

CASE STUDY

Turning institutional knowledge into trusted AI context

How Scout24 built Collate into the governance and context catalog at the base of its AI platform

**<10 weeks**

From signed contract to production

10+

Subsidiaries brought under a single data context layer

50%

Targeted reduction in analyst time spent managing data quality

Founded in 1998 and listed on Germany's DAX index, **Scout24** is the country's leading real estate platform, with roughly 1,100 employees and serving 500,000+ B2C and 26,000+ B2B customers across a wide variety of subsidiaries. These different subsidiaries left data scattered and hard for AI to reason over. To give AI trustworthy context at scale, Scout24 replaced a self-managed legacy catalog with **Collate**, going from a 1.5-week proof of concept to production in under 10 weeks, and building Collate into the governance and context catalog at the base of its AI platform.

"We all talk about AI, but the real product is not AI. Context is the real product. AI just multiplies it."

Angelita Frozza Sanches, Head of Core Data Platform, Scout24

INDUSTRY

Real Estate Technology

HEADQUARTERS

Munich, Germany

DEPLOYMENT

Collate SaaS with a hybrid agent on Kubernetes

TECHNOLOGIES

Starburst/Trino, Delta Lake, AWS Glue, Airflow, PostgreSQL/Aurora, DynamoDB, MicroStrategy, MCP

Making Scattered, Multi-Subsidiary Data Understandable to AI

Scout24's growth through subsidiaries means its data estate spans companies in different sectors, each with its own platform, technologies, and formats. To deliver trusted data and AI at scale, the data platform team had a clear vision: empower people to get insights with no training, no technical barriers, and no need to know SQL. But simply pointing an LLM at that data without context led to many complications for the team.

SCATTERED DATA ACROSS SUBSIDIARIES

Each subsidiary posts data to a shared lake with data contracts that describe team, table, fields, types, and PII. But the assets themselves still live in different technologies and formats. Without a common context layer, AI had no reliable way to reason across the organization.

DOCUMENTATION BUILT FOR HUMANS, NOT AGENTS

Producers and owners had been expected to document their sources, but doing it by hand and keeping it current was time-consuming — and this documentation was designed for human users, not agents. Fed to agents, it produced seemingly trustworthy responses that were incorrect.

LOW ADOPTION OF A SELF-MANAGED CATALOG

The previous self-managed catalog required engineers to write descriptions manually, driving low adoption beyond a small engineering audience. The broader data community stayed disengaged, with description coverage thin and falling behind as the estate grew.

NO SELF-SERVICE DATA QUALITY

Data quality was uncoordinated across the organization. Analysts spent significant time chasing freshness issues, tracking down owners, and resolving discrepancies — with no central model and no safe way to create and monitor checks at scale.

NO TRUSTED CONTEXT FOR AI

Scout24 had made AI-led product innovation a priority, with MCPs running on Starburst so users could query without SQL. But without curated, trustworthy context, those answers couldn't be relied on — nothing made the estate contextual, transparent, or governed for agents.

An AI Virtuous Cycle, Anchored on Collate as the Context Catalog

Scout24 selected **Collate** after a 1.5-week proof of concept, having evaluated it against its incumbent catalog's cloud version and other leading data-quality and governance tools. It deployed on SaaS with a hybrid agent on Kubernetes for secure connectivity, following Collate's Launchpad framework for rapid onboarding. Scout24 calls the result an "AI virtuous cycle": using AI to enrich the context that AI then consumes, with Collate as the governed source of truth at its base.

MCP-PUBLISHED KNOWLEDGE SOURCES

Rather than ask people to document more, Scout24 published every source it deemed legitimate — team communication tools, docs and knowledge bases, code repositories, database systems, file storage, even external sources — as an MCP on a central, approved MCP registry, including Collate's own read/write MCP.

UNIFIED, GOVERNED DISCOVERY

A single Collate instance spans Scout24's 10+ subsidiaries, connecting Starburst/Trino, Delta Lake, AWS Glue, PostgreSQL/Aurora RDS, DynamoDB, and MicroStrategy under one discovery layer. RBAC and tag-based policies control visibility by team — analysts see only consumption-ready assets while ETL-internal tables stay restricted — without breaking lineage.

AI-ENRICHED DOCUMENTATION, HUMAN-VERIFIED

An LLM reads from those sources and, through the Collate MCP, generates and enriches documentation inside Collate for assets that had none: table and column descriptions, PII classification, lineage, certification, and usage. Every AI-generated entry carries a visible "AI generated" tag; when an owner reviews it, they remove the tag and it becomes original, human-verified documentation.

CONTEXT-FIRST AI ANSWERS WITH CONFIDENCE SCORES

Scout24's internal data agents are instructed to look for context in Collate before generating SQL. When an agent answers, the user sees a confidence score based on how much trusted context it found versus how much it had to improvise — the less context, the lower the score, so users can evaluate the answer for themselves.

Trusted Context at Scale, for Both AI and Humans

Scout24 made Collate the governed context layer beneath its AI platform, turning scattered institutional knowledge into a single source of truth that serves people and agents alike, with transparency built into every answer.

SELF-SERVICE ANSWERS, NO SQL REQUIRED

Scout24 set out to let anyone in the business get the insights they needed to drive decisions. Now they can: people ask in plain language and the agents answer from trusted context in Collate rather than guessing — answers directly across the company, without queuing for them.

DOCUMENTATION THAT KEEPS PACE WITH THE ESTATE

Assets that never had documentation now get it automatically, drafted by AI and verified by owners, and scattered tribal knowledge is centralized in one place. Coverage grows as the data grows instead of falling behind it, and every entry is either human-verified or clearly flagged as AI-generated.

GOVERNANCE BUILT FOR AI AND HUMANS

Governance and context were always needed, but now they serve a new outcome. AI and humans have different needs, and not every source of institutional knowledge is equally trustworthy or well-structured for AI. Collate gives the team one place to govern, tag, and expose only trusted context — whether the consumer is an analyst or an agent.

A COMPOUNDING CONTEXT LOOP IN PRODUCTION

The AI virtuous cycle is live: AI enriches context, humans verify it, and verified context sharpens the next answer. Accuracy climbed once Collate was placed first in the agents' answer hierarchy — more accurate answers, higher confidence, transparency, and governance from a single source of truth.

LEGACY CATALOG RETIRED

Within weeks of go-live, Scout24 announced to executive leadership the deprecation of its self-managed legacy catalog. The team also consolidated other data tools into Collate as the new unified context layer for both its human and AI data teams.



getcollate.io

THE AI FOR DATA PLATFORM