TAXONOMY · ONTOLOGY · GOVERNANCE · AI

From Taxonomy to Governed Ontology

What changes when ontology, governance, and AI meet, and how to get through it without starting over.

A practical guide for data and knowledge leaders who already have an enterprise taxonomy and now need it to carry production AI: six moves, a place for the migration, four worked examples, and a readiness checklist.

The taxonomy was the easy part. That is not a failure.

Most large organizations stand up an enterprise taxonomy in a few months. Then the program meets three new demands at once: a richer model, a way to govern change, and AI systems that read the model directly. That is where timelines slip.

A taxonomy answers one question: **where does this belong?** It is a hierarchy of broader and narrower terms. A small team can build a good one from spreadsheets and existing category trees. The tooling is mature, the standard (SKOS) is simple, and the output is easy for a business reviewer to read.

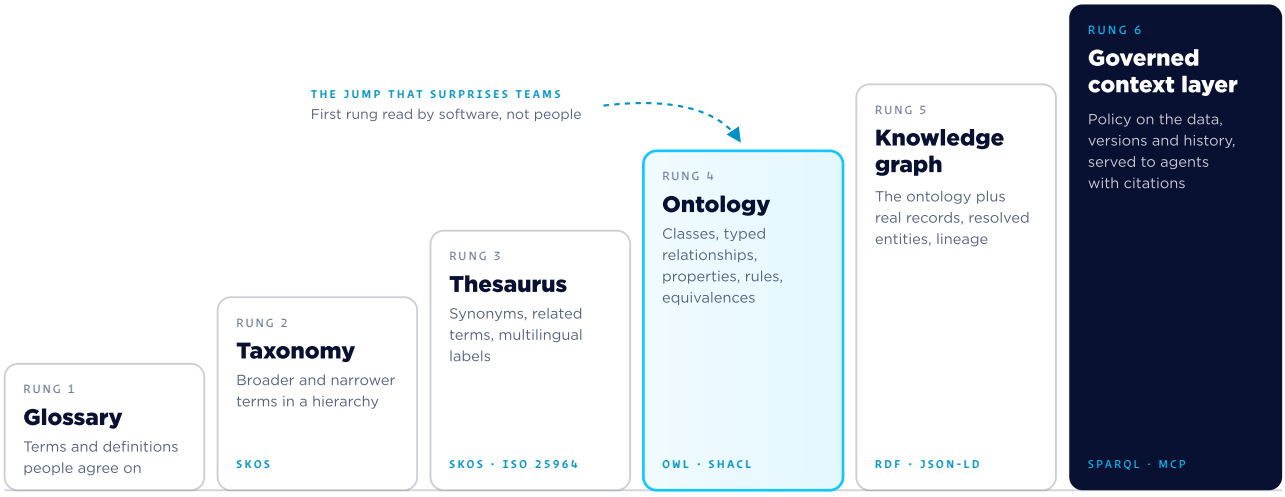
An ontology answers a harder question: **what is this, how does it relate to everything else, and which rules apply?** It adds classes, typed relationships, properties, constraints, and equivalences. It has to agree with how the whole business operates, not with how one department files things. And once AI agents read it, every ambiguity in the model becomes an ambiguity in an answer.

The good news: nothing built for the taxonomy is wasted. Concepts, preferred labels, hierarchies, and synonyms become the vocabulary layer of the ontology. What changes is the discipline around them: stable identifiers, formal relationships, validation, versioning, ownership, and a publishing path into the systems and agents that rely on the model.

This guide lays out that discipline in six moves, shows where a data migration fits into it, gives four worked examples, and closes with a readiness checklist you can run with your own team in an afternoon.

The ladder, and what each rung adds

Each step keeps everything below it. The jump that surprises teams is the one from thesaurus to ontology, because it is the first rung that software, not people, consumes.



Three things change when ontology, governance, and AI arrive together.

Each is manageable alone. Programs stall when all three land in the same quarter and the taxonomy team is expected to absorb them with the same tools and operating model.

