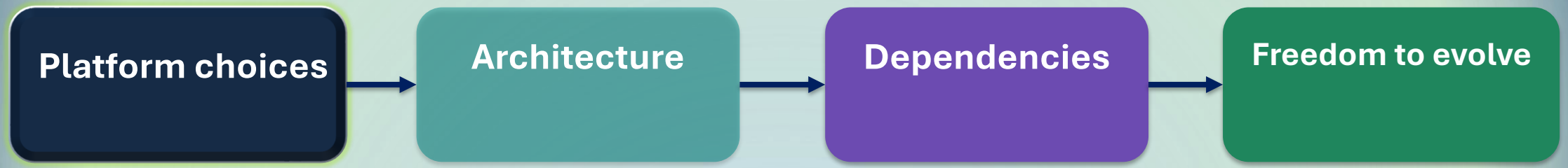
THE CDAO'S FOUNDATIONAL BET

Building an AI-ready enterprise without creating tomorrow's lock-in

CDAO Fall Keynote • Manajit Barman, Data & AI Executive

Every CDAO is making a bet

Technology choices become strategic constraints—or strategic options.



If your vendor changes tomorrow, what stays yours?

We have heard “unified platform” before

Warehouses → lakes → lakehouses → data mesh → data fabric → AI platforms

Each wave promised:

One place. One truth. Less integration.

The reality in many enterprises:

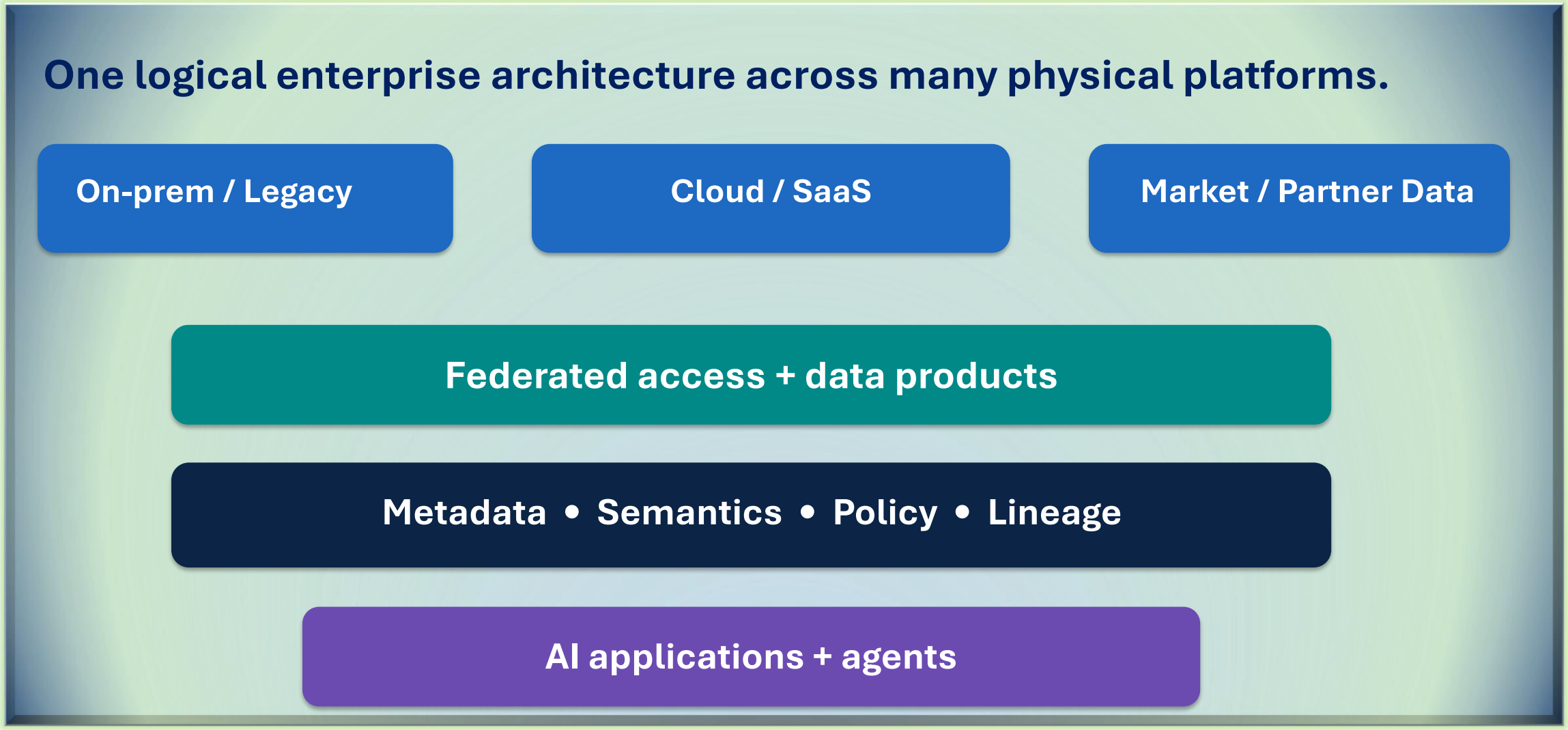
More copies • More pipelines • More silos

