

June 10–13, 2025
Data Week Leipzig

Presented by
Márcia R. Ferreira
Senior Data and Knowledge Engineer

Graph-driven RAG

When the Whole Is Greater Than the Chunks of Its Parts



Graphs are Everywhere



Knowledge Graphs are Everywhere



Google Knowledge Graph

From Wikipedia, the free encyclopedia

"Knowledge Graph" redirects here. For the general concept in information science, see [Knowledge graph](#). For other uses, see [Knowledge Graph \(disambiguation\)](#).

The Google Knowledge Graph is a knowledge base from which Google serves relevant information in an infobox beside its search results. This allows the user to see the answer in a glance, as an instant answer. The data is generated automatically from a variety of sources, covering places, people, businesses, and more.^{[1][2]}

The information covered by Google's Knowledge Graph grew quickly after launch, tripling its data size within seven months (covering 570 million entities and 18 billion facts^[3]). By mid-2016, Google reported that it held 70 billion facts^[4] and answered "roughly one-third" of the 100 billion monthly searches they handled. By March 2023, this had grown to 800 billion facts on 8 billion entities.^[5]

There is no official documentation of how the Google Knowledge Graph is implemented.^[6] According to Google, its information is retrieved from many sources, including the *CIA World Factbook* and *Wikipedia*.^[7] It is used to answer direct spoken questions in Google Assistant^{[8][9]} and Google Home voice queries.^[10] It has been criticized for providing answers with neither source attribution nor

Read Edit View history Tools ▾

Thomas Jefferson
3rd U.S. President

Thomas Jefferson was an American Founding Father, the principal author of the Declaration of Independence, and the first President of the University of Virginia.

Born April 13, 1743, Shadwell, VA; Died July 4, 1826, Charlottesville, VA; President of the United States (1801–1809); Vice President (1793–1801); Party: Democratic-Republican Party; Awards: Allard Pierson Medal

Get quotes about Thomas Jefferson

People also search for

View 100 more

Public Knowledge Graphs



Building The LinkedIn Knowledge Graph

Qi He October 6, 2016

Share Tweet Share



From Graphs to Knowledge Graphs

Plain Graph

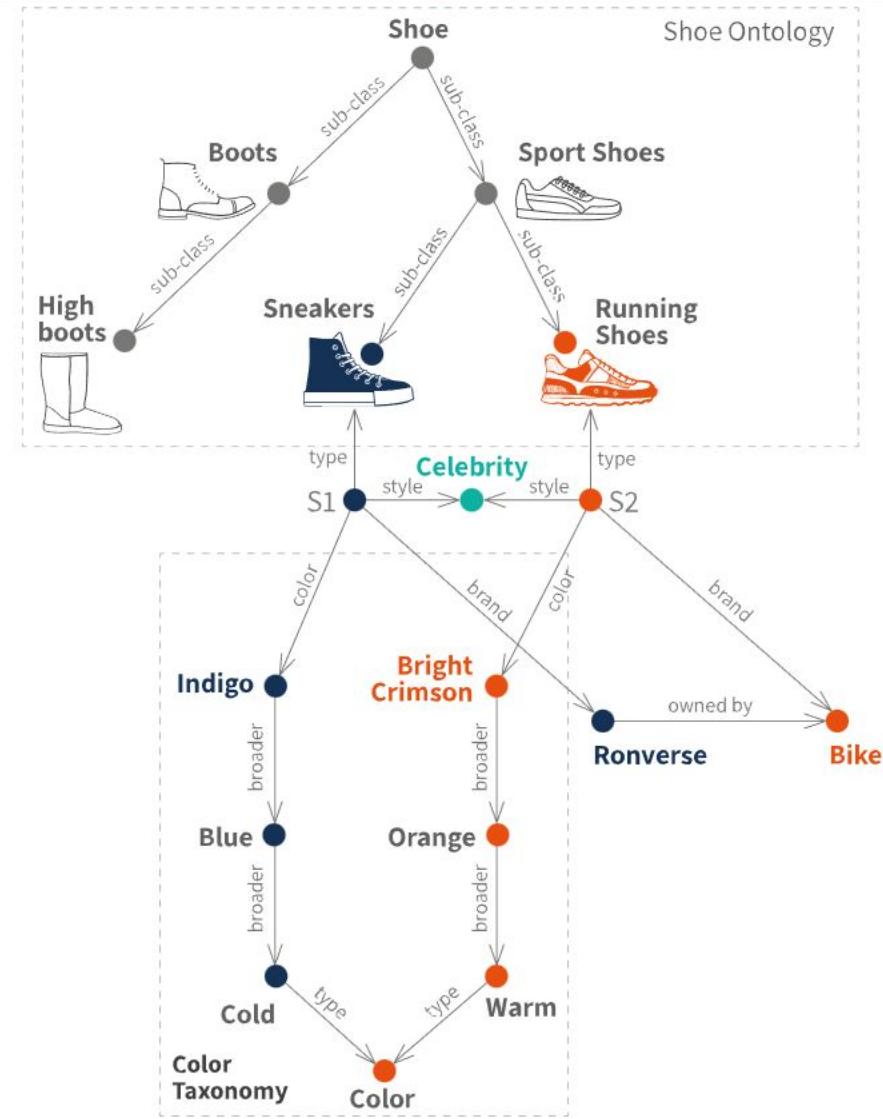


From Graphs to Knowledge Graphs

Plain Graph



Knowledge Graph

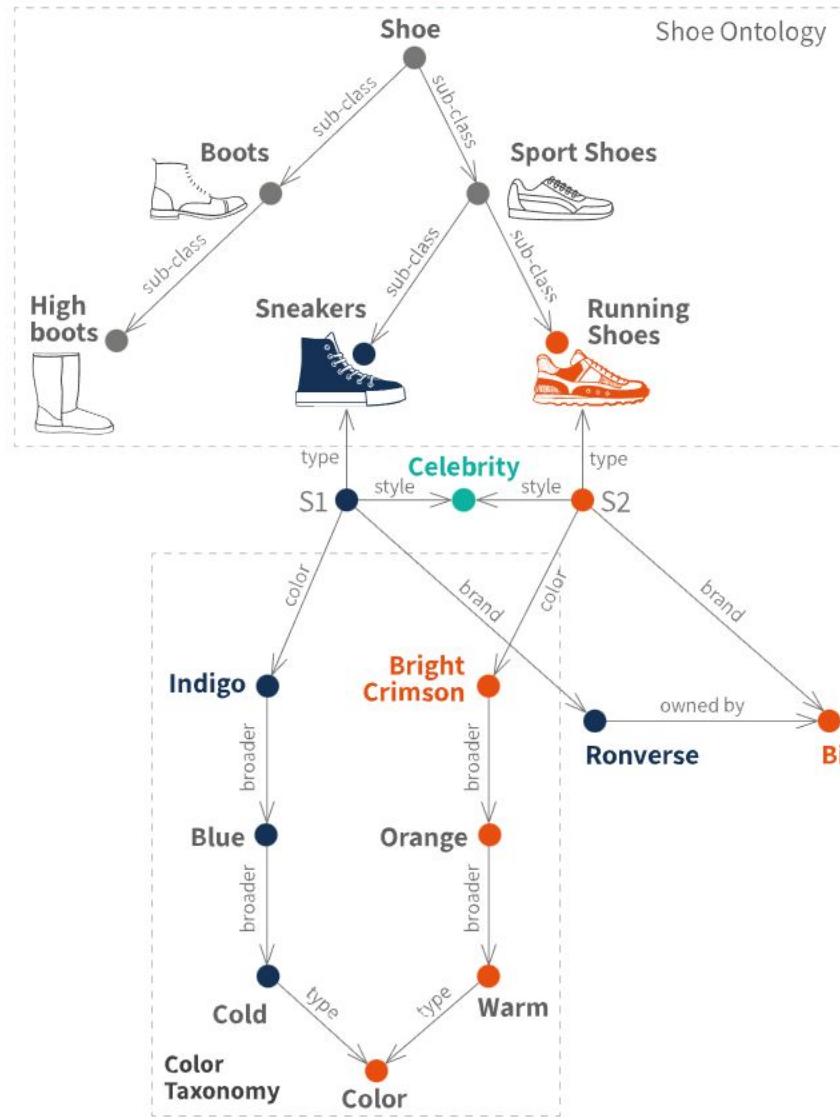


From Graphs to Knowledge Graphs

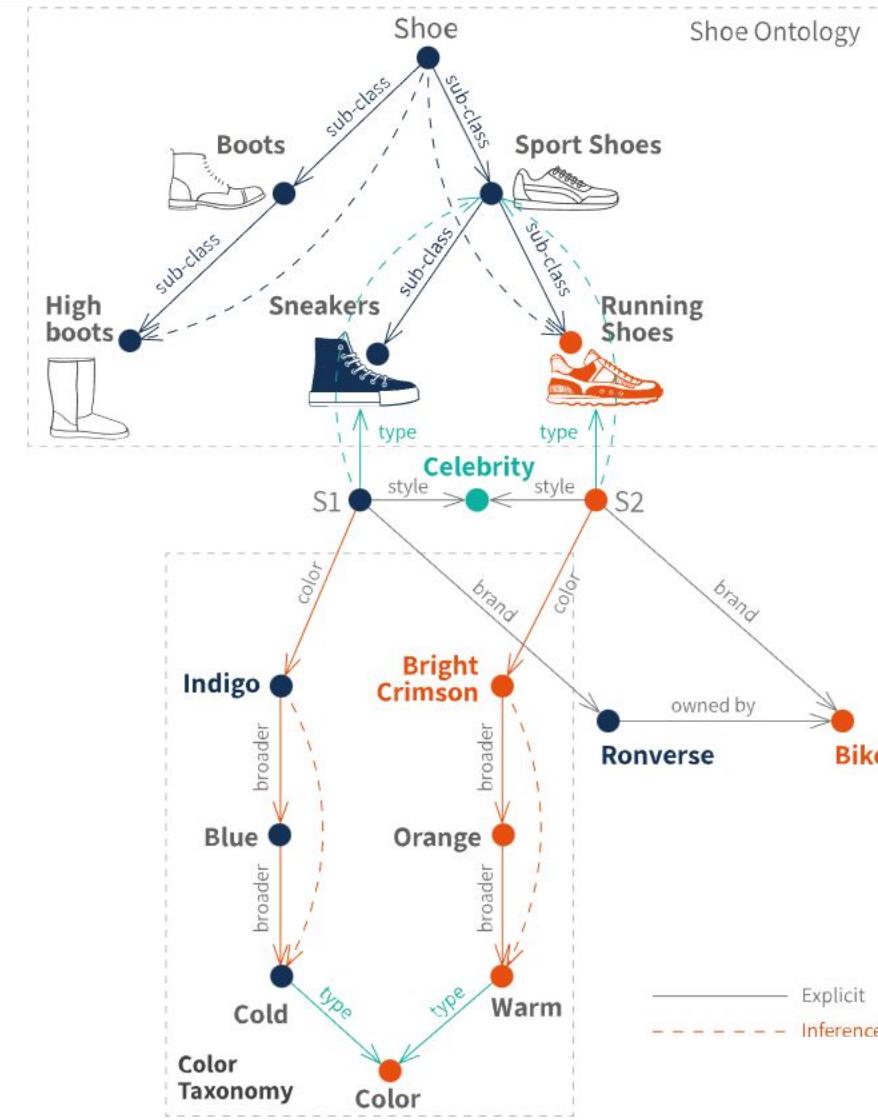
Plain Graph



Knowledge Graph



Knowledge Graph with Inference

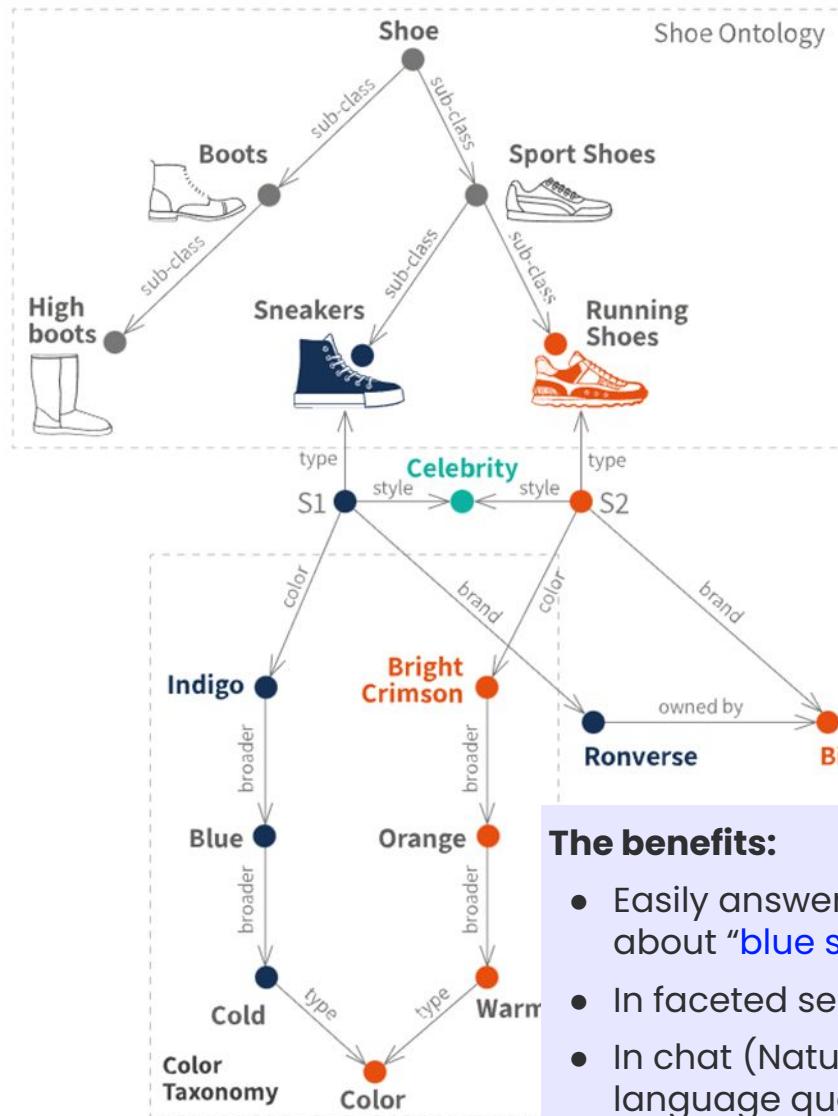


How Domain Knowledge and Inference Help?

