EXECUTIVE BRIEF

# Governed Migration, Lineage, and Catalog Reconciliation

How Theom supplies the evidence layer that keeps a modernizing enterprise data platform governed and its catalog accurate

| PREPARED BY | DOCUMENT | DATE |
| --- | --- | --- |
| Theom | Public reference version | August 2026 |

## Executive perspective

Enterprises modernizing a data platform have to keep a large downstream analytics community running while the ground moves underneath it. The estate typically spans relational and microservice transactional sources, change-data capture and continuous-load ingestion staged through object storage, transformation and orchestration tooling on top, and a cloud warehouse or lakehouse on a medallion pattern serving BI, analytics teams, and a growing population of AI agents.

Theom provides the continuous evidence layer for that transition: what data exists, which data is sensitive, where it came from, who still uses raw or transitional paths, whether each identity's access remains appropriate, how identity carries through BI and agent workflows, and where live platform evidence no longer matches the enterprise catalog.

> **Recommended starting point:** Establish governance and migration evidence in the warehouse serving environment first, then extend the lineage graph across the lake and the upstream ingest and orchestration layers using the available connector and pipeline metadata.

## Architecture context

### The stack this brief assumes

- Relational and microservice transactional systems — Oracle, SQL Server, Postgres, MySQL — are the operational sources of record.
- Replication and ingest run through change-data capture and continuous load — GoldenGate, Debezium, Fivetran, Snowpipe, Kafka Connect — plus object storage, increasingly carrying documents and images that pipelines turn into structured tables.
- A cloud warehouse or lakehouse on a medallion pattern — raw, silver, gold — is the serving layer: Snowflake, Databricks, BigQuery or Fabric, often more than one per estate.
- Transformation runs in dbt, Spark, or in-warehouse SQL, with Airflow, Dagster, or a managed scheduler orchestrating it.
- Consumption spans BI tools such as Tableau, Power BI and Looker, analytics teams working directly in SQL and notebooks, and AI agents consuming data programmatically.
- An enterprise catalog — Alation, Collibra, Atlan — holds business context, and collaboration platforms are where data comes to rest once it leaves the estate.

### Where the governance gap opens

- The medallion pattern is rarely finished. Consumers keep reading raw ingested tables directly, so no one can say which raw objects are safe to retire.
- Evidence is scattered across the platform, the transformation tool, the orchestrator, and the catalog, and none of those is a complete record on its own.
- Catalog entries are authored top-down and drift as tables, pipelines, copies, and consumers appear between reviews.
- Identity is obscured wherever a shared service account or a BI tool sits between a person and the data they reached.
- A table derived from a document or an image has no obvious path back to its source, so some of the most sensitive content in the estate is also the least traceable.

## 01 Raw-table migration tracking

### The situation

Enterprises want users and applications to move from raw ingested data toward curated, governed products. Direct raw-table consumption makes it difficult to determine which paths are still active, which dashboards or jobs depend on them, and when a raw object can safely be deprecated.

### How Theom contributes

- Continuously inventory and classify raw, curated, external, and open-table-format objects across the agreed monitoring scope.
- Use observed query and access activity to identify active consumers of raw tables, including people, service identities, transformation jobs, and BI workloads using the available identity context.
- Connect raw objects to downstream derivatives, dashboards, and data products so platform owners can see migration dependencies before making a change.
- Add freshness, ownership, access, and sensitivity context to prioritize which raw paths should move first and which require additional controls during transition.
- Track declining raw usage and growing curated usage over time, producing evidence for deprecation decisions rather than relying on surveys or static documentation.

### Practical outcome

- A current list of raw-table consumers and the downstream assets they support.
- A measurable migration baseline and progress view for raw-to-curated adoption.
- Evidence that a raw path is unused or that remaining dependencies must be remediated before retirement.
- Reduced risk of breaking analytics or application workflows during the architecture transition.

> **Delivery approach:** The initial phase establishes the warehouse and BI baseline, then extends coverage across the lake and any near-real-time paths as the program expands.

---

## 02 Lineage and controls across ingestion and serving paths

### The situation

Data reaches consumers through multiple concurrent paths. The governance requirement is not simply to draw a lineage diagram, but to understand how sensitivity, ownership, freshness, access rights, masking, and actual usage change as data is copied, transformed, cataloged, and served.

### How Theom contributes

- Combine warehouse object lineage and query activity with transformation artifacts, pipeline and job metadata, catalog context, and supported source or lake integrations.
- Trace one representative change-capture path and one representative continuous-load path from source through ingestion into the warehouse.
- Extend the map through streaming, batch or micro-batch processing, the lake, the warehouse, the BI layer, and the consuming identity using connector, job, and application metadata.
- Carry classification, ownership, freshness, authorization, and policy context along the lineage graph to reveal copied sensitive data, stale downstream products, new consumers, and missing controls.

- Compare access and masking posture across derivatives so teams can identify control drift when a protected source is copied into a less-protected table or serving path.

