The Federated Data Governance Program

Making data trustworthy enough to put AI on top of it

Publication Date 9/28/2026

Version 1.0

Author Jeremy Taylor

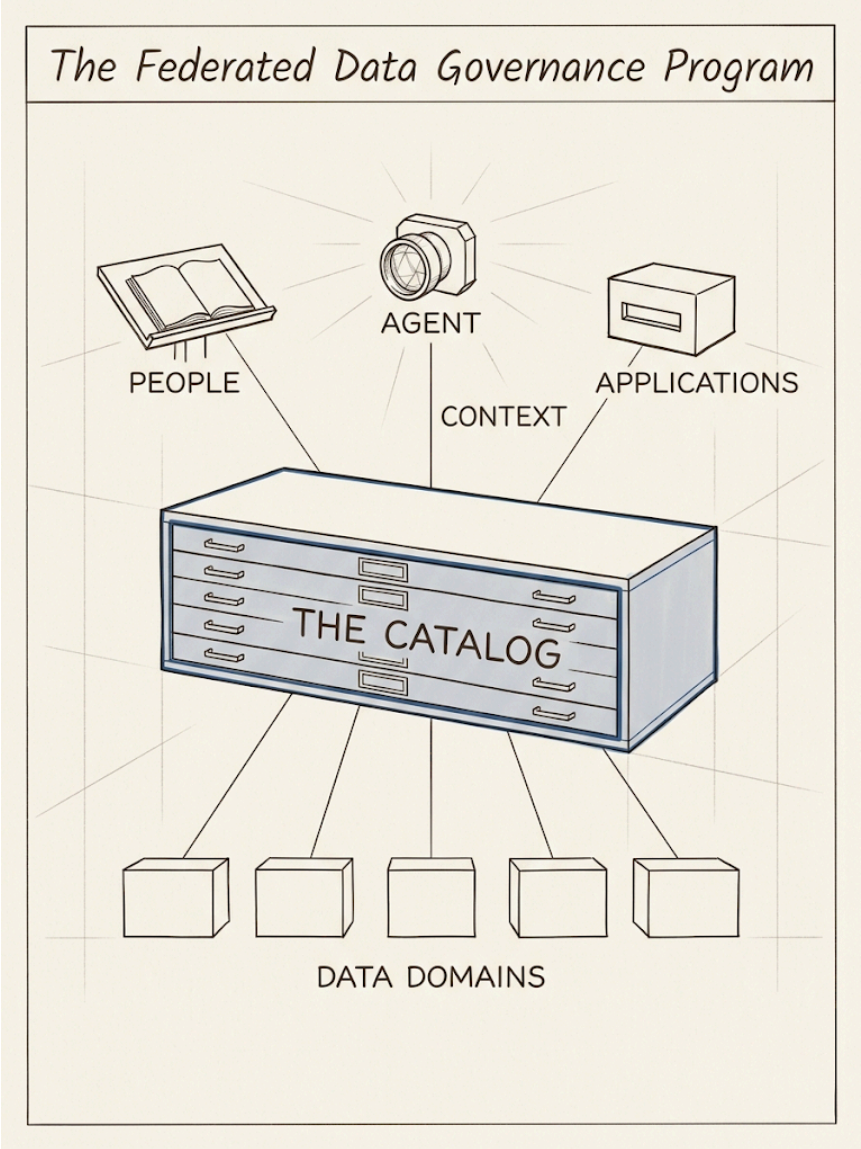

Table of Contents

Introduction

There's no shortage of discussion that AI needs context, and that the context needs governance. Both are true. People want to make it work.

The value is in tying three things together: an architecture where agents are required to use approved data, a process to manage that data, and tooling to implement it. That's what this program is. It's the minimum structure that makes data trustworthy enough to put AI on top of it.

Governance fails when it isn't attached to value and a project requirement. People want to be recognized for delivery and impact, and governance for the sake of governance stalls. So this program is limited, focused, and built directly into the technology implementation. The desire to be part of the deployed agent meets the reality of governing the data the agent uses.

I wrote The Federated Data Governance Program from two directions: the Carnegie Mellon University Chief Data and AI Officer executive program, and my own responsibility for standing up enterprise data governance. This is the companion to Pillar 3, Data Management & Intelligence, of The Enterprise AI Operating Model. It turns that pillar's catalog, ownership, and grading into a program you can run.

The program builds on work by Robert S. Seiner, Zhamak Dehghani, and DAMA. Sources and Influences credits each one and lists what is original here.

How to use this

This is for a mid-size business, described by its shape rather than its revenue: a central data team, data owned by the business, and data created across many teams, systems, and locations. Every role and cadence is sized for that shape. A larger organization can add to it, and a smaller one can take that first step in Where to Start.

Three readers will get the most from it:

- a data leader asked to get the company's data ready for AI, with no governance program yet
- a leader whose existing program has stalled and needs resetting around AI
- a practitioner (an architect or catalog owner) asked for trusted data, with no mechanism to achieve it or declare it

Part 1 explains why the program exists and what it rests on. Part 2 sets out who decides what. Part 3 is how data earns certification. The appendices are reference for the people running the program, and Where to Start is a first step that needs no platform.

The model names no products so everything specific is a placeholder. Appendix A says what to swap and what to keep. Appendix C maps each capability to whatever you run.

Part 1: Program Overview

1.1 Why governance, and what it covers

Data underpins nearly every meaningful decision and product the organization ships, and the volume, complexity, and regulatory expectations around that data keep growing. Without clear ownership, shared standards, and a shared language, teams duplicate work, lose confidence in reports, create compliance exposure, and slow down the analytics and AI initiatives that should be accelerating.

Data is managed the way any company asset is managed: to protect its value and to grow it. For a business putting AI to work, that comes down to one test. Is the data trustworthy enough to put AI on top of it? The program exists to pass that test with the least structure that will do it, and it breaks down into seven goals.

1. **Clear accountability** with a named Owner for every domain and every shared asset
2. **Shared language** of one definition and name per business concept
3. **Trusted data** includes a certification that tells consumers how far to rely on it
4. **Secure by default** means classification, privacy, and access are built in from the start
5. **Explorable** so people and AI tools can find what exists and what it means
6. **Lifecycle status** on every asset so consumers know what is active and what is going away
7. **Scale** by automating the metadata capture but not the judgment

1.2 Fundamentals

Five ideas underpin the program. Anyone reading further should be comfortable with all five.

The catalog

The catalog is the curated inventory of the enterprise data assets, and the heart of the program. Anything registered in it is an asset.

Three audiences use it.

- **People** browse it to see what exists, or query it through AI tools connected over the Model Context Protocol (MCP) to test an idea for a new product or capability.
- **AI agents** take their context from it. They find the right data by concept name, Tier, and certification.
- **Specialized applications** scale from it through its integration options: cached data, exported optimizations, and published contracts.

Everything else in this model exists to make what the catalog says true.

The catalog is the context layer, not the data layer. It describes the data and points to it; it does not hold or serve it. This model assumes a separate data layer that people and agents can query: a data lake or warehouse, or the operational databases themselves, exposed to agents through MCP or an API under the requesting person's access. Building that layer is outside the scope of this model.