1

The model becomes an operating contract

In a taxonomy program the vocabulary is documentation: people read it and tag against it. In an ontology program, software reads it: forms, APIs, validation rules, pipelines, and agent instructions derive from the model. A change to the model is a change to production behavior, so it needs the care of a code release.

2

Change needs a release process, not an admin task

Taxonomies change slowly and mostly by addition. Ontologies change by refactoring: a relationship gets retyped, two concepts merge, a constraint tightens. Without candidate workflows, versions, approvals, audit trails, and diff reports, the model drifts, and nobody can say what the business meant last quarter.

3

AI consumes the model directly, at machine speed

A human tagger notices a missing term and works around it. An agent does not. It answers with whatever the model allows, thousands of times a day, and the error compounds. Governance moves from a monthly committee to controls that run at query time: policy on the data, validation on every write, and a record of what the model said when.

What the analysts are saying

Two Gartner statements, fifteen months apart, bracket the problem: first that AI projects fail without AI-ready data, then that AI-ready data is a semantics problem.

> **“Gartner predicts that through 2026, organizations will abandon 60% of AI projects unsupported by AI-ready data.”**
>
> GARTNER NEWSROOM · Q&A WITH ROXANE EDJLALI · FEBRUARY 26, 2025 · SOURCE 1

> **“Proving AI readiness of the data is a process and a practice based on the availability of metadata to align, qualify and govern the data.”**
>
> ROXANE EDJLALI, SENIOR DIRECTOR ANALYST, GARTNER · FEBRUARY 2025 · SOURCE 1

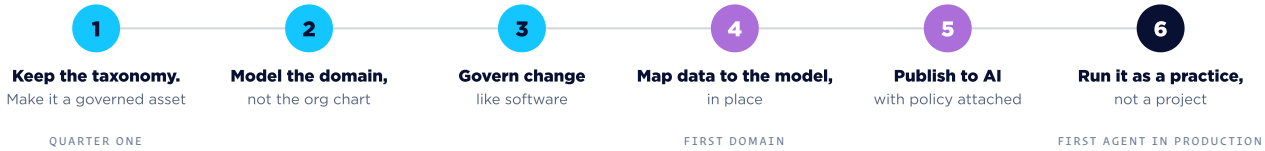
> **“Without governance, the context graph becomes a write-only log.”**
>
> CHARLES BETZ, FORRESTER · APRIL 10, 2026 · SOURCE 5

> **“Context with semantic coherence will become a cost-control and trust strategy, not a nice-to-have.”**
>
> RITA SALLAM, DISTINGUISHED VP ANALYST, GARTNER · DATA & ANALYTICS SUMMIT, LONDON, MAY 11, 2026 · SOURCE 2

Gartner’s May 2026 prediction: by 2027, organizations that prioritize semantics in AI-ready data will increase agentic AI accuracy by up to 80% and reduce costs by up to 60%. The same release advises data and analytics leaders to establish a context layer as a core component of their infrastructure, and expects regulators to demand greater semantic transparency (source 2).

Six moves from taxonomy to governed ontology.

The order matters less than the discipline. Most teams run moves 1 to 3 in the first quarter, prove move 4 on one domain, and add moves 5 and 6 once the first agent is in front of real users.



1 Keep the taxonomy. Make it a governed asset.

Import what exists (spreadsheets, category trees, glossaries) into a SKOS vocabulary with a stable identifier for every concept, preferred and alternative labels, and language tags for every market you serve, so one concept carries its Dutch, French, and English labels instead of three concepts. Record who owns each branch. Resist the urge to clean up the hierarchy first. The point of this move is to get the vocabulary under version control so that every later change is visible and reversible. If the glossary already lives in a data catalog, keep it there: the catalog inventories the data assets, the ontology defines what they mean, and the two link by identifier.

WHAT GOOD LOOKS LIKE

Every concept has one identifier that never changes, even when it is renamed. Synonyms are alternative labels, not duplicate concepts. Imports are repeatable jobs rather than one-off migrations, so the next spreadsheet lands the same way as the first.

