or Excel file, a Confluence page, or even a dedicated web app built as a data catalog. Each had its own definition of what a data catalog was, so there was no consistent way to see what data existed across the group.

INCONSISTENT, UNCLEAR GOVERNANCE

Because each team did things its own way, governance was unclear across the group. From the user's perspective, the relationship between data on one platform and another was opaque, and there was no standard classification, taxonomy, or domain.

SLOW, UNCERTAIN DISCOVERY

When a data consumer arrived with a business question, they often didn't know which team could answer it or where to look. A user survey found users needed two to three months to find the useful data for a problem.

WASTED PIPELINES, LOW UTILIZATION

Many pipelines produced data that was underutilized, giving it low business impact while Rakuten still paid to compute, store, and maintain it for little return.

NO AI-READY FOUNDATION

The fractured catalogs offered no path to the AI use cases the team wanted to enable: automatic description generation across thousands of tables, and agents that query lineage and data characteristics to debug pipelines.

A Single Group-Wide Platform on OpenMetadata

After a requirement definition exercise, evaluation of solutions on the market, and a POC to validate fit, Rakuten selected **OpenMetadata** as its group-wide platform. The team ran that selection with the first two business platforms rather than for them, listing the problems each needed solved and the features they wanted before making a choice. That paid off in adoption: the businesses recognized the result as the answer to problems they had helped define.

ONE CONSOLIDATED PLATFORM

OpenMetadata replaces the patchwork of Excel, Confluence, and custom catalog apps with a single group-wide platform: a place to capture any data and the processes that move it, from operational databases through pipelines to the data platforms that serve it.

FAST SEARCH & DISCOVERY

Search and discovery let consumers find relevant data by filtering on data product, domain, or tags, then judge its usefulness in the platform instead of hunting across teams for months.

BUILT-IN, STANDARD GOVERNANCE

Rather than per-team, ad-hoc models, OpenMetadata provides standard governance the team can apply consistently, using data products, a glossary, tagging, and rules and policies to manage access.

KUBERNETES-NATIVE ORCHESTRATION + OPENLINEAGE

The release of a Kubernetes-native orchestrator lets Rakuten deploy OpenMetadata without a dedicated Airflow cluster, which the team cited as a significant cost saver. The team is also working on OpenLineage integration to smooth ingestion from its various databases and processes.

A Foundation for Data Teams and AI

Rakuten is continuing to invest in its OpenMetadata deployment, expanding the footprint across more business groups and use cases. With cataloging in place, the work shifts from building a foundation to adopting more capabilities, deepening observability, and going further into AI capabilities.

DISCOVERY: 2-3 MONTHS → UNDER AN HOUR

With a single searchable platform, Rakuten expects determining whether a dataset is useful to drop from two-to-three months to under an hour. That turns a quarter-long wait into a single working session, so data scientists and analysts can answer business questions faster.

UTILIZATION: 30-60% → 80%

As businesses onboard and data becomes discoverable, Rakuten expects utilization to rise from a 30-60% baseline toward 80%. Cross-business visibility and initiatives are a key driver for the deployment.

OBSERVABILITY FOR PIPELINE OPERATIONS

Data quality, profiling, and lineage are next on the roadmap. Together they show how data behaves at runtime rather than just how it is structured, which is what the team needs to smooth day-to-day pipeline operations.

BUILDING TOWARD AI USE CASES

Beyond human discovery, Rakuten sees OpenMetadata as the foundation for AI. Automatic description generation and MCP integration are on the roadmap, so agents can query lineage and data characteristics to help engineers debug pipelines.

"OpenMetadata offers the technical metadata, business metadata, quality, and observability that we are looking for. It's easy to use the search and discovery features, it has the standard governance features which fit well with our use cases — with the notion of domain, products, tags — and it's an open source platform with a growing community. We see that OpenMetadata is a complete package solution for Rakuten Group."

Muqtafi Akhmad

Assistant Manager, Data Pipeline Platform Team, Data Platform Department, Rakuten Group, Inc.



open-metadata.org

THE OPEN STANDARD FOR DATA CONTEXT