Data domain

A data domain is a logical grouping of data owned by the part of the business that knows it best. Domains follow how the business is organized (customer, product, finance, orders) instead of systems or technologies. Each domain has a named Owner.

Data product

A data product is a curated, documented, and supported dataset that other parts of the organization can find, access, and trust. It comes with a data contract: a published agreement describing the schema, refresh cadence and SLA, quality expectations, access rules, and example queries. A dataset shared beyond its origin team without these things is not yet a data product.

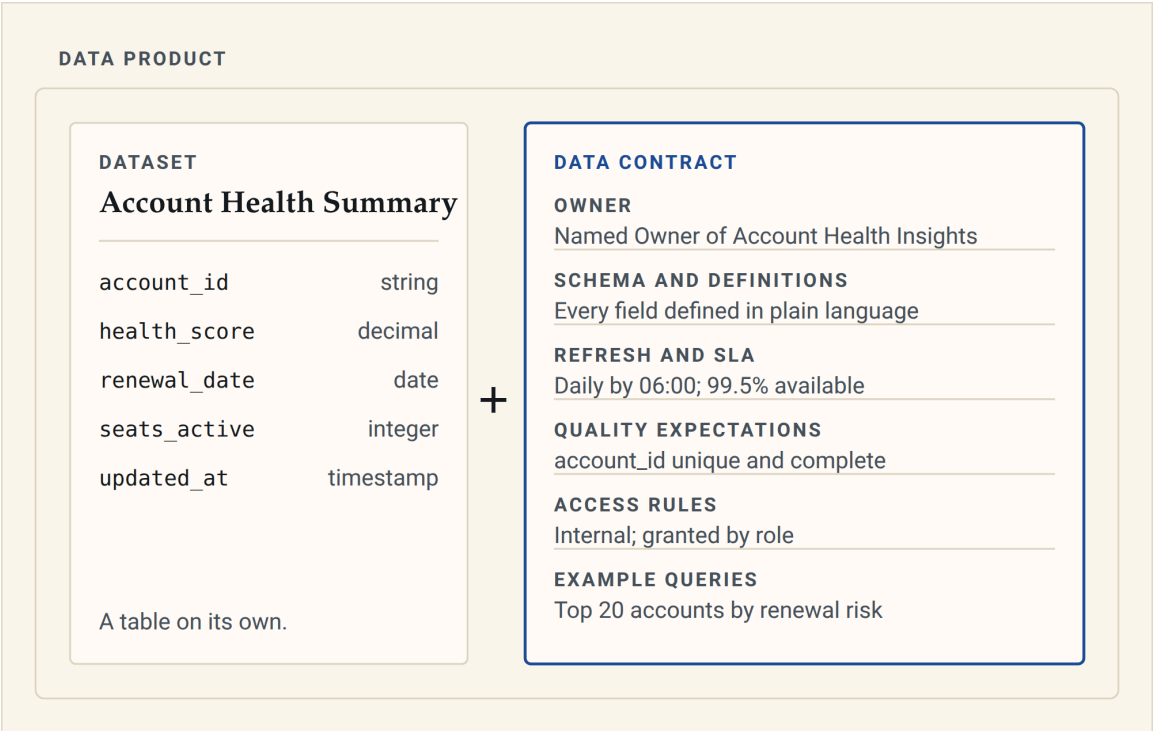

Figure 1. A data product is a dataset plus its contract. Field names and SLA values are illustrative; the Owner defines the real contract.

Federated ownership

The teams that create data own it. A technology team that builds an application owns the datasets that application produces. A business sponsor who buys an application owns the datasets it produces, even when a vendor runs it. The central data team provides the platform,

the standards, and the catalog, and builds shared and aggregate products until the domains can.

The idea comes from data mesh, a sociotechnical approach: an organizational agreement about who owns data, with architectural consequences. This model takes the ownership half on day one and the engineering half as capacity allows. Workloads move to the domains as soon as they can carry them.

Some data originates outside the organization's control: franchisees, partners, vendors, customers. Ownership still lands on whoever brings it in. That Owner chooses between system-enforced rules at intake and cleanup after the fact, and records the choice in the contract.

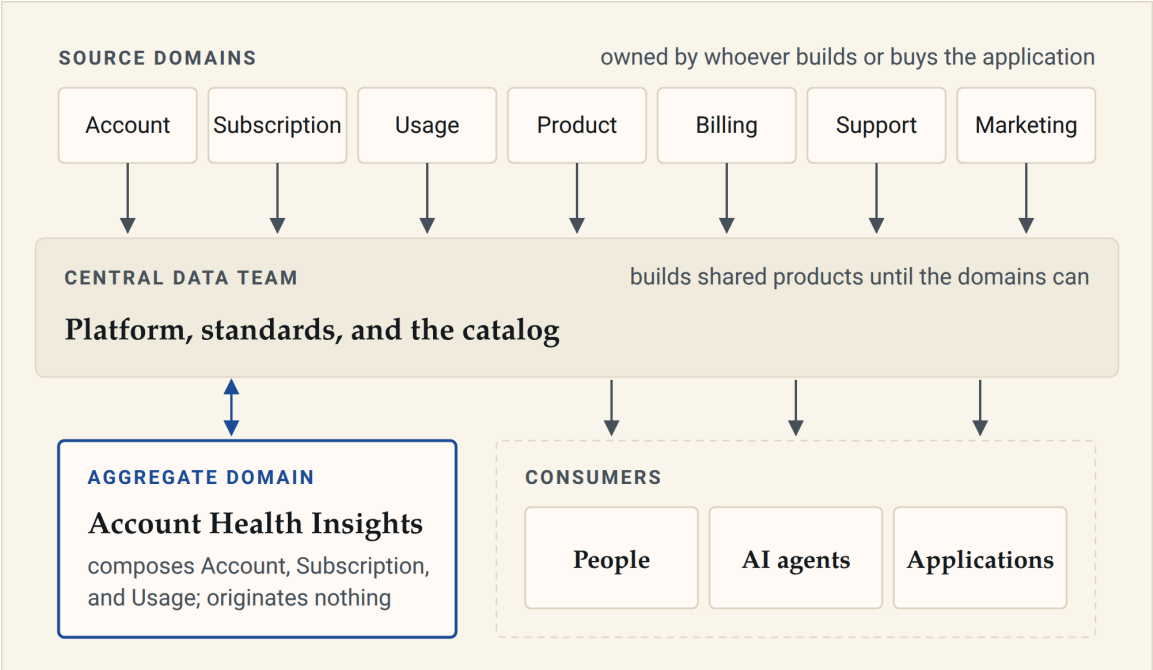

Figure 2. Seven source domains plus one aggregate domain. Account Health Insights composes data from Account, Subscription, and Usage instead of originating any of its own.

Substitute. The eight domains are illustrative. Use your own set, and keep the distinction between source domains, which align to the systems that create data, and aggregate domains, which compose from others. Aggregate domains consume definitions instead of creating them, and they are where overlapping concepts get reconciled (§3.7).

