

DATA STRATEGY TRACKS

Metadata strategy

Metadata is the data about your data: what it means, where it came from, who owns it, how good it is and who uses it. A data strategy without a metadata strategy cannot prove that any objective was met.

THE SO WHAT OF THIS TRACK

Fund metadata as the traceability layer of the data strategy: start from the few critical measures on your scorecard, make their meaning and lineage visible end to end, and grow coverage only where a business sponsor can show the value.

Part of stage 4: Design

Every level of the cascade below the strategy map depends on being able to find, trust and trace data. Data objectives become measurable only when the measures behind them have certified definitions and visible lineage back to source. Metadata is also the context that AI and agents need to act safely, so it now decides how far automation can go. Treat it as a strategic asset with an owner, a business case and a scorecard, not as a tool the data team buys.



QUESTIONS THIS TRACK ANSWERS

- Which measures on our scorecard must have a certified definition and lineage to source, and by when?
- Who owns business meaning, and who owns the technical description of each critical data set?
- Which business outcome will the metadata programme be accountable for, and what is its baseline?
- How much of our metadata is captured passively, and where should it start to trigger action?
- What context will AI assistants and agents need before we let them answer questions or act on data?
- How will we prove the value to sponsors each quarter, in their language rather than ours?

1

The method

Catalogue, glossary, lineage and active metadata

Four kinds of metadata, one picture

Each kind answers a different question. Value comes from joining them, so a business term leads to the tables, the pipelines, the quality score and the people who rely on it.

Business metadata

What does it mean? Terms, definitions, business rules, owners, sensitivity and the decisions a data set supports.

Technical metadata

What is it physically? Systems, schemas, tables, columns, types, keys and the code that transforms them.

Operational metadata

What happened to it? Load times, freshness, volumes, failures, quality results and lineage from source to report.

Usage metadata

Who relies on it? Queries, reports, downstream consumers, ratings, comments and how often each asset is used.

SO WHAT

Start with the business kind for your critical measures; technical and operational metadata can largely be harvested automatically, meaning cannot.

From passive record to active metadata

Most organisations document metadata after the fact. The strategic shift is to let metadata drive what happens next.

Passive metadata

- Captured in a catalogue and read by people when they remember to look
- Refreshed by periodic scans or manual documentation drives
- Answers where data is and what it means
- Value depends on adoption, which often fades after launch

Active metadata

- Continuously harvested and analysed as systems and usage change
- Triggers action: alerts owners on drift, blocks a release that breaks a certified measure, tunes pipelines
- Feeds policies at the point of access, such as masking sensitive columns
- Supplies grounded context to AI assistants and agents at run time

SO WHAT

Plan the move in steps: catalogue first, then alerts on critical lineage, then policy enforcement, then context for agents.

Six capabilities the strategy should name

Decide which of these you need for your themes, and in what order. Few organisations need all six on day one.

Data catalogue

A searchable inventory of data assets with owners, descriptions, ratings and access paths.

Business glossary

Agreed terms and definitions, linked to the physical data that implements them.

Lineage

The path from source system through every transformation to the report, model or agent that uses it.

Semantic consistency

One meaning for a concept such as customer or margin, reconciled across systems and reports.

Rules management

Quality, privacy and retention rules recorded once, attached to data, and measured.

Active metadata

Signals from usage, quality and change that trigger alerts, policies and automation.

Tell a story a sponsor will fund

A metadata programme gets funded when it is framed as a business outcome with a named sponsor, not as a catalogue.

Business story	Metadata it relies on	Natural sponsors	Outcome to measure
Personalise the customer experience	Automated discovery of customer data, end-to-end lineage, modelled relationships between customer entities	Customer experience lead, CDO, CFO	Time to launch a personalised journey
Grow sales with predictive analytics	A common semantic view across sources, reuse of master data definitions	Sales leadership, CDO, CIO	Forecast accuracy and campaign conversion
Reduce regulatory and conduct risk	A catalogue of reported data, lineage readable by business users, workflow evidence	CEO, CDO, data protection officer	Effort and elapsed time to answer a supervisory request
Make analysts more productive	A catalogue, lineage and managed rules for trusted data sets	Analytics centre of excellence, CDO, CFO	Time spent finding and validating data per project

SO WHAT Pick one or two stories that match your strategic themes and let them set the scope of the first release.

