



Data Meaning

If You Can't Explain Where Your Numbers Come From, Neither Can Your AI

Executive Resource Guide
February 2026

Executive Summary



Executive Summary

Most agencies produce numbers they rely on but cannot fully explain. When leadership asks "where did this number come from?" the answer often requires hours of research, or worse, "I'm not sure."

Today, this creates delays and erodes trust. Tomorrow, when you try to use AI, it becomes a crisis. If you can't explain where your data comes from, you can't explain or defend AI outputs.

This guide shows you how to:

Assess your current
lineage maturity

Document lineage for
critical metrics
without replatforming

Build the foundation for
explainable analytics
and AI readiness

Bottom line: You don't need perfect lineage for everything. You need sufficient lineage for the numbers that matter most.

Executive Summary

If You Can't Explain Where Your Numbers Come From, Neither Can Your AI

Someone in your agency sends a dashboard to leadership. The dashboard shows a critical number—enrollment, revenue, compliance rate, caseload, whatever matters for your mission.

Leadership looks at it and asks: **"Where did this number come from?"**

And the person who sent it doesn't know. They know which system generated the report. They might even know which database it pulled from. But they can't explain the full path from source to screen. They can't tell you how the number was calculated, what transformations happened along the way, whether it includes or excludes certain populations, or why this month's number looks different from last month's.

This happens constantly in government. Numbers get reported, decisions get made, and nobody can fully explain where the data came from or how it was processed.

That's a problem today. It becomes a crisis when you try to use AI.



The Legislative Inquiry

A state legislator questions an agency's caseload numbers during a budget hearing. The analyst who built the dashboard left six months ago. Nobody can explain the methodology or business rules. The hearing is delayed. Trust erodes.



The Failed AI Pilot

A county tries to deploy AI to prioritize service requests, but the team cannot explain how "resolution time" was calculated in the training data. Without clear lineage, they cannot validate the AI learned the right patterns. The pilot stalls.



The Migration Nightmare

During a system migration, the team discovers that dozens of reports rely on undocumented logic. Nobody knows what calculations are performed. Replicating them becomes a multi-year project.

The Explainability Gap

The Explainability Gap

Here's why data lineage-the ability to trace data from its source through every transformation to its final use-matters more now than ever before.

Today, when a human analyst produces a report, they might be able to explain their work even if the underlying data lineage is unclear. They remember pulling data from three systems, doing some cleanup in Excel, applying business rules they learned from a colleague, and producing the final numbers. It's not documented. It's not repeatable by someone else. But the analyst can explain what they did, at least in general terms.

When AI produces a recommendation, there's no analyst to ask. The AI applied logic it learned from training data. If that training data's lineage is unclear-if you can't explain where it came from, how it was transformed, what it includes or excludes-then you can't explain why the AI made the recommendation it did.

And in government, unexplainable recommendations don't get used. They get questioned, audited, rejected.

If you can't explain where your numbers come from today, you won't be able to explain AI outputs tomorrow. The explainability gap you have in your current analytics will carry forward-and get amplified-when you introduce AI.

What Data Lineage Actually Means

Q What Data Lineage Actually Means

Data lineage isn't a technical abstraction. It's the ability to answer a straightforward question: **"Where did this number come from, and how was it calculated?"**

A complete answer includes:

The Source

1. Which system or systems did the data originate from?
2. When was it last updated?
3. Who owns that system?

The Transformations

1. What happened to the data between source and destination?
2. Was it filtered, aggregated, joined with other data, cleaned, or adjusted?
3. What business rules were applied?

The Logic

1. How was the final calculation performed?
2. What definitions were used?
3. What populations were included or excluded?

The Validation

1. How do we know this number is correct?
2. Was it validated against another source?
3. Are there quality checks in place?

When agencies have strong data lineage, someone who didn't create the report can still explain where the numbers came from. When lineage is weak or missing, only the person who built the report-if anyone-can explain it, and their explanation often amounts to "I pulled it from the system and it looked right."

Why Lineage Gaps Exist

? Why Lineage Gaps Exist

Most agencies don't lack lineage because they're careless. Data systems evolved organically, and documentation didn't keep pace.

- **Data moves through many hands.** Each handoff from source system to data engineering to warehouse to analytics to decision-maker creates a potential gap.
- **Transformations live in undocumented tools.** Excel, Python scripts, SQL queries. Transformations happen in analysts' heads or local files, not formal documentation.
- **Business rules are tribal knowledge.** Metric logic lives in email chains or "how we've always done it," not written documentation.
- **Legacy systems don't track lineage.** Older systems weren't built to document transformations.