Minimally invasive governance

Governance should feel like a tailwind, not a tax. Minimally invasive means four things. The first two adapt principles from Robert S. Seiner's Non-Invasive Data Governance™; see Sources and Influences.

- **Inherent responsibility.** Stewardship is a relationship to data. If you can see, update, or define data, you already carry responsibility for it. That is where the role model in Part 2 starts, and it means most of the program describes people who are already doing the work.
- **Apply, don't reinvent.** Most people already do the right things informally. The program codifies that, fills gaps, and reduces friction. It adds no approval layers for their own sake.
- **Federation over central control.** The center defines what good looks like; domains decide how to achieve it within those guardrails.
- **Automation over policy enforcement.** Wherever possible, the catalog, CI/CD, and quality tooling make the right thing the easy thing.

1.3 Guiding principles

These principles apply across all data governance activities. Each is a rule a reasonable person could argue against, which is what makes it useful when a decision is close.

- **Minimally invasive**. Codify what people already do; add friction only where it pays back.
- **A person decides**. Named Owners may use automation to support decisions.
- **Protected by default**. In new data and AI applications, unclassified means Restricted.
- **Visible**. Definitions, lineage, quality, certification, Tier, and lifecycle status live in the catalog, where data consumers (humans or machines) look.
- **Govern in proportion to value**. Effort follows importance. Tier sets the target; not every asset needs Gold.

Part 2: Roles and Decision Rights

Who is accountable for what, mapped to the five program layers. This part reconciles enterprise governance roles with data mesh execution roles into a single taxonomy. Every role lives at exactly one layer, assigned by the work it does instead of its seniority.

2.1 Everyone is already a steward

Before any titles, the role model rests on one idea.

- If you can see the data, you are responsible for how you use it.
- If you can update or enter the data, you are responsible for the quality of inputs.
- If you define data used by your part of the organization, you are responsible for keeping that definition consistent with the enterprise standard.

Being a data steward describes a relationship to data. It is not a job title or a box on an org chart. Most people in the organization are already implicit stewards; the program makes that explicit and gives them the tools to do it well.

AI agents use data too, but an agent is not a steward. Responsibility for what an agent does with data rests with the person it acts for, which is why agents act under that person's access.

2.2 The five program layers

The program runs in five layers. The **executive layer** provides authority and air cover, rarely involved day to day. The **strategic layer** is the Council: it sets standards and priorities, owns the certification rubric, and resolves cross-domain conflicts on a periodic cadence. The **tactical layer** is where most of the recurring work happens: owning the domains and certifying their data. The **operational layer** is everyone who creates, updates, or uses data, the implicit stewards, most of whom carry no governance title. The **supporting layer** is central services (program management, compliance, security, the data platform).

None of the roles in these layers is a full-time position. Each makes explicit a decision someone already makes, and at a mid-size business one person often holds several: one Owner may own several domains, and the Program Manager is often the head of the central data team.

2.3 Unified role taxonomy

The taxonomy merges enterprise governance roles with data mesh execution roles. Every role lives at one program layer.

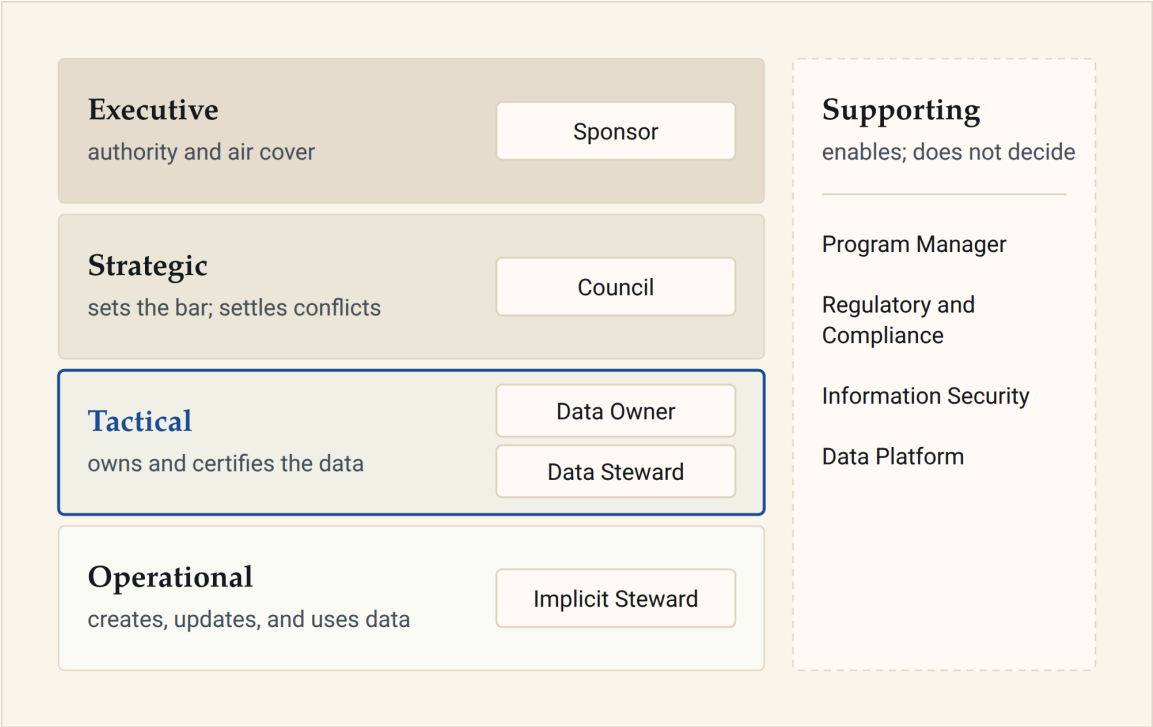

Figure 3. Each role sits in the layer where its work happens. The darker the band, the rarer the decision. The supporting layer runs alongside the other four and enables them without deciding.

Appendix B lists each role's typical holder and primary accountability, and the RACI for common activities.

2.4 Decision rights and escalation

- **Inside a domain,** the Owner decides: definitions, quality thresholds, access lists, Tier, and certification.
- **Across domains,** the Council decides: enterprise standards, the certification rubric, and any concept two domains name or define differently. Domains decide whether their assets meet the bar; they do not set the bar.
- **Regulatory exposure or material risk** escalates through the Program Manager to the Sponsor, with legal and compliance consulted.

Escalation should be the exception. The program is designed so that most decisions are made at the lowest reasonable layer.

Part 3: Certifying Data

How data earns Bronze, Silver, and Gold certification, whether it's a single table or an entire domain.

3.1 Purpose

Certification is what makes the catalog trustworthy enough to put AI on top of it. An agent choosing between two sources reads the Tier to find which one is authoritative (§3.6), then reads the certification to decide how far to rely on it. A person deciding whether to build a report on a table does the same.

The unit of certification is an asset: whatever you register in the catalog. The level you register it at is its **grain**, a term borrowed from dimensional modeling. Four grains are useful in practice: a column, a table or dataset, a data product, and a whole domain (§3.7).