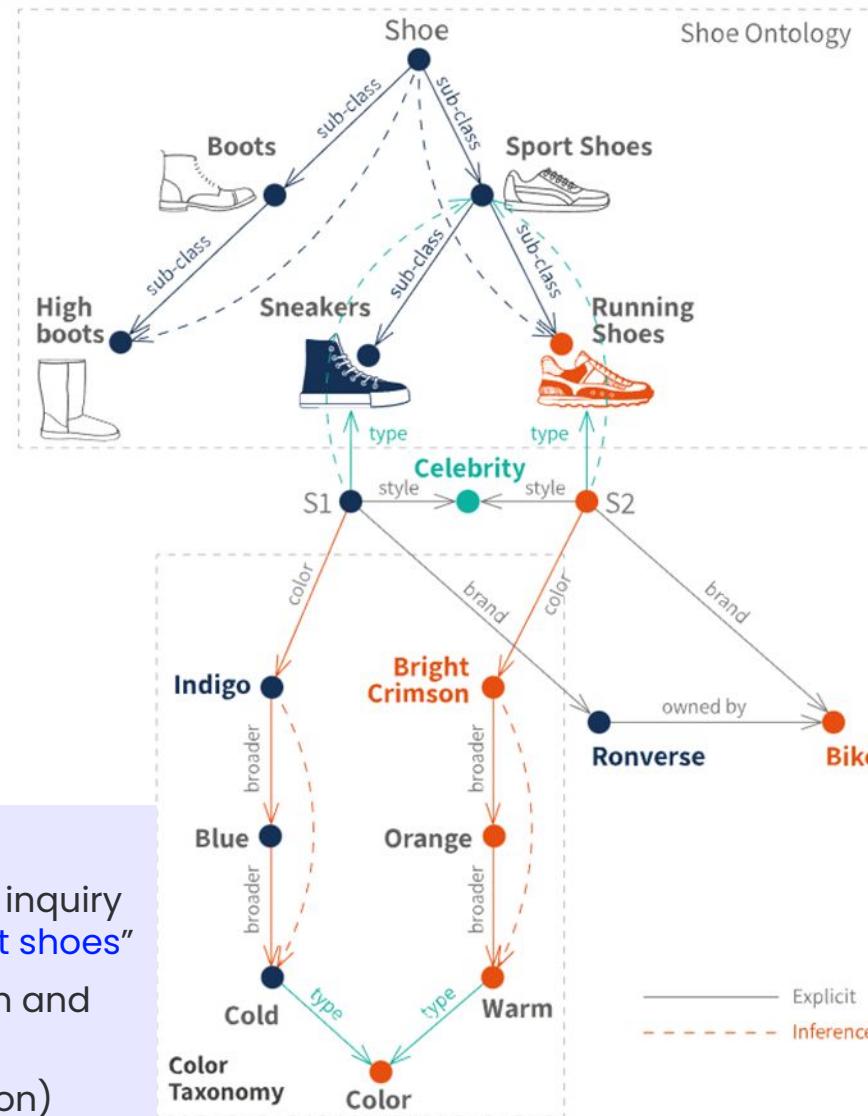
Plain Graph



Knowledge Graph



Knowledge Graph with Inference



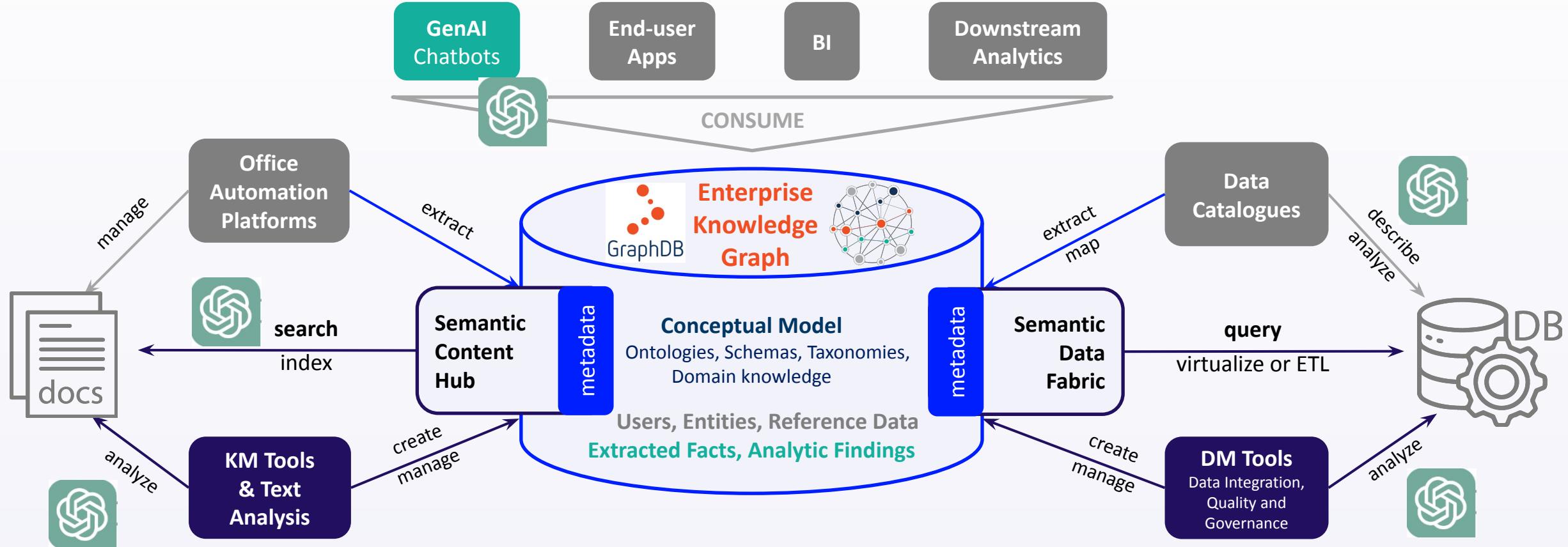
The benefits:

- Easily answer an inquiry about “blue sport shoes”
- In faceted search and
- In chat (Natural language question)

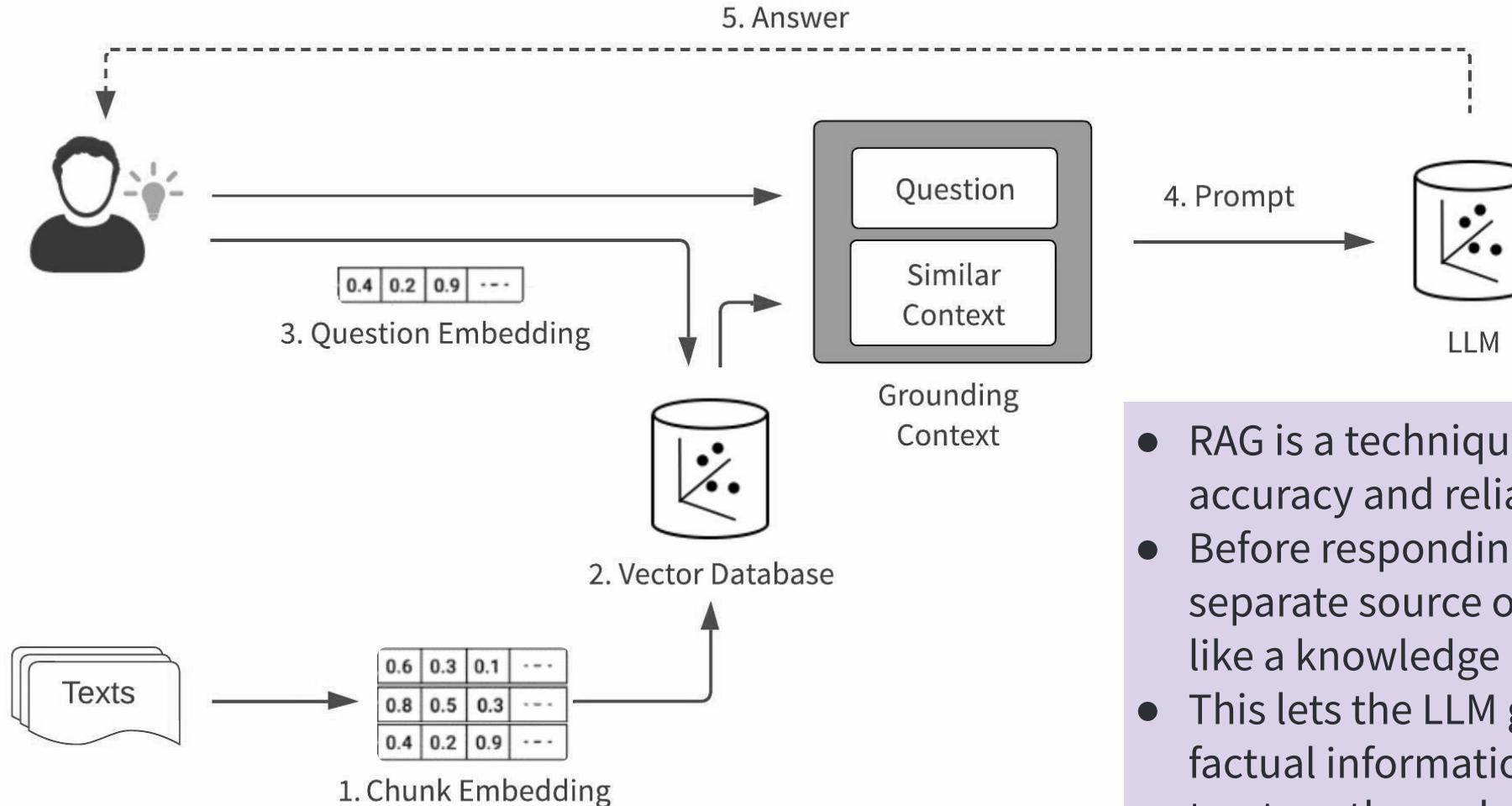
Every Graph is NOT a Knowledge Graph

- Graph of a LAN network is not a KG if there is no schema and semantics
- KGs are NOT about searching paths between nodes
- KGs are about putting data in context via linking and semantic metadata
- KGs are structured around interlinked metadata of entities and concepts
- Descriptions have formal semantics lets human & machines process
unambiguously
- Includes Rules with Inferencing, Reasoning from Relationships
- Entity Resolution and Extraction to deal with messy data

Knowledge Graph & LLMs in the Enterprise

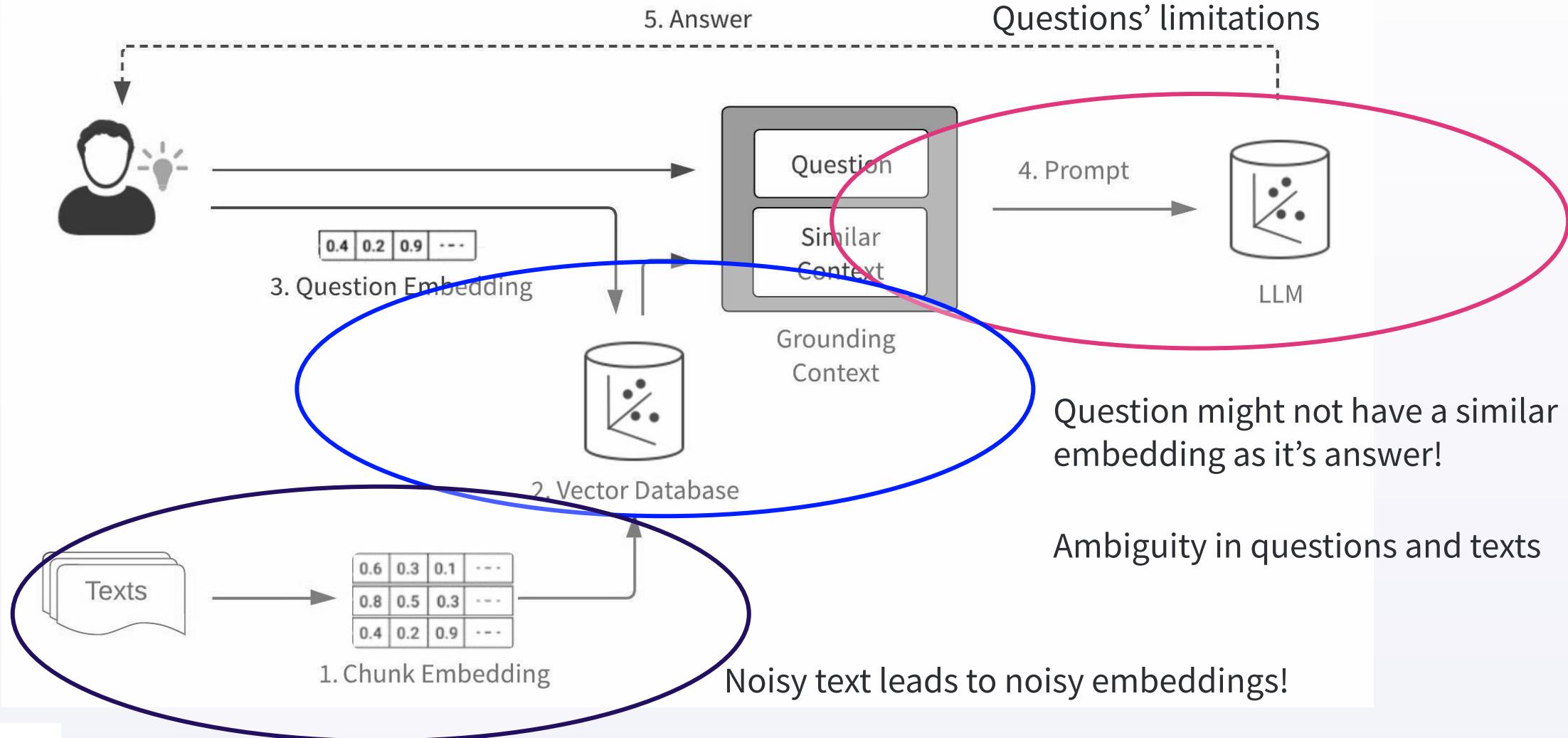


RAG is Everywhere Lately



- RAG is a technique used to improve the accuracy and reliability of LLMs
- Before responding, the LLM consults a separate source of reliable information, like a knowledge base
- This lets the LLM ground its response in factual information, making it more trustworthy and up-to-date.