### Practical outcome

- Impact analysis before schema changes, pipeline retirement, or raw-table deprecation.
- Faster investigation of stale data, stopped updates, unexpected consumers, or new integration paths.
- A shared lineage view for platform, governance, security, privacy, and audit teams.
- Evidence that governance controls remain appropriate as data moves from source to lake to serving layer.

> **Delivery approach:** Theom will combine directly observed lineage with approved pipeline and transformation metadata, giving the customer a clear view of coverage across each hop.

## 03 Catalog reconciliation and continuous evidence

### The situation

Most enterprises are running or evaluating an enterprise catalog while the platform underneath it is being rebuilt. A traditional catalog can hold trusted business definitions, ownership, stewardship, and policy intent, but those records can become incomplete or stale as new tables, pipelines, copies, and consumers appear.

### How Theom contributes

- Continuously compare live discovery, classifications, lineage, permissions, freshness, and observed usage against catalog records and required tags.
- Identify assets missing from the catalog, stale ownership, conflicting classifications, undocumented derivatives, and data products whose actual consumers differ from documented expectations.
- Highlight permission drift and data-contract exceptions using live datastore evidence rather than periodic manual review.
- Feed validated findings into the reconciliation workflow of the selected catalog.
- Preserve the catalog as the system of record for business context while Theom supplies the bottom-up operational evidence needed to keep it accurate.

### Practical outcome

- Higher catalog coverage and confidence during a period of rapid architectural change.
- Less manual effort spent comparing catalog entries with warehouse and lake reality.
- A reviewable queue of missing tags, stale ownership, lineage gaps, and policy exceptions.

- Clear separation between business definitions owned by the customer and technical evidence continuously produced by Theom.

> **Delivery approach:** Theom will reconcile live evidence with the selected enterprise catalog while preserving the customer's ownership of business definitions and approval workflows.

## 04 Identity and access assurance

### The situation

As a data platform expands, access assurance must go beyond confirming that an identity has permission. Data, audit, and compliance teams also need evidence of whether a human, service account, BI workflow, or agent should be using sensitive data in the observed way, including when the initiating user is obscured behind a shared application or service identity.

### How Theom contributes

- Establish per-identity baselines for people, service accounts, applications, and agents using observed queries, data access, permissions, timing, and peer behavior.
- Build compound-identity traces that connect the initiating user, the BI dashboard, the application or service account, the query, and the data accessed into one reviewable chain.
- Identify service-account activity that departs from its established purpose, normal query fingerprint, expected datasets, operating window, or downstream consumers.
- Join entitlement, sensitivity, ownership, policy, and observed-use context to answer the audit question: should this identity be using this data in this way?
- Use query fingerprinting and access patterns to surface activity consistent with unauthorized AI or automation usage for investigation.
- Trace sensitive-data access into downstream collaboration and sharing paths, including Microsoft 365, OneDrive, and SharePoint, to reveal potential exfiltration chains and unexpected propagation.

### Practical outcome

- A defensible baseline of normal activity for each human, service account, application, and agent identity.
- End-to-end evidence from user or agent to BI tool, service identity, query, data, and downstream destination.
- Faster detection and investigation of service-account misuse, unauthorized AI usage, and unusual data movement.
- Audit and compliance evidence that explains not only what access was permitted, but whether the observed use was appropriate.

- Prioritized access-review and remediation actions grounded in actual behavior rather than entitlement snapshots alone.

> **Delivery approach:** Begin with the warehouse and the priority BI identity paths, then extend the same evidence model to agents and collaboration destinations.

## Proposed first engagement

1. Connect one representative warehouse account and agree the raw, curated, external, transformation, and BI scope.
2. Create an initial inventory, classification, activity, permission, freshness, and consumer baseline.
3. Select two representative lineage paths — one change-capture, one continuous-load — and document the available evidence at each hop.
4. Select a representative catalog domain and compare its records with live warehouse evidence.
5. Model one compound-identity path through a BI tool and one service-account scenario, including established and anomalous activity.
6. Review the resulting migration dependencies, lineage gaps, catalog exceptions, and identity findings, then agree the next integrations for the lake, streaming and orchestration layers, AI workflows, and Microsoft 365.

## Evidence to expect

- A ranked inventory of raw objects and their active consumers.
- Representative lineage views carrying sensitivity, ownership, freshness, identity, access, and policy context.
- Migration progress and deprecation evidence for selected raw paths.
- A catalog reconciliation report covering missing assets, conflicting classifications, stale ownership, and lineage gaps.
- Per-identity baselines, compound BI identity traces, and prioritized evidence of service-account misuse or unauthorized AI activity.
- A coverage roadmap for extending evidence across the lake, streaming and orchestration layers, AI workflows, and Microsoft 365.

> **How Theom fits:** Theom supplies the continuous evidence and assurance layer across data platforms, working alongside transformation and orchestration tooling and the enterprise catalog to strengthen operational governance.