This is a data maturity model instead of a domain maturity model, deliberately. The same rubric applies at every grain. Maturity belongs to the data at whatever grain you register it.

Certification is assigned by the Owner, recorded in the catalog, and shown to consumers there. The catalog is the system of record; the rubric in §3.4 is what the Owner uses to certify.

3.2 The three certification levels

Not medallion layers. Bronze, silver, and gold are also the layer names in the widely used medallion architecture, where they describe how far through a pipeline data has been processed: raw, cleaned, curated. That is a different axis. Here the levels describe how well an

asset is described and how much its Owner stands behind it. A raw landing table can be certified Gold if it is fully described, owned, monitored, and classified, and a curated serving table can sit at Bronze if nobody has documented it. If your platform already uses medallion names for pipeline stages, say so when you publish the rubric, or rename these levels (Appendix A).

Each level is a promise, and each includes everything in the level below it.

Before Bronze: registered, uncertified

Anything the catalog discovers, or anyone registers, enters the inventory at once and without a certification. Inventory does not wait on classification: you cannot classify what you have not found. In new data and AI applications, an uncertified asset is treated as **Restricted** until its Owner confirms a classification. Limited access is the cost of being unclassified, and classifying is how an Owner lifts it. Registering existing data does not revoke anyone's current access.

Bronze: known

Someone answers for it. The asset has a named Owner, a confirmed sensitivity classification, its source system, and a line on what it is for. People and agents can find it and know whom to ask; nobody has yet vouched for its meaning or its quality.

Silver: described

It is explained and measured. A Steward handles day-to-day questions, the fields that matter are defined in plain language, upstream lineage is traced, core quality measures are checked regularly, and the refresh cadence is stated. Consumers can use it with care and know where to go when something looks wrong.

Gold: committed

The Owner stands behind it. Quality thresholds and a freshness commitment are published and monitored, lineage runs end to end, every field a consumer relies on is defined, access is reviewed on a schedule, and consumers get notice before a breaking change. Consumers, human or agent, can depend on it without checking with the Owner first.

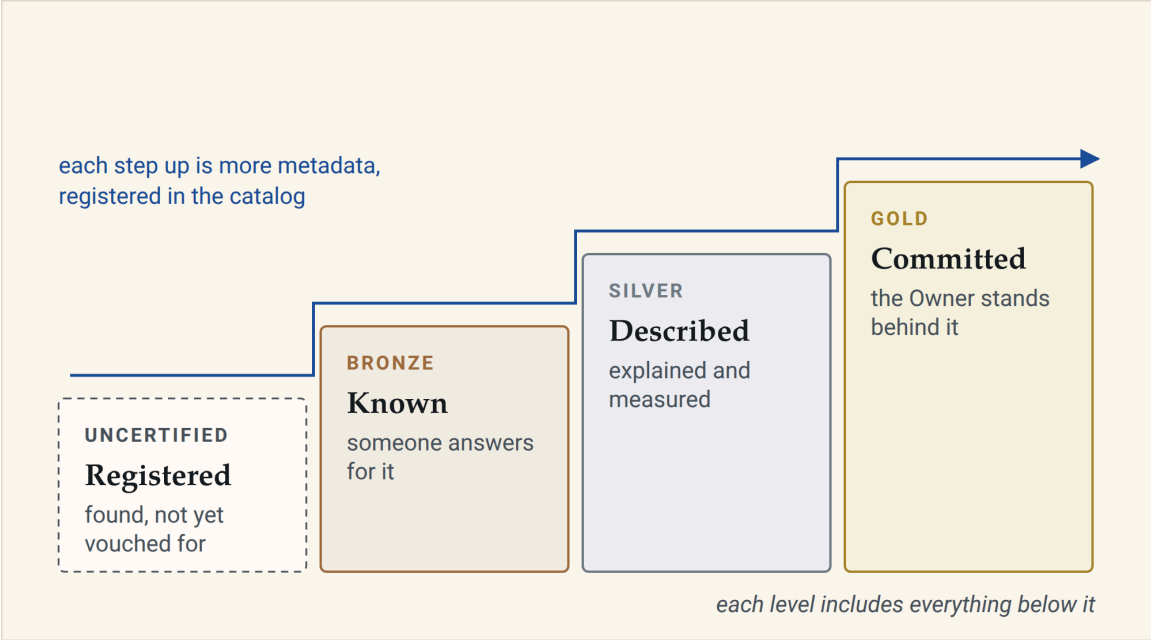

Figure 4. Gold is where business-critical assets aim. Silver is the target for anything shared beyond its team. Bronze is the floor every registered asset should reach within a quarter.

The path from Bronze to Silver to Gold is the path of registering more metadata in the catalog. Once an asset's gaps are visible against the six dimensions, the next piece of work is concrete.

3.3 The six dimensions

Owners use the rubric in §3.4 to certify each asset they own, at any grain, with the **best fit**: the level that best describes the asset as a whole. The rubric has six dimensions, each named for a

question a consumer asks before relying on an asset. The set is deliberately short, so an Owner can hold the whole rubric in mind while judging an asset.

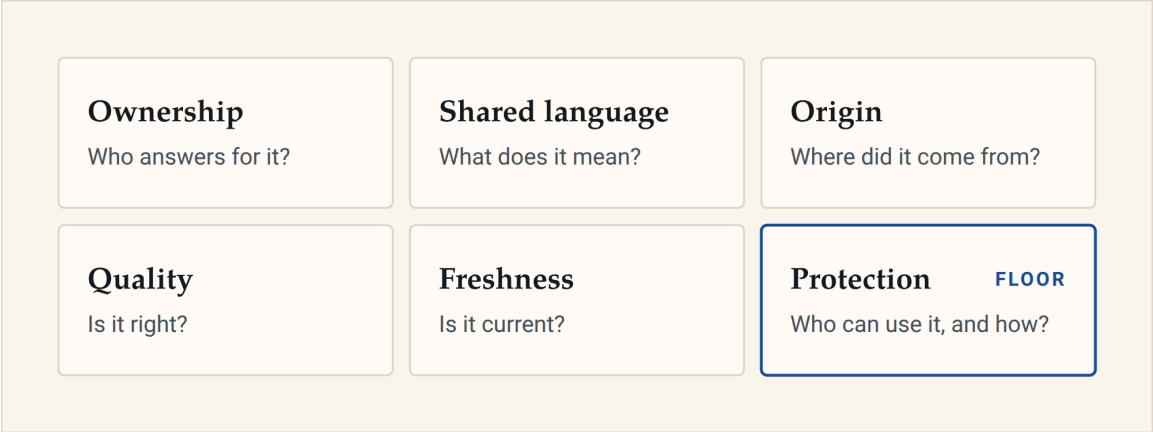

Figure 5. Each dimension is named for a question a consumer asks before relying on an asset. No dimension outranks another, except Protection: a floor that no weighting can offset.