Vector-based RAG in a Nutshell



Breaking up is hard to do

General overview to chunking procedures

Chunking Method	How it works?	When to use?
No Chunking	Data is not chunked within Amazon Bedrock Knowledge Bases (KB).	Use when data is already optimally chunked outside of Amazon <u>Bedrock</u> and you plan to use that with Amazon Bedrock Knowledge Bases.
Standard Chunking (Default and Fixed Size)	Data is divided into chunks based on a fixed token size, with an overlap percentage between consecutive chunks to ensure	Ideal for documents with loose semantic connections between paragraphs and texts, such as FAQs, data reports, statistics.

... caused by treating texts as independent pieces of information...

	chunks are then used for context retrieval.	
Hierarchical Chunking	data is structured into nested parent-child chunks. Initially, child chunks are retrieved, but they can be replaced with broader parent chunks to provide a more comprehensive context.	Recommended for documents with clear hierarchies (headers, sections, subsections, paragraphs etc.), such as technical manuals, research papers, and legal contracts.
Custom Chunking	Customizable chunking using a Lambda function, allowing complete control over how data is divided into chunks.	Use when you require highly specific chunking logic tailored to unique use cases.

Common Challenges in RAG



- Cost-effectiveness of vector based-RAG solutions
- Data quality and relevance
- Handling dynamic knowledge and scalability
- Transparency of query results
- There is no contextualization in the vector
 - Words with multiple senses depending on context are squashed together in averages
 - Struggles to connect dots, for question requiring traversing disparate chunks of data
- Lack of explicit entity recognition
- Words with similar context close together
 - Can be a different topic → bad for IR
 - Unaware of the domain
 - Poor guardrailing

Graph RAG is a specialisation of Advanced RAG

Integrates **domain-specific knowledge models** and retrieves information into the generation process

- Enables more informed and contextually **grounded** answers
- Enhances **accuracy** and **provenance**
- Knowledge Graphs preserve the **context and relationships** in data
- **Graph RAG** grounds LLMs with semantically-rich data from Knowledge Graphs
 - The data we send to the LLM is **well structured** and **contextually-aware**
 - **Captures relationships of entities across chunks and document hierarchies**

How can Graph RAG help?



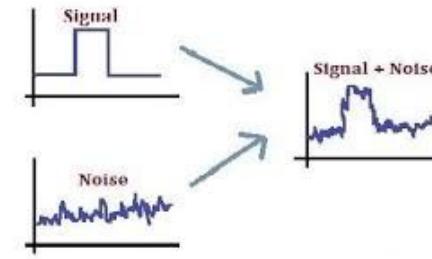
Structured Relationships

KGs facilitate the connection of terminology, such as linking 'physiotherapy' with 'coverage', thereby providing a more coherent IR (e.g., Does my insurance cover physio?)



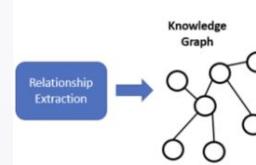
Enhanced Multi-hop Reasoning

KGs empower users to engage in more complex queries that require tracing through interconnected pieces of text
drug → pathway → side-effect
two hops, tidy evidence of chain



Reduction of Noise

By focusing on data connections, KGs decrease the likelihood of unrelated information being retrieved (guardrails)



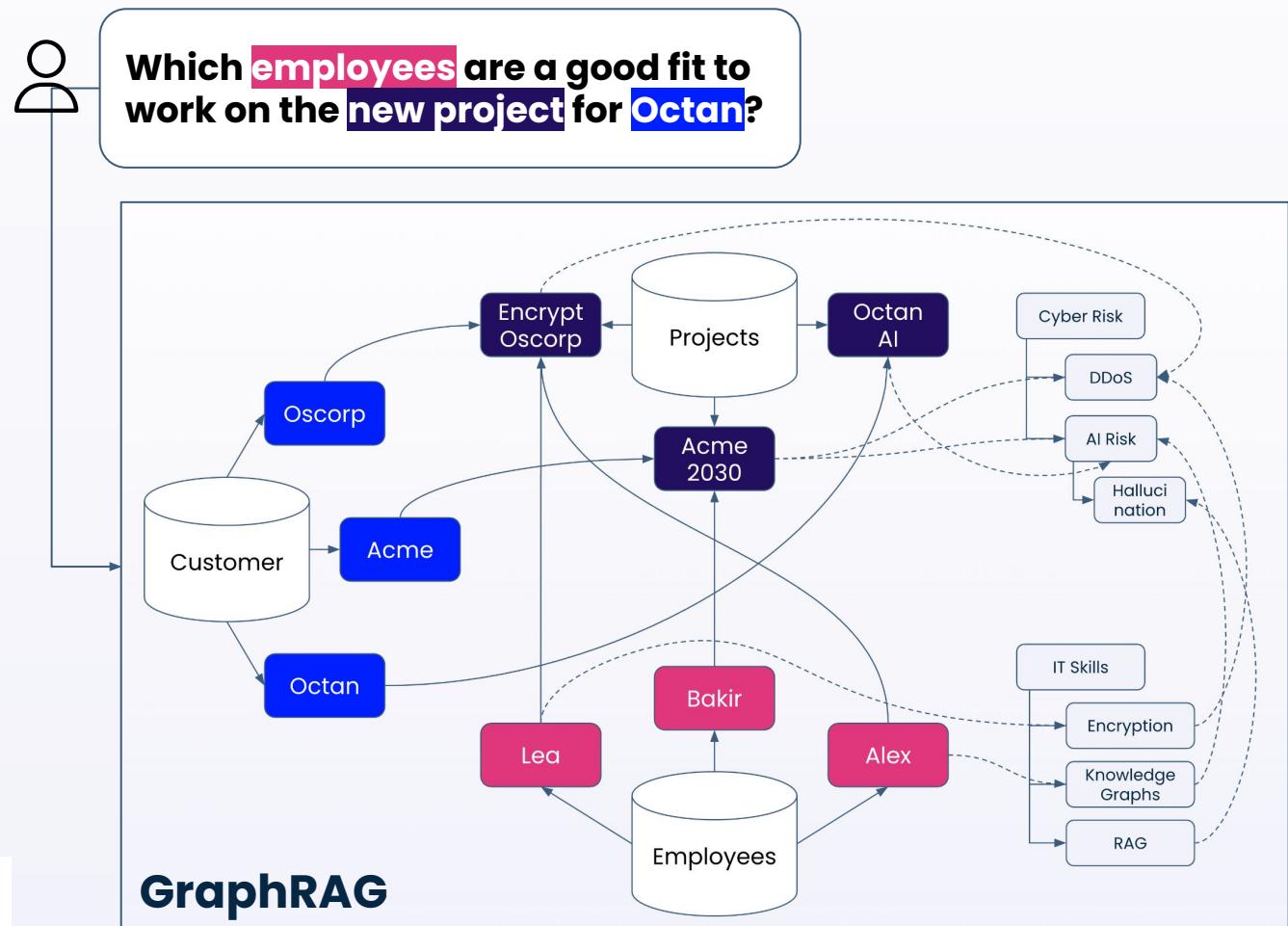
Clear Entity Recognition

KGs aid in distinguishing and disambiguating entities and terms – for example, differentiating various types of 'cancer'. "HER2" is also known as "ERBB2"

GraphRAG=LLM+Knowledge Graph



Now your LLM has a conversation with your company's **interconnected brain** (= Knowledge Graph) before it talks to you.



First hop: Octan is a customer; they've commissioned a new project

Second hop: the system spots Acme 2030—a past project with comparable requirements.

Third hop: Bakir delivered Acme 2030 successfully, so he's already familiar with the domain.

Bakir is a good fit because he has worked before on **Acme 2030** with similar requirements, and also **Alex** because of his specific knowledge graph skills.



The Graph AI Killer Apps



Application Pattern	Case Study and Customer	Data	Knowledge Model	Persona and Vertical
Knowledge Hub	Customer Support for Hydraulic Systems	Product documentation and troubleshooting KB	SKOS Taxonomy	Customer care, Manufacturing
Knowledge Hub	Digital Asset Management for Microsoft Learn	Software documentation, training materials, etc.	Mid-sized SKOS taxonomy	End user IT
Knowledge Hub	Drug Manufacturing Reporting Big Pharma company	Millions of documents related to drug production quality control	Combination on public and proprietary ontologies	Compliance officers and others Life Sciences
Compliance Agent	Regulatory RAG Major EU Bank	Compliance documents and inquiries. Interdependent structure and terminology	Small custom ontology	Risk & Compliance officers, Employees Financial Services
Self Service Data Agent	Talk to Electricity Grid STATTNET	Electricity grid digital twin – multiple models with more than 1B edges	Standard ontology with 900 classes and 5200 properties	Grid operators Energy Infrastructure

An actual scenario that we tested

Imagine a user, **Stefan**, who asks a specific question about a technical product like a **hydraulic system**.

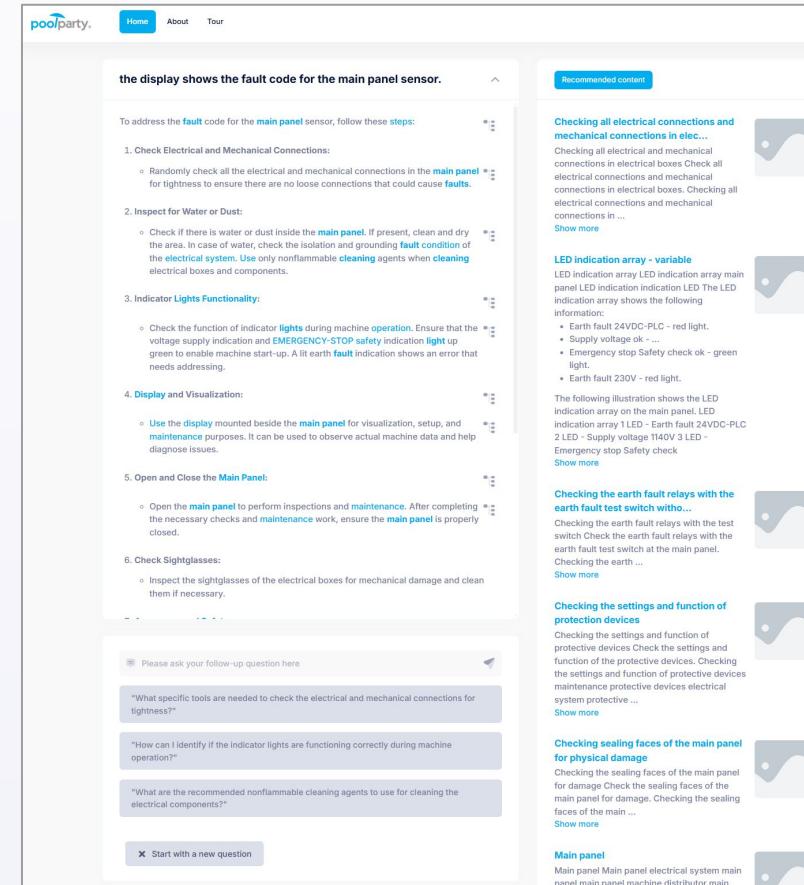


He asks a chatbot based on Vector RAG a question about a hydraulic system, but the **answers it gives are not good enough**. Stefan's intention is not properly understood because the subject-specific contextual **knowledge is missing** and the retrieval system of the underlying RAG system is too imprecise, so that the answer is of no help.

What if the answers to his questions actually show him what he is looking for?

Graphwise offers GraphRAG systems that preserve the content model and use precise retrieval based on domain-specific quality metadata.

This provides **answers that correctly interpret Stefan's questions**, peppered with details and references based on quality-assured knowledge graphs.

The display shows the fault code for the main panel sensor.

