



Building Data Lineage That Actually Works

How HSBC, LSEG, M&T Bank, Royal London Asset Management address organisational barriers to prove compliance, predict impacts, and protect data use.

Executive Summary

Large financial institutions have invested millions in data lineage technology. HSBC, M&T Bank, and Royal London Asset Management did too. Yet, when these institutions began their lineage initiatives, they faced a pattern familiar to most CDOs and IT leaders. Catalog tools captured technical metadata, but business teams couldn't use them. Governance policies documented standards, but data engineers ignored them. Consultant-led projects produced copious amounts of detailed documentation, but became outdated within weeks or months. Spreadsheets manually tracking lineage worked great at first, until they became unwieldy and unmaintainable. And finally, OpenLineage frameworks captured technical flows, but regulators rejected them for lacking the critical business context required. Why does the same pattern repeat across different technologies, different vendors, and different implementation approaches?

The pattern repeats because most organizations misdiagnose the root cause. In addition to the technical limitations of incomplete data lineage solutions, data lineage fails due to three organizational barriers that technology alone cannot solve. These include misaligned incentives, language gaps, and trust deficits. Misaligned incentives reward delivery speed over documentation quality, while language gaps prevent business and technical teams from understanding each other. Trust deficits make people work around systems rather than use them. The institutions that succeeded addressed these barriers through three strategic decisions that enable them to prove

compliance, predict impacts, and protect data use. The successful organizations are able to:

Prove:

Align incentives to ensure business and technical lineage exists and remains current, creating audit-ready evidence that regulators can verify instantly.

Predict:

Connect technical flows to business context, enabling impact analysis to model consequences before changes are deployed.

Protect:

Establish trust mechanisms with bi-temporal lineage and continuous validation to prevent data misuse and policy conflicts.

The results are measurable. In less than six months, with two people, HSBC spent 90% less on lineage work that competitors estimated would cost over \$5 million, take 18 months, and require 25 people.¹ M&T Bank eliminated 200 hours of manual processing per data product while building federated governance with over 100 data stewards.² Organizations now complete change assessments in minutes rather than months.

¹ "HSBC Case Study," Solidatus, <https://www.solidatus.com/wp-content/uploads/2024/01/Case-study-HSBC.pdf>

² Andrew Foster, interview by Solidatus, "Unifying the Enterprise: How M&T Bank is Rewriting the Rules of Data Governance with Solidatus," Solidatus webinar, September 25, 2025. <https://www.solidatus.com/resource/unifying-the-enterprise-how-mt-bank-is-rewriting-the-rules-of-data-governance-with-solidatus/>

The European Central Bank's May 2024 BCBS 239 update explicitly requires "advanced lineage" at the "data attribute level." Most catalog tools, spreadsheets, and open source frameworks don't meet this standard.

This ebook examines the three organizational barriers that prevent most institutions from proving compliance, predicting impacts, and protecting data use. More importantly, it shows the strategic decisions CDOs and IT leaders must make to overcome each barrier and build sustainable lineage capabilities.

Why Lineage Initiatives Keep Failing

Large banks don't lack data lineage technology, but many lack the organizational capability to use it successfully.

Four failure patterns repeatedly appear across institutions—from catalog tools that business teams can't use to consultant projects that become outdated within weeks or months. Each pattern reveals the same root cause: organizational barriers that prevent fine-grained lineage with business context, block communication between technical and business teams, and undermine system adoption. Understanding why these patterns repeat is the first step toward breaking the cycle.

The Four Failure Patterns

Most institutions have attempted data lineage multiple times. The patterns reveal why technology alone cannot solve the problem.

1. The catalog tool

Banks invested in data catalogs with bolt-on lineage capabilities that scanned systems and produced detailed technical diagrams. Six

months later, adoption stalled as business teams struggled to interpret technical column names and risk officers got lost in system-level visualizations. When compliance asked where counterparty credit exposure originates, the tool displayed 47 transformation steps across 12 systems—technically accurate but operationally useless. In most cases, the lineage capabilities of data catalog tools just don't provide the level of detail and business context needed to satisfy regulators.