The result: agencies have numbers they rely on but can't fully explain.

📍 Where Does Your Agency Stand?

Level 1: Undocumented	Lineage exists only in analysts' heads. When someone asks "where did this number come from?" the answer is "let me check" or "I'll ask Sarah."
Level 2: Partially Documented	Some high-visibility metrics have basic documentation, but it's scattered and often out of date.
Level 3: Systematically Documented	Priority metrics have complete lineage documentation with clear ownership and regular review.
Level 4: Automated	Lineage is captured automatically through pipeline tooling and integrated with governance systems.

Most agencies should target Level 3 for critical metrics. You don't need expensive automation. You need clear documentation of the numbers that matter most.

The Cost of Poor Lineage



The Cost of Poor Lineage

When lineage is weak, agencies pay in multiple ways.

Decisions get delayed

When someone questions a number, tracking down its source and logic can take hours or days. Meetings get postponed. Decisions wait. The person who originally built the report might not be available, and nobody else knows how to explain it.

Trust erodes

When leadership or oversight asks "where did this come from?" and the answer is uncertain, trust in the data declines. Decision-makers start relying on instinct instead of analytics because they don't trust the numbers.

Compliance risk increases

Audits, legislative inquiries, and public records requests often require agencies to explain how numbers were calculated. If lineage is unclear, agencies can't provide confident answers, and that creates legal and political risk.

The Cost of Poor Lineage



The Cost of Poor Lineage

Change becomes expensive

When business rules or source systems change, agencies can't easily identify everything downstream that might be affected. Reports break. Dashboards show incorrect numbers. Fixing the problem requires hunting through undocumented logic.

AI initiatives fail

When agencies try to use AI, the lack of lineage becomes a blocker. AI can't be trained on data that can't be explained. AI outputs can't be defended if the input data's lineage is unclear.

The Hidden Cost of Poor Lineage

The Hidden Cost of Poor Lineage

Poor lineage creates recurring costs that most agencies don't track:

Staff time on detective work

Every time leadership asks "where did this number come from?" and it requires hours of research, that's a lineage cost you're already paying.

Extended project timelines

When systems change and you can't identify downstream impacts, projects take longer and cost more.

Failed technology investments

AI pilots and modernization projects stall when data can't be explained or defended.

Audit findings and compliance risk

Inability to explain calculations creates legal and political exposure.

Documenting lineage for your most critical metrics is a one-time investment that eliminates recurring costs. Most agencies see ROI within months as they avoid repeated research cycles and reduce project delays.

What Good Lineage Looks Like

🔍 What Good Lineage Looks Like

Good data lineage doesn't require perfect documentation of every transformation. It requires enough clarity that someone other than the original creator can understand and defend the numbers.

At minimum, good lineage includes:

- Transformation logic: Major transformations, filters, and business rules applied between source and destination.
- Calculation definitions: Plain-language descriptions of how metrics are calculated and what they include/exclude.
- Ownership: Who owns each piece and how often data refreshes.
- Validation: How accuracy is confirmed.

When documented at this level, anyone can explain where a number came from, even if they didn't build the report.

Example: What Good Lineage Looks Like

Here's documented lineage for **"Monthly Active Caseload"** in a human services agency:

Definition:

Count of unique cases with at least one service interaction during the calendar month, excluding closed/pending closure cases.

Source:

Case Management System (CMS v4.2) → CMS.Cases + CMS.ServiceInteractions tables → Nightly refresh at 2AM

Business Rules:

- "Service interaction" includes: in-person visits, phone contacts, service deliveries, case plan updates
- Excludes: automated notifications, internal notes without client contact
- Cases that reopen during month count as active for full month
- Multi-program cases counted once (not per program)

Calculation:

Count unique CaseIDs where status is active, interaction occurred in current month, and client record is valid

Known Limitations:

- **Used In:** Executive dashboard, legislative reports, federal compliance, workload allocation
- **Owner:** Director of Program Operations

This level of documentation allows anyone (technical or non-technical) to explain and defend the metric.

Building Lineage Incrementally



Building Lineage Incrementally

You don't need to replatform or rebuild your infrastructure. Start with your most important numbers and work backward.

Step 1	Identify 5-10 critical metrics. Pick numbers that, if questioned, would create operational or political problems.
Step 2	Map backward from report to source. Document where it appears, how it's calculated, what feeds it, what transforms it, and who owns it. A spreadsheet is often sufficient.
Step 3	Fill the gaps. Talk to the people who built the reports. If logic lives in someone's head, write it down.
Step 4	Validate. Review with creators and users—does it match reality and their understanding?
Step 5	Establish ownership. Assign clear owners and quarterly review cycles.