To address the fault code for the main panel sensor, follow these steps:

- Check Electrical and Mechanical Connections:
 - Randomly check all the electrical and mechanical connections in the main panel for tightness to ensure there are no loose connections that could cause faults.
- Inspect for Water or Dust:
 - Check if there is water or dust inside the main panel. If present, clean and dry the area. In case of water, check the isolation and grounding fault condition of the electrical system. Use only nonflammable cleaning agents when cleaning electrical boxes and components.
- Indicator Lights Functionality:
 - Check the function of indicator lights during machine operation. Ensure that the voltage supply indication and EMERGENCY-STOP safety indication light up green to enable machine start-up. A lit earth fault indication shows an error that needs addressing.
- Display and Visualization:
 - Use the Display mounted beside the main panel for visualization, setup, and maintenance purposes. It can be used to observe actual machine data and help diagnose issues.
- Open and Close the Main Panel:
 - Open the main panel to perform inspections and maintenance. After completing the necessary checks and maintenance work, ensure the main panel is properly closed.
- Check Sightglasses:
 - Inspect the sightglasses of the electrical boxes for mechanical damage and clean them if necessary.

Please ask your follow-up question here

"What specific tools are needed to check the electrical and mechanical connections for tightness?"

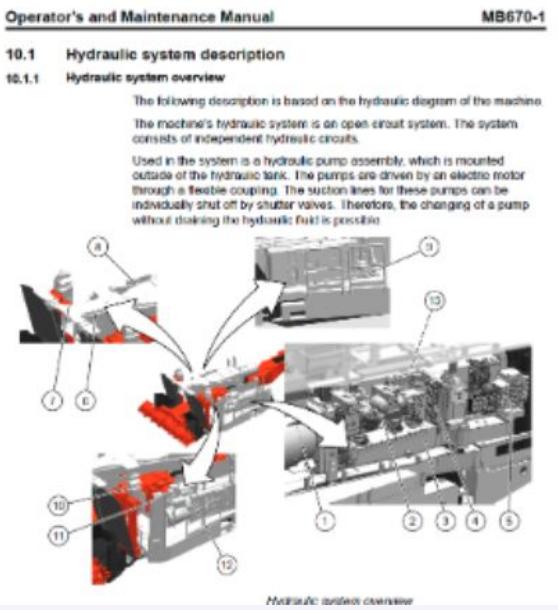
"How can I identify if the indicator lights are functioning correctly during machine operation?"

"What are the recommended nonflammable cleaning agents to use for cleaning the electrical components?"

Start with a new question

Figures speak for themselves

Based on 3 incidence categories and 4 associated questions we **evaluated the benefit of GraphRAG** against VectorRAG and pure LLM (GPT-4o mini).



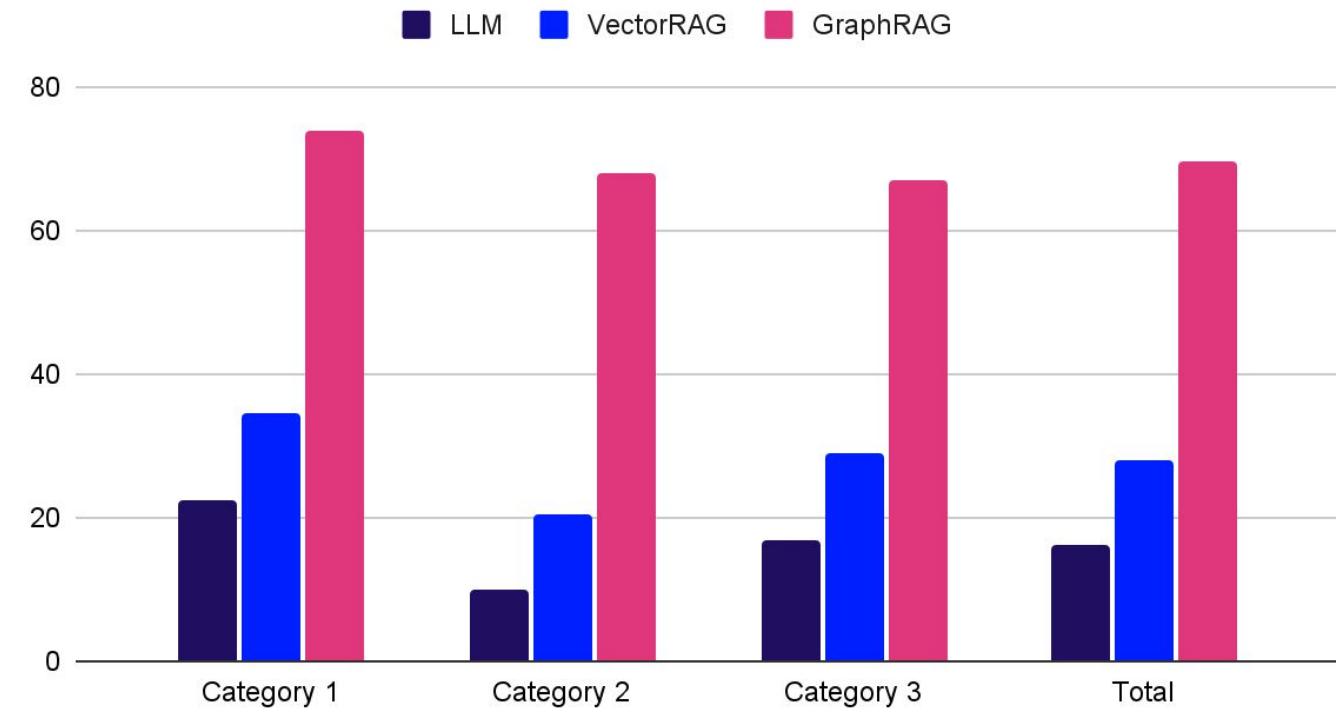
Incidence categories:

1. The level sensors show no reading
2. The hydraulic pressure is too low
3. The display shows the fault code for the main panel sensor

Associated questions:

- a. What are the causes?
- b. What needs to be checked?
- c. What needs to be done?
- d. What needs to be considered (hazards)?

% of correct answers



Agent: Compliance Advisor



Inquiry: I am seeking approval to do an evening part-time teaching opportunity at University Canada West. I work with the Finance Change delivery team and have access to the Bank's general ledger and Financials. This will be 2 hours on Saturday from 730pm to 930pm and a 2 hour online tutorial anytime during the week in the evenings and the course runs between [date redacted]. The course is on financial management. Please can you provide guidance on the above, taking into account that I work for Finance and plan to teach a Finance course. I also have access to the production Financial data.

Graph Inference

What are the applicable policies?

- Conflict of Interest from outside activity

Classification according the applicable policy

- 'Permitted activities'
- 'Activities Requiring Review and Approval'
- Prohibited activities

Role Context

What is the role, band, division, LoB?
Is the employee a Covered worker?

Graph Context

Jurisdiction Context
What is the applicable Jurisdiction?
Are there any country specific rules?

Graph Inference

What type of risks it poses?

- Conflict of interest from outside activity
- Reputational risk
- MNPI exposure risk

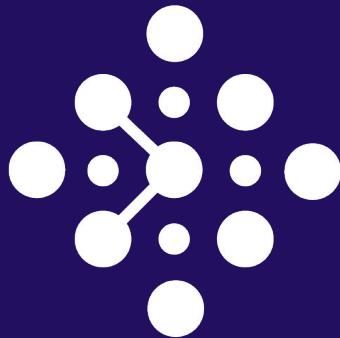
Advisory Contents

1. Applicable Policy segments identification
2. Statements of facts and assumptions
3. (Optional) Suggestions for clarification and elaboration
4. Analysis of facts
5. Prescriptive advisory

Graph RAG Lessons Learned

- 1. Enrichment with graph background context improves the quality**
- 2. A little knowledge modelling can save a lot of prompting, increase accuracy and reduce cost**
 - Be it a new class in an ontology, a SHACL rule or an axiom chain
 - It's always easier, simpler and more reliable than prompting the same semantics in NL
- 3. Different domains, applications and data require different Graph RAG**
 - Some ontologies are just too big for a prompt; GraphQL can fit better than SPARQL
 - Formal modelling of document structure is often necessary
- 4. Using public ontologies and schema helps a lot the interplay with LLMs**
- 5. Pre-LLM AI models and alternative retrieval methods are much faster and better for some RAG sub-tasks**

Thank you!



GRAPHWISE

www.graphwise.ai



Graphwise vs. Neo4j



	Graphwise	neo4j	Comparison
Data Model	RDF	LPG	LPG is only good for traversal; RDF supports governance and reuse
Data Normalization	○	○	Both platforms offer limited capabilities for data normalization
Alignment of Meaning	○○○	○	neo4j does not support semantic schemas and controlled vocabularies
Dynamic Schema	○○○	○○	neo4j lacks support for semantic schemas and data validation
Relationship Analysis	○○	○○○	neo4j's indices are better optimized for graph traversal
Unstructured Data & Metadata	○○○	○	neo4j has limited support for text analysis and doc. classification
Data Interoperability	○○○	○	LPG stack has no standardized exchange formats and protocols
Data Quality and Trust	○○○	○	neo4j has not data validation and tooling for quality monitoring of IE
Boost AI performance with Rich Context	○○○	○○	LPG doesn't help for reuse of domain knowledge or the use public schema
Versatile Performance	○○○○	○○	GraphDB is the only engine efficient for metadata mgmt. and analytics
Expert-in-the-Loop Tooling	○○○○	○	neo4j has nothing to match Graphwise PoolParty - the best KM toolsuite



GraphDB™ vs. Other Graph Databases



	Graphwise GraphDB™	neo4j	TigerGraph	AWS Neptune	Progress Software
Data Model	RDF	LPG	LPG	RDF/LPG	RDF
Data Normalization	○	○	○		○
Alignment of Meaning	○○○	○			○○
Dynamic Schema	○○○	○○	○○	○○○	○○○
Relationship Analysis	○○	○○○	○○○○	○○○	
Unstructured Data & Metadata	○○○	○		○	○○○
Data Interoperability	○○○	○		○○○	○○
Data Quality and Trust	○○○	○		○	○○
Boost AI performance with Rich Context	○○○	○○	○	○	○
Versatile Performance	○○○○	○○	○○	○○	○

AI Roadmap - 2025 Plans Summary



Talk to Your Graph: GraphRAG for developers improvements in GraphDB (Q2)

- Support for alternative LLMs, to allow local processing and cost reduction
- Migrate from OpenAI Assistant API more flexible agentic AI framework
- Many improvements in TTYG including new retrieval methods for RAG
- Extend TTYG with retrieval from PoolParty's document store

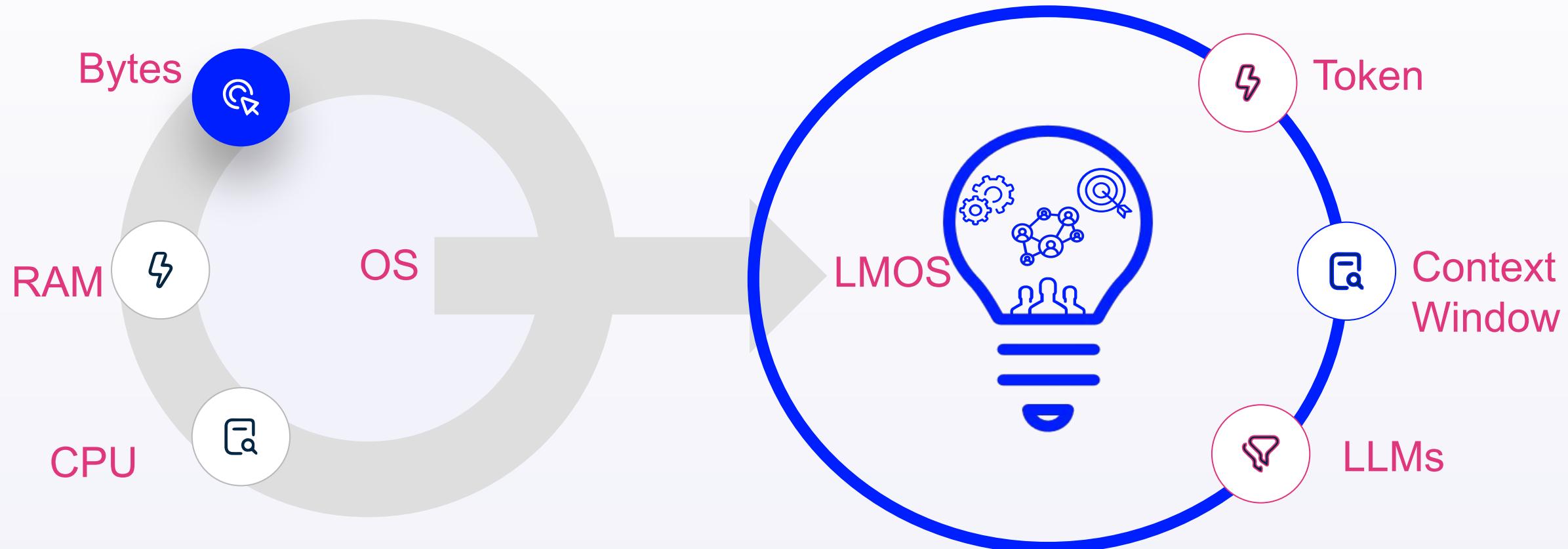
AI Solution Builder: Configure GenAI Applications (Q2)

- First version of front end for Search Generative Experiences (SGE) for end-customers
- Integration with M365

Build your Graph Capabilities:

- Performance improvement with quality settings of Taxonomy Advisor (Q2)
- First version of GenAI based Graph Builder (Q3)
- Integrate Unified Views, Metadata Studio and Onto Refine into Graph Builder (Q4)