What changed in the last 18 months?

Five shifts make platform coherence materially more valuable.

1. Agents consume data and invoke tools
2. Unstructured and multimodal data matter
3. Metadata becomes runtime context
4. Models evolve faster than platforms
5. Governance must operate at machine speed

Unified does not mean centralized



Build, buy, or partner?

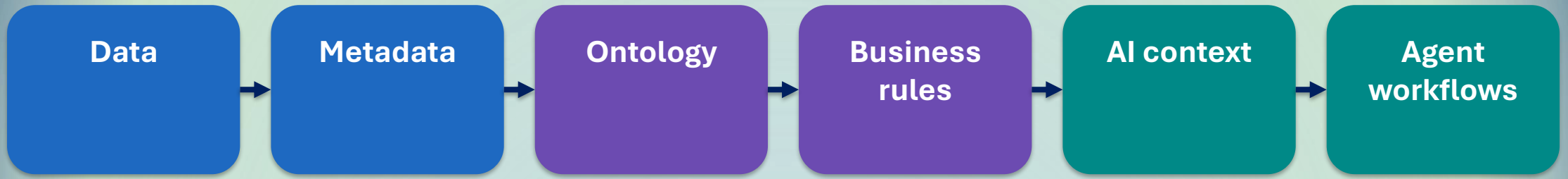
Protect what makes your enterprise uniquely intelligent.

BUY	PARTNER	OWN
Commodity infrastructure	Specialist capabilities	Semantics and ontology
Compute and general tools	Accelerated integration	Critical rules and policy

Buy technology. Partner for acceleration. Own enterprise meaning.

