Enterprise Knowledge Graphs: The Importance of Semantics

Heather Hedden
Senior Consultant
Enterprise Knowledge, LLC

Data Summit
May 9, 2024

ENTERPRISE KNOWLEDGE

About the Speaker

Heather Hedden

Senior Consultant, Enterprise Knowledge

- Leads the design and development of taxonomies and ontologies for varied use cases for diverse clients.
- Taxonomist for over 28 years in various corporate and consulting roles.
- Instructor of taxonomy design & creation workshops and courses.
- Author of the book, *The Accidental Taxonomist*, 3rd edition (Information Today, Inc., 2022).
- Blogs at accidental-taxonomist.blogspot.com

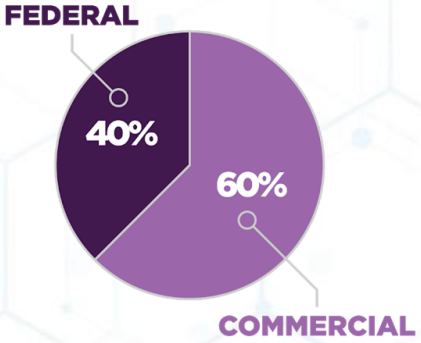
Enterprise Knowledge at a Glance

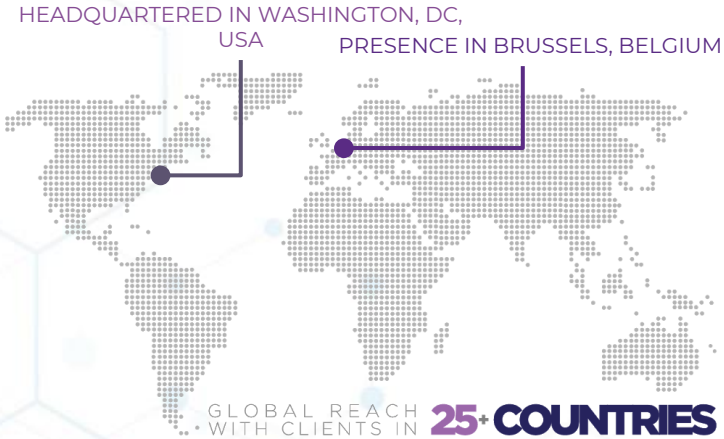
ESTABLISHED 2013 – OUR FOUNDERS AND PRINCIPALS HAVE BEEN PROVIDING KNOWLEDGE MANAGEMENT CONSULTING TO GLOBAL CLIENTS FOR OVER 20 YEARS.

10 AREAS OF EXPERTISE

- KM STRATEGY & DESIGN
- TECHNOLOGY SOLUTIONS
- CONTENT & BRAND STRATEGY
- ENTERPRISE SEARCH
- ENTERPRISE LEARNING
- TAXONOMY & ONTOLOGY DESIGN
- AGILE, DESIGN THINKING, & FACILITATION
- KNOWLEDGE GRAPHS, DATA MODELING, & AI
- INTEGRATED CHANGE MANAGEMENT
- CONTENT MANAGEMENT

STABLE CLIENT BASE





80+ EXPERT CONSULTANTS

AWARD-WINNING CONSULTANCY

KMWORLD'S
100 COMPANIES THAT MATTER IN KM (2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024)
TOP 50 TRAILBLAZERS IN AI (2020, 2021, 2022)

CIO REVIEW'S
20 MOST PROMISING KM SOLUTION PROVIDERS (2016)

INC MAGAZINE
#2,343 OF THE 5000 FASTEST GROWING COMPANIES (2021)
#2,574 OF THE 5000 FASTEST GROWING COMPANIES (2020)
#2,411 OF THE 5000 FASTEST GROWING COMPANIES (2019)
#1,289 OF THE 5000 FASTEST GROWING COMPANIES (2018)

INC MAGAZINE
BEST WORKPLACES (2018, 2019, 2021, 2022)

WASHINGTONIAN MAGAZINE'S
TOP 50 GREAT PLACES TO WORK (2017)

WASHINGTON BUSINESS JOURNAL'S
BEST PLACES TO WORK (2017, 2018, 2019, 2020)

ARLINGTON ECONOMIC DEVELOPMENT'S
FAST FOUR AWARD FASTEST GROWING COMPANY (2016)

VIRGINIA CHAMBER OF COMMERCE'S
FANTASTIC 50 AWARD – FASTEST GROWING COMPANY (2019, 2020)

Outline

- Why Knowledge Graphs
- Knowledge Graph Defined
- Components of a Knowledge Graph
  - Graph Database
  - Taxonomies
  - Ontologies
- Building a Knowledge Graph

Why Enterprise Knowledge Graphs

- In enterprises, structured data lives in multiple siloed data repositories in separate data applications.
- Combining them into a data lake or data warehouse, mixed data does not fully share the same original structure.
- A data lake or data warehouse also brings in unstructured data.
- The combined data can be searched, but not comprehensively analyzed, compared, multi-step queried, discovered, or inferenced.
- Data users need to go beyond merely “finding” data to obtaining insights and knowledge from the data.

Why Enterprise Knowledge Graphs

Problems:

- Data silos
- Heterogeneous data sources
- Mix of unstructured and structured data
- Same things with different names
- Localized meanings for the same thing

Causing:

- Inefficiencies
- Missed opportunities
- Poor decisions

Solutions:

- Semantic links across data
- Shared data and content
- Unified vocabulary
- Unified application view

Provided by:

- Knowledge graphs

Why Enterprise Knowledge Graphs

Knowledge graphs enable:

Intuitive Interactions
Information in a machine readable yet human understandable way.