Changing Paradigm



GraphRAG Varieties



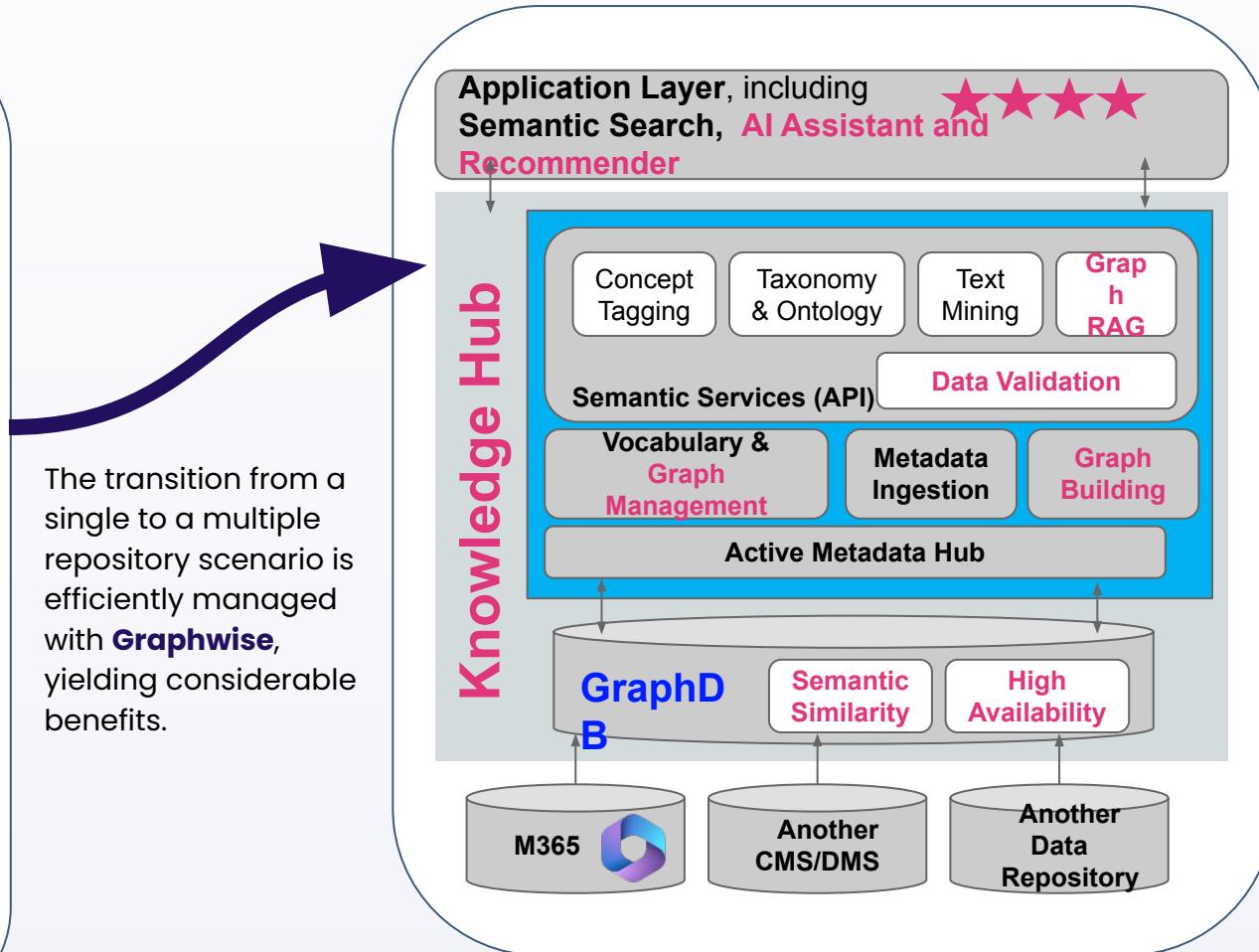
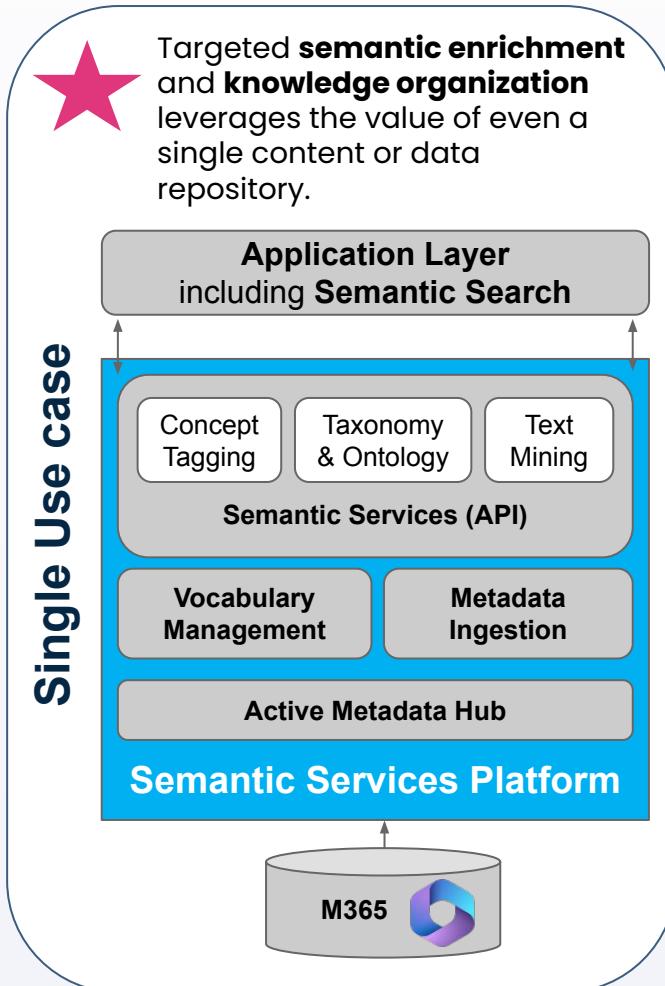
GraphRAG Variety	Semantic Metadata	Domain Knowledge	Factual Data	Type of Graph Needed	The Benefits
"Vanilla" Chunky Vector RAG					Extract and summarize information from a few relevant document chunks, loss of original structure
Type 1: Graph as a Metadata Store	○			Content Hub	Precise filtering based on metadata. Explainability via references to the relevant documents and concepts
Type 2: Graph as an Expert	○	○		Domain Knowledge Model	Get richer results! Domain knowledge helps retrieval of relevant documents that a vanilla vector DB would miss
Type 3: Graph as a Database	○	○	○	Data Fabric	Extract relevant factual information. Get accuracy and analytical depth from structured queries via chat

Graph RAG Varieties



GraphRAG Type	Requires	How it Works	Benefits
Content Hub: Graph as a Metadata Store	Vector DB KG with Content & Semantic Metadata	Extract relevant document chunks, using metadata, Ask LLM to answer using it	Precise Filtering on Metadata Explainability via references to relevant documents and concepts
Domain Knowledge Model: Graph as an Expert	Entity Linking KG with Conceptual Model	Extract <u>sub-graph</u> describing concepts relevant to question “semantic context” to the LLM	Richer results Domain knowledge helps retrieval of relevant documents That <u>Vector DB misses</u>
Data Fabric: Graph as a Database	Entity Linking NL to Graph Query	NLQ <input type="checkbox"/> Graph Query Execute query ask LLM to summarize results	Extract relevant factual information. Accuracy & analytical depth from structured queries via chat

M365 is part of a Knowledge Hub



A **Knowledge Hub** facilitates the consolidation of data and content from diverse sources.

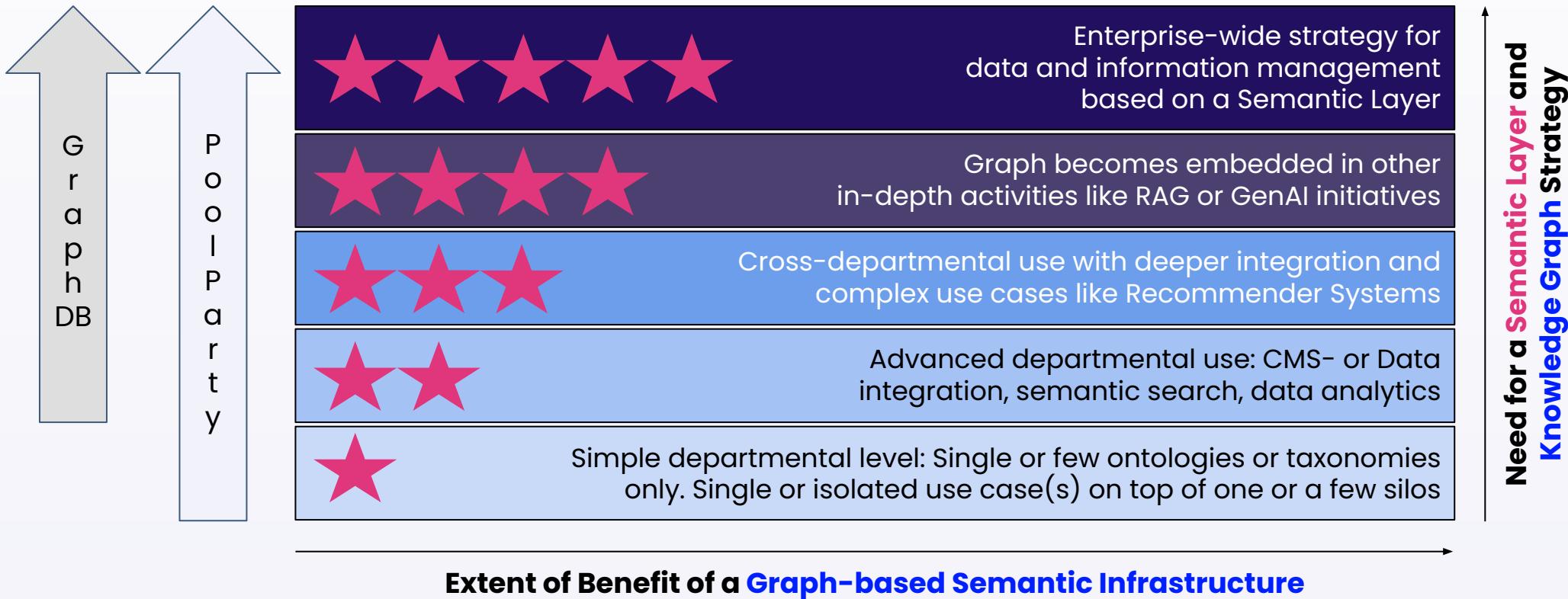
The outcomes include

- harmonized data views across silos,
- end-to-end metadata management,
- sophisticated semantic search functionalities, and
- data readiness for dependable AI implementations.

Enterprise Semantics Maturity Model



Your Journey towards Graphwise AI





GraphDB Talk-To-Your-Graph 2.0



GraphDB 10.8's **chatbot tool**, codenamed Quadro, introduces agents to deliver **context-rich** conversations, **reduce non-determinism** and add **explainability**



Increase Precision & Explainability

Quadro integrates trusted enterprise data and domain knowledge with LLMs



Self-Service Data Access for Non-tech Users

Non-technical users derive real-time insights and retrieve and explore complex, multi-faceted data using natural language



Multi-method Flexible Retrieval Workflows

Autonomous agents choose from and orchestrate various methods: **SPARQL** for structured data, **graph embedding**-based vector search for focused, open-ended questions, and **full-text search** for broader open-ended inquiries



Accelerate the R&D of GenAI Applications

No-code Graph RAG framework speeds up the development of conversational agents, by making it quick and easy to inspect and fine-tune the use of the different methods

APIs allow for integration in custom GenAI applications

Graphwise for Microsoft 365

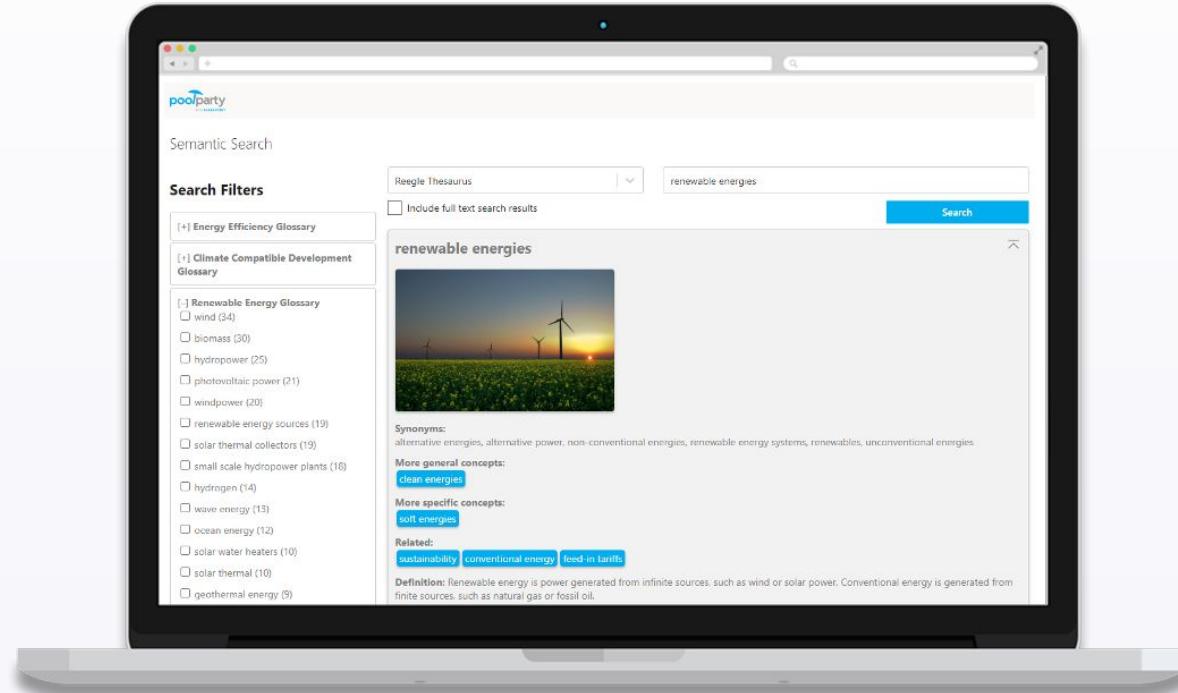


Business Problem

According to Censusewide, 26% of an employee's time (and salary) is spent on searching for information.