Best fit, not a mechanical score. An asset that is strongly Silver in most respects and Bronze in one dimension may still be certified Silver if the Owner judges that the overall state warrants it. The gap becomes the next thing to address.

3.4 The rubric

Each level assumes the one below it: Silver assumes Bronze is met, and Gold assumes Silver. Owners certify with the best fit.

Dimension	Bronze	Silver	Gold
Ownership (Who answers for it?)	A named Owner on record	A Steward handles day-to-day questions; consumers know where to ask	Advance notice to consumers before breaking changes; a defined escalation path

Dimension	Bronze	Silver	Gold
Shared language (What does it mean?)	A concept name and a one-line statement of purpose	The fields that matter are defined in plain language and tied to the shared business glossary; a few example queries answer common questions correctly	Every field consumers rely on is defined; known caveats, intended uses, and example queries with their expected results are written down
Origin (Where did it come from?)	Source system named	Upstream sources and major transformations traced	End-to-end lineage, including downstream consumers, so the impact of a change is visible before it is made
Quality (Is it right?)	Not yet measured; known issues noted where they exist	Completeness, validity, and uniqueness checked on a regular schedule	Thresholds published and monitored; consistency reconciled against the source; breaches reach the Owner
Freshness (Is it current?)	Time of last update visible	Expected refresh cadence stated	Freshness commitment published; late or missed loads raise an alert
Protection (Who can use it, and how?)	Sensitivity classified and confirmed by the Owner. Required for Bronze.	Sensitive fields flagged individually; access granted by role	Applicable regulatory obligations mapped; access reviewed on a schedule

Quality criteria follow DAMA's data quality dimensions; see Sources and Influences.

Protection is a floor, not a ladder. Best fit lets an Owner certify Silver while one dimension still sits at Bronze. Protection is the exception at the bottom of the scale: no asset holds any certification until its classification is confirmed. Sensitive data cannot wait for Silver to be handled properly, and the Restricted default for uncertified assets in new applications keeps discovery from outrunning protection.

3.5 How certification works

The Owner reviews each asset against the six dimensions and assigns the level that best describes its overall state. The certification is recorded in the catalog and visible to consumers.

Certification is a manager-level responsibility. An Owner may delegate the assessment to a trusted person, often the Steward, but still signs for the result. When a Steward registers metadata that meaningfully improves an asset, they raise it with the Owner for re-certification.

This is judgment-based and deliberately manual. The rubric tells the Owner what to look for; the Owner decides which level fits.

If your platform offers rule-driven certification. Some catalogs can evaluate a rule set and apply a level automatically. That capability reports dimension coverage for the Owner's review, and leaves the assignment to the Owner. An automated rule can confirm metadata is present; it cannot confirm an asset is trustworthy, and switching it on quietly moves accountability from a named person to a configuration. If you do adopt automated assignment, say so explicitly, because it changes what a certification means.

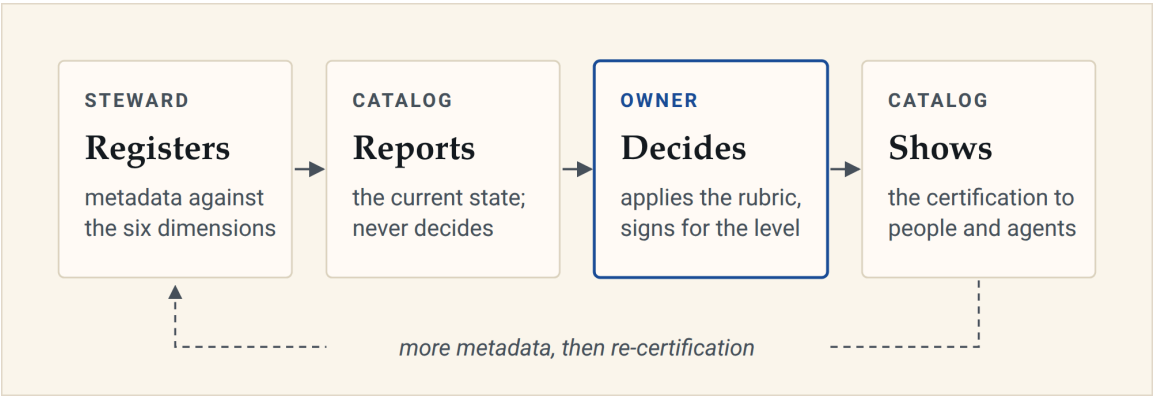

Figure 6. The catalog reports; the Owner decides. A certification is a standing commitment by a named person, which is why it is never computed.

3.6 Tier, targets, and cadence

Certification says how mature an asset is. Tier says how much it matters. The two are separate, and Tier sets the target that certification is measured against.

Tier	What it means	Target certification
Tier 1	Business-critical. An outage or a wrong answer reaches customers, money, or regulators.	Gold
Tier 2	Shared beyond the team that owns it.	Silver
Tier 3	Local to one team.	Bronze

- The Owner sets Tier, and the Council settles disputes. Every registered asset should reach Bronze within a quarter of registration; beyond that, Tier decides where effort goes. A Tier 1 asset certified Bronze is a visible gap and goes to the top of its Owner's list.

- Only Tier 1 is limited: **one Tier 1 asset per concept name.** Tier 2 and Tier 3 assets may overlap freely. For an agent, the two labels answer different questions. Tier finds the authoritative source for a concept; certification says how far to trust it.

- Two cadences keep this current. Each quarter, the Owner reviews the domain: certify newly registered assets, re-certify those that have advanced, check each asset against its Tier target, and decide where to invest next. Twice a year, the Program Manager brings the program measures and the cross-domain patterns to the Council.

3.7 What gets certified

Each of the four grains in §3.1 carries a slightly different claim.

- **Column or field**. Rarely certified on its own. Its metadata is assessed as part of the table that holds it, chiefly under Shared language and Protection.
- **Table or dataset**. The most common grain.
- **Data product**. The claim covers the product and its published contract.
- **Domain**. The claim covers the domain's own metadata: charter, boundaries, ownership, glossary, and classification profile. At this grain, Shared language is the domain's glossary and Origin is its source systems.

Certifying a product or a domain

A product's or domain's certification is a judgment about its own metadata, informed by its members and never computed from them. It is neither the weakest member's level nor an average, so a domain can be Silver while some of its tables are still Bronze. A computed rule would hold every large domain at its weakest table indefinitely and turn best fit back into a mechanical score.

Consumers will still read a coarse certification as a promise about everything inside it. So the catalog shows the two together, for example Account domain: Silver (3 Gold, 8 Silver, 14 Bronze), and people and agents rely on the certification of the asset they use, never the one above it.

Concept names

Each business concept has one name in the catalog at Tier 1. When two sources overlap, they are different concepts, and they are named for what they are. CRM accounts and billing accounts are not the same thing (a billing system holds customers the CRM never saw), so the catalog holds `crm_account` and `billing_account`, and each declares how it relates to

the other, usually through a shared key. Two assets named “accounts” is a collision, and it goes to the Council.