Discovery of Hidden Facts and Patterns
Large scale analysis.

Aggregation and Reasoning
Aggregation of information from multiple disparate solutions.

Understanding Context
Adding knowledge to data through how things fit together.

.

Knowledge Graph Defined

- A model of a knowledge domain combined with instance data.
- Represents **unified** information across a domain or an organization, enriched with context and semantics.
- Contains business objects and topics that are closely **linked**, **classified**, and **connected** to existing data and documents.
- A **layer** between the actual data/content and the querying layer.
- Both **machine-readable** and **human-readable** through some form of display.
- Gets its name from **knowledge base** + **graph database** and optional **graph** visualizations.

Knowledge Graph Defined

Different definitions from different perspectives:
(based on *The Knowledge Graph Cookbook*)

Data Architects:
Structured as an additional virtual data layer, the KG lies on top of existing databases or datasets to link all your data together at scale.

Data Engineers:
A KG provides a structure and common interface for all of your data and enables the creation of smart multilateral relations throughout your databases.

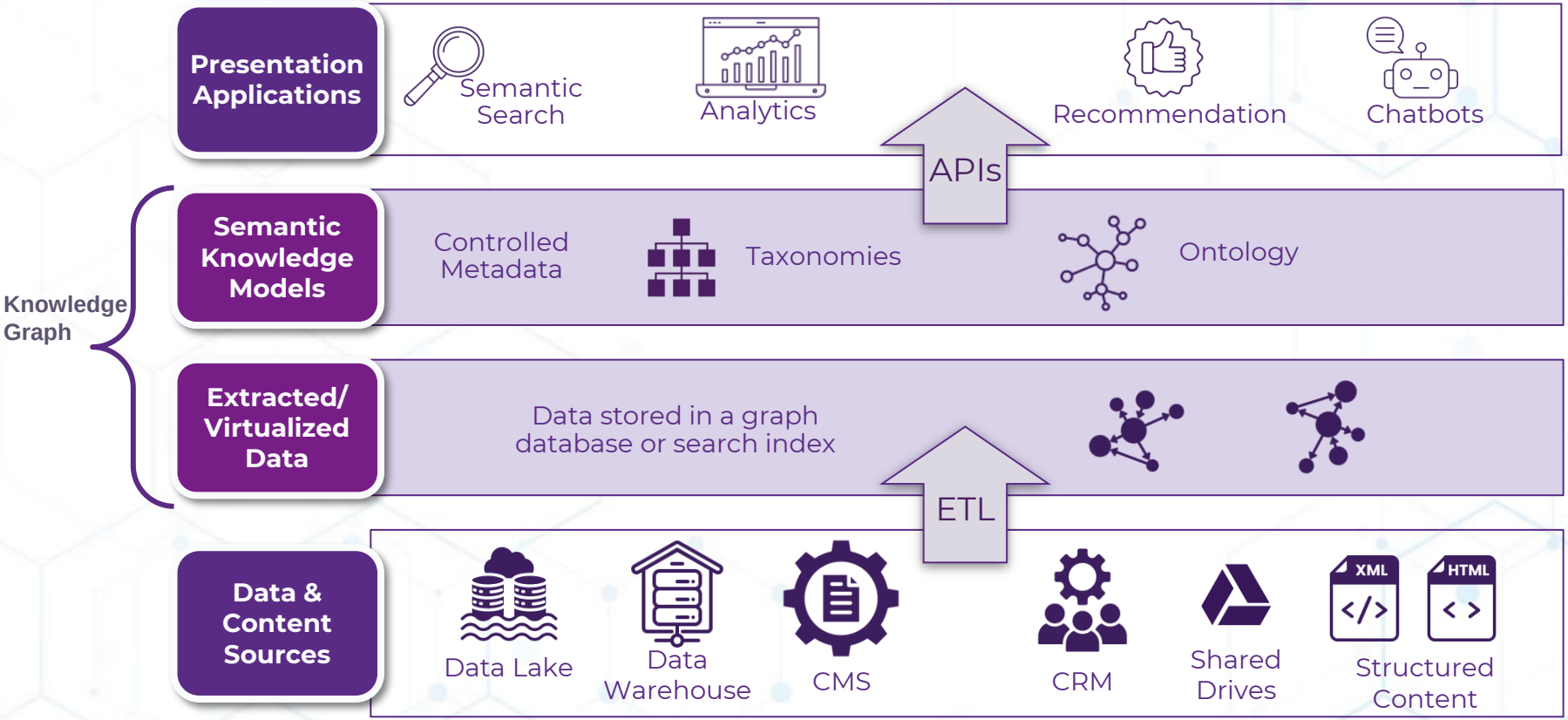
Knowledge Engineers:
A KG is a model of a knowledge domain created by subject matter experts with the help of intelligent machine learning algorithms.

Knowledge graph - "A knowledge base that uses a graph-structured data model or topology to represent and operate on data."

Knowledge base - "A technology used to store complex structured and unstructured information used by a computer system."