2 Model the domain, not the org chart.

Start from competency questions: the 20 to 30 questions the business needs answered that the taxonomy cannot answer today. “Which category is this product in?” is a taxonomy question. “Which live promotions include products from a supplier that sent a force majeure notice?” is an ontology question. Model only what those questions need. Borrow before you build: an upper ontology such as gist for general business concepts, the GS1 Web Vocabulary for product, party, and location in retail and consumer goods, schema.org for anything that touches the public web. Use AI-assisted discovery to draft classes and relationships from existing schemas and data dictionaries, then let domain experts refine the draft.

WHAT GOOD LOOKS LIKE

Every class and relationship traces back to a competency question. The model reuses public vocabularies wherever they exist. Equivalences (“account” in the CRM, “client” in the ERP, “customer” in billing) are explicit, scored mappings rather than tribal knowledge.

3 Govern change like software.

Treat the ontology as a versioned release. Run two workflows: maintenance for internal edits and candidate for proposals from outside the core team. Keep the stable and the modified version side by side so reviewers can compare them, approve or reject in bulk, and nothing becomes official without an audit trail that records the actor, the timestamp, and the before and after. Validate at entry: required fields, naming conventions, cardinality, data types, and pattern rules enforced before a non-compliant change can save. Sweep the whole vocabulary on a schedule for orphan concepts, circular references, and missing language tags. Express data quality rules as SHACL shapes so the rules travel with the model into every system that loads it.

WHAT GOOD LOOKS LIKE

A diff between any two versions of the model takes minutes, not a meeting. A business steward, not a developer, approves changes. Every downstream system can say which version of the model it is running, and the model is listed as an asset, with an owner, in the data governance register.

GOVERNANCE CONTROL	IN A TAXONOMY PROGRAM	IN A GOVERNED ONTOLOGY PROGRAM
Change request	Email to the taxonomist	Candidate workflow with reviewer, status, and decision record
Versioning	Dated spreadsheet copies	Stable and modified versions side by side; diff report on demand
Validation	Manual review before publishing	Rules enforced on save; SHACL shapes shipped with the model
Audit	Change log, if kept	Every create, update, and delete tied to an actor and a timestamp
Release	Export and send around	Published to the graph, pipelines, and agents through one governed path

4 Map data to the model. Leave the data where it lives.

Classify structured sources (ERP, PIM, CRM, warehouse tables) against the ontology. Resolve duplicate entities into golden records that keep lineage back to every contributing source. Extract entities and relationships from documents, images, and audio against the same model, so a contract clause and a product record land in one graph. Keep exceptions in a review queue for subject matter experts and let the models learn from their decisions. Measure progress with the competency questions from move 2, not with row counts. Moving the data is optional; agreeing on what it means is not.

WHAT GOOD LOOKS LIKE

One product, customer, supplier, or location has one identity across systems. Every golden record can show its sources. Coverage against the competency questions is tracked as a number, and the number goes up every sprint.

5 Publish to AI with policy attached.

Expose the graph to agents and applications through standard interfaces: SPARQL and JSON-LD for developers, REST for applications, the Model Context Protocol (MCP) for agents. Evaluate access policy at query time, per entity, relationship, and property, so an agent sees only the slice it is permitted to see, however it asks and whichever model it runs on. Version the data and the ontology together and keep the history queryable, so any answer can be reproduced as of the moment it was given. Require citations back to source in every answer, and reject answers that cannot produce them.

WHAT GOOD LOOKS LIKE

An agent's answer names its sources. A compliance reviewer can rerun last quarter's question against last quarter's model and data. Sensitive fields never reach an embedding store or a prompt, because the policy is enforced before retrieval rather than filtered afterwards.

6 Run it as a practice, not a project.

Give the model an owner on the business side and a release cadence. Track three numbers: answer accuracy on the competency questions, coverage of the questions the business is asking this month, and the time from a vocabulary change to its availability in production. Add new sources, new markets, and new agents against the same model instead of starting over. The second use case should cost a fraction of the first; if it does not, the model is not being reused, and that is the problem to fix.

WHAT GOOD LOOKS LIKE

Quarterly model releases with release notes that business users can read. A steward community across editorial, product, compliance, and regional teams that shares one model with typed mappings between their vocabularies. An accuracy number that leadership recognizes and asks about.

The accuracy threshold the whole effort is aimed at

Fluree's April 2024 study, *GraphRAG for GenAI Accuracy*, measured the same enterprise questions against different data foundations. Retrieval over fragmented relational data plateaued at roughly 80%; retrieval grounded in a governed knowledge graph reached 95% and above, with every answer traceable to source (source 7). That gap separates an assistant people double-check from one they act on.

~80%
GENERIC RAG ON FRAGMENTED DATA

95%+
GRAPH-GROUNDED, WITH CITATIONS

Three ways the program stalls, and the move that unsticks it

1 · Modeling everything before answering one question

Months of workshops and no agent in production. Competency questions (move 2) bound the first release and give every class a reason to exist.

2 · Governing by committee instead of by workflow

A monthly board reviews slides while pipelines run last quarter's export. Workflows and versions (move 3) let stewards approve in hours.

3 · Letting the target system become the model

The new platform's schema becomes the ontology by default. Mapping to your own model first (move 4) keeps the platform a consumer, not the definition.

The migration you have to do anyway can fund the ontology you need next.

A data migration is a forced inventory. Every source gets profiled, every field gets mapped, and every duplicate gets found. The only question is whether that knowledge is kept or thrown away when the cutover is done.

Most programs spend the mapping effort on source fields to target schema, field by field. It works, once. The mapping lives in a migration tool or a spreadsheet, the duplicates are resolved by hand, and when the next system arrives, or the first AI agent asks what a “customer” is, the work starts again.

The alternative costs little more and returns far more. Map the sources to the semantic model first, then generate the target mapping from the model. The migration lands on time, and the enterprise keeps three things it did not have before: a governed ontology of the domain, resolved entities with lineage back to every source, and a set of mappings that exist as data rather than as project artifacts.

Four rules for a migration that leaves an ontology behind

Profile before you map. Let the classification run across every source before anyone opens a mapping sheet; the duplicate rate is usually the first surprise.

Resolve entities before you load. The target system should receive one supplier, one product, one location, with lineage attached.

Keep the mapping as data. Source-to-concept mappings belong in the vocabulary, versioned and governed, not in a tool that is switched off at cutover.

Do not let the target vendor’s data model become your enterprise model. Their schema is one consumer of your ontology, not its definition.

60 days

INSTEAD OF A TYPICAL 18 MONTHS

In a Global 500 manufacturer’s SAP S/4HANA program, semantic mapping compressed the legacy data migration from the usual 18-month plan to 60 days, as documented in Fluree’s 2023 e-book on the project (source 9).

48%

OF “UNIQUE” PRODUCT RECORDS WERE DUPLICATES

A Global 500 consumer goods company found the duplicate rate only when five product sources were classified against one model. The project left behind a governed product ontology the company still runs on (source 8).

What this looks like when it works.

Four published Fluree engagements, chosen because each one turned a taxonomy or category framework into a governed model that outlived the project.

CONSUMER GOODS · PRODUCT CATALOG

An ontology, not a rulebook

97.5%	48%	3 wks
CATALOG ACCURACY	DUPLICATES FOUND	TO PRODUCTION

A Global 500 consumer goods company with more than 20 billion-dollar brands needed one trusted catalog across five retailer and third-party sources. Rather than hard-coding matching rules, the master category framework was formalized into a working product ontology: classes, attributes, and equivalence rules that every classification decision referenced. Missing attributes such as GTIN-14, packaging, and flavor were inferred. The durable asset is a governed, versioned vocabulary that new sources and new agents classify against instead of starting over.

flur.ee/case-studies/cpg-ai-ready-product-catalog

FINANCIAL SERVICES · RESEARCH PORTAL

Ontology, taxonomies, and entities as one source of truth

500K	10x	Live
DOCUMENTS INDEXED	KNOWLEDGE BASE GROWTH	TOPIC DISCOVERY

A global financial institution's analysts produced nearly half a million documents and emails, but manual tagging made them hard to find. The enterprise ontology, topical taxonomies, and named-entity datasets were unified into one reference graph. AI tagging with inference, disambiguation, and relevance scoring was wired into the publishing chain, and continuous topic mining fed validated new concepts back into the model. Error-prone tagging was replaced, the knowledge base grew tenfold, and portal users' trust returned.

flur.ee/case-studies/financial-services-data-portal

GOVERNMENT · ENVIRONMENTAL SCIENCE

A golden ontology for 30 years of data

6M+	60+	30 yrs
SPECIES RECORDS	APPLICATIONS SERVED	OF HISTORY UNIFIED

A government department depended on a biodiversity knowledge base spread across Oracle, Access, and spreadsheets, plus decades of camera-trap images, audio, video, and scanned reports invisible to analytics. Model, in Fluree AI, became the governed home of the taxonomic models: harmonized terms, version control, and full audit history across every ontology change. Resolve mapped the structured legacy data into the model; Extract pulled species, locations, dates, and relationships from the unstructured evidence; GraphRAG gave researchers natural-language access with sensitive-species controls intact.

flur.ee/case-studies/government-biodiversity-ai-ready-platform

BEAUTY AND COSMETICS · EUROPE

Knowledge model stewardship as a job, not a phase

5	4	MCP
OPERATING FUNCTIONS	AGENT KPIS	MULTI-MODEL ACCESS

A global beauty brand wanted AI agents that could answer questions about ad spend, product positioning, and engagement across internal systems, agencies, and channels. Instead of point solutions it built an AI Agent Factory: a repeatable operating model with five functions, the first of which is knowledge model stewardship, the ontology work that gives every agent a shared understanding of the business. Agents are tuned against competency questions and measured on time to market, cited accuracy, timeliness, and role-based privacy. The second agent cost a fraction of the first.

flur.ee/case-studies/marketing-intelligence-sovereignty

Twelve questions to ask before the ontology program starts.

Answer honestly with the people who will do the work. A “no” is not a blocker; it tells you which move to start with.

MODEL

1. Does every concept in our taxonomy have a stable identifier that would survive a rename or a reorganization?
2. Can we write down the 20 business questions the taxonomy cannot answer today, in the words the business uses?
3. Which public vocabularies (GS1, schema.org, gist, our industry’s standards) will we reuse rather than reinvent?

GOVERN

4. Who on the business side owns each branch of the model, and who approves a change to it?
5. Can we produce a diff between two versions of the model, with the actor and timestamp of every change?
6. Are validation rules enforced when a change is saved, or discovered weeks later by the people who consume the model?

CONNECT

7. Do we know which systems and pipelines consume the model today, and which version each of them is running?
8. Can one product, customer, supplier, or location be resolved to one identity across systems, with lineage to each source?
9. Are documents, images, and media classified against the same model as structured data, or against a separate one?

AI

10. Is access policy evaluated at query time, wherever the question comes from, or reapplied by each application and agent?
11. Can we reproduce an agent’s answer as of the date it was given, with the model and data as they were then?
12. Does every answer cite its sources, and do we measure accuracy against the competency questions rather than by anecdote?

FEWER THAN SIX “YES”

Start with moves 1 to 3. Get the vocabulary under version control and pick the competency questions before touching pipelines or agents.

SIX TO NINE

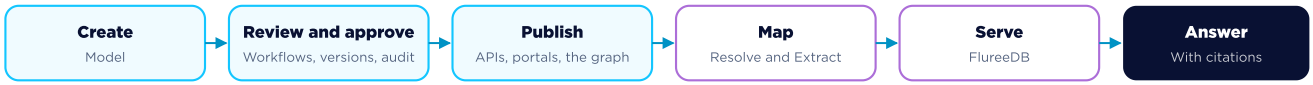
The model is in decent shape. Prove move 4 on one domain with a measurable duplicate rate and a coverage number, then expand.

TEN OR MORE

You are ready for moves 5 and 6: put an agent in front of real users with policy attached and citations required, and start the release cadence.

One governed workflow, from vocabulary to cited answer.

Fluree AI covers the six moves with three capabilities that share one model. Every change is created in a governed workspace, reviewed and approved, then published to the knowledge graph and the agents that depend on it.



MOVES 1 TO 3 · MOVE 4 · MOVES 5 AND 6

MODEL AND GOVERN

Model

Taxonomies, ontologies, and controlled vocabularies in tree, whiteboard, and search-first views, built for taxonomists and domain experts rather than developers. Language-tagged labels and notes. SKOS, RDF, OWL, and ISO 25964, with guided import from Excel, RDF, and SKOS and export to eight formats. Maintenance and candidate workflows, stable versus modified versions, lifecycle statuses, complete audit trails, and Excel diff reports between any two versions. Validation at entry and scheduled reasoning sweeps for orphans, cycles, and missing language tags. Its alignment tools create typed, scored mappings between the vocabularies different teams already use.

MAP

Resolve and Extract

Resolve classifies structured data from ERPs, PIMs, CRMs, and warehouses against the published model, resolves duplicates into golden records, infers missing attributes, and keeps lineage to every source, with exception queues for subject matter experts. Extract pulls entities and relationships from PDFs, images, audio, and video against the same model, so unstructured evidence lands in the same graph as the records it describes. Both publish knowledge-graph-ready data as JSON-LD.

SERVE

FlureeDB

A W3C-native knowledge graph database: RDF and JSON-LD throughout, queried with SPARQL, JSON-LD, or Cypher. Access policy is evaluated at query time per entity, relationship, and property, so agents inherit permissions with no application-side filtering. Every transaction is an immutable commit: query any past state, branch off any commit, and merge back. SHACL validation at transaction time, OWL and RDFS inference and Datalog rules in the query engine, and MCP for agents. GraphRAG answers arrive with citations to source.

Time to value. With AI-assisted ontology discovery, automated classification, and entity resolution, most teams move from fragmented schemas to a working governed knowledge graph in four to eight weeks, starting with one domain and expanding from there.

Gartner COOL VENDOR

Fluree was named a **2024 Gartner Cool Vendor in Data Management** in the report *Cool Vendors in Data Management: GenAI Disrupts Traditional Technologies* (May 30, 2024). See source 3 for the report citation and Gartner's disclaimer.

Go deeper.

Everything below is public on flur.ee. The first three take fifteen minutes and cover the vocabulary; the rest are for the team that will do the work.

CONCEPTS

Taxonomies Versus Ontologies: A Short Guide
The distinction in one page, with examples of each.
flur.ee/blog/taxonomies-versus-ontologies-a-short-guide

Controlling LLMs with Enterprise Taxonomies
Five non-negotiables for taxonomy systems that feed AI.
flur.ee/blog/controlling-llms-with-enterprise-taxonomies

Model, Map, Connect
The framework behind the six moves in this guide.
flur.ee/blog/model-map-connect

What is SHACL?
Validation rules as a W3C standard, and why they belong with the model.
flur.ee/blog/what-is-shacl

How to Build a Semantic Layer for Enterprise AI
The five components of a semantic layer and a three-stage build.
flur.ee/blog/how-to-build-a-semantic-layer-for-enterprise-ai

REPORTS, GUIDES, AND REPLAYS

So You Think You're Ready for a Knowledge Graph?
A readiness whitepaper on data quality, interoperability, and governance.
flur.ee/resources/enterprise-knowledge-graph-readiness

GraphRAG for GenAI Accuracy
The April 2024 study behind the 80% and 95% numbers.
flur.ee/resources/graphrag-genai-accuracy-report

Migrating Legacy Data to SAP S/4HANA in 60 Days
The e-book on semantic mapping as a migration method.
flur.ee/resources/sap-s4hana-migration-60-days

A Practical Approach to Implementing Semantic Data Layers
Webinar replay with Enterprise Knowledge: drivers, barriers, and case studies.
flur.ee/resources/videos/practical-approach-to-semantic-data-layers-webinar

Model: Taxonomy and Ontology Management in Fluree AI
Product page, with the governance features referenced in move 3.
flur.ee/platform/model

Twelve terms, defined the way this guide uses them

Taxonomy	Concepts in a hierarchy of broader and narrower terms. Answers "where does this belong?"
Thesaurus	A taxonomy plus synonyms, related terms, and multilingual labels (SKOS, ISO 25964).
Ontology	A formal model of classes, typed relationships, properties, and rules. Answers "what is this, and how does it relate?"
Knowledge graph	An ontology populated with real records, resolved entities, and lineage, queryable as a graph.
Context layer	A knowledge graph with policy, versions, and history, served to applications and agents.
Competency question	A business question the model must answer. The unit of scope for a release.
SKOS	W3C standard for taxonomies and thesauri: concepts, labels, broader, narrower, related.
OWL	W3C ontology language for classes, properties, and inference.
SHACL	W3C language for validation rules on graph data: required properties, types, cardinality, patterns.
Golden record	One resolved record per real-world entity, with lineage to every contributing source.
Policy at query time	Access rules stored with the data and evaluated on every query, before retrieval.
MCP	Model Context Protocol: the open standard agents use to call tools and data sources.

Sources and notes.

1. Gartner, “Lack of AI-Ready Data Puts AI Projects at Risk,” Q&A with Roxane Edjlali, Senior Director Analyst, Gartner newsroom, February 26, 2025. gartner.com/en/newsroom/press-releases/2025-02-26-lack-of-ai-ready-data-puts-ai-projects-at-risk
2. Gartner, “Gartner Says Lack of Semantics Causes Inaccurate AI Agents and Wasted Spending,” press release, London, May 11, 2026, quoting Rita Sallam, Distinguished VP Analyst. gartner.com/en/newsroom/press-releases/2026-05-11-gartner-says-lack-of-semantics-causes-inaccurate-artificial-intelligence-agents-and-wasted-spending
3. Gartner, “Cool Vendors in Data Management: GenAI Disrupts Traditional Technologies,” Nina Showell, Ehtisham Zaidi, and Aaron Rosenbaum, May 30, 2024. Fluree announcement: flur.ee/blog/fluree-named-a-2024-gartner-cool-vendor-in-data-management. GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally, and Cool Vendors is a registered trademark of Gartner, Inc. and/or its affiliates and is used herein with permission. Gartner does not endorse any vendor, product or service depicted in its research publications and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner’s research organization and should not be construed as statements of fact. Gartner disclaims all warranties, express or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.
4. Gartner, “Emerging Tech Impact Radar: Generative AI,” 2024, in which knowledge graphs are designated a critical enabler of generative AI. gartner.com/en/documents/6179023 (Gartner client access)
5. Charles Betz, Forrester, “Context Graphs Are A Convergence, Not An Invention,” April 10, 2026. forrester.com/blogs/context-graphs-are-a-convergence-not-an-invention/
6. Cloudera and Harvard Business Review Analytic Services, March 5, 2026: only 7% of enterprises say their data is completely ready for AI. cloudera.com/about/news-and-blogs/press-releases/2026-03-05-only-7-percent-of-enterprises-say-their-data-is-completely-ready-for-ai-according-to-new-report-from-cloudera-and-harvard-business-review-analytic-services-reveals.html
7. Fluree, *GraphRAG for GenAI Accuracy*, April 2024. Accuracy figures are averages from the study’s benchmark of enterprise questions across data foundations. flur.ee/resources/graphrag-genai-accuracy-report
8. Fluree customer case studies: consumer goods product catalog, financial services data portal, government biodiversity platform, and marketing intelligence AI Agent Factory. flur.ee/case-studies
9. Fluree, *Migrating Legacy Data to SAP S/4HANA in 60 Days*, e-book, 2023. flur.ee/resources/sap-s4hana-migration-60-days
10. Fluree AI Model product page and FlureeDB documentation (policy, time travel, branching, SHACL, inference, MCP). flur.ee/platform/model · labs.flur.ee/docs/db
11. Standards referenced: SKOS and ISO 25964 (thesauri), OWL (ontologies), SHACL (validation), RDF and JSON-LD (representation), SPARQL (query), GS1 Web Vocabulary (retail and consumer goods), gist (upper ontology, Semantic Arts), schema.org.

NEXT STEP

Run the twelve questions with your team, then bring the answers to a working session.

Fluree will map your answers to the six moves and scope a first domain, including where your current migration fits. flur.ee/contact