Build the business case in seven moves

A business case for metadata is a chain from a sponsor's goal to a number finance will sign. Skip a link and the case collapses at review.

1. Anchor in a sponsor's vision

Find the executive whose goal depends on trusted, traceable data, and agree whether the effort is top-down, bottom-up or both.

2. Choose business measures

Select the measures that sponsor already reports, not metadata activity counts.

3. Measure the baseline

Record today's value of each measure, how it was measured and by whom. Without it, no improvement can be claimed.

4. Agree the target

Negotiate a realistic improvement with the sponsor and the people who run the process.

5. Translate into money

Convert the improvement into revenue gained, cost avoided, risk reduced or capital released, using finance's assumptions.

6. Cost the whole thing

Model total cost of ownership: software, people, stewardship time, integration, training and run costs over several years.

7. Show the return and payback

Compare value with cost, show when it pays back, and state the assumptions you will test each quarter.

SO WHAT

Build the case with the sponsor and finance in the room; a case written by the data team alone is rarely believed.

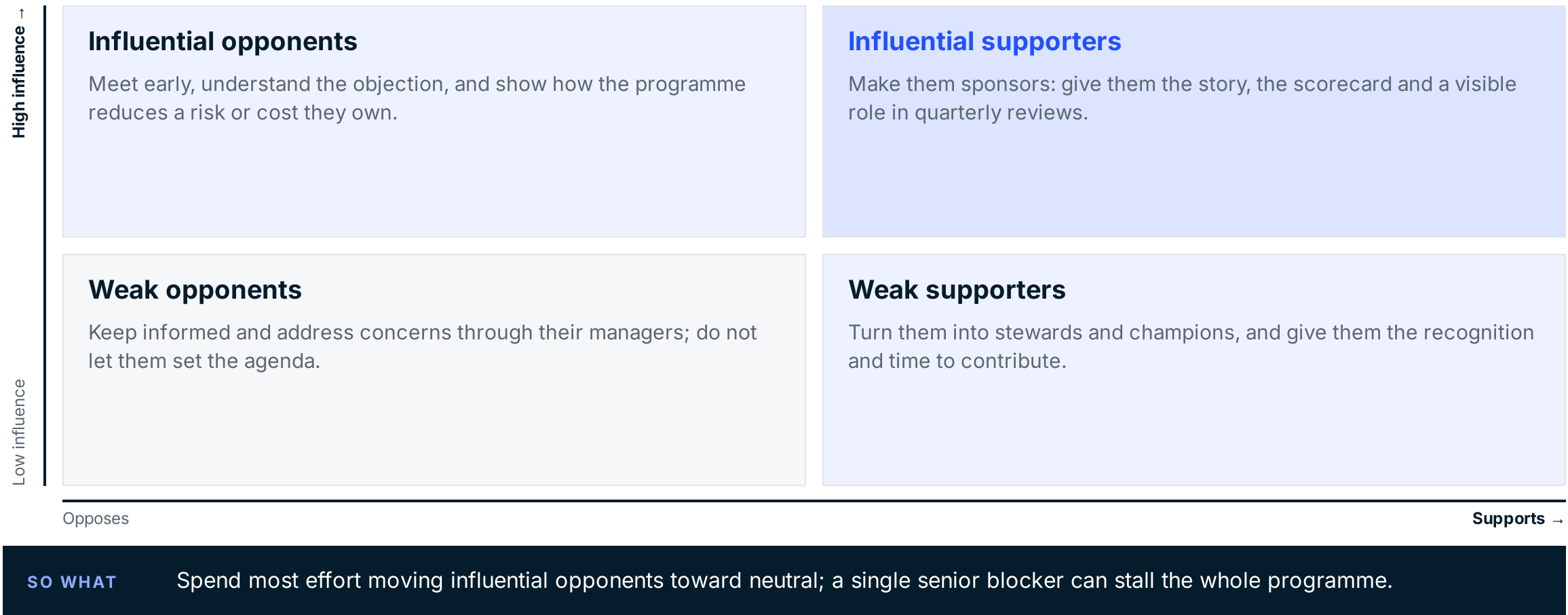
Leading and lagging measures for metadata

Lagging business measures prove the value; leading measures tell you early whether you will get there.