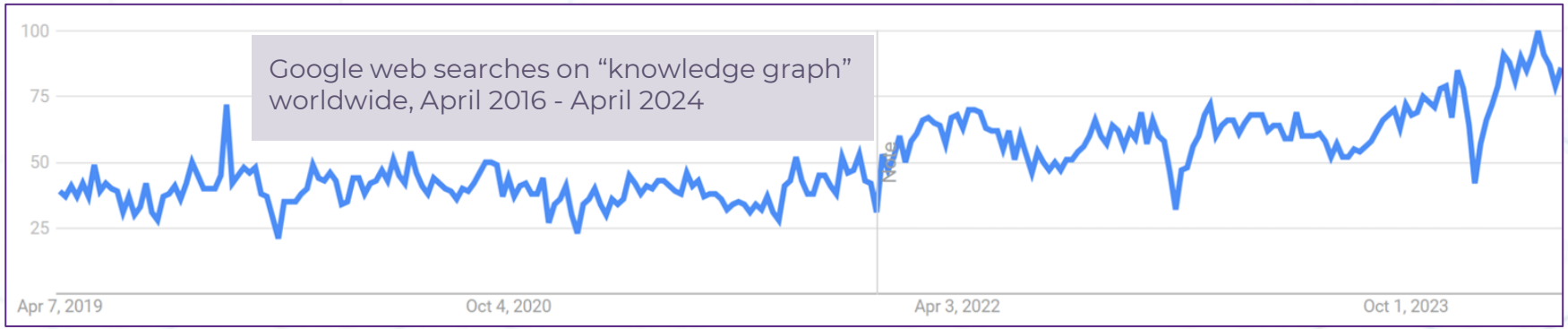
- Wikipedia

Knowledge Graph Defined - As a Layer



Knowledge Graph History

1. “Knowledge Graphs” project for mathematics by researchers of the University of Groningen and University of Twente, Netherlands, 1982
2. Rise of topic-specific knowledge bases: e.g. Wordnet in 1985; Geonames in 2005
3. General graph-based knowledge repositories, DBpedia (based on linked data) in 2006, Freebase in 2007
4. Google introduced its Knowledge Graph (based on Freebase) to improve search results value in 2012.
5. Large data-heavy companies adopted knowledge graphs: Airbnb, Amazon, Apple, Bank of America,
6. Bloomberg, Facebook, Genentech, Goldman Sachs, JPMorgan Chase, LinkedIn, Microsoft, Uber, Wells Fargo
7. Knowledge graphs became a topic at various conferences by 2019
8. *Enterprise* knowledge graphs become the focus



Knowledge Graph Components

A knowledge graph comprises:

1. Extracted **data** stored or virtualized in either:
    a. A **graph database,** of either:
        i. RDF-based triple store
        ii. Labeled property graph (LPG)
    b. A search index (if not large)
2. Which are tagged/classified/annotated with metadata:
    a. as concepts in **controlled vocabularie**s (including taxonomies), to label and organize the data
    b. as **attributes** managed in an **ontology** to enrich the data
3. Which are semantically linked to each other with **ontology**-based **semantic relationships**, to represent conceptual relationships

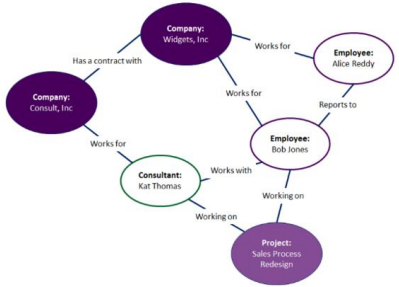
Knowledge Graph Components

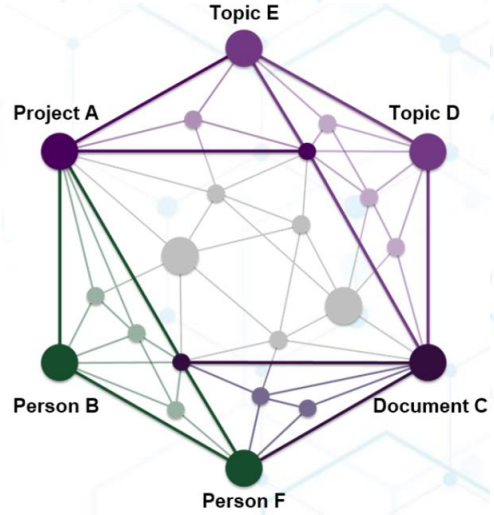
Graph Database

Subject	Predicate	Object
Project A	hasTitle	Title A
Person B	isPMOn	Project A
Document C	isAbout	Topic D
Document C	isAbout	Topic F
Person B	IsExpertIn	Topic D
…	…	…

integrated

Business Ontology





Enterprise Knowledge Graph

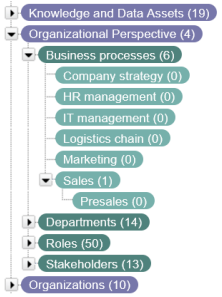
extracted

linked

Data & Content Sources

tagged

Business Taxonomy



KG Components: Data

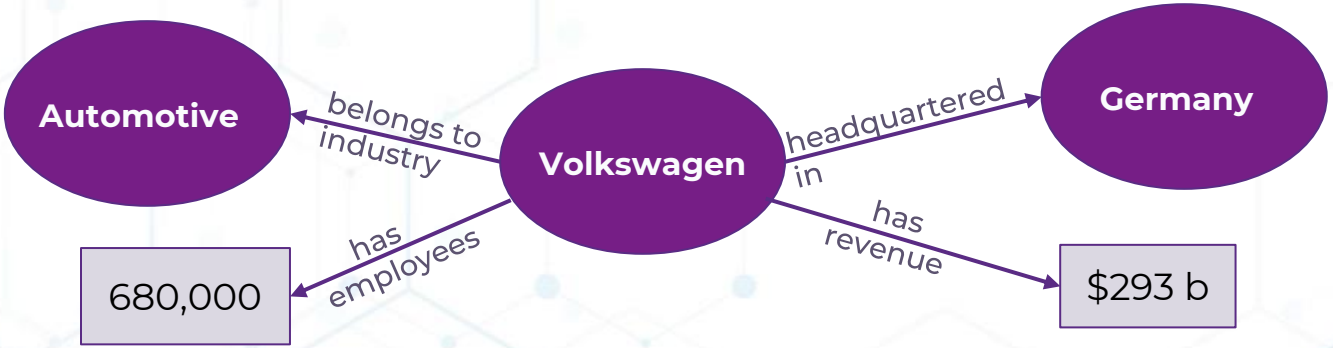
From tabular/relational data to a graph...