2. The consulting project

Firms manually documented critical flows for BCBS 239 compliance and delivered comprehensive documentation using end-user tools like Visio and Excel. Within months, systems changed, pipelines evolved, and nobody maintained the documentation because there was no organizational process to keep it current.

At M&T Bank, employees could recount seeing four or more failed data initiatives before CDO Andrew Foster arrived in 2023. "I presented to our banking services senior leadership—about 300 people. I said, Can you put your hand up if you've seen a failed data initiative at the bank? Everyone put their hand up. Two failed initiatives, three failed initiatives, four failed initiatives. By the time I got to four and stopped, about 50% of the hands were still up."

3. DIY/Build it yourself

Institutions attempted to build their own catalog and lineage solutions rather than buy commercial tools. In simpler cases, someone built an Excel tracker to manually map data flows—working well for a single

domain with a dedicated owner until that person left or the scope expanded beyond spreadsheet capabilities. More sophisticated attempts involved purchasing graph databases like Neo4j and building custom lineage tools on top, developing their own scanners, user interfaces, and integration layers. These efforts initially showed promise, often championed by talented engineering teams who understood the technical requirements. However, as the original scope, scale, and investment needed to connect and integrate disparate sources increased, maintaining the custom-built solution became unsustainable. Documentation for the custom code was incomplete. Business requirements evolved faster than the internal team could adapt, leaving the organization to support homegrown software without dedicated resources.

Building basic cataloging and lineage capabilities is relatively easy and initially would show promise. However, building advanced, fine-grained

end-to-end lineage across a landscape of new and old technologies is hard, and the Devil is certainly in the details. Whether spreadsheets or custom-built graph database solutions, these DIY approaches eventually sit abandoned on shared drives or internal repositories—outdated and untrustworthy. Teams revert to asking colleagues who might remember or manually tracing through code.

4. Open source (and the allure of free)

Banks adopted OpenLineage or similar tools to automatically capture technical lineage. Data engineers appreciated the ability to track data flows across pipelines with minimal manual effort. Then, regulators audited and asked for the business context. What does this transformation mean? Why does this calculation exist? Who owns this data? The technical lineage couldn't answer. The audit revealed gaps that necessitated months of manual documentation effort to remediate.

“ The 240-Year-Old Data Problem

Bank of New York operates in over 100 markets with \$52 trillion in assets under custody. BNY's Head of Corporate Engineering, Eric Hirschhorn: “When you think about a 240-year-old legacy, we have a little bit of everything. Documents in boxes. Super advanced computer systems.”³

Every large bank has homegrown applications that have been built over decades. Automated scanners cannot read them. These gaps prevent proving compliance, predicting impacts, or protecting data use. ”

³ Eric Hirschhorn, interview by Solidatus, “Pioneering Data Strategies: How Bank of New York is Shaping Business Success in the Age of AI,” Solidatus webinar, 2025. <https://www.solidatus.com/resource/how-bank-of-new-york-is-shaping-business-success-in-the-age-of-ai/>.

What do these four patterns have in common?

In every case, institutions couldn't **prove** compliance because documentation either didn't exist, wasn't maintained, or lacked the business context regulators require. They couldn't **predict** impacts because technical flows weren't connected to business consequences. The tools showed system dependencies without explaining what breaks when changes occur. They couldn't **protect** data use because governance requirements were aspirational rather than enforceable, leading people to work around systems they didn't trust or understand.

Four different technologies. Four different implementation approaches. One common failure: treating data lineage as a technology problem when it's fundamentally an organizational challenge.

The Root Cause

These aren't four different problems, but rather, they're symptoms of the same organizational dysfunction. Companies keep focusing on technology when the problem is organizational.

One super-regional bank invested in three different lineage platforms over five years. Not surprisingly, each failed—not because the technology didn't work, but because organizational dynamics prevented consistent use. Technical teams found the tools too business-focused, while business teams found them too technical. And the risk and compliance teams? They simply didn't trust the accuracy, and no one had the incentive to maintain documentation.

M&T Bank's CDO, Andrew Foster, is direct about the challenge. "A lot of information comes from historic tribal knowledge. You have to find ways to make that sustainable. Modern cloud systems like Snowflake allow fast automated scanning. Legacy systems require patient knowledge capture."