Where a presentation layer reconciles the two, that conformed asset belongs to an aggregate domain and carries the concept name. It is usually the Tier 1 asset for the concept (§3.6), and it is where an agent asked “how many accounts do we have?” should look first.

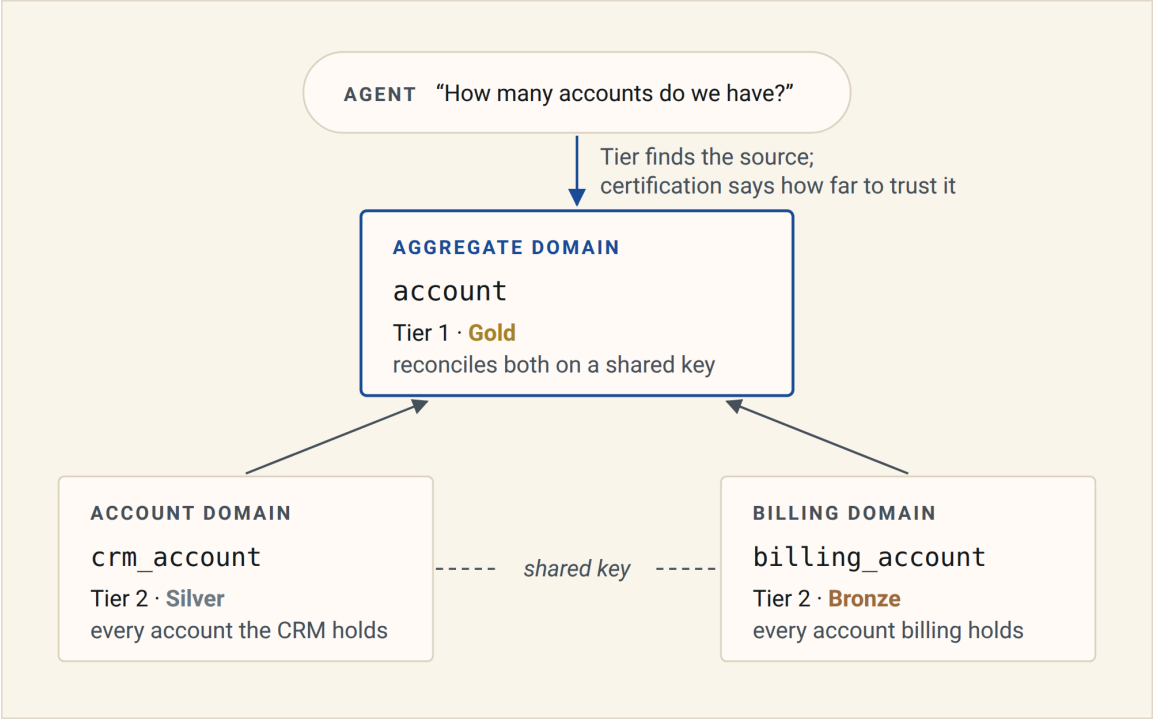

Figure 7. Two overlapping sources are two concepts, named for what they are and related by a shared key. The conformed asset carries the concept name and is its one Tier 1 asset, so an agent finds it by Tier and decides how far to trust it by certification.

Certification is separate from lifecycle

Every registered asset also carries a lifecycle status, maintained by its Owner. Lifecycle answers *is this still here, and should I build on it*; certification answers *how far can I rely on it*. An active asset can be Bronze, and a Gold asset can be deprecated.

- **In development.** Being built. Expect change; do not depend on it yet.
- **Active.** In service and supported at its certified level.
- **Deprecated.** Still works; do not build on it. A deprecated asset always carries a sunset date, and its known consumers are told.
- **Retired.** No longer served. The catalog entry remains so lineage and history stay readable.

Four states are enough. Retention periods, archiving, legal holds, and disposal are records management, owned with the regulatory and compliance function, and sit outside this model.

How Success Is Measured

Most measures below can be counted from the catalog itself, so nobody has to run a survey or a time study. The headline tests the thesis directly, and it needs one more source: the data access layer's query log.

Headline: the share of AI agent queries answered from Silver-or-better assets. The data access layer logs which assets each agent query touched; the catalog says what each was certified at when it ran. This says how much of what agents actually used someone has vouched for. Before that log exists, sample it by hand.

Clear accountability: share of registered assets at Bronze or better, and share of Tier 1 and Tier 2 assets their Owner has reviewed in the last quarter.

Shared language: share of Tier 1 and Tier 2 assets linked to the business glossary, and the number of cross-domain term conflicts open at the Council.

Trusted data: the certification mix of shared assets, the share of Tier 1 assets certified Gold, and the share of Silver-or-better assets passing their quality checks.

Secure by default: uncertified assets in new applications (target: zero).

Explorable: catalog use, counting people and the AI tools and agents connected over MCP.

Lifecycle status: share of deprecated assets with a sunset date and notified consumers.

Scale: median time from registration to Bronze and the share of certifications signed by Owners outside the central data team.

Where to Start

Start with one domain and one document.

Pick a domain whose data is well understood, used beyond the team that creates it, and matters to the business. Write data contracts as a formal, machine-readable spec: what each dataset is, who owns it, what each field means, how often it refreshes, what quality it promises, who may use it, and a few example queries that answer common questions correctly. An open standard such as the Open Data Contract Standard (ODCS) gives the spec a structure people and tools already recognize. No platform is required.

The contract is useful the day it's written. Added as context to any AI session, it lets people and agents work with that dataset right away: an hour, a day, or a week to value, depending on the dataset's complexity. It's also the first catalog entry, and writing it forces the decisions the rest of this model formalizes: a named Owner, a confirmed classification, and a shared definition.

Then work outward in three moves.

Register. Record the rest of the domain's datasets, each with an Owner, a classification, its source, and a line on what it is for. That is Bronze, and it is most of the value of the first quarter: an inventory people and agents can trust to be complete.

Certify. Take the one asset others depend on most, set its Tier, and bring it to Silver. It becomes the worked example every later domain copies.

Connect. Put the contracts where people already work with AI. Add them as context in an enterprise AI tool, have it generate queries, and run those queries against the dataset under the person's own read access. A Silver contract should produce correct queries without a subject-matter expert in the room. Where it doesn't, the gap is the next thing to write down. That is the thesis working at a small scale.

Keep the contracts in one shared, versioned repository so every session reads the current copy. Move them into a catalog exposed over MCP when files stop scaling: usually when agents need to choose between sources by Tier and certification, or when a second domain arrives.

Then add the next domain. The Council forms when a second domain produces the first cross-domain conflict worth resolving, usually two teams using the same word for different things.

What Changes

In practice, a lot changes.

Teams stop pulling time-constrained subject-matter experts into meetings to rediscover what has been known and discussed many times. The context is written down once, has an Owner, and is read as often as it's needed.

Every role works from that established context. Product ideation, architecture drafts, development work, and support all start from the same definitions, lineage, and contracts.

AI can self-serve answers that have historically required someone to run ad hoc BI or analysis. The agent answers from certified data, so the person asking can see what it relied on and who

stands behind it. The headline measure (the share of agent queries answered from Silver-or-better assets) is how you watch that shift happen.