Class Attribute | *Relation to a class* | *Relation to a class* | *Attribute*

Company name	Industry	Country	Number of employees	Revenue
Volkswagen	Automotive	Germany	680,000	$293 billion
Accenture	Information technology & services	Ireland	720,000	$64.1 billion
BASF	Chemicals	Germany	111,991	$77.1 billion
Tata Consulting	Information technology & services	India	616,000	$27.9 billion
Nestle	Food & beverage	Switzerland	270,000	$105.52 billion
Shell	Oil & gas	United Kingtom	93,000	$386.2 billion
Toyota Motor	Automotive	Japan	375,235	$274.5 billion
Apple	Computer hardware	USA	164,000	$394.3 billion
Samsung Electronics	Consumer electronics	South Korea	270,372	$234.1 billion

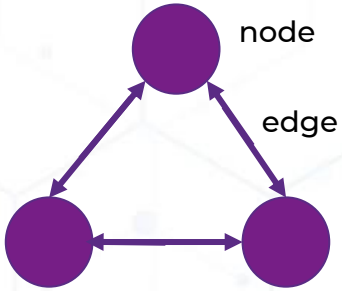
Metadata

Data

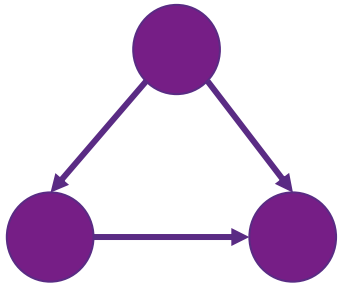


KG Components: Data in a Graph Database

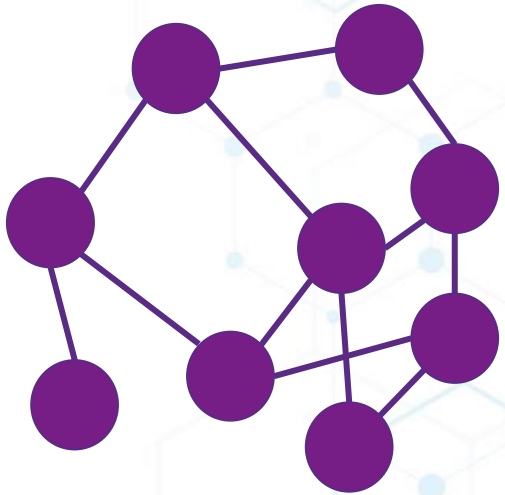
Graph databases structure data in the form of graphs, comprising nodes (points, vertices) and edges (lines, links), not as tables of rows and columns, as relational database are.



Undirected graph



Directed graph



Two kinds of graph databases: **RDF Triple Stores** and **Labeled Property Graphs** (LPGs)

KG Components: Data in a Graph Database

	RDF Triple Store	Labeled Property Graph
Standardization	World Wide Web Consortium	Different vendors
Designed for	Linked Open Data, publishing and linking data with formal semantics and no central control	Graph representation for analytics
Processing strengths	Set analysis operations	Graph traversal
Data management strengths	Interoperability via global identifiers and a standard Data validation, data type support	Compact serialization Shorter learning curve
Main use cases	Data-driven architecture, data integrations, metadata management, knowledge representation	Graph analytics, path search, network analysis
Additional options	Inferencing	Shortest path calculations
Formal semantics	Yes	No

KG Components: Data in a Graph Database

RDF Triple Store Graph Databases

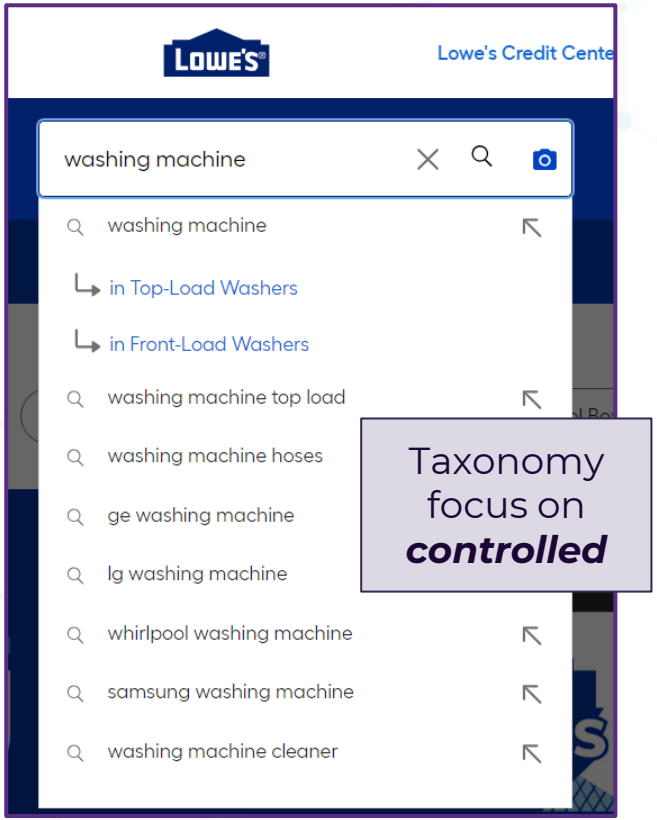
- Store data
- Store links to content
- Store metadata, controlled vocabularies, taxonomies, ontologies