Prove that documented lineage reduces risk and delays, then expand.

Lineage as Defense, Not Documentation

Lineage as Defense, Not Documentation

The real value of lineage isn't documentation for its own sake. It's defense.

When leadership questions a number, you can explain it confidently and quickly. Decisions don't get delayed. Trust doesn't erode.

When auditors or oversight request details, you have clear, defensible answers. Compliance risk decreases.

When systems or business rules change, you can identify everything affected downstream and validate that changes were applied correctly. Fewer errors. Less firefighting.

When you evaluate AI, you have the lineage needed to explain training data and defend AI outputs. AI initiatives have a foundation they can actually build on.

Lineage transforms data from "numbers someone pulled from a system" to "traceable, explainable, defensible information that supports decisions and withstands scrutiny."

Why Stakeholders Should Care

CIOs	CDOs	CFOs
Faster incident response, controlled change management, reduced technical debt, AI enablement.	No more "let me get back to you" when leadership questions a number. Knowledge survives staff turnover.	One-time investment eliminates recurring costs from delays, audit findings, and failed projects.
Compliance	Program Leaders	
Immediate answers to auditor questions instead of weeks of research.	Complete documentation of how you measured success when outcomes are questioned.	

Why AI Makes Lineage Non-Negotiable

🔗 Why AI Makes Lineage Non-Negotiable

Training Data Must Be Explainable

AI learns from historical data. If you can't explain where that data came from or how it was prepared, you can't validate that the AI learned the right patterns. Without lineage, bias detection is guesswork.

AI Outputs Must Be Auditable

When an AI denies a benefit application and the applicant appeals, you need to trace the recommendation back to source data. If lineage isn't documented, you can't defend the decision—and it may be overturned.

Model Drift Requires Lineage Monitoring

When AI performance degrades, you need to know if it's a model issue or a data issue (changed source systems, modified business rules, data quality degradation). Without lineage, you're guessing.

Compliance Will Require It

Proposed regulations consistently require agencies to explain AI training data, document sources and transformations, and provide audit trails. All of this depends on lineage.

The bottom line: AI readiness is a data readiness question. If you can't explain your data today, you won't be able to explain AI outputs tomorrow.

Agencies that pursue AI without lineage find themselves stuck: the AI works in demos but can't be deployed because nobody can explain or defend its outputs.

Agencies that build lineage first create an environment where AI can actually succeed.

Lineage Documentation Template

Lineage Documentation Template

Use this template for each critical metric:

[METRIC NAME]

Definition: [Plain-language description]

Source System(s):

- System: [e.g., Case Management System v4.2]
- Tables/Fields: [e.g., CMS.Cases, CMS.ServiceInteractions]
- Refresh: [e.g., Nightly at 2:00 AM]
- Owner: [Name, role]

Transformation Logic:

1. [Step-by-step description]
2. [...]

Business Rules:

- [What's included/excluded]
- [How edge cases are handled]

Calculation: [SQL, formula, or plain-language]

Validation: [How accuracy is confirmed]

Known Limitations: [Data quality issues, lags, edge cases]

Used In: [Dashboards, reports, compliance filings]

Contacts:

- Technical: [Team/person]
- Business: [Team/person]

Last Reviewed: [Date]

Start With One Number

You don't need to document everything. Start with one critical metric, something that, if questioned, would create problems.

Map its lineage backward. Document where it comes from and how it's calculated. Validate with users. Assign ownership. Prove value. Then expand.

If you can't explain where your numbers come from today, you won't be able to explain AI outputs tomorrow.

Lineage Documentation Template

Lineage Documentation Template (Blank)

Use this template for each critical metric:

[METRIC NAME]

Definition:

Source System(s):

- System:
- Tables/Fields:
- Refresh:
- Owner:

Transformation Logic:

- 1.
- 2.

Business Rules:

-
-

Calculation:

Validation:

Known Limitations:

Used In:

Contacts:

- Technical:
- Business:

Last Reviewed:

Start With One Number

You don't need to document everything. Start with one critical metric, something that, if questioned, would create problems.

Map its lineage backward. Document where it comes from and how it's calculated. Validate with users. Assign ownership. Prove value. Then expand.

If you can't explain where your numbers come from today, you won't be able to explain AI outputs tomorrow.



Data Meaning

Data Meaning

Data Meaning partners with public sector agencies to build sustainable data and analytics foundations that support mission outcomes and enable AI readiness without requiring expensive replatforming.

Learn more:

- **[Visit Data Meaning on The Trading Post](#)**
- **[Take the AI Readiness Assessment](#)**