Informatica from Salesforce | Deloitte.

White Paper

Building the Semantic Data Layer for Agentic AI

Without semantics, AI may learn the words – but it will never speak the language of your business

By Marc Beierschoder, Jasa Andrensek and Tobias Rebele

Where data & AI come to LIFE™

Building the Semantic Data Layer for Agentic AI

A new breed of AI is emerging inside enterprises: agentic systems that produce more than insights. They can make decisions, take actions, make recommendations, and help humans collaborate to achieve a common objective.

But the extent to which they can deliver real impact depends on how well each agent understands the business it serves. Agentic systems demand data that is semantically explicit and machine-interpretable. These are fundamentally different capabilities. Bridging them starts – or perhaps ends – with semantics.

Most Chief Data Officers (CDOs) are rewarded for availability and control: delivering access to data, enforcing governance, and improving maturity scores across quality, lineage, and compliance. These are necessary achievements, but not enough for meaningful agentic AI use cases.

In this paper, we examine how Chief Data Officers (CDOs) can build the semantic backbone for AI agents to reason, cooperate and act responsibly throughout the enterprise.

It sets out a framework for constructing a semantic layer that incorporates business context into data architecture that's scalable, explainable and fit for autonomous operations. The objective for CDOs? Make sure your data isn't just usable, but intelligible.

Meet the authors

Marc Beierschoder
Partner and AI & Data Lead,
Deloitte

Jasa Andrensek
Senior Manager,
Deloitte Consulting

Tobias Rebele
Data Platform
Sales Leader DACH,
Middle East and LATAM
Informatica

The rise of agentic AI and the value of semantics

Agentic AI represents a paradigm shift in enterprise intelligence. Instead of sitting back and passively analyzing business information, agentic systems act. They open workflows, read messages from your business, and collect feedback.

Close to **25% of enterprises** deployed AI agents in 2025, rising to 50% in 2027. But without semantics, all these agentic AI systems will fail – at scale.

Semantics refers to the data about your data, the vital logic that describes essential attributes like "customer," "order" or "risk exposure". Too often that information is buried in application code or kept in spreadsheets – accessible only to people who know where to look. For AI agents, that kind of opacity is deadly.

Business context is needed:

Definition of business entities	Real-world relations between business entities	Mapping between business entities and underlying data
What is sales? What is car insurance product? ...	B2C Customer buys Car insurance, ...	What are B2C customers + addresses stored in DB? ...

Sample data, sample queries	*Technical metadata, joins*	*Metrics calculations*	*Synonyms*	*Cached user queries & responses*

As organizations scale from pilot AI experiments to systems that continuously deploy, evaluate, and improve AI, semantics becomes a new form of data infrastructure. It's the connective tissue between raw data and intelligent behavior.

AI accuracy improves dramatically with a well-constructed semantic layer. A **2024 study** by data.world[1] looked at property and casualty data and compared the accuracy of AI answers both with a semantic layer – and without. Researchers found that adding a semantic layer boosted overall accuracy from 16% to 54%, rising to 90% for some questions.

Model quality is no longer the bottleneck. Meaning is. The question for CDOs isn't whether to build semantics into their data ecosystems, but how fast, and how systematically.

[1] data.world | Nov 2023: A BENCHMARK TO UNDERSTAND THE ROLE OF KNOWLEDGE GRAPHS ON LARGE LANGUAGE MODEL'S ACCURACY FOR QUESTION ANSWERING ON ENTERPRISE SQL DATABASES

Challenges and constraints

Every enterprise talks about data as an asset, yet most treat metadata as an afterthought. That means most AI data platforms are structurally incapable of supporting autonomous systems. In the age of AI, metadata and business context are essential. Without them, AI systems are technically powerful but contextually blind:

Fragmented semantic
In most organizations, data meaning is scattered. Application developers define logic locally. Analysts maintain their own glossaries. Line-of-business users rely on tribal knowledge, where business meaning resides in the heads of SMEs rather than being systematized. When different systems define "customer value" or "churn" differently, AI models trained on those inputs will behave inconsistently.

Metadata debt
Metadata may be available, but in practical terms it's often inert: stored in static catalogs and rarely accessible to the models that need it. Some call this the "metadata debt" effect – an invisible technical burden that builds up when documentation lags behind reality. That debt erodes AI performance, governance and trust in its capabilities.