Horizon	Measure	What it tells you
Lagging	Revenue or cost impact attributed to faster, trusted insight	Whether the programme paid for itself
Lagging	Effort to answer audit and supervisory requests	Whether lineage and definitions reduced risk work
Leading	Time to complete the data selection phase of an analytics project	Whether people find and trust data faster
Leading	Reuse of certified data sets and measures	Whether teams stop rebuilding the same data
Leading	Critical measures with a certified definition and lineage	Whether the scorecard is becoming traceable
Leading	Active users and searches in the catalogue and glossary	Whether adoption is real or fading

Map stakeholders by influence and support

Metadata touches every function, so expect both champions and blockers. Place each stakeholder, then plan a move for each box.



2

Tools and evidence

Score it, measure it, see it in six industries

Metadata business case calculator

Defaults are placeholders, not benchmarks. Replace them with your measured baseline and finance's cost assumptions. Time recovered only counts as value if it is redeployed to work the sponsor cares about.

Input	Placeholder	Unit
People who search for and validate data	[40]	people
Hours each spends finding and validating data per week	[6]	hours
Working weeks per year	[46]	weeks
Share of that time a catalogue, glossary and lineage could recover	[30]	%
Audit and supervisory data requests per year	[25]	requests
Hours of effort per request today	[80]	hours
Share of request effort lineage could remove	[40]	%
Loaded cost per hour	[60]	per hour

How the case is calculated

Output	Calculation
Hours recovered per year	$([People\ who\ search\ for\ and\ validate\ data] \times [Hours\ each\ spends\ finding\ and\ validating\ data\ per\ week] \times [Working\ weeks\ per\ year] \times [Share\ of\ that\ time\ a\ catalogue,\ glossary\ and\ lineage\ could\ recover] / 100 + [Audit\ and\ supervisory\ data\ requests\ per\ year] \times [Hours\ of\ effort\ per\ request\ today] \times [Share\ of\ request\ effort\ lineage\ could\ remove] / 100)$
Value of time recovered per year	$[Hours\ recovered\ per\ year] \times [Loaded\ cost\ per\ hour]$
Net value per year	$[Value\ of\ time\ recovered\ per\ year] - [Programme\ cost\ per\ year\ (software,\ stewards,\ run)]$
Payback on set-up	$[One-off\ set-up\ cost] / max([Net\ value\ per\ year] / 12, 1)$
Three-year return multiple	$([Value\ of\ time\ recovered\ per\ year] \times 3) / max([Programme\ cost\ per\ year\ (software,\ stewards,\ run)] \times 3 + [One-off\ set-up\ cost], 1)$

SELF-CHECK

Is this in place today?

Score each statement 1 (not at all) to 4 (fully, evidenced).

#	Statement	Score
1	Every measure on our data scorecard has a certified business definition with a named owner.	
2	We can trace each critical report from source system to number, and business users can read that lineage.	
3	Technical and operational metadata is harvested automatically rather than documented by hand.	
4	Our metadata programme has a sponsor, a business measure, a baseline and a target.	
5	Metadata triggers action, such as alerts on breaking changes or policies at the point of access.	

ILLUSTRATIVE

How it looks in six industries

Industry	Situation	The move	Measure
Manufacturer	Plants describe the same part, defect and downtime reason differently, so quality and delivery measures cannot be compared across the five sites.	A shared glossary for part, defect and downtime codes, linked to each plant's systems, with lineage from shop floor to the delivery scorecard.	Plant measures on the scorecard with one certified definition
Retail bank	The supervisory review found risk reports could not be traced to source or reconciled quickly.	Business-readable lineage and certified definitions for every regulatory risk measure, with owners and quality rules attached.	Elapsed time to answer a supervisory data request
Retailer	Stores and online plan stock separately and each team defines sell-through and margin its own way.	One semantic definition of product, stock and margin across channels, published in the catalogue and reused by pricing and range teams.	Reuse of certified product and stock data sets
Hospital group	A privacy incident and a coding audit exposed unclear ownership and sensitivity of patient data.	Sensitivity classification and ownership recorded for patient data, with lineage behind coded activity and access policies driven by metadata.	Patient data sets with owner, classification and access policy
Telecom operator	Mobile, broadband and business run on separate billing and CRM stacks, and network inventory drifts from the field.	A catalogue across the three stacks with a common customer and service definition, and change alerts when inventory records drift.	Time for an analyst to assemble a single customer view
Insurer	Regulatory and actuarial reporting relies on manual reconciliations between policy, claims and finance systems.	Lineage from policy and claims systems to reserving and regulatory measures, with reconciliation rules recorded once and monitored.	Manual reconciliation effort per reporting cycle