Banks betting everything on automated scanning never achieve the completeness needed to prove compliance, predict impacts accurately, or protect against misuse.



What Regulators Actually Require

The European Central Bank's May 2024 update to BCBS 239 guidance explicitly requires "advanced lineage" at the "data attribute level." Column-by-column, field-by-field, and system-to-system, with granularity that shows not just where data flows, but what it means and why transformations exist.

When regulators audit, they test whether front-line users actually use documented lineage. If teams manually trace flows during an audit despite having lineage technology, they fail the audit.

Without this capability, you cannot prove to regulators that the data is accurate or

trustworthy. You cannot predict which downstream systems will break when you make changes. You cannot protect against data being used in ways that violate policy or regulation, and ultimately, you cannot show which controls are in effect or missing.

HSBC, M&T Bank, Royal London Asset Management, Deutsche Bank, and BNY faced these same challenges. They succeeded by making different strategic decisions about how to build lineage as an organizational capability. But first, they had to understand the three organizational barriers standing in their way.

The Three Organizational Barriers to Proving, Predicting, and Protecting

Three barriers prevent organizations from building enterprise-wide, up-to-date, and fully contextual data lineage capabilities. Without proper incentives, you cannot prove anything because documentation doesn't exist. Without a business context, you cannot predict impacts because technical teams can't communicate with business teams. Without trust mechanisms, you cannot protect data use because nobody believes the lineage is accurate.

BARRIER 1: **The Incentive Gap—Why Organizations Can't Prove Compliance**

Data teams are measured on delivery speed. Governance teams are evaluated based on their policy creation, while business analysts are measured by the insights they produce. Not surprisingly, nobody is measured on lineage quality and completeness.

When a data engineer faces the choice between delivering an analytics feature

on schedule or delaying delivery to properly document lineage, they choose delivery—every time. Documenting lineage might add a week to a project. That delay is visible, immediate, and blamed on the engineer. Skipping lineage creates problems later, often after the engineer has moved into a different role.

Without incentives to document, critical knowledge accumulates in individuals rather than systems. One major bank couldn't reconstruct critical risk calculations after a senior data engineer

departed. The systems still ran, but nobody understood the business logic. Six months of effort and expensive consulting support were required to reverse-engineer the calculations.

When regulators asked to see documented lineage, the bank had nothing to show. They couldn't prove the accuracy of their risk calculations, couldn't prove that the data came from authoritative sources, and couldn't prove that the transformations were correct.

“ The Partnership Model

M&T Bank's Chief Data Officer Andrew Foster:

“Whenever we talk about good governance, someone's got to do work they didn't have to do yesterday.”

His response:

The Data Academy trained 2,000 employees with marketable skills. Federated governance assigned 100-120 data stewards with clear accountability. The CEO meets with them annually. One data product eliminated 200 hours of manual processing.

”

Banks that embedded lineage metrics into performance management saw documentation rates improve substantially. One tier-1 global bank achieved audit readiness four times faster than its previous process.

BARRIER 2:

The Language Gap—Why Organizations Can't Predict Impacts

Technical teams document tables, columns, and ETL jobs while business teams need to understand metrics, KPIs, and regulatory requirements. When regulators ask where a number comes from, business teams cannot translate technical lineage diagrams into answers.

Many banks document lineage at the table or dataset level. Regulators don't

accept this because it's inherently inaccurate. Fundamentally, summary lineage represents someone's best guess. Column-level lineage shows which systems actually do what. The European Central Bank's May 2024 guidance explicitly mandates lineage "to the data attribute level."

Impact analysis requires understanding both technical dependencies and business consequences. When you change transformation logic in System A, you need to know not just that it affects

Systems B, C, and D, but also which business metrics become unreliable, which regulatory reports fail validation, and which customer-facing applications display incorrect data.

Bank of New York's Eric Hirschhorn describes what happened when they visualized data flows: "When we started to show people the big picture, they were horrified. 'Why are you running it like this?' We ended up seeing system A feeding

system B, and system B feeding system A. Things that were very counterintuitive."