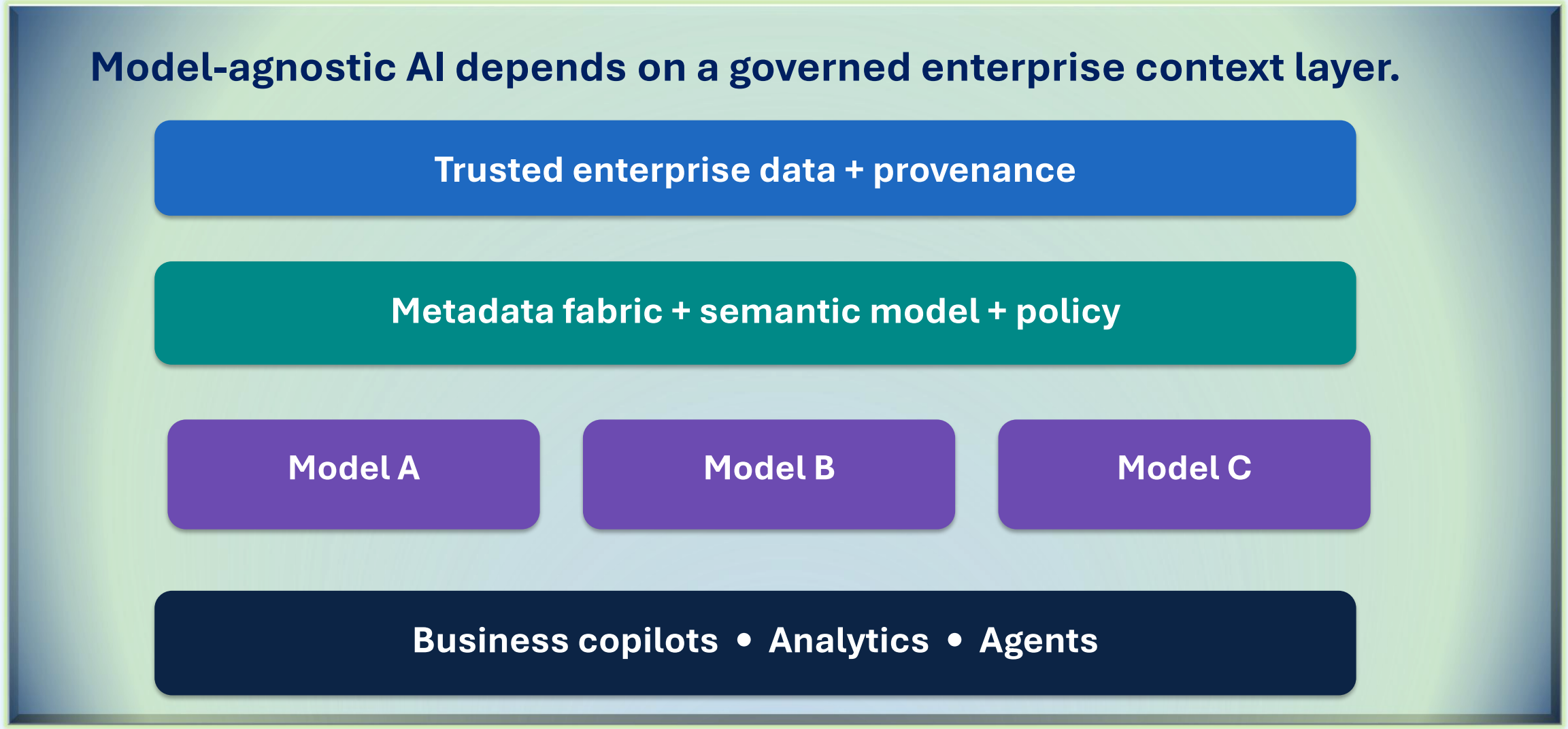
The new lock-in is semantic

Open files alone do not guarantee an open enterprise.



Can you export, explain, govern, and rebuild every layer?

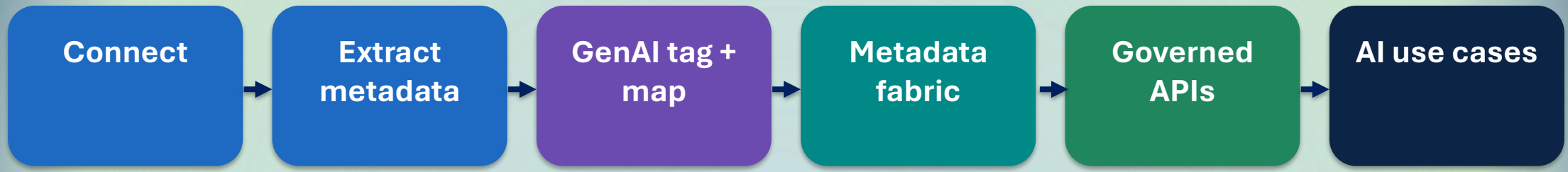
Own the context, not the model



Financial asset management: from sources to AI

An illustrative end-to-end metadata fabric pattern.