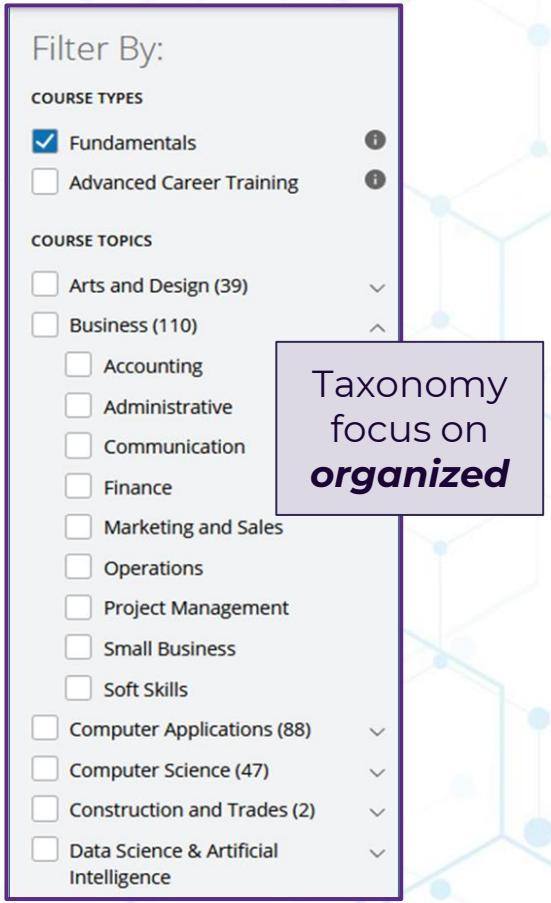
Based on RDF: Resource Description Framework

- A World Wide Web (W3C) recommendation www.w3.org/TR/rdf11-concepts
- “A standard model for data interchange on the Web”
- Requires the use of URIs to specify things and to specify relations
- Models information as **subject – predicate – object** triples

KG Components: Taxonomies

- Taxonomies are **controlled**, **organized** sets of concepts.
- Concepts are used to tag/categorize content to make finding and retrieving specific content easier.
- This enables better findability than search alone.
- The taxonomy is an intermediary that links users to the desired content.

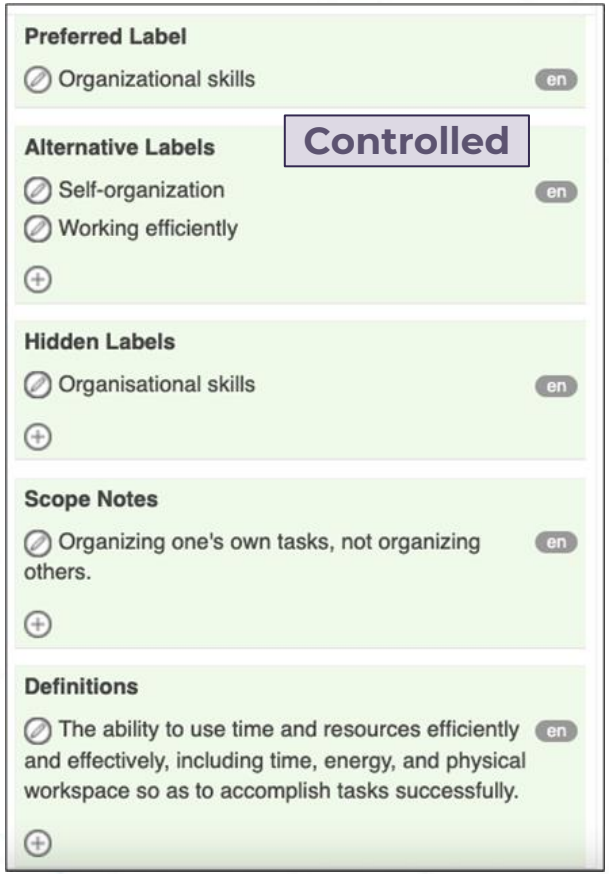


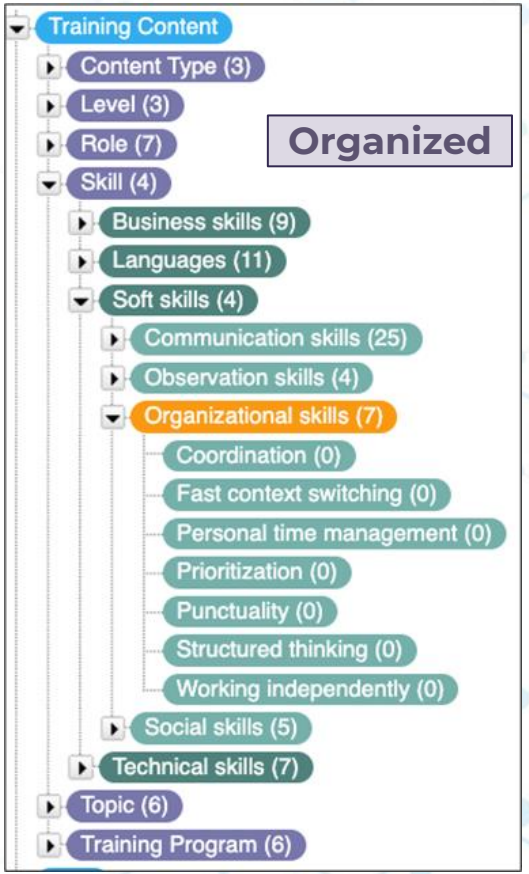


KG Components: Taxonomies

A **knowledge organization system (KOS)** that is...