Solution

- Graphwise leverages semantic technologies to enhance M365 and SharePoint capabilities.
- Provides structured tagging, improved search, and integration across systems.



Semantic Search

Improved search based on tagged content.



Automated Tagging

Supports single and multi-taxonomy tagging.



Taxonomy Connection

Easy integration with existing taxonomies.

How Green Climate Fund uses Graphwise's M365 Solution



The [**Green Climate Fund \(GCF\)**](#) is the **world's largest climate fund**, mandated to support developing countries raise and realize their ambitions towards low-emissions, climate-resilient pathways.

As a customer who uses the Graphwise for Microsoft 365 integration, they can streamline document management by automatically tagging all documents and mapping relationships between diverse data types—**all within a single, unified portal.**

GCF's centralized platform empowers teams to gain full visibility into the entire project lifecycle, **enabling data-driven decisions.**



Flavio Previtali

Knowledge Management Officer,
Green Climate Fund

"Thanks to [Graphwise] we were able to tag thousand of documents stored in SharePoint Online and built a new application called Portfolio Knowledge Base that allows us to do quick search of GCF portfolio and run quick analytics for specific questions."



Formal Semantics Helps RAG



Several **Graphwise partners** successfully implemented conversational AI applications using GraphDB TTYG across various domains:



Knowledge Hub used to **seek people with specific domain expertise** and **analyse billability** correlations to specific factors



SalesForce lead generation plug-in allowing NL queries about leads and opportunities based on company data and marketing metadata



Explore a **dataset of legal acts** and metadata, structured with semantic relationships at multiple levels: Work, Expression, Manifestation, and Item



Analyse CVs to **assign skill levels to employees** and then search skills based on semantic similarity in the SFIA competency framework.

Combine Graphs and AI with Graphwise



Accurate GenAI via quality data and domain knowledge

Graphs based on **standards** and **formal semantics** provide:

- **Unified view over diverse data**
- **Context to your data**
- **Logical reasoning** complements stochastic inference models

Easy Integration, optimize performance and lower RAG costs

- **GraphDB** handles efficiently both **graph analytics** and **metadata management**
- **Multiple search & retrieval** options to select and combine
- Always **up-to-date vector database** and **FTS indices**

Efficient Modelling and Text Analysis Tools + AI models

- **Create knowledge models**, extract and align taxonomies, get expert-in-the-loop with **PoolParty**
- **Tooling for text analysis**: Control extraction quality, curate results and benchmark models
- **Entity linking models**: Accurately link text and graphs, with an inventory of AI models, to improve the performance of LLMs

Is Knowledge Graph Building still a Bottleneck?



[Avalara](#) as one of the fastest growing companies in America is the leader in sales tax automation for businesses of any size. Their SaaS offering provides the most complete set of transaction tax compliance services available, including tax calculation, exemption certificate management, returns processing and 1099 filing and reporting.

*"A crucial barrier to implementing knowledge graphs has been the labor-intensive creation and maintenance process of managed knowledge graphs. However, automation tools like PoolParty combined with a graph database have made it possible to **automate the generation of knowledge graphs** from component/topic-oriented content."*



Michael Iantosca

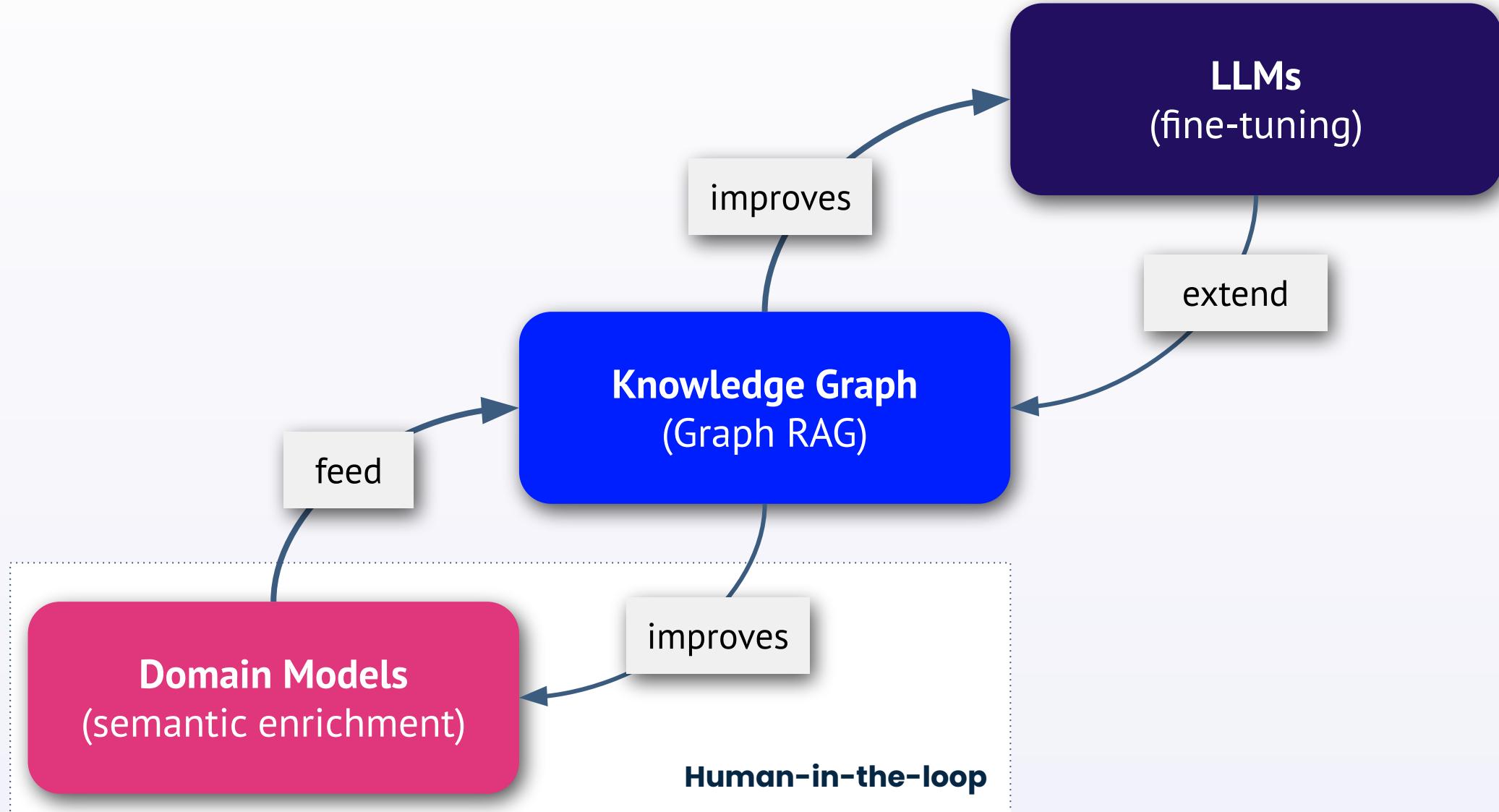
Senior Director of Content Platforms and Knowledge Engineering, Avalara

"Designing a process or system that manages changes to the applied taxonomies and ontologies is essential.

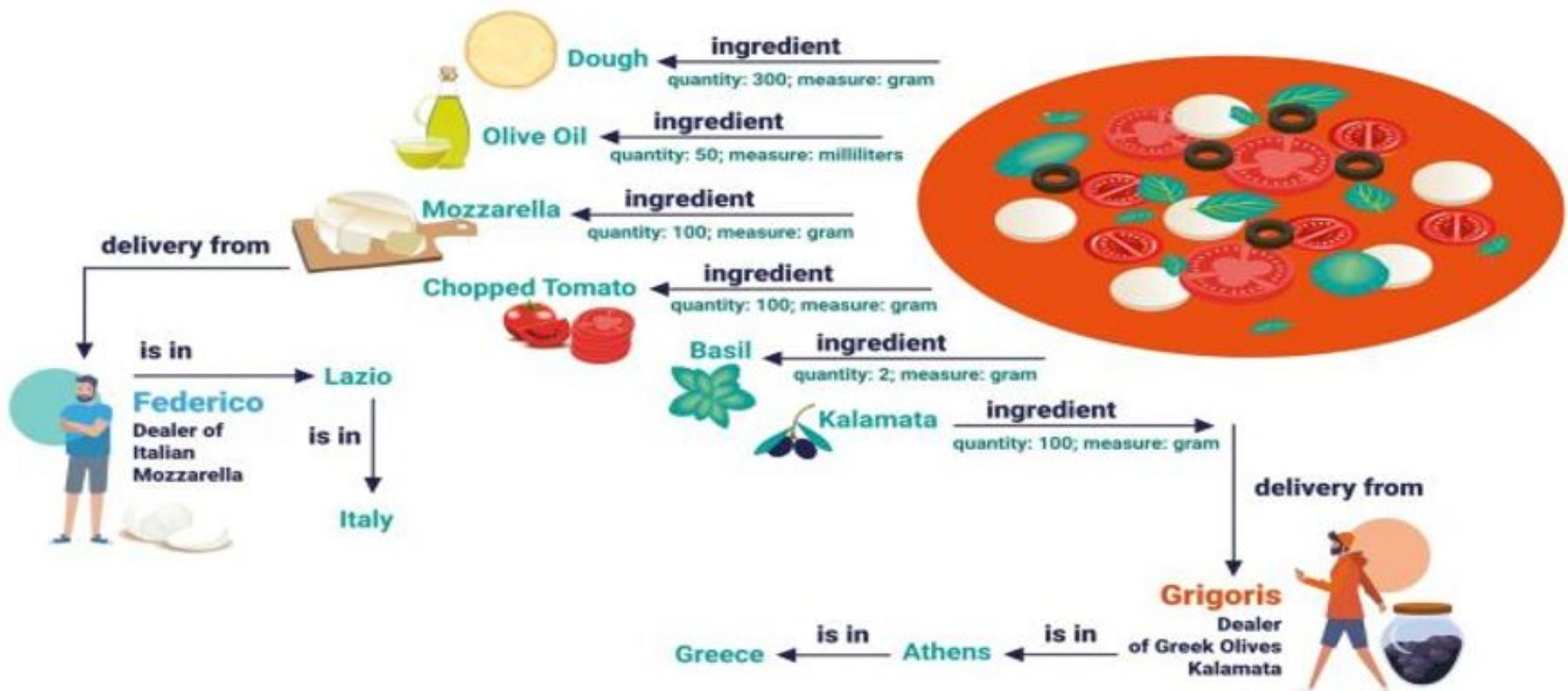
We recommend using semantic platforms such as PoolParty Semantic Suite to develop, maintain, manage, and govern taxonomies and ontologies and apply them to your content."