Legacy custody • Cloud portfolios • Market feeds • Research PDFs



Example: “AAPL” → Securities domain | Equity | ISIN / ticker | Source owner | Quality | Sensitivity | Provenance

The journey: fragmented → AI-ready

Modernize the architecture before migrating the enterprise.

1	2	3	4	5
Discover	Tag & classify	Align semantics	Govern & federate	Expose & measure

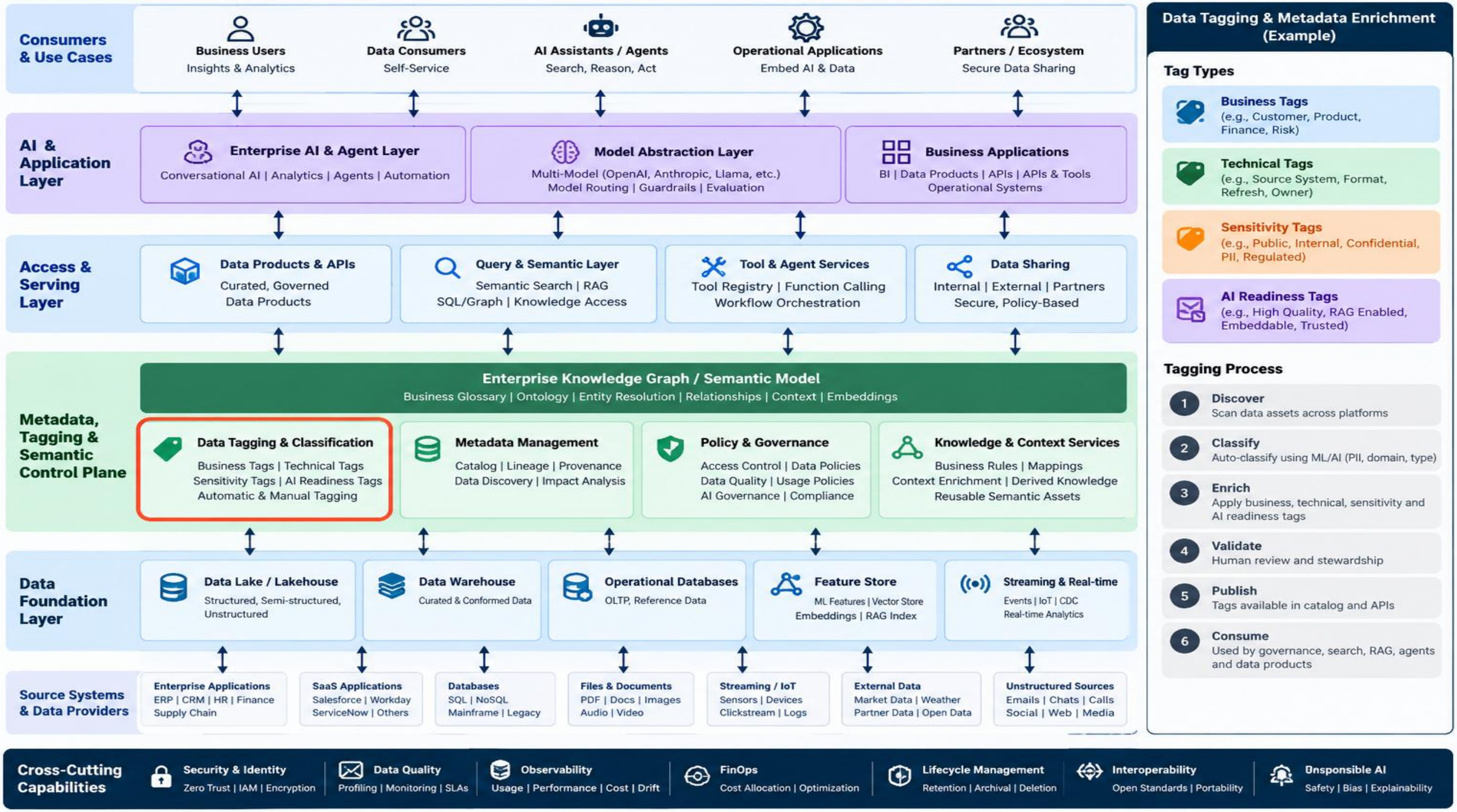
Pilot one valuable data product → prove reuse → scale domain by domain