Measures for the data scorecard

Measure	Definition	Lead or lag
Certified critical measures	Share of scorecard measures with an approved definition, owner and lineage to source	Lead
Lineage coverage	Critical reports and models traceable end to end, source to output	Lead
Time to find trusted data	Median time for an analyst to locate and validate the data for a new question	Lead
Catalogue adoption	Active monthly users and searches as a share of the target audience	Lead
Certified asset reuse	Number of downstream uses per certified data set or measure	Lead
Breaking changes caught	Schema or logic changes flagged before they reached a certified report	Lead
Audit request effort	Hours and elapsed days to answer an audit or supervisory data request	Lag
Realised value	Benefit confirmed with finance against the business case baseline	Lag

3

Making it happen

Traps, the first 90 days and where SCIKIQ fits

Traps to avoid

Catalogue as the goal

A tool is bought and populated, adoption peaks at launch and then fades. Counter-move: Tie every release to a sponsor's business story and a measure they already report.

Boiling the ocean

The team tries to document every table before anyone gets value. Counter-move: Start with the critical measures on the scorecard and the data behind them; grow by demand.

Meaning left to IT

Technical metadata is rich but business definitions are missing or disputed. Counter-move: Give business owners accountability for definitions, with stewards and a fast approval path.

No baseline

The programme claims benefits nobody can verify, and funding is cut at the next review. Counter-move: Measure the baseline before go-live and report against it every quarter with finance.

Stale metadata

Documentation drifts from reality and users stop trusting the catalogue. Counter-move: Harvest technical and operational metadata automatically and alert owners on drift.

AI without context

Assistants answer from raw tables and give confident, wrong numbers. Counter-move: Ground AI in certified definitions, lineage and policies, and log the context used for each answer.

The first 90 days

Days 1-30: Anchor and baseline

- Agree the sponsor and the one or two business stories that set the scope
- List the critical measures on the data scorecard and their current owners
- Measure the baseline for the chosen business measures and for time to find data
- Map stakeholders by influence and support

Days 31-60: Prove it on the critical path

- Harvest technical and operational metadata from the systems behind the critical measures
- Certify definitions for those measures with business owners
- Publish end-to-end lineage for the first critical reports
- Complete the business case with finance, including total cost of ownership

Days 61-90: Make it active and report

- Switch on alerts for breaking changes to certified measures
- Open the catalogue and glossary to the first analyst community
- Report leading measures and early wins to the sponsor
- Agree the next domains to cover, ranked by business value

Where SCIKIQ fits

Enterprise Discovery Engine

Harvests technical and business metadata, classifies data automatically, traces lineage including SQL lineage from views and procedures, and records usage, owner and criticality.

Semantic and Ontology Construction

Holds the taxonomy, glossary terms and business definitions, concepts and rules in a governed editor with maker-checker approval and provenance on every object.

Metrics and Knowledge Fabric

Parses BI reports into a governed metrics layer with formulas, definitions, lineage from BI to metric and report usage, reconciled across systems.

Context Engine and Governance Gate

Assemble context packs of data, rules, metrics and policies for AI and agents, and enforce classification, access, audit and lineage before anything acts.

Take it further

- The SCIKIQ Data Strategy Framework: the cascade this track plugs into
- Strategy Cascade and Scorecard Workbook: add this track's objectives and measures
- Run the interactive tool and self-check at scikiq.org/data-strategy#track-metadata

scikiq.org/data-strategy