The problem wasn't complexity. Nobody could predict what would happen when they changed anything because they couldn't see how technical flows connected to business outcomes.

“ Storytelling Over Spreadsheets

BNY's Head of Corporate Engineering, Eric Hirschhorn: "When we showed them a glossary or catalog, their eyes glazed over. But when we put Solidatus on top, we were able to tell the story."

He uses a simple metaphor: "X-ray the pipe as it goes from copper to plastic to steel. Watch the data flow, annotate with transformations, integrate with data quality tools." ”

Royal London Asset Management manages £162.3 billion through migration to BlackRock Aladdin. Their Head of Data Management, Lynn Watts: "Data leaders want to know precisely what data exists, where it resides, and how to access it physically within tables."⁴ They modeled 175,000 fields across 41 source systems with column-level granularity connected to business definitions.

M&T Bank's Andrew Foster uses an anatomical metaphor: "Lineage gives us a spine that runs through the organization. Then you get to layer things on top of it." Data quality scores. Governance roles. Instead of throwing information at people that is hard to consume, you make it as easy as possible."

⁴ "RLAM Case Study," Solidatus. <https://www.solidatus.com/wp-content/uploads/2024/06/RLAM-Case-study.pdf>.

BARRIER 3:**The Trust Deficit—Why Organizations Can't Protect Data Use**

Within months of initial lineage documentation, portions become outdated as systems are upgraded or migrated to the cloud, pipelines are modified, and new data sources are integrated. Most banks lack systematic processes to update lineage documentation when changes occur, leading to an inevitable drift between documented flows and operational reality.

A business user consults the documented lineage. They verify one connection against reality and discover it's wrong. They attempt to verify another, also wrong. They stop trusting the documentation and revert to asking people or manually tracing through code.

You cannot enforce data governance policies if people don't trust that documented lineage reflects reality. You cannot prevent data misuse if users

work around the system, and you cannot identify policy conflicts before they cause problems if the system shows inaccurate flows.

Bank of New York's Eric Hirschhorn recognized this: "We have 50,000 people. I've got 1,000 people with 1,000 questions, and I've got 100 people that have access to the pictures."

His solution was to integrate lineage into the software development lifecycle. "Every time you change code, every time you add a data element, we refresh these scans. The worst thing is you're really good on a Monday, and by Wednesday, it wasn't telling you the truth."

Without trust mechanisms that ensure accuracy, organizations cannot protect data use through governance policies because no one believes the policies reflect reality. They cannot prevent data misuse if users work around systems they don't trust. They cannot identify policy conflicts before they cause problems if documentation shows inaccurate flows.

Why So Few Have Succeeded

David Thomas, LSEG's Group Head of Data Trust, Data & Analytics & Commercial Data, has worked at seven different institutions in the data space for over 15 years. His assessment of why data lineage remains unsolved across the industry is direct:

"There's a reason no one has tackled data lineage. It's bloody difficult to get right."

Thomas points to the scale of the challenge:

"It's bloody difficult to get right, and even when you do get it right, you then have to maintain the data."⁵



⁵ Shibli, Aliya. "LSEG's Journey from Regulation to Revenue through Data." *The Banker*, September 25, 2024. <https://www.thebanker.com/content/18fd41d8-f187-4b3e-81fin-a-corporate-hierarchy-534797e24f8f>.

One global investment bank automated data quality rules that alert owners immediately when issues are detected. Clear ownership drives sustainable accuracy. M&T Bank's federated model assigns 100-120 data stewards, each with clear accountability for their domain.

These three barriers—misaligned incentives, language gaps, and trust

deficits—prevent most organizations from proving compliance, predicting impacts, and protecting data use. The institutions that succeeded didn't find easier problems to solve. They made different strategic decisions about how to address each barrier.

Strategic Decisions to Prove, Predict, and Protect

HSBC, M&T Bank, Royal London Asset Management, Deutsche Bank, BNY, and LSEG succeeded by making three strategic decisions. Each decision addresses one barrier and enables one capability.

DECISION 1: Align Incentives to Prove Compliance

Banks achieving measurable success have embedded lineage metrics into performance management, giving them substantial weight in affecting bonuses and promotion decisions.