Five architectural bets worth protecting

Design for change, not for the current vendor landscape.

- Own your enterprise semantics
- Separate storage, compute, and intelligence
- Keep models replaceable
- Enforce policy at runtime
- Build federated, reusable data products

Are we building an AI platform or the freedom to evolve?


Consumers & Use Cases
Business Users
Insights & Analytics
Data Consumers
Self-Service
AI Assistants / Agents
Search, Reason, Act
Operational Applications
Embed AI & Data
Partners / Ecosystem
Secure Data Sharing
AI & Application Layer
Enterprise AI & Agent Layer
Conversational AI | Analytics | Agents | Automation
Model Abstraction Layer
Multi-Model (OpenAI, Anthropic, Llama, etc.)
Model Routing | Guardrails | Evaluation
Business Applications
BI | Data Products | APIs | APIs & Tools
Operational Systems
Access & Serving Layer
Data Products & APIs
Curated, Governed
Data Products
Query & Semantic Layer
Semantic Search | RAG
SQL/Graph | Knowledge Access
Tool & Agent Services
Tool Registry | Function Calling
Workflow Orchestration
Data Sharing
Internal | External | Partners
Secure, Policy-Based
Metadata, Tagging & Semantic Control Plane
Enterprise Knowledge Graph / Semantic Model
Business Glossary | Ontology | Entity Resolution | Relationships | Context | Embeddings
Data Tagging & Classification
Business Tags | Technical Tags
Sensitivity Tags | AI Readiness Tags
Automatic & Manual Tagging
Metadata Management
Catalog | Lineage | Provenance
Data Discovery | Impact Analysis
Policy & Governance
Access Control | Data Policies
Data Quality | Usage Policies
AI Governance | Compliance
Knowledge & Context Services
Business Rules | Mappings
Context Enrichment | Derived Knowledge
Reusable Semantic Assets
Data Foundation Layer
Data Lake / Lakehouse
Structured, Semi-structured,
Unstructured
Data Warehouse
Curated & Conformed Data
Operational Databases
OLTP, Reference Data
Feature Store
ML Features | Vector Store
Embeddings | RAG Index
Streaming & Real-time
Events | IoT | CDC
Real-time Analytics
Source Systems & Data Providers
Enterprise Applications
ERP | CRM | HR | Finance
Supply Chain
SaaS Applications
Salesforce | Workday
ServiceNow | Others
Databases
SQL | NoSQL
Mainframe | Legacy
Files & Documents
PDF | Docs | Images
Audio | Video
Streaming / IoT
Sensors | Devices
Clickstream | Logs
External Data
Market Data | Weather
Partner Data | Open Data
Unstructured Sources
Emails | Chats | Calls
Social | Web | Media
Data Tagging & Metadata Enrichment (Example)
Tag Types
Business Tags
(e.g., Customer, Product, Finance, Risk)
Technical Tags
(e.g., Source System, Format, Refresh, Owner)
Sensitivity Tags
(e.g., Public, Internal, Confidential, PII, Regulated)
AI Readiness Tags
(e.g., High Quality, RAG Enabled, Embeddable, Trusted)
Tagging Process
1 Discover
Scan data assets across platforms
2 Classify
Auto-classify using ML/AI (PII, domain, type)
3 Enrich
Apply business, technical, sensitivity and AI readiness tags
4 Validate
Human review and stewardship
5 Publish
Tags available in catalog and APIs
6 Consume
Used by governance, search, RAG, agents and data products
Cross-Cutting Capabilities
Security & Identity
Zero Trust | IAM | Encryption
Data Quality
Profiling | Monitoring | SLAs
Observability
Usage | Performance | Cost | Drift
FinOps
Cost Allocation | Optimization
Lifecycle Management
Retention | Archival | Deletion
Interoperability
Open Standards | Portability
Responsible AI
Safety | Bias | Explainability