Continuing the Conversation

If you're working through any of this or think part of the program is wrong, I'd like to hear about it: jeremy.taylor@ossnetworks.com or linkedin.com/in/jeremytaylor00. The Enterprise AI Operating Model, which this program expands from Pillar 3, and additional articles are at ossnetworks.com.

About the Author

I lead global data and AI organizations with twenty years in technology. I've spent the past eleven years building and transforming cloud data platforms. Most recently, I've focused on adding an AI platform layer and agents in production, and I've had responsibility for standing up enterprise data governance. I completed the Carnegie Mellon University Chief Data and AI Officer executive certificate in 2026. This program came out of that work.

Appendix A: Adapting This to Your Organization

The document is built to be taken and made your own.

It names no products. Where the model needs a capability (a catalog, a quality framework, lineage capture), it describes the capability and leaves the tool to you. Appendix C is a worksheet for mapping each capability onto whatever you run.

Everything specific is a placeholder. Organization names, domain names, role titles, named resources, field names, and example data are all generic. The lists below say what to swap and, more usefully, what to keep.

The certification colors carry meaning. Bronze, silver, and gold are the only saturated colors in the document. If you re-skin it for your own brand, keep that discipline and give the certification levels the loudest colors you have.

Substitute throughout

- **The catalog:** the name of your platform, or your written data contracts if you have no platform yet (Where to Start).
- **The eight domains** (Account, Product, Subscription, Usage, Billing, Support, Marketing, Account Health Insights): supply your own domains. These come from a generic subscription business, purely to have something concrete on the page. Keep at least one aggregate domain if you have one; it governs differently.
- **Account Health Summary,** the worked example in Figure 1: a real data product of yours, with its real fields and SLA.
- **Enterprise Data Council:** whatever your governance body is called.
- **Named resources** in Appendix B: actual names and titles. An unnamed role is an unowned role.
- **Targets** for the measures in How Success Is Measured: your own baselines and targets.
- Split out a Data Product Owner if a domain holds many products with separate consumers and roadmaps. Accountability for the domain stays with its Owner.

Keep, unless you have a deliberate reason

- **Certification is a human judgment.** Automating the assignment removes the accountability that makes a certification mean anything. The catalog reports; the Owner decides.
- **One named Owner per domain.** One person may own several domains; a committee may own none.
- **Domains decide whether assets meet the bar; the Council sets the bar.** Collapse that and federation turns back into either central control or a free-for-all.
- **One name per concept, and one Tier 1 asset per name.** Overlapping sources are named for what they are and declare how they relate.
- **Certification and Tier stay separate.** Maturity and importance are different questions, and an agent needs both answers.
- **In new applications, unclassified means Restricted.** Register everything; restrict what nobody has classified. Reverse that default and sensitive data reaches the wrong hands before anyone knew it was there.
- **The supporting layer enables and does not decide.** Security, compliance, program management, and the data platform give the domains capability. If Infosec starts classifying a domain's data instead of the domain classifying against Infosec's standard, the program has quietly recentralized, and nobody announced it.

Adjust freely

- The wording of the six dimensions and their criteria, if your catalog or your consumers frame them differently. Keep Protection as the floor.
- Which grains you certify at. Drop any of the grains listed in §3.7 your platform can't hold a certification against, or that you don't need.

- Review cadences, to fit rhythms your teams already keep.
- Certification level names. Bronze, silver, and gold travel well, but if your platform uses medallion names for pipeline stages, rename these instead of running two meanings of the same three words. *Known, described, committed* works and is unambiguous.
- Which figures you keep. Each one stands alone.

Appendix B: Role Reference

Role accountabilities

Substitute. The named resource column describes the kind of person who typically holds each role. Replace it with the actual names and titles in your organization. An unnamed role is an unowned role.

Layer	Role	Named resource	Primary accountability
Executive	Sponsor	Chief data or technology executive	Advocate for the program; provide resources and authority; align with company strategy; resolve escalations.
Strategic	Enterprise Data Council	The Owners of the most-used domains, the head of the central data team, the regulatory and compliance lead	Set standards and priorities; own the certification rubric; resolve cross-domain conflicts and competing definitions; review program measures.
Tactical	Data Owner	Manager level: the lead whose team builds the	Value of data within the domain; approve

Layer	Role	Named resource	Primary accountability
		application, or the business sponsor who bought it. One person may own several domains.	domain policies and access; set Tier and certify the domain's assets (the work may be delegated to a trusted person; the accountability may not). Publish and maintain data contracts for the domain's products; see quality issues through to resolution.
Tactical	Data Steward	Assigned by the Owner, typically an analyst or data engineer who already knows the data	Define schemas, definitions, lineage, quality rules, and example queries for the domain; keep catalog entries current; triage quality issues to the Owner.
Operational	Implicit Steward	Everyone who sees, updates, or defines data	Use data within policy; judge fitness for use from the catalog before relying on it; raise gaps with the Owner or Steward; take responsibility for what they enter or define.
Supporting	Program Manager	Often the head of the central data team, part time	Coordinate roles and teams; run the Council's cadence; track the measures; develop training and data literacy.
Supporting	Regulatory and Compliance	Legal	Monitor regulatory obligations; identify

Layer	Role	Named resource	Primary accountability
			risk; audit governance processes.
Supporting	Information Security	Security	Partner on classification, personal data, and access policy.
Supporting	Data Platform	The central data team, with DevOps and cloud operations	Run the catalog, warehouse, and quality tooling the program depends on: uptime, configuration, access controls, backup and recovery.

RACI for common activities

R = responsible · A = accountable · C = consulted · I = informed · blank = not involved

SPN Sponsor · CNL Council · DO Data Owner · STW Data Steward · PM Program Manager · REG Regulatory and Compliance · SEC Information Security · PLT Data Platform

Activity	SPN	CNL	DO	STW	PM	REG	SEC	PLT
Set standards, priorities, and the certification rubric	I	A	C	I	R	C	C	C
Set classification and privacy standards	I	A	C	I	I	C	R	C

Activity	SPN	CNL	DO	STW	PM	REG	SEC	PLT
Register an asset and confirm its classification			A	R			C	C
Define and maintain the data contract			A	R				C
Set Tier and certify an asset			A	R	I			
Approve access to domain data			A				C	R
Remediate data quality issues			A	R				C
Set lifecycle status and deprecate assets			A	R				C
Resolve cross-domain conflicts and Tier disputes		A/R	C	C	C			
Escalate regulatory exposure or material risk	A	I	C		R	C	C	
Run the catalog and program tooling			I		C		C	A/R
Track and report program measures	I	A	C		R			C
Audit governance processes		I	C		C	A/R	C	

Implicit Stewards are not shown; their responsibilities are in §2.1. The matrix holds for built and bought applications alike. What changes is who holds the Owner role (§1.2). For a bought application, fixes at the source go through the vendor, and the Owner manages that relationship.

Appendix C: Platform Capability Mapping

This model names no products. Use the worksheet to record which of your systems provides each capability and where the gaps are. A gap is no blocker: it tells you which dimensions will need manual effort, and therefore where automation pays back fastest. One platform may cover many of these rows, or all of them.

Capability needed	**Dimension**	**Your system**	**Automated?**	**Gap notes**
Data access: the lake, warehouse, or databases that hold the data, queryable by people and agents under their own access	assumed			
Asset registry and search: the catalog itself	all			
Access for AI tools and agents over MCP or an API	all			
Integration options: cached data, exports, published contracts	all			
Ownership and escalation records	Ownership			
Schema registration and drift detection	Shared language			
Field-level annotation	Shared language			
Business glossary with term linking	Shared language			
Concept relationships: shared keys between overlapping concepts	Shared language			
Lineage capture, upstream and downstream	Origin			

Capability needed	Dimension	Your system	Automated?	Gap notes
Quality rule definition and execution	Quality			
Data profiling	Quality			
Freshness monitoring and alerting	Freshness			
Sensitive data detection, suggesting a class for the Owner to confirm	Protection			
Default-restricted policy for uncertified assets in new applications	Protection			
Access policy enforcement and audit logs	Protection			
Lifecycle status and sunset notices	lifecycle			
Certification and Tier recording, including uncertified and the date of last review	all			
Consumption telemetry: who and what uses each asset, and how often, from the data access layer's query logs	Origin, measures			

A minimum viable platform is a written contract spec. One domain's contracts, in a formal machine-readable format, cover Ownership, most of Shared language, and the Protection floor, and they record certification and Tier. That is enough to certify Bronze across a domain and start Silver. A catalog tool, profiling, quality execution, and freshness monitoring are what Silver at scale requires. Gold adds no new platform capability beyond Silver; its extra requirements are documentation and commitment, which are human.

Appendix D: Glossary

Aggregate domain. A domain whose data is composed from other domains instead of originating in operational systems. It consumes upstream definitions and is where overlapping concepts are reconciled.

Asset. Anything registered in the catalog, and therefore certifiable: a table, view, file, stream, dashboard, feature set, data product, or domain. Grain is a cataloging choice; see §3.1.

Catalog. The curated inventory of the enterprise data assets. People browse it and query it through AI tools; agents take their context from it; applications scale from its integration options.

Certification. The Bronze, Silver, or Gold level an Owner assigns to an asset: a judgment of how far a consumer can rely on it. “Certified Silver.”

Concept name. The single name a business concept carries in the catalog. Overlapping sources are distinct concepts, named for what they are and related by a shared key.

Council. The Enterprise Data Council: the single cross-domain governance body. Sets standards, owns the rubric, and resolves conflicts between domains.

Data contract. A published agreement describing an asset’s schema, refresh cadence and SLA, quality expectations, access rules, and example queries. Written first as a machine-readable spec.

Data domain. A logical grouping of data owned by the part of the business that knows it best, with a named Owner.

Data mesh. A sociotechnical approach that distributes data ownership to business domains while a thin central function provides standards, platform, and catalog. The source of this model's federated ownership.

Data product. A curated, documented, supported dataset with a data contract, discoverable and trustable across the organization.

Dimension. One of the six questions the rubric asks of an asset: Ownership, Shared language, Origin, Quality, Freshness, Protection.

Federated ownership. The rule that the teams creating data own it: the team that builds an application, or the sponsor who buys one, owns its datasets.

Grain. The level at which an asset is registered and certified: column, table, data product, or domain. The rubric reads the same at every grain.

Implicit steward. Anyone who sees, updates, or defines data, and therefore carries responsibility for it without a formal governance title.

Inherent responsibility. The rule that access to data creates responsibility for it, independent of job title.

Lifecycle status. Whether an asset is in development, active, deprecated, or retired. Separate from certification. A deprecated asset always carries a sunset date.

Minimally invasive. The principle that governance should support the work: codify what teams already do well, and automate instead of policing.

Owner. The named, manager-level person accountable for a domain or asset. A part-time responsibility that makes an existing decision explicit. May delegate certification work, never the accountability.

Program layer. One of five bands (executive, strategic, tactical, operational, supporting) to which every role is assigned by the work it does.

Tier. An asset's importance to the business. Tier 1 is business-critical, Tier 2 shared, Tier 3 local. Tier sets the target certification; only one Tier 1 asset may carry a given concept name.

Uncertified. A registered asset with no certification yet. In new applications, treated as Restricted until its Owner confirms a classification.

Sources and Influences

This document is a synthesis. Where it builds on other people's work, that work is named here. Everything not named below is my own.

Governance philosophy. The stewardship philosophy in Part 1 and the idea in §2.1 build on Robert S. Seiner's Non-Invasive Data Governance™ (KIK Consulting & Educational Services): that stewardship describes an existing relationship to data, that a program should formalize what people already do, and that accountability is layered across the organization. The five program layers in §2.2 follow his operating model of roles and responsibilities. Read the source directly: Seiner, *Non-Invasive Data Governance: The Path of Least Resistance and Greatest Success* (Technics Publications, 2014), and *Non-Invasive Data Governance Strikes Again* (2023).

Trademark notice. Non-Invasive Data Governance is a registered trademark of Robert S. Seiner and KIK Consulting & Educational Services. The *minimally invasive* framing in this document is the author's own adaptation and extension of those ideas. It says *minimally* because that is the more honest framing: any program asks something of the people in it, and the promise here is to keep that small. It is neither affiliated with nor endorsed by Robert S.

Seiner or KIK Consulting, and it should not be read as a restatement of the Non-Invasive Data Governance method.

Data architecture. Federated ownership, data products, and aggregate domains follow Zhamak Dehghani, *Data Mesh: Delivering Data-Driven Value at Scale* (O'Reilly, 2022). This model does not require a mesh implementation.

Data quality. The measures in the Quality dimension (completeness, validity, uniqueness, and consistency) and the treatment of timeliness as a dimension of its own follow the data quality dimensions set out by DAMA: the DAMA UK Working Group, *The Six Primary Dimensions for Data Quality Assessment* (2013), and DAMA International, *DAMA-DMBOK: Data Management Body of Knowledge*, 2nd edition (Technics Publications, 2017).

Data contracts. The Open Data Contract Standard (ODCS) is an open standard from Bitol, a Linux Foundation AI & Data incubation project, published under the Apache 2.0 license.

Common practice. Grain is Ralph Kimball's dimensional-modeling term. Tier for importance and bronze, silver, and gold for certification are conventions found across enterprise catalog products. The names of the other dimensions (ownership, lineage, classification) are common practice. RACI and decision rights are ordinary management practice.

Agent context. The Model Context Protocol (MCP) is an open standard published by Anthropic.

What is original here

The parts are established. What I've put together is a minimal governance program joined to a data catalog, so the subject-matter experts who know the data curate the context AI works from: they designate what is trusted and how much, and a named person's judgment is required where it counts. The mechanisms that make it work are mine: the six-question rubric

with Protection as a floor, one Tier 1 asset per concept name, and certification at any grain without a roll-up formula.

License