M&T Bank provides visible benefits before imposing requirements. Their Data Academy trained 2,000 employees, equipping them with marketable skills. They built federated governance with 100-120 data stewards across the organization, each with explicit responsibilities for their domain.

Andrew Foster is direct: "Data and AI have to be core skill sets and core accountabilities across the enterprise, not just within a central team."

People are often many layers down from a corporate hierarchy perspective. One early data product eliminated 200 hours of manual processing through

better organizational design, not better technology.

Foster uses the iceberg metaphor: "1/10 above the water is the business value—your life is better today than yesterday. The 9/10 below is lineage, data quality, observability, catalog—the foundation that executives and data stewards care about."

With aligned incentives, documentation exists and stays current. When regulators audit, teams have audit-ready evidence available on demand, maintain historical records through bi-temporal lineage, ensure clear accountability by showing who owns each data flow, and measure quality through metrics tied to compensation.

DECISION 2:**Build Lineage with Business Context to Predict Impacts**

Banks must invest in business lineage that enables impact analysis, not just technical lineage that satisfies engineers.

HSBC modeled its entire Wholesale Credit Lending book in 6 months with two people, at 90% less than competitor bids. It serves over 500 internal users with self-service access across 40 source systems globally.

The key was building multiple interconnected views for the different teams. Technical lineage for data engineers, business lineage for analysts

and domain experts, regulatory lineage for compliance officers, and operational lineage for teams managing daily changes.

Former Head of Wholesale Credit & Lending Data Sid Mubashar: "With Solidatus, I can visualize the data model for each business outcome, manage the requirements of key stakeholders, and provide greater clarity over the intricacies involved in addressing their requests."

Generative AI Requires Governed Data

M&T Bank deployed Microsoft Copilot to 16,000 of its 22,000 employees. Andrew Foster:

"Data and AI come very tightly coupled. It's quite hard for AI deployment to be successful without the trusted data you need."⁶

The bank uses retrieval-augmented generation to limit AI training to internal, governed data stored in their Edison repository. Gen AI "gets you 60% of the way, then a human reviews it and takes it the other 40%."

With data lineage, you can model consequences before changes are deployed because simulation capabilities can test modifications before production. Impact analysis spans technical and business dimensions, enabling cross-functional collaboration. One global insurer reduced full impact analysis to under eight weeks.

⁶ Penny Crosman, "How M&T Bank Ensures Data Quality as It Implements Gen AI," *American Banker*, September 18, 2025.

DECISION 3:
Establish Trust Mechanisms to Protect Data Use

Banks must ensure lineage stays accurate and trusted over time, not just at launch.

Leading institutions use bi-temporal approaches to capture past, present, and

future states. During regulatory audits, examiners ask where the data came from on specific dates. Banks without bi-temporal lineage must manually reconstruct historical states, which is often impossible if people have left or systems have changed.

“ **Bi-Temporal Lineage: The Foundation of Trust**

Leading institutions capture three temporal states to maintain accuracy and build trust:

Past: During regulatory audits, examiners ask where the data came from on specific dates—December 31, 2024, when the annual risk report was produced—not just where it comes from today. Banks without bi-temporal lineage must manually reconstruct historical states, which is often impossible if people have left or systems have changed.

Present: Real-time validation compares documented flows against live system metadata daily, flagging drift immediately when documentation diverges from operational reality.

Future: Simulation models test impacts before changes go live, allowing teams to validate whether modifying transformation logic affects downstream systems. This builds trust through validation before deployment. ”

HSBC governs and maintains over 200,000 transitions across its Wholesale Credit Lending book. "Ongoing maintenance is distributed, federated, and democratized. Solidatus provides the ability to manage a federated workflow, with tasks being

assigned to SMEs where appropriate." The platform serves over 500 users without overwhelming the central team.

With trust mechanisms, organizations can enforce governance and prevent

misuse. Regulators can verify accuracy through instant historical reconstruction during audits, while policy conflicts are identified before data is used. One major bank saved millions in remediation and manual control mapping by establishing these trust mechanisms before scaling its data governance program.

federated model assigns specific individuals with the authority to enforce lineage quality standards in their domain, the responsibility for investigating discrepancies, the resources allocated to maintain accuracy, and the consequences tied to lineage quality.