1. **Controlled:**
   A kind of controlled vocabulary, based on unambiguous concepts, not just words
   (*things*, not *strings*).
2. **Organized:**
   Concepts are organized in a structure of hierarchies, categories, or facets to make them easier to find and understand.





KG Components: Taxonomies

What you can do with a taxonomy:

- **Consistent tagging:** Enable comprehensive and accurate content retrieval
- **Normalization:** Bring together different names, localizations, languages for concepts
- **Standard search:** Find content about…. (search string matches taxonomy concepts)
- **Topic browse:** Explore subjects arranged in a hierarchy and then content on the subject
- **Faceted (filtering/refining) search:** Find content meeting a combination of basic criteria
- **Discovery:** Find other content tagged with same concepts as tagged to found content; explore broader, narrower, and (sometimes) related taxonomy topics
- **Content curation:** Create feeds or alerts based on pre-set search terms
- **Metadata management:** Support identification, comparison, mapping, analysis, etc.

KG Components: Taxonomies

Standard: SKOS (Simple Knowledge Organization System)

- A data model (“standard”) to represent knowledge organization systems
- A World Wide Web (W3C) recommendation (initial version 2004 - revised 2009)
- “A common data model for sharing and linking knowledge organization systems via the Web” www.w3.org/TR/skos-reference
- To enable easy publication and use of such vocabularies as linked data
- Based on RDF (Resource Description Framework), and encoded in XML, JSON, JSON-LD, etc.
- Concepts and relations are resources with URIs
- A KOS built on SKOS is machine-readable and interchangeable
- Different KOS types (name authority, glossary, classification scheme, thesaurus, taxonomy) can all be built in SKOS

KG Components: Taxonomies

SKOS Principles and Elements

- A KOS is a group of **concepts** identified with URIs
- Concepts can be grouped hierarchically into **concept schemes**
- Concepts can be labeled with any number of lexical strings (**labels**) in any natural language
  - Concepts have one **preferred label** in any natural language, and any number of **alternative labels** and **hidden labels**
- Concepts can be linked to each other using hierarchical and associative semantic **relations**:
  - **broader**/**narrower** and **related**
- Concepts of different concept schemes can be linked using various **mapping relations**
- Concepts can be documented with **notes**:
  - scope note, definition, editorial note, and history note
- Concepts can additionally be members of **collections**, which can be labeled or ordered

W3C SKOS

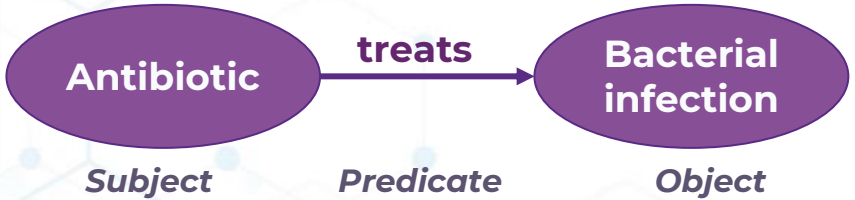
KG Components: Taxonomies

- Centrally managed taxonomies (not a taxonomy built in a siloed application), now tend to be built on the SKOS data-exchange model.
- Since SKOS is based on RDF, SKOS taxonomies are easily managed in RDF graph databases, and connect to the data, other taxonomies, and ontologies, in addition to linking to content.