From Any Source to AI-Ready Data Foundation

Connect → Pull Metadata → Tag & Enrich with GenAI → Build Metadata Fabric → Expose as APIs → Enable Next Gen AI Use Cases

1. Data Sources
Any System, Any Format

- **On-Prem / Legacy** — Mainframe, Oracle, SQL Server, Files
- **Cloud Platforms** — AWS, Azure, GCP, SaaS Applications
- **Unstructured Data** — Documents, Emails, Audio, Video, Images
- **Streaming / IoT** — Market Feeds, Sensors, APIs
- **External / Third Party** — Market Data, ESG, Regulatory Data

2. Ingest & Pull Metadata
Connectors, APIs, Direct Access

- Pre-built Connectors
- APIs
- Database Access
- File & Document Ingestion
- Streaming Connectors
- Custom Connectors

3. GenAI Classification & Tagging
Automatically Understand, Classify and Tag

GenAI Models
LLM + Domain Models (Finance Ontology)

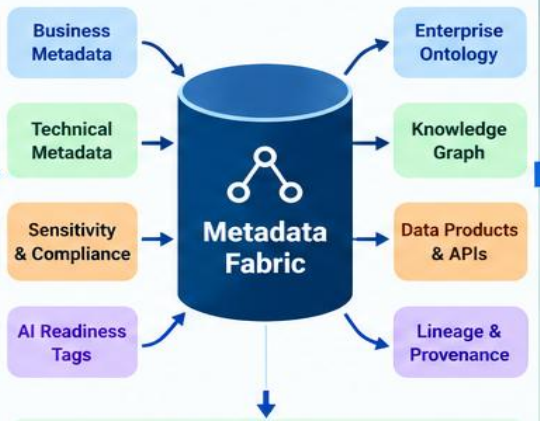
Auto-Tag Data

- Identify data domain
- Extract key metadata
- Detect entities & relationships
- Apply business, technical, sensitivity and compliance tags
- Map to enterprise ontology

Example: Financial Asset Management Data

Domain Tag	Securities
Entity Tag	AAPL (Equity)
Data Type	Reference Data
Sensitivity	Internal
Regulatory Tag	FINRA
Business Tag	Trading, Portfolio

4. Autonomous Metadata Fabric
Enterprise Catalog, Semantic Model & Data Products



Single Source of Truth

- Unified Metadata
- Common Semantics
- Trusted & Governed

5. Expose & Enable AI
APIs, Services and Data Products

- **Data Product APIs** — REST / GraphQL
- **Streaming APIs** — Real-time Data
- **Self-Service Portal** — Search & Discover
- **AI / Agent Tools** — MCP, Function Calling
- **Secure Data Sharing** — Internal / External
- **Governed Access** — Identity & Policy

Data Foundation Layer
Physical Data, Virtualized Access and Processing

- **Data Lake / Lakehouse** — Raw, Curated, Conformed
- **Data Warehouse** — Structured Data
- **Vector Store** — Embeddings for RAG
- **Event Store** — Real-time Data
- **Data Virtualization** — Federated Access

Next Generation AI-Ready Use Cases (Example for Financial Asset Management)

- Investment Research Copilot
- Portfolio Optimization
- Risk & Compliance Monitoring
- Market Intelligence & Insights
- Automated Reporting
- AI Agents for Client Advisory

QUESTIONS ??