Siloed governance
Data governance is still trapped in a compliance mindset where governance teams manage policies while data scientists work on projects, leaving the connection between metadata quality and AI accuracy undefined. For agents to distinguish between compliant and non-compliant data, governance needs to be operationalized.

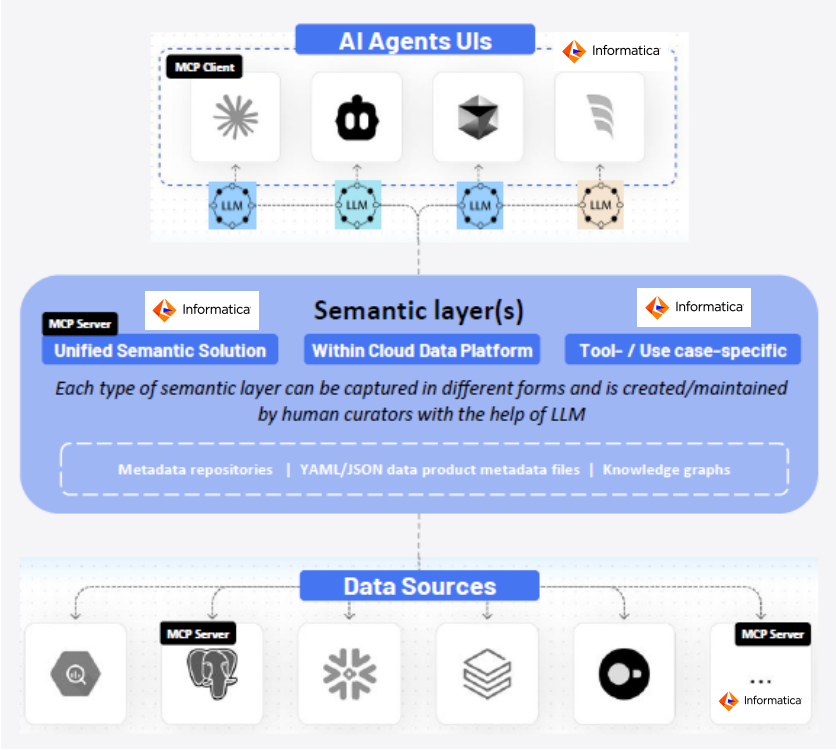
Context loss
Lack of clarity about why a data set was created can also impact agentic performance. For example, an AI model might know that a past transaction had occurred, but not that it was part of a refund process. Without provenance and context, even the most sophisticated model can output brittle or biased outputs.

Metadata allows AI to reason, trace lineage, and learn responsibly.

Emerging semantic architectures for AI agent

Enterprises moving toward agentic AI need a semantic layer that can deliver consistent context, interoperable metadata, and transparent reasoning across the full data estate. Three models are emerging:

1. **Unified semantic solutions** that provide strong cross-platform consistency, giving data teams a single source of truth for business definitions and metadata.
2. **Cloud-provider semantic layers** that are embedded within the major platforms offering native performance and seamless integration for data already resident in those environments, which is particularly useful as "talk-to-data" agents become more standard.
3. **Tool-specific semantic layers** that build lightweight semantics directly into analytics, cataloging, and integration products, targeting use cases pragmatically and cost-effectively.



Each of these options excels in scenarios that demand multi-hop reasoning, rely on both structured and unstructured sources, or need auditable chains of inference. They also help make AI agent behavior more transparent by making clear the links that underpin an agent's conclusions.

Regardless of which model you choose, semantic layers will need human governance. Experts must define and maintain key metadata relationships, supported by LLM-based automation that speeds-up mapping, enrichment, and validation.

Informatica's tooling can operate across each one of these architectural tiers, providing a consistent foundation for enterprises scaling toward live deployments of agentic AI.

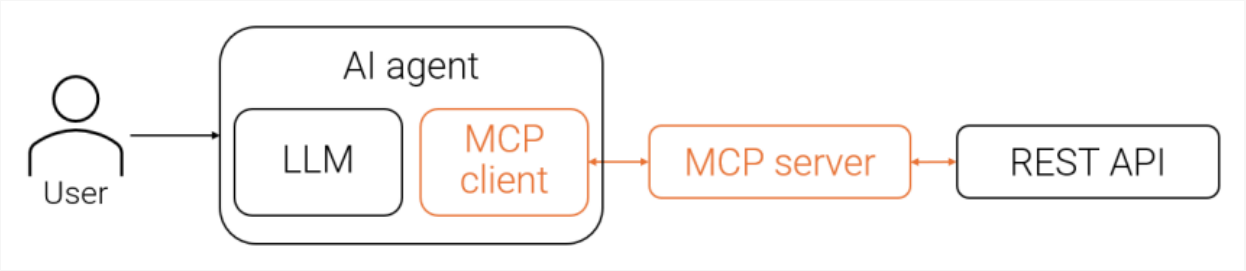
MCP: From theory to practice

For AI agents to be effective, they need to read data from – and perform actions on – live data systems.

Model Context Protocol (MCP) is an open protocol that solves this problem by providing a standardized way for large language models (LLMs) to interact with tools, such as data systems and processes. You can leverage MCP to expose Application Integration processes as tools that make your AI agents more accurate and contextual.

For example: You might want to enable your AI agent to create customer records in Informatica Customer 360 SaaS. If you already have a process in Application Integration that creates customer records, you can use MCP to bridge the gap from the process's REST API endpoint to the agent.

To use MCP to connect a REST API endpoint to an agent, you first create an MCP server and then add a reference to the server in your agent's MCP client:



MCP connects AI agents to REST API systems with a client-server architecture. An MCP server defines how agents can interact with a specific REST API, and an MCP client creates a 1:1 connection from an agent to an MCP server.

This is how a metadata layer built on Informatica's Intelligent Data Management Cloud (IDMC) can be operationalized. IDMC maintains the system of record for business semantics while the MCP server translates that context into a form consumable by large language models and AI agents.

Agents can then retrieve authoritative definitions, entity relationships, and governance constraints on demand. This allows the same semantic layer to be reused across a wide range of GenAI and agentic use cases, ensuring that every interaction is grounded in shared business meaning.

The vital role of AI-ready data

Leveraging MCP and achieving meaningful outcomes with agentic AI systems demands AI-ready data, but there are critical gaps:

- Data quality is still the major constraint, as model outputs will only be as reliable as the underlying information assets that feed them.
- Data access is equally important if users are going to receive results aligned with their permissions and entitlements.
- Clear data ownership is also required so that stewards can set the rules for how an AI agent may use data and under what conditions.

Strong semantics connect business language and system representation, improving the relevance, accuracy and interpretability of any LLM-generated output needed for an agent to operate.

Designing the semantic foundation for agentic AI

If data is the fuel for AI, semantics makes the engine run. CDOs must now create a system where meaning – not just data – flows through to the enterprise. Four interlocking principles form the foundation.

1

Define an enterprise semantic layer

Step one is to create a shared ontology that connects business concepts like "customer", "policy" or "asset" with their technical representations across multiple systems. This layer is a living schema, not a static glossary, one that describes connections, dependencies, and intent.

In practical terms it means merging business glossaries, data catalogs, and AI-driven mappings into a single unified semantic layer. When agents query the data, they can interpret the meaning and apply reasoning logic, rather than just relying on pattern recognition.

2

Metadata as an OS

In modern AI architecture, metadata powers critical data management capabilities: data lineage, explainability, access control, and trust scoring. A semantic metadata layer creates a machine-usable context for agents by clarifying who owns the data, how it was transformed, and which policies apply.

Informatica's CLAIRE AI engine automatically discovers metadata patterns, classifies data sensitivity, and updates catalogs dynamically. Paired with an AI governance framework, metadata insights can be integrated into model risk assessments and ethical compliance layers.

3

AI-driven data modelling

Generative AI can now help create that vital semantic layer. Using large language models and graph techniques, CDOs can automate schema discovery, relationship inference, and ontology maintenance.

Informatica's tools already apply generative AI to build knowledge graphs that map business entities across disparate systems. Imagine an AI assistant that reads through your data warehouse, discovers hidden relationships, and proposes a semantic model that aligns with business logic, then validates it against governance policies.

4

Active governance

Governance needs to evolve from a compliance bolt-on to a functional layer of AI's underlying data infrastructure. In an agentic environment, policies can't be imposed after the fact; they need to be integrated into the logic that tells agents how to behave.

This means agents only use trusted, compliant and up-to-date data sources. Deloitte's "Responsible AI" model fits perfectly here. Governance as the enabler of responsible autonomy, not the gatekeeper.

Success factors

The principles of successful semantic layer metadata design are still being established, but companies that dominate agentic AI adoption follow at least four common patterns:

1. **Semantic data fabric**
They build adaptive data fabrics that consolidate business and technical metadata into a single, query-able layer. This enables AI agents to cross systems without losing the meaning (for example, moving between ERP and IoT data streams).

2. **Trusted AI model**
In **Deloitte's AI Factory** model, metadata serves as an audit trail for decision-making. When an AI agent acts, the mechanism can list every single step from the data that informed this decision. This turns governance from reactive documentation into active assurance.

3. **Accelerated onboarding**
Semantic discovery enables new AI use cases to launch in weeks rather than months. Instead of needing to invent new definitions for each project, it allows team members to plug into pre-defined business concepts and use established ontologies.

4. **Operational efficiency**
Semantic mapping – driven by AI – eliminates thousands of hours of spreadsheet reconciliation, allowing data teams to focus on higher-value work, such as model tuning, policy design and innovation.

In **Deloitte's AI Factory** model, metadata serves as an audit trail for decision-making. When an AI agent acts, the mechanism can list every single step from the data that informed this decision.

Industry use cases

Real-world progress can start quietly, perhaps from a single domain, a limited pilot or a data team intent on grounding AI in trusted information. The following examples show how organizations in different sectors are already improving agentic systems with semantics, via measurable gains in accuracy, reasoning and governance:

Life sciences

A global pharmaceutical company set out to build a generative AI assistant for its Clinical Operations team that could answer ad-hoc questions from structured data without relying on technical staff.

The initial prototype used a SQL-generating agent. It worked, up to a point. Queries grew increasingly complex; the schema became more tangled, and overall accuracy quickly degraded.

The team then re-architected the system around a knowledge graph model, linking trial data, study protocols and operational metrics through shared semantics. The result? Across six tiers of query complexity, accuracy exceeded 95 percent for the first three, which accounted for the vast majority of user questions.

What changed wasn't so much the model but the meaning layer beneath it. By giving the agent a semantic map of the domain, they turned data access into contextual reasoning..

Public sector

A government agency had a highly complex resource planning process involving numerous stakeholders. Its data spanned departments, contractors, and external registries over a fragmented landscape that complicated efforts to output traditional reporting.

The agency deployed multiple AI agents to assist planners, but each needed to interpret different tabular sources. Rather than force-fitting a central model, the data team built an interpretation layer comprised of lightweight semantic middleware that could reformulate user prompts into structured queries.

The new hybrid system produced immediate gains in reporting accuracy and decision speed. The full semantic layer is still being rolled out, but the pilot demonstrated a critical principle: you don't need full metadata maturity to see value.

Consumer goods

A global consumer-products brand has taken a dual approach to semantic design, balancing scale and pragmatism.

First, the company developed a knowledge graph defining ontologies across business units and departments. This serves as the foundation for "talk-to-data" interfaces, enterprise search and analytics that effectively transform scattered datasets into a connected, durable knowledge asset.

In parallel, the brand implemented semantic views within its core data platform. These store business context directly with data products to create a simpler, but still effective, model for smaller domains. Together, the two approaches blend to create a layered fabric of meaning. The knowledge graph enables reasoning across the enterprise while the semantic views deliver local agility.

The payoff has been both technical and human. Data scientists and business users alike now query the same language of the enterprise – not just its data.

CDO Checklist: 5 Steps to Semantic Readiness

The business case for investing in better metadata and semantics is clear, but where do you start. A full enterprise rollout may not be necessary at the outset. Instead, CDOs should map out a focused, time-boxed blueprint that sets direction.

1. **Anchor on a Single Domain and Use Case**
   Select one business domain and one agentic AI use case where lack of context is already limiting accuracy, trust or automation. Avoid enterprise-wide scope. If you need full rollout approval to start, you are already too late. The goal is to expose semantic gaps quickly, not to perfect the data estate.

2. **Define the Core Business Semantics**
   Identify the critical business entities, definitions, and relationships the AI agent must understand to reason correctly. Make semantics explicit and machine-interpretable — not embedded in code, dashboards, or human knowledge.

3. **Assess Metadata Fitness for AI Consumption**
   Evaluate whether existing metadata is complete, current, and accessible to AI systems. Focus on lineage, ownership, policy constraints, and semantic alignment – not catalog coverage alone.

4. **Embed Governance into the Semantic Layer**
   Define ownership, stewardship and policy enforcement for semantic assets. Ensure AI agents consume only trusted, compliant, and explainable data by design, not by exception.

5. **Produce a Scalable Semantic Blueprint**
   Deliver a domain-level semantic target state, governance model and prioritized rollout plan. Use this blueprint to guide future investments and expand agentic AI capabilities across additional domains.

Timeframe: 6–8 weeks

Outcome: A clear, defensible path from governed data to AI systems that can reason, justify and act responsibly.

Now is the time to build

The value of good data has gone up, and the role of data management is now to provide AI-ready data that will enable agents to do their work. Semantic information provides the business context agents need to accurately reason over it.

Deloitte and Informatica have the integrated expertise to deliver this transformation: Informatica makes semantics executable at scale. Deloitte ensures it is designed, governed and adopted as enterprise capability. Together, we can help CDOs create a blueprint that transforms data into intelligence.

Ready to start your enterprise semantic blueprint?
Begin with one domain.

Key takeaways for CDOs

- Agentic AI needs contextual data. The autonomous systems can only reason, act and learn responsibly when they understand what that data means, not just what it says.
- Semantics is the new data infrastructure. The next competitive advantage in AI will not come from bigger models, but smarter data with interpretable business semantics.
- Metadata needs to start to "work". Consider metadata as a live, executable part of your architecture that powers trust, governance, and reasoning.
- AI governance must be proactive and forward thinking. Placing compliance and policy intelligence at the semantic level lets AI act safely at scale.

Get in touch

To find out more, contact Deloitte and Informatica's experts:

Marc Beierschoder

Partner and AI &
Data Lead, Deloitte
msbeierschoder@deloitte.ch

Jasa Andrensek

Senior Manager,
Deloitte Consulting
jandrensek@deloitte.ch

Tobias Rebele

Data Platform Sales Leader DACH,
Middle East and LATAM Informatica
trebele@informatica.com

Or contact us **here**.

About Deloitte

Deloitte has been around for 175 years now, and in an ever more complex world our purpose is to make an impact that matters. To our clients, people and society. Whether it's for providing assurance, digital transformation, adaptive business, M&A advice or the great societal challenges in the fields of Energy, Food, Health, Mobility, Security or Work – at Deloitte we create solutions. Together with our clients.

Through multi-disciplinary collaborations we connect our clients with strong partners and ecosystems. We develop new ideas, insights and solutions for the most challenging and complex issues of our time. It's how we accelerate breakthrough outcomes.

Learn how Deloitte's approximately 415,000 people worldwide make an impact that matters at **www.deloitte.com.**

About Informatica

Informatica from Salesforce is a leader in AI-powered enterprise cloud data management. Its Intelligent Data Management Cloud (IDMC) platform enables organizations to connect, manage and unify AI-ready data across the enterprise. With capabilities spanning data cataloging, integration, governance, quality, privacy, metadata management and master data management, Informatica supports a broad partner ecosystem and helps customers unlock the full value of their data and AI initiatives.

About Salesforce

Salesforce is the #1 AI CRM, empowering companies to connect with their customers in a whole new way through the power of artificial intelligence, data, and trust. For more information about Salesforce (NYSE: CRM), visit: **www.salesforce.com**.

Worldwide Headquarters
2100 Seaport Blvd.
Redwood City, CA 94063, USA
Phone: 650.385.5000
Fax: 650.385.5500
Toll-free in the US: 1.800.653.3871

informatica.com
linkedin.com/informatica
x.com/Informatica

CONTACT US

Where data & AI come to LIFE™

Informatica

IN09-0326-5340