KG Components: Ontologies

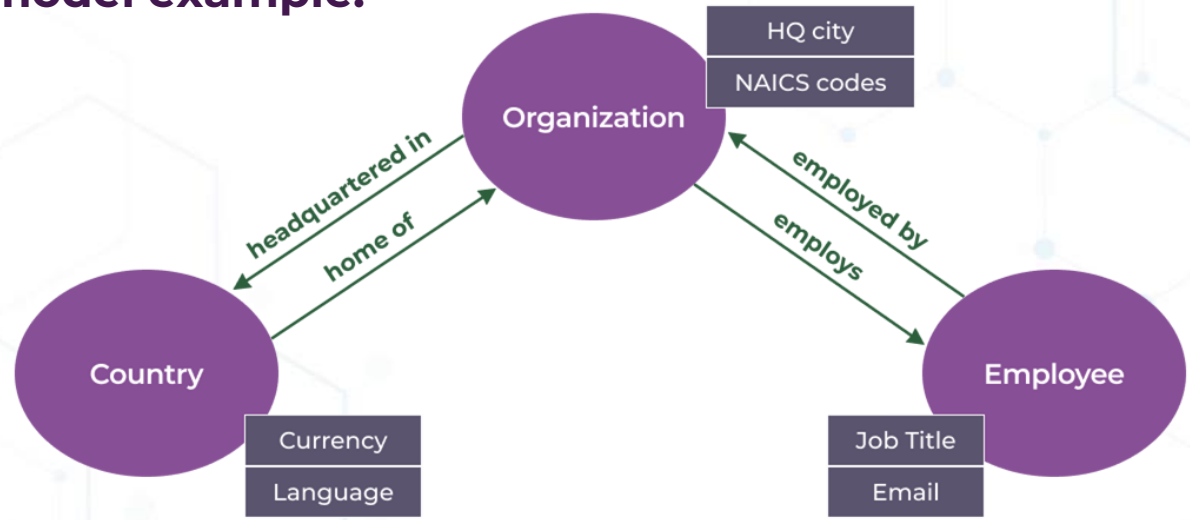
Ontology

- A model of a knowledge domain
- Similar to (most of) a knowledge graph, but doesn't include all actual instance data
- A formal naming and definition of the types (classes), attribute properties, and interrelationships of entities in a particular domain
  - Relations contain meaning, or are "semantic"
  - Properties are customized attributes of entities
- Standards provided by W3C: Web Ontology Language (OWL) and RDF-Schema
- A set of of precise descriptive statements about a particular domain
- Statements are expressed as subject-predicate-object triples
- Comprises classes, relations, and attributes, which are linked in statements of triples



KG Components: Ontologies

Ontology model example:



Classes: Employee, Country, Organization

Relations: headquartered in < > home of
employed by < > employs

Attributes: Email address, Job title, HQ city, NAICS codes, Currency, Language

KG Components: Ontologies

W3C Standards and Guidelines for for Ontologies

RDF (Resource Description Framework)
www.w3.org/TR/rdf11-concepts
"A standard model for data interchange on the Web" modeled in triples

RDFS (RDF-Schema)
www.w3org/2001/sw/wiki/RDFS
"A general-purpose language for representing simple RDF vocabularies on the Web"

- Goes beyond RDF to designate classes and properties of RDF resources, as ontology basics

OWL (Web Ontology Language)
www.w3.org/OWL
"A Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things"

- An extension of RDFS

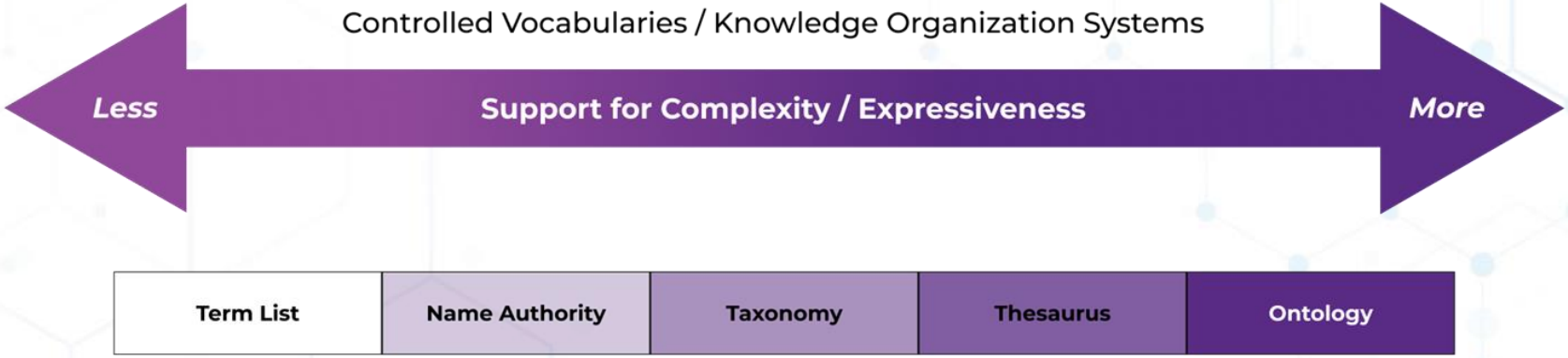
SPARQL (SPARQL Protocol and RDF Query Language)
https://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/
Language to query and updated RDF data

KG Components: Ontologies

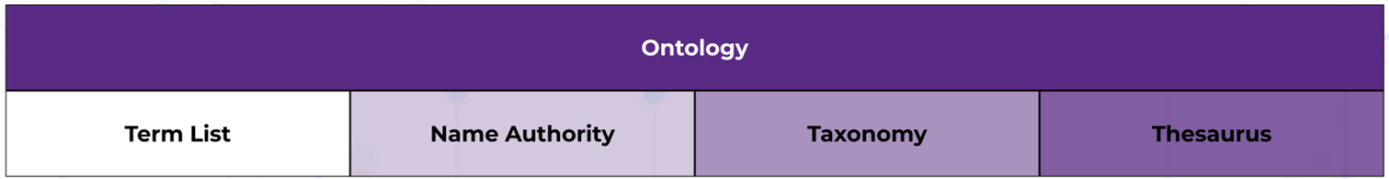
OWL-Specific Ontology Components

- **Entities** – subjects (domains) or objects (ranges) of triples *- graph nodes*
  - Classes
    - Named sets of concepts that share characteristics and relations
    - May group subclasses or individuals (instances of the class)
  - Individuals
    - Members or instances of a class (may be managed in a linked taxonomy)
- **Properties** – predicates of triples, about individuals *- graph edges*
  - Object properties
    - **Relations** between individuals
    - May be directed, symmetric, or with an inverse
  - Datatype properties
    - **Attributes** or characteristics of individuals
    - The object of a datatype property is a value
- **Literals** – values of attributes (metadata values)

KG Components: Ontologies + Taxonomies



An ontology is a semantic layer that links to and enhances other controlled vocabularies.