Clear ownership drives this. M&T Bank's

Prove It

Most large financial institutions have attempted data lineage multiple times using different approaches. Tools, policies, consultants. The pattern repeats because the underlying organizational barriers remain unaddressed.

The institutions breaking this cycle didn't find better technology. They made different organizational decisions first.

- **To prove compliance:**

They aligned incentives so documentation exists and stays current. HSBC did this in six months with two people for 90% less than competitor bids. They can now prove to regulators that data is accurate, transformations are correct, and controls are in place.

- **To predict impacts:**

They built translation infrastructure connecting technical flows to business context. M&T Bank eliminated 200 hours of manual processing per data product. They can now predict which downstream systems and business processes will be affected by any change before it is deployed.

- **To protect data use:**

They established trust mechanisms with bi-temporal lineage and continuous validation. One major bank saved millions in remediation and manual control mapping. They can now protect against data misuse and identify policy conflicts before they cause problems.

The European Central Bank's May 2024 BCBS 239 update mandates "advanced lineage" at the "data attribute level." Most catalog tools, spreadsheets, and open source frameworks don't meet this standard because complete, enterprise-wide data lineage is hard and requires the right tool for the job.

Regulatory pressure continues to increase. Banks face BCBS 239, Model Risk Management, and SOX requirements. Insurers deal with Solvency II and IFRS 17. All sectors must address GDPR, CCPA, ESG reporting, and compliance with the EU AI Act.

Consider what happened to institutions that delayed addressing these barriers. One global bank failed its BCBS 239 audit when regulators discovered lineage documentation didn't reflect actual system behavior. The remediation effort required 18 months, external consulting support, and millions in compliance costs. During that period, they couldn't prove data accuracy for regulatory reporting, couldn't predict impacts of necessary system changes, and couldn't protect against data being used in ways that violated policy.

LSEG's Group Head of Data Strategy & Management Debbie Lawrence recognized this risk when the stock exchange began its data lineage journey. Without a lineage tool, LSEG would be "flying a little bit blind," Lawrence said. "I'm not entirely sure how we survived up until this point."⁷

Meanwhile, institutions that addressed organizational barriers first gained competitive advantages:

- **Launch new products faster:** Impact analysis takes days instead of months
- **Pass regulatory audits without scrambling:** Documentation exists and stays current
- **Deploy AI safely:** They can prove data provenance and predict downstream effects

The strategic question isn't whether to build these capabilities. Regulators and competitive pressures have already decided that. The question is timing.

Sustainable, enterprise-wide data lineage is hard—legitimately hard. Most tools aren't built for this challenge. Catalog tools lack the business context regulators require, and spreadsheets and DIY solutions often lack scale and become unmaintainable. Consulting projects deliver documentation

with no process to keep it up to date, and open source frameworks capture technical flows but can't answer the business questions regulators ask during audits. These approaches fail because they treat an organizational problem as a technology problem.

The institutions succeeding—HSBC, M&T Bank, BNY, LSEG—addressed organizational barriers first, then implemented purpose-built lineage capabilities. The choice is clear: make those strategic decisions now, or continue repeating patterns that have not succeeded while competitors pull ahead.

⁷ Shibli, Aliya. "LSEG's Journey from Regulation to Revenue through Data." *The Banker*, September 25, 2024. <https://www.thebanker.com/content/18fd41d8-f187-4b3e-81fa-534797e24f8f>.

The Next Step

Download the comprehensive guide: "BCBS 239: A Solution to the Challenge"

This whitepaper provides:

- Detailed regulatory requirements and ECB May 2024 updates
- Implementation frameworks from HSBC, M&T Bank, and Royal London Asset Management
- Technical specifications for "data attribute level" lineage
- Audit preparation checklists
- Organizational readiness assessments

Or contact Solidatus directly to discuss your specific challenges:

Understand how to build organizational capabilities that enable you to prove compliance, predict impacts, and protect data use at scale.

Visit: solidatus.com

Email: hello@solidatus.com