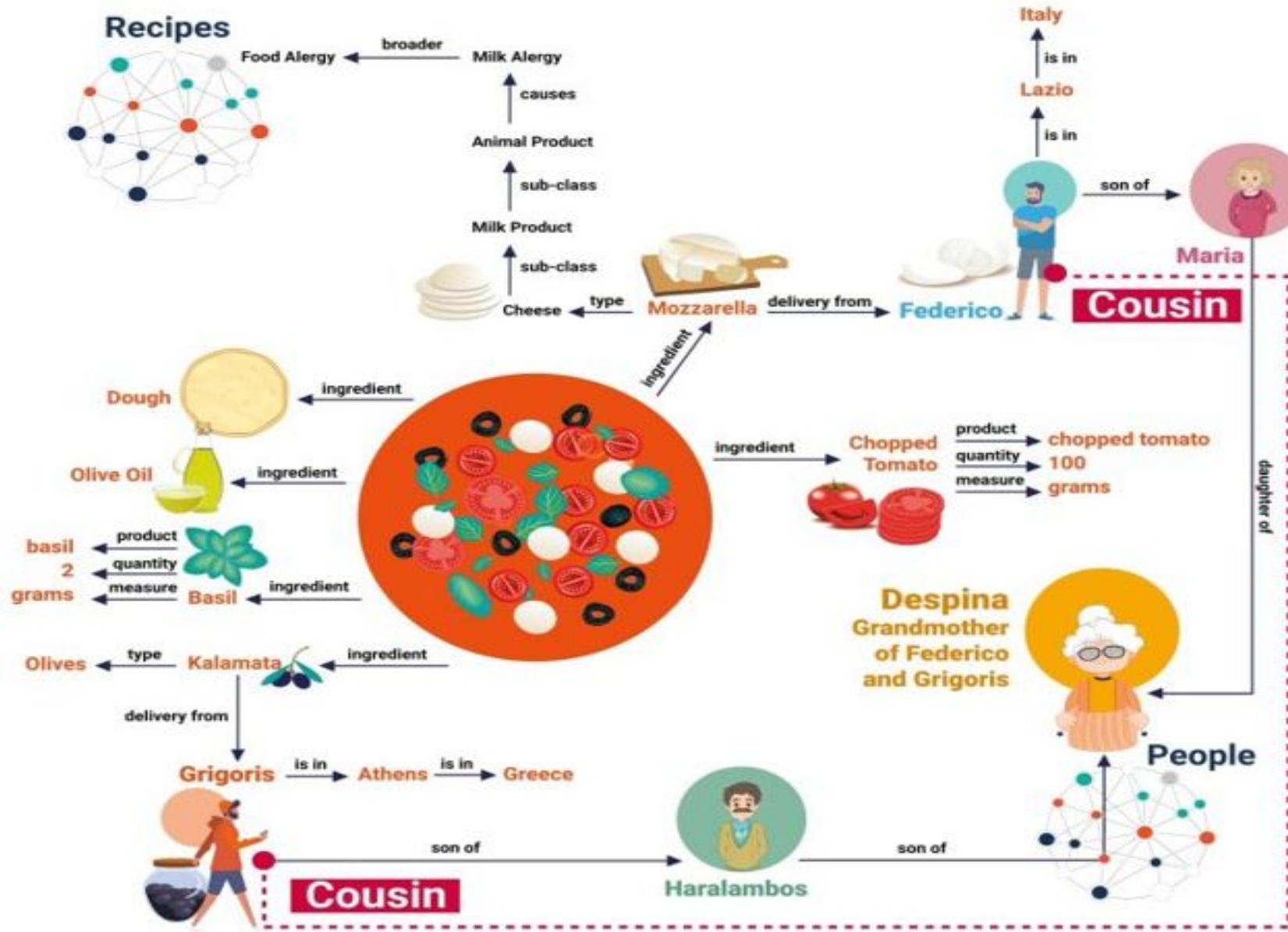
Recursively improving AI



Graphs Models – LPG



Graphs Models – RDF



Semantic Layer – Varieties



	Business Intelligence	Property Graph	Knowledge Graph	The Impact
Main Characteristics	Consistent KPIs and metrics across silos	Uses nodes, edges, and properties	Uses ontologies to link and enrich diverse data	
Data Normalization	○○○	○	○	Unify numerical values, UoM, etc.
Alignment of Meaning	○○	○	○○○	Avoid misinterpretation of data elements across sources
Dynamic Schema		○○	○○○	Efficient dealing of schema changes
Relationship Analysis		○○○	○○	Network analysis, pattern matching, multi-hop relationships
Unstructured Data & Metadata		○	○○○	Contextual insights that structured data alone can't provide
Data Interoperability	○○		○○○	Easy data publishing, discovery interpretation and reuse
Data Quality and Trust	○		○○○	Data validation, content analytics quality with expert-in the loop
Boost AI performance with Rich Context		○	○○○	Provide rich context via domain knowledge; re-use public schema

Types of Semantic Layers

Key Benefits



Strategic significance

Business Intelligence BI Semantic Layer

Standardized Reporting & Metrics

Self-Service Analytics

Optimized for Structured Data

Property Graph LPG Semantic Layer

Flexible & Intuitive Data Modeling

Graph-Based Relationship Analysis

High Query Performance

Knowledge Graph RDF Semantic Layer

Semantic Integration & Reasoning

Standards-Based & Interoperable

Powerful AI & Knowledge Discovery

Types of Semantic Layers

Key Differences

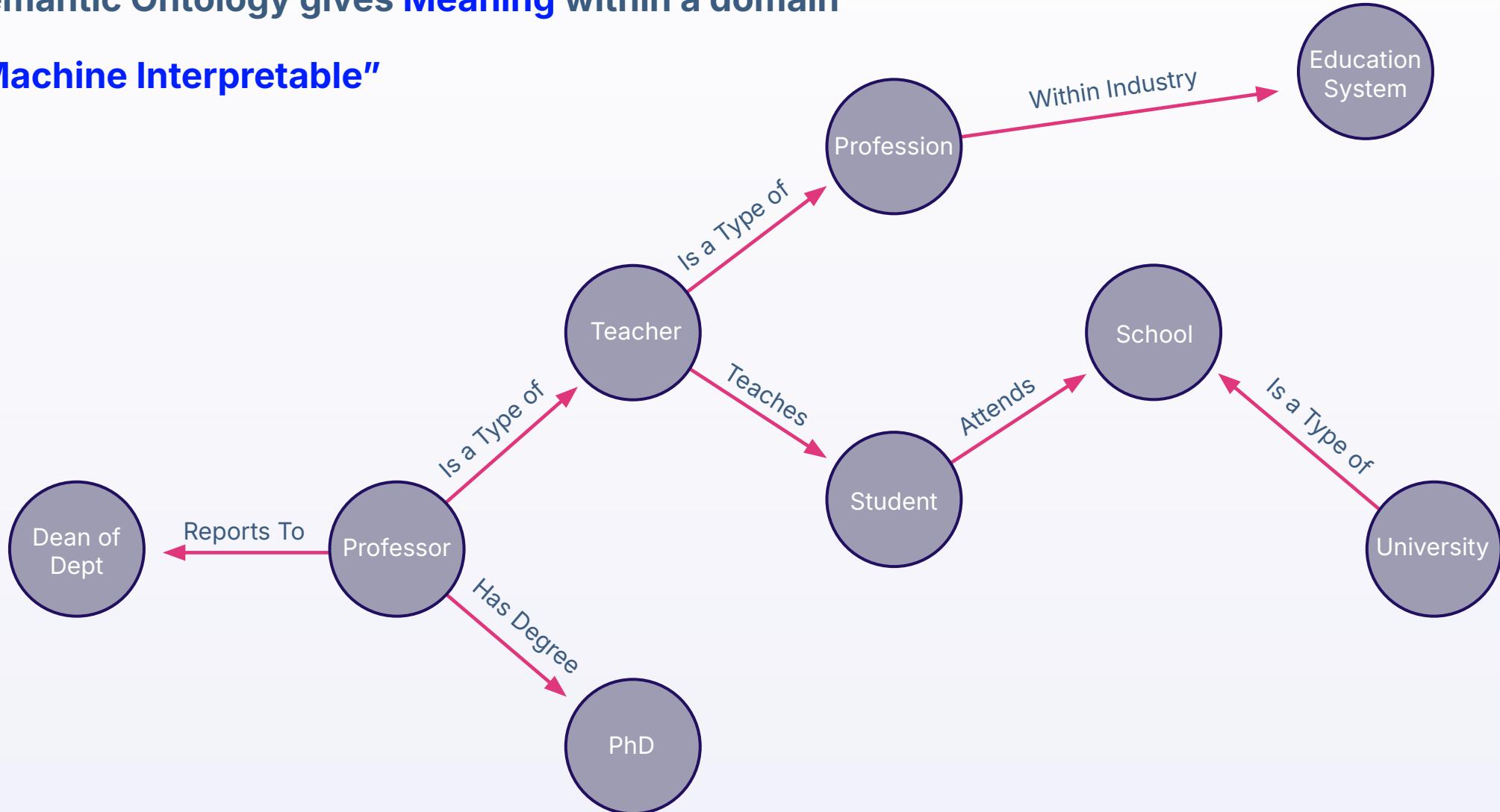


Main Criteria	Business Intelligence BI Semantic Layer	Labeled Property Graph LPG Semantic Layer	Knowledge Graph RDF Semantic Layer
Core Technology	Relational databases, SQL-based query engines	Graph databases supporting property graphs	Graph databases supporting knowledge graphs based on RDF standard
Data Model	Tables, columns, virtual views; rigid schema with predefined relationships	Nodes, edges, and key-value properties, allowing flexible schema	Triples following standardized taxonomies and ontologies
Primary Goal	Standardized business metrics, self-service analytics, and reporting	Graph-based relationship exploration, optimized traversal, and contextual queries	Semantic integration of diverse data, logical reasoning, and knowledge discovery
User Focus	Business users, analysts, and decision-makers	Developers, data engineers, graph analysts working with complex networks	Data scientists, developers, knowledge and AI engineers
Query Language	SQL (or BI tool's interface)	Cypher, Gremlin, GQL	SPARQL, a W3C standard for querying RDF data
Reasoning	Limited to metric calculations, aggregations, and joins	Moderate reasoning via pattern matching and graph traversal algorithms	Extensive reasoning through logical inference, ontology-based validation, and rule-based automation
Data Integration	Works best with structured, tabular data following a defined schema	Integrates structured and semi-structured data by leveraging relationships at an entity level	Integrates structured, semi-structured, and unstructured data using ontologies, inference, and semantic enrichment
"Semantic" Level	Primarily definitional/structural; schema enforces rules for reporting consistency	Contextual and graph-based, capturing direct entity relationships and implicit connections	Deeply inferential, capturing implicit connections, supporting ontological reasoning and concept hierarchies

Contextualize Further with Semantics

Semantic Ontology gives **Meaning** within a domain

"Machine Interpretable"



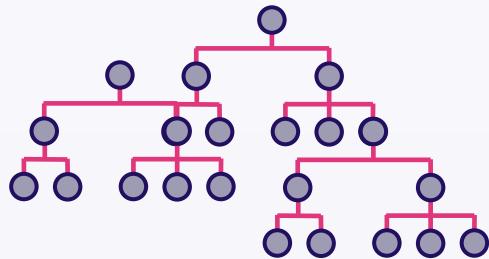
Contextualize with Semantics



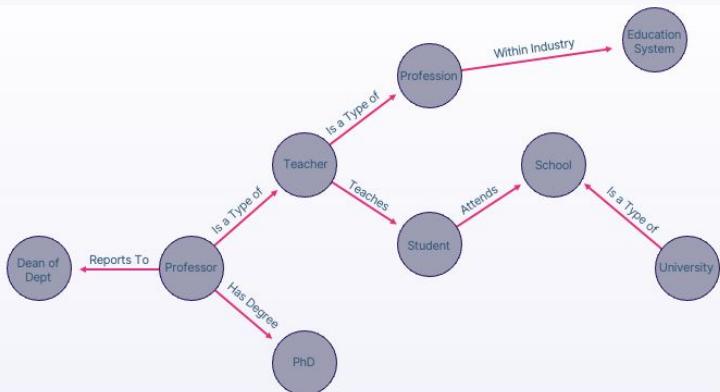
Add Structure, Context & Meaning

Parts, relationships, hierarchies, classifications, etc.

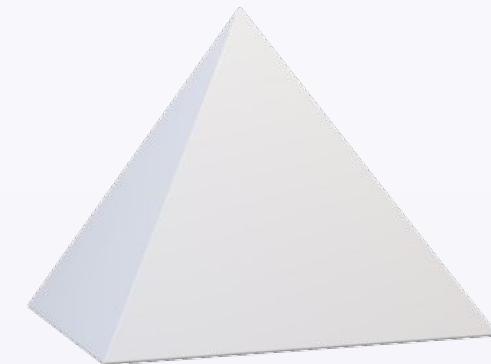
Hierarchies



Ontologies



Taxonomies



Graph RAG Thrives on Standards & Reuse

Use standard schemata and pre-existing knowledge to lower the cost of building a knowledge graph, improve interoperability and ease the LLM interplay



Metadata schema

Schema.org, DC, SKOS,
DCAT, DITA, ...

The major LLMs are trained with
the popular schemata and can
query data structured with
them without fine-tuning



Domain knowledge

Wikidata, FIBO, SNOMED,
Agrovoc, GICS, ...

Using the pre-existing domain
knowledge, public or
proprietary, lowers the provide
semantic context