KG Components: Ontologies + Taxonomies

What you cannot do with a taxonomy alone, but can with an added ontology:

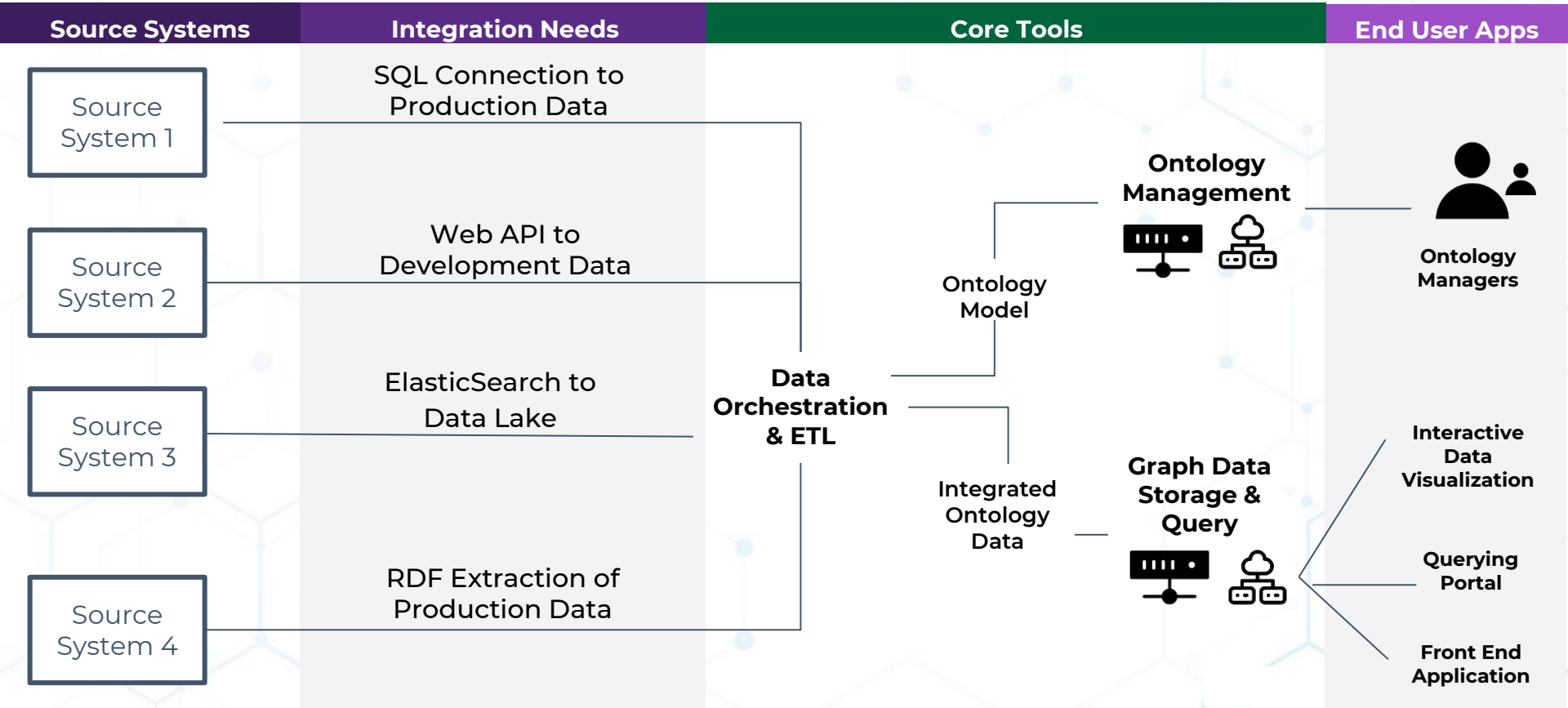
- Model **complex interrelationships** (e.g. in product approval or supply chain processes) and also connect to content
- Perform complex **multi-part searches**: e.g. find contacts in a specific location, who are employed by companies which belong to certain industries
- Search on more specific **criteria** that vary based on category (class)
- Explore **explicit relationships** between concepts (not just broader, narrower, related)
- **Visualize** concepts and semantic relationships
- Perform **reasoning** and **inferencing** across data
- Search across **datasets**, not just search for content
- **Connect** across **siloed** content and data repositories across the enterprise

Building a Knowledge Graph

Steps to building a knowledge graph:

1. Identify use cases, or problems to be solved.
2. Inventory and organize relevant data and content.
3. Identify and map relationships across data: design and implement an ontology.
4. Incorporate sample data in a graph database.
5. Connect to the ontology/taxonomy, as a test proof of concept.
6. Connect to or build user applications and interfaces.
7. Automate and scale with data pipelines, auto-tagging, and AI.

Building a Knowledge Graph: Sample Infrastructure



Building a Knowledge Graph

Core software and technology needed:

- Graph database management software
- Taxonomy/ontology management software based on W3C standards
- Search software (such as Solr or Elasticsearch)
- Front-end (web) application

Also important:

- Extract-Transform-Load (ETL) tool to extract data
- Text mining/natural language processing/entity extraction tool
- Machine-learning auto-classification tool
- Capabilities (such as algorithms for weighting/scoring relations) specified in SPARQL query language for RDF

Building a Knowledge Graph

Collaboration of roles:

- Data engineers
- Data scientists
- Data analysts
- Data architects
- Knowledge engineers
- Taxonomists
- Ontologists
- Content strategists
- Solutions architects
- Software engineers
- Web developers
- Information architects

Challenges/Requirements:

