

EXECUTIVE PAPER · DATA

Your Business Needs One Shared Model of Itself

Every report, integration and AI tool in your company carries its own private definition of a customer, an order and a site. Executives who fund one shared model of the business stop paying to reconcile them.

AUTHOR

Chris Watkins

SERVICE

Data

FOR

CEO, COO, CIO, CDO · 12 min read

AT A GLANCE

The problem. Ask finance, sales and operations how many active customers the company has and you will often get three numbers, each correct by its own rules. The same split runs through every dashboard, every system integration and now every AI tool, because each one was built with its own copy of the company's definitions. In a 2024 survey, 67% of data professionals said they did not completely trust the data they use for decisions.

Why it persists. Definitions get written inside projects. A report, an integration or an AI pilot needs to know what an order is, so its builders decide, and the decision stays buried in that project's code. Nobody owns the meaning of the business's core objects across all of them.

The recommendation. Fund one shared model of the business: an agreed definition of each core object, how the objects relate, which actions are allowed on them, and where each lives in the source systems. Give the business ownership of the definitions, start with the handful of objects that carry money, and require every new report, integration and agent to use the model.

The Decision in Front of You

Most companies have spent heavily on data in the past decade: a warehouse or lakehouse, a reporting tool, an integration platform, and now AI pilots. Each of those purchases came with a project, and each project defined the business's core objects for itself.

The decision here is whether to keep paying for meaning one project at a time, or to fund a single model of the business that everything shares. That model is usually called an ontology. In plain terms, an ontology is a written, machine-readable agreement about what the business's things are (customer, order, asset, site, contract), how they connect to each other, and what may be done to them. The software that holds those definitions between the source systems and the tools that read them is called a semantic layer.

Two earlier papers in this series cover preparing data for one AI agent and funding AI work in the right order; this paper concerns the model the whole business shares, whether or not AI is involved.

What's Happening

Leaders do not trust their own numbers

The Precisely and Drexel University LeBow College of Business survey of more than 550 data and analytics professionals, published in November 2024, found 67% did not completely trust the data they use for decision-making, up from 55% the year before [1]. Data quality was the top data integrity

challenge, cited by 64% [1]. Salesforce's State of Data and Analytics report, drawn from nearly 8,000 executives, found data and analytics leaders estimate 26% of their data is untrustworthy [2]. Both firms sell data software.

Ownership is the recurring gap. dbt Labs' 2026 State of Analytics Engineering report, a survey of 363 data practitioners and leaders published in April 2026, found 41% named ambiguous data ownership as a governance obstacle, and 71% named incorrect or hallucinated outputs reaching stakeholders as a top concern [3].

Regulators now expect a dictionary of the business

The Basel Committee's principles for risk data aggregation, known as BCBS 239, expect banks to produce accurate, consistent risk numbers across the whole group [4]. In May 2024 the European Central Bank published a guide on how it expects banks to meet them [4]. The ECB reported that in its review of 25 significant banks, none had fully followed the principles, including the global systemically important ones, and it judged progress since then "generally insufficient" [4].

The guide asks for a dictionary of the main business definitions and a metadata repository, uniform definitions with clear ownership, validation rules, and lineage, meaning a record of where each figure came from and how it was transformed, down to individual data attributes [4]. Outside banking, that list is a fair description of a shared model.

The tools have converged on the same idea

The World Wide Web Consortium's Web Ontology Language, OWL 2, became a W3C Recommendation in 2009, with a second edition in December 2012 [5]. It describes classes of things, their properties and the individual instances, and it is exchanged as RDF, the W3C's standard format for linked data [5].

dbt Labs' semantic layer lets data teams define a metric such as revenue once and have every downstream tool use that definition; when the definition changes, it changes everywhere it is used [6]. Palantir describes its ontology as an operational layer that maps existing data sources into objects, properties and links, and adds action types that capture decisions and write them back to existing systems [7]. A model that also defines allowed actions is what an integration or agent needs before it changes anything.

In September 2025, Snowflake, Salesforce, dbt Labs, BlackRock and RelationalAI launched the Open Semantic Interchange, with a dozen partners, to create a vendor-neutral format for semantic definitions [8]. In January 2026 dbt Labs reported the first version of the specification was published under the Apache 2.0 license, covering datasets, metrics, dimensions and relationships, with Databricks also named as a partner [9].

Gartner placed the rising importance of semantic modeling among its top data and analytics trends for 2025 [10]. According to TechTarget's coverage, Gartner predicted that organizations prioritizing semantic modeling would improve the accuracy of their AI tools by 80% and cut the cost of building and

maintaining them by 60% [10]. Treat those as analyst predictions without published measurement behind them.

AI raises the cost of disagreement

Spider 2.0, a benchmark of 632 real enterprise database tasks presented at ICLR 2025, found that an agent built on OpenAI's o1-preview model solved 17.0% of tasks, against 91.2% on the older, simpler Spider 1.0 benchmark [11]. Enterprise schemas are large and cryptic, and the business meaning lives nowhere in them, so an AI tool reading raw tables guesses what "active" or "shipped" means.

Two Cases

Citigroup, 2024. In 2020 the Office of the Comptroller of the Currency fined Citi $400 million for shortcomings in risk management, data governance and internal controls, and the Federal Reserve ordered it to fix its data quality practices [12]. On July 10, 2024, the Federal Reserve fined Citigroup $60.6 million for insufficient progress remediating its data quality management problems and for lacking controls to manage the risk those problems created [13]. The OCC added about $75 million, for a combined total near $136 million [12][13]. Four years of remediation had not made the bank's data consistent. Fixing definitions after decades of separate systems takes years, so the work should start before a regulator sets the deadline.

WM Technology, 2024. WM Technology, the parent of the Weedmaps cannabis marketplace, reported "monthly active users" as a headline operating metric, and its SEC filings described those users as people who had engaged with its site [14]. On September 24, 2024, the SEC charged the company and its former CEO and CFO with negligent misrepresentations about the metric for May 2021 to May 2022, a period that included the company's listing through a SPAC merger [14]. According to the SEC, a large and growing share of the counted users had been shown the site through pop-under advertisements on other websites, and most of them never clicked or engaged [14]. Reported users grew while engagement was flat or falling [14]. The company agreed to a $1.5 million penalty, and the two executives paid $175,000 each and accepted three-year bars from serving as public company officers, without admitting or denying the allegations [14].

The written definition said one thing and the counting rule in the systems did another, and nobody reconciled the two before the number went to investors. In a shared model, "active user" carries one definition with an owner, and the mapping states exactly which traffic counts, so a change in what gets counted becomes visible as a change to the definition.

Why the Usual Approach Fails

Definitions are written where the work is funded. A dashboard project needs "revenue," so the analyst writes a rule. An integration project needs "order," so the developer picks a status code. An AI pilot needs "customer," so someone writes it into a prompt. Each rule is reasonable, and each lives where

only its builder can see it. After a few years a mid-size company has dozens of definitions of its five most important nouns.

Consolidating systems leaves the problem in place. Reports, integrations and AI tools still sit outside the new ERP, and each still needs its own logic for what its fields mean.

Large enterprise data models failed in a different way. IT teams modeled the whole business before delivering anything, the business never signed off, and the model aged before it was used.

Tool vendors each ship their own semantic layer. Buying three tools gives the company three places to define "customer" that do not talk to each other, the problem the Open Semantic Interchange was formed to address [8].

Nobody is accountable for meaning. The definition of an active customer or a completed order usually belongs to no one, so disputes get settled meeting by meeting and never written down [3].

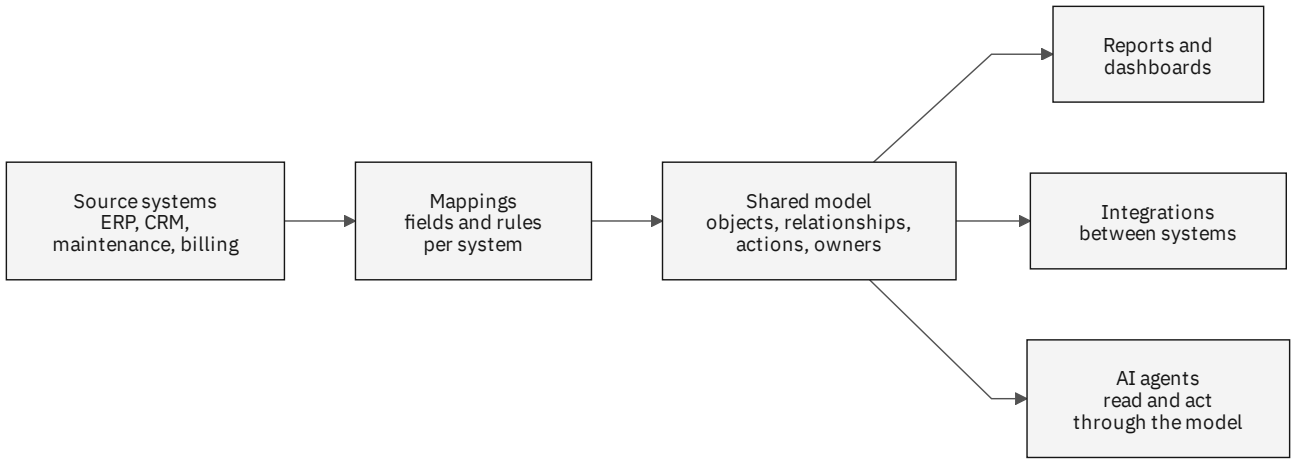
The Framework: One Model, Four Parts

Exhibit 1. Each part of the shared model replaces a decision that projects now make on their own

Part	What it records	Example	What goes wrong without it
Objects	Each core business thing, its definition, its identifier, and its owner	"Active customer: an account with an invoiced order in the last 12 months. Owner: CFO."	Finance, sales and the AI assistant report different customer counts
Relationships	How objects connect, and how many of each	An order belongs to one customer and ships from one site; an asset sits at one site	Integrations double-count orders or attach assets to the wrong site
Actions	What may be done to each object, by whom, with what checks	"Cancel order: allowed before shipment, by customer service or the order agent, with a refund check"	Agents and integrations change records under rules nobody approved
Mappings	Where each object and field lives in each source system, with the conversion rules	"Revenue: invoice date in the ERP, booking date in the CRM; finance's rule and owner stated"	Systems join fields that share a name and differ in meaning

Source: author's framework, informed by W3C OWL 2 [5], the ECB's expectations for data taxonomies [4], and published semantic layer and ontology designs [6][7][9].

Exhibit 2. Every consumer reads the business through one model, and the model maps to every source



Source: author's framework.

Five rules make the model work in practice.

The business owns the definitions. Each core object has a named executive owner who settles disputes about its meaning. The COO is often the right sponsor, because operations crosses every department that argues about definitions.

Start with the objects that carry money. Most companies can name the five to ten objects behind most of their revenue, cost and risk: customer, order, product, asset, site, contract, employee. Model those first and stop.

New work must use the model. Every new report, integration and AI agent reads core objects through the shared model, and the architecture review rejects anything that defines them privately.

Actions are part of the model. Agents and integrations change things, so the allowed actions, the roles that may take them, and the checks that apply belong in the same place as the definitions [7].

Keep the model portable. Store definitions in an open format, such as OWL and RDF or the Open Semantic Interchange specification, so the company can change vendors without rewriting it [5][9].

Costs, Risks, and Tradeoffs

The model costs executive time more than software. Settling what "customer" means requires a decision from someone with authority over the departments that disagree, and that decision has political cost.

Exhibit 3. Tradeoffs of funding a shared model of the business

Area	Cost or risk	Mitigation
Executive time	Owners must settle definition disputes their departments have avoided for years	Limit the first round to five to ten objects; give each owner a deadline and a tie-break rule
Standing team	A small permanent team of modelers and data engineers, plus part of each owner's time	Fund it as an operating cost, the way finance funds the chart of accounts
Over-modeling	Teams try to define everything before shipping anything	Every modeled object must have a named report, integration or agent using it within 90 days
Vendor lock-in	A semantic layer built inside one product is hard to move	Keep definitions in an open format and export them on a schedule [5] [9]
Model drift	The business changes and the model falls behind	Owners review their objects quarterly; source schema changes alert the model team
When this is wrong	A small firm on one SaaS suite with a handful of reports gets little from a separate model; a business about to be sold or split should wait for the new structure	Adopt the suite's own data model and write down the few definitions that matter

Source: author's analysis.

How to Act

Exhibit 4. A one-year roadmap

Window	Actions	Owner	Milestone
First 30 days	Name an executive sponsor. Choose five to ten core objects. Name one executive owner for each. Collect every existing definition of them.	CEO names COO as sponsor; CDO runs the inventory	Owners confirmed in writing; inventory shows how many definitions exist per object
First 90 days	Owners sign one definition per object. Write the relationships and the allowed actions. Map each object to its source systems with conversion rules. Choose a semantic layer and an open storage format.	CDO with object owners; CIO for platform	Signed definitions and mappings for every core object; definitions exportable in an open format
First 90 days	Pick the most disputed number, rebuild it from the model and retire the competing versions.	CDO with CFO	One number, used in the executive report, matching finance
First 365 days	Require all new reports, integrations and AI agents to use the model. Move the twenty most-used reports onto it. Put agent actions behind the model's action rules. Hold quarterly owner reviews.	CIO enforces in architecture review; CDO runs the team	No new work defines core objects privately; top 20 reports on-model; every agent action listed in the model

Source: author's recommendation.

Questions to Ask Your Team

1. How many different definitions of "customer" are in use across our reports, integrations and AI tools today?
2. Who has the authority to decide what an active customer or a completed order means, and has anyone used it?
3. Which number did we last argue about in an executive meeting, and how long did it take to settle?
4. Which actions can our integrations and AI agents take on customer, order or asset records, and where are those rules recorded?
5. If we changed our reporting or AI vendor next year, would our definitions come with us?
6. What would a regulator or auditor find if they asked for our dictionary of business definitions and the lineage of our top ten figures?

Closing

Every company already pays for its definitions. It pays in each project that rewrites them, in each meeting that reconciles them, and in each AI tool that guesses at them. Fund one shared model of the business instead: a small set of core objects, their relationships and allowed actions, mapped to the source systems, owned by the executives accountable for the numbers, and used by every report, integration and agent that comes after.

Sources

[1] Precisely. "2025 Planning Insights: Data Quality Remains the Top Data Integrity Challenges." Findings from "2025 Outlook: Data Integrity Trends and Insights," with Drexel University's LeBow College of Business. November 5, 2024. https://www.precisely.com/blog/data-integrity/2025-planning-insights-data-quality-remains-the-top-data-integrity-challenges.

[2] Salesforce. "The State of Data and Analytics." 2025. https://www.salesforce.com/analytics/state-of-data-and-analytics/.

[3] dbt Labs. "New dbt Labs report finds AI-driven acceleration is outpacing trust and governance." 2026 State of Analytics Engineering. April 14, 2026. https://www.getdbt.com/blog/new-dbt-labs-report-finds-ai-driven-acceleration-is-outpacing-trust-and-governance.

[4] European Central Bank, Banking Supervision. "Guide on effective risk data aggregation and risk reporting." May 2024. https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.supervisory_guides240503_riskreporting.en.pdf.

[5] World Wide Web Consortium. "OWL 2 Web Ontology Language Document Overview (Second Edition)." W3C Recommendation. December 11, 2012. https://www.w3.org/TR/owl2-overview/.

[6] dbt Labs. "dbt Semantic Layer." Product documentation. 2026. https://docs.getdbt.com/docs/use-dbt-semantic-layer/dbt-sl.

[7] Palantir Technologies. "Overview." Foundry Ontology documentation. 2026. https://www.palantir.com/docs/foundry/ontology/overview/.

[8] Open Source For You. "Snowflake, Salesforce, dbt Labs, BlackRock Introduce Open Semantic Interchange." Coverage of the Open Semantic Interchange launch. September 24, 2025. https://www.opensourceforu.com/?p=90309.

[9] dbt Labs. "The OSI spec updates." January 29, 2026. https://www.getdbt.com/blog/the-osi-spec-updates.

[10] TechTarget. Avidon, E. "Agents, semantic layers among top data, analytics trends." Coverage of Gartner's 2025 data and analytics trends. July 1, 2025. https://www.techtarget.com/data-technologies/feature/Agents-semantic-layers-among-top-data-analytics-trends.

[11] Lei, F., et al. "Spider 2.0: Evaluating Language Models on Real-World Enterprise Text-to-SQL Workflows." arXiv:2411.07763, ICLR 2025. 2025. https://arxiv.org/abs/2411.07763.

[12] Claims Journal. "Citi Fined $136 Million for Failing to Fix Regulatory Issues." July 12, 2024. https://www.claimsjournal.com/news/national/2024/07/12/325007.htm.

[13] Board of Governors of the Federal Reserve System. Press release on enforcement action against Citigroup Inc. July 10, 2024. https://www.federalreserve.gov/newsevents/pressreleases/enforcement20240710a.htm.

[14] U.S. Securities and Exchange Commission. "SEC Charges Public Company WM Technology, Inc. and Two Former Senior Executives with Making Negligent Misrepresentations Regarding the 'Monthly Active Users'." Litigation Release No. 26127. September 25, 2024. https://www.sec.gov/enforcement-litigation/litigation-releases/lr-26127.

The views in this paper are the author's own and do not represent any employer. Scenarios are hypothetical. This paper is not legal, financial, or compliance advice.