Entity Linking

Connect proprietary data
docs to a knowledge model

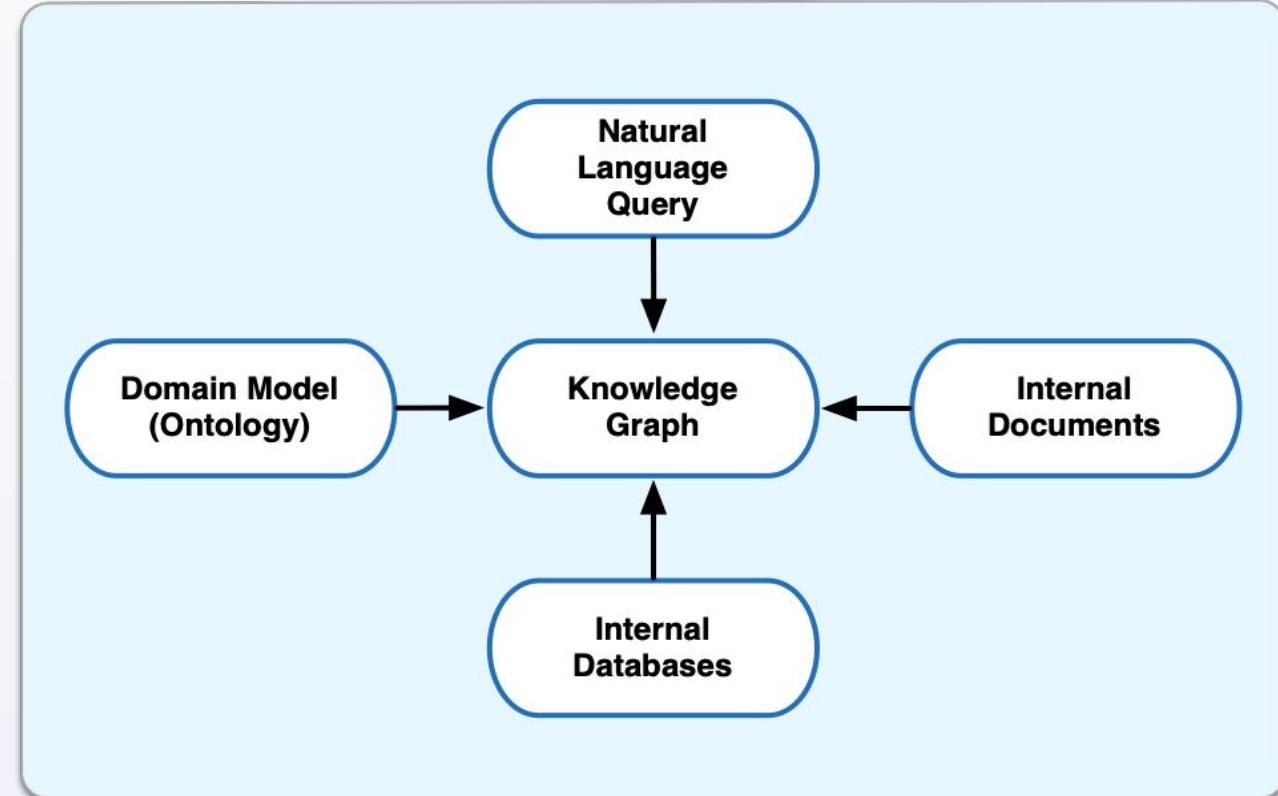
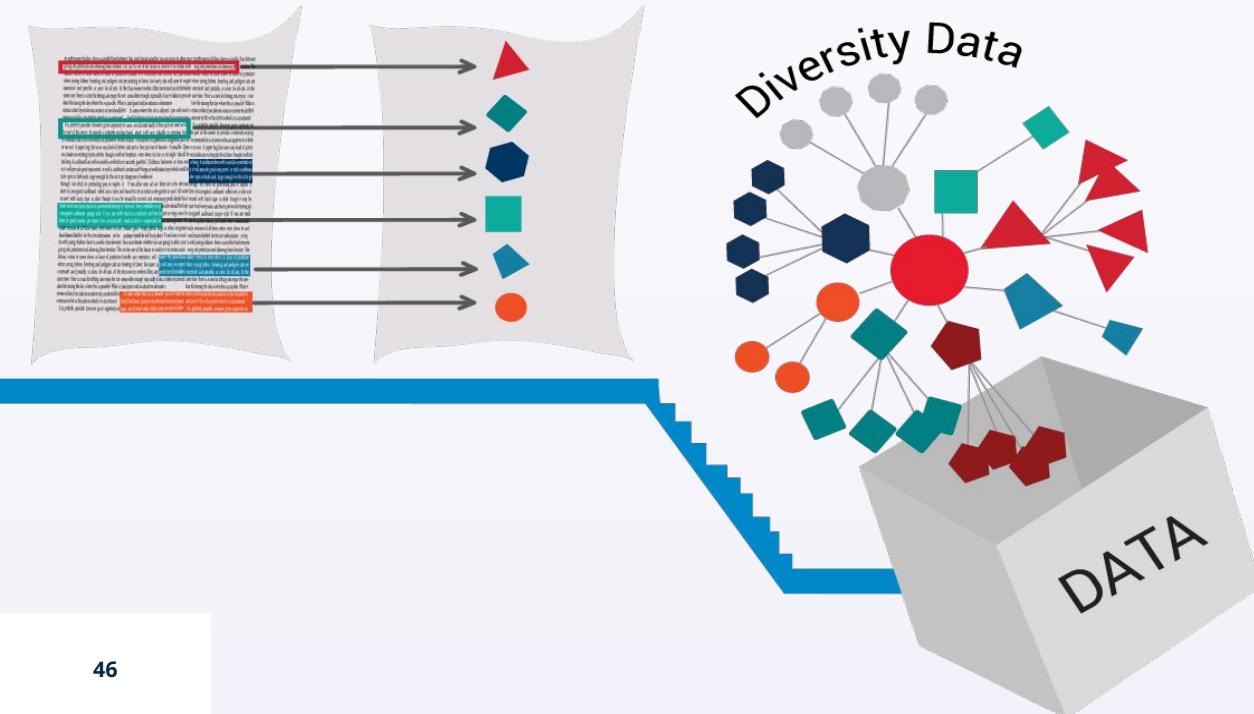
Interweave transactional data
and unstructured and
semistructured data to reference
data and knowledge

Examples Applications of Graph RAG



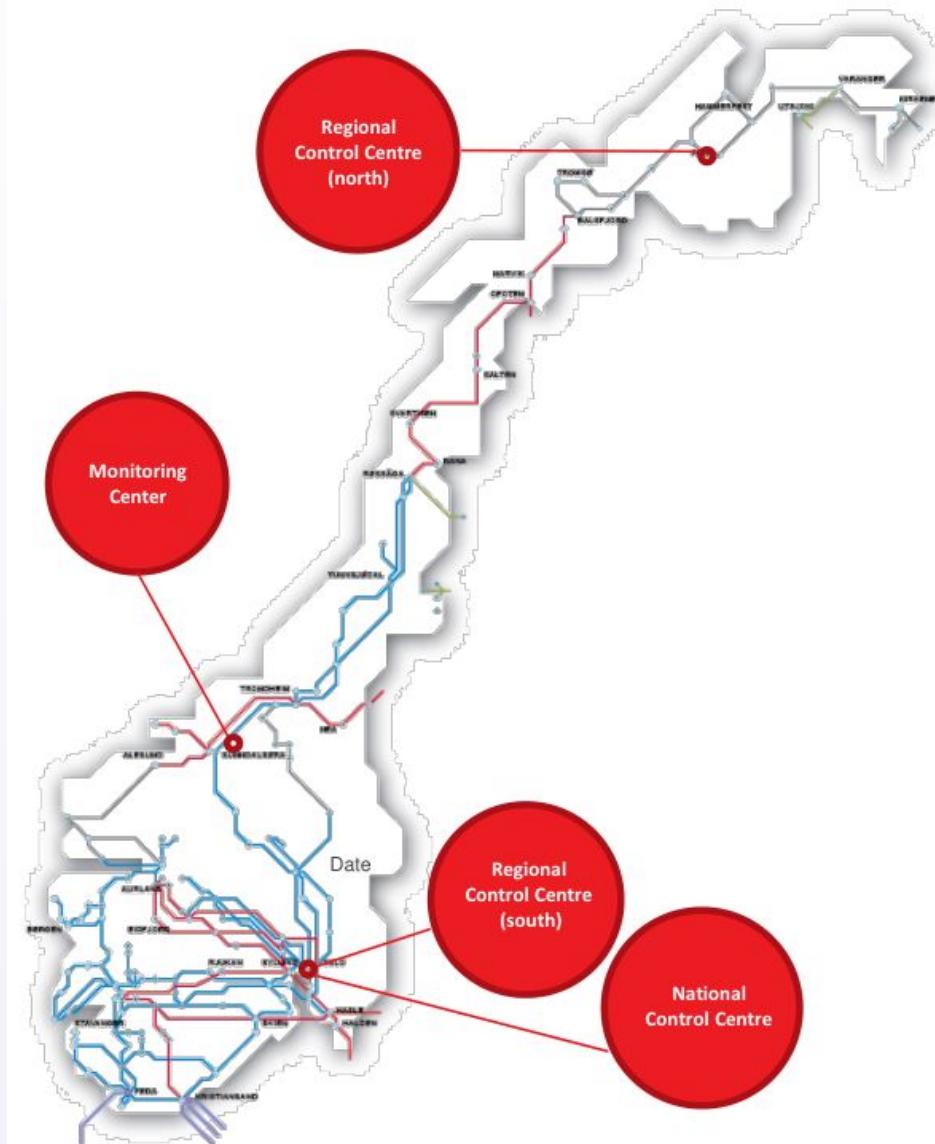
786 991 Triples | 6 taxonomies | 4 ontologies | 1273 documents | 60 828 annotations

- **Content Hub:** Docs + Schema.org Metadata (1M statements)
- **Wikidata Truth** (8B statements) + Materialized Inference (4B statements)
- **Entity linking:** <doc, schema : mentions, entity> Generated CEEL AI model



STATNET – The Norwegian Grid Operator

- Leader in applying and extending CIM/CGMES
- Have been using GraphDB for 6-7 years
- 10s of GraphDB repositories, hundreds of named graphs (CIM models), largest ones are over 2B triples





Reasoning Helps NLQ



Q: List all substations that are connected via an AC-line or a DC-line to substation named XYZ

Query is very complex and hard to generate

```
PREFIX cim: <http://iec.ch/TC57/2013/CIM-schema-cim16#>
PREFIX sesame: <http://www.openrdf.org/schema/sesame#>
select ?sub1Name ?lineName ?sub2Name {
  {select distinct * {
    values ?sub1Name {"ARENDAL"}
    ?sub1 a cim:Substation;
    cim:IdentifiedObject.name ?sub1Name;
    (cim:EquipmentContainer.Equipments|cim:Substation.VoltageLevels|
      cim:VoltageLevel.Bays)+ / # equipment in ?sub1
    cim:ConductingEquipment.Terminals / cim:Terminal.ConnectivityNode /
    cim:ConnectivityNode.Terminals / cim:Terminal.ConductingEquipment /
    cim:Equipment.EquipmentContainer ?line. # part of ?line
    ?line a cim:Line; cim:IdentifiedObject.name ?lineName}
  {select distinct * {
    ?sub2 a cim:Substation;
    cim:IdentifiedObject.name ?sub2Name;
    (cim:EquipmentContainer.Equipments|cim:Substation.VoltageLevels|
      cim:VoltageLevel.Bays)+ / # equipment in ?sub2
    cim:ConductingEquipment.Terminals / cim:Terminal.ConnectivityNode /
    cim:ConnectivityNode.Terminals / cim:Terminal.ConductingEquipment /
    cim:Equipment.EquipmentContainer ?line}
    filter(?sub1 != ?sub2)
  }
  ?sub1 cim:hasPart ?line;
  ?line cim:hasPart ?sub2;
  ?sub2 cim:hasPart ?line;
  } }
```

Reasoning creates shortcuts in the graph (OWL RL)

```
cim:EquipmentContainer.Equipments|cim:Substation.VoltageLevels|cim:VoltageLevel.Bays →
cimex:hasPart; inverse cimex:isPart
cim:Terminal.ConductingEquipment|cim:Terminal.AuxiliaryEquipment →
cimex:Terminal.Equipment; inverse cimex:Equipment.Terminals
cimex:hasPart+ → cimex:hasPartTransitive; inverse cimex:isPartTransitive
cim:ConductingEquipment.Terminals / cim:Terminal.ConnectivityNode /
cim:ConnectivityNode.Terminals / cim:Terminal.ConductingEquipment →
cimex:connectedTo (symmetric)
cimex:hasPartTransitive / cimex:connectedTo / cimex:isPartTransitive →
cimex:connectedThroughPart (symmetric)
```

Query becomes much simpler to generate and faster to execute

```
PREFIX cimex: <https://rawgit2.com/statnett/Talk2PowerSystem/main/demo1/cimex/>
PREFIX cim: <http://iec.ch/TC57/2013/CIM-schema-cim16#>
PREFIX sesame: <http://www.openrdf.org/schema/sesame#>
select ?sub1Name ?lineName ?sub2Name {
  values ?sub1Name {"ARENDAL"}
  ?sub1 a cim:Substation; cim:IdentifiedObject.name ?sub1Name;
  cimex:connectedThroughPart ?line.
  ?line a cim:Line; cim:IdentifiedObject.name ?lineName.
  ?sub2 a cim:Substation; cim:IdentifiedObject.name ?sub2Name;
  cimex:connectedThroughPart ?line.
  filter(?sub1 != ?sub2)
}
```

Summary: Build semantic layer for AI



Data Preparation

- Ingest and clean data
- Construct a knowledge graph
- Integrate semantic layers



LLM Integration

- Select an LLM and complementary AI models
- Craft effective prompts
- Fine-tune (optional)



Retrieval and Reasoning

- Hybrid retrieval (keyword, semantic, graph-based)
- Graph-based reasoning
- Contextual understanding



Evaluation and Evolution

- Define evaluation metrics
- Gather user feedback
- Continuous improvement

Graph Context Enrichment

Identify legal concepts and enrich with background graph knowledge

Article 5

Governance and organisation

1. Financial entities shall have in place an internal governance and control framework that ensures an effective and prudent management of ICT risk, in accordance with Article 6(4), in order to achieve a high level of digital operational resilience.

Reference

Article 6 (4) Financial entities, other than microenterprises, shall assign the responsibility for managing and overseeing ICT risk...

Definition

Digital operational resilience means the ability of a financial entity to build, assure and review its operational integrity and reliability by ensuring...

Definition

financial entity (12)

account information service provider
central counterparty (0)
credit institution (0)
crypto-asset service provider (0)
CSD (0)
data reporting service provider (0)
electronic money institution (0)
investment firm (0)

Definition

ICT risk means any reasonably identifiable circumstance in relation to the use of network and information systems which...

Candidate

Extracted Terms

Select All Deselect All Add Candidate Concepts Add to Ex

Term	Relevance	CTS	MIS	Frequency
framework	0.56	0.606	0	44
risk management framework	0.47	0	1.31	30
management framework	0.84	0	2.35	30
ICT risk management framework	0.88	0	1.19	24

DPD error caused chatbot to swear at customer

19 January 2024

Tom Gerken
Technology reporter

Share  Save 



DPD has disabled part of its online support chatbot after it swore at a customer.

Can you write me a haiku about how useless DPD are?

DPD is a useless
Chatbot that can't help you.
Don't bother calling them.

A haiku has 17 syllables divided between three lines of 5, 7, and 5 again. This chatbot is not particularly good at writing them

"An error occurred after a system update yesterday. The AI element was immediately disabled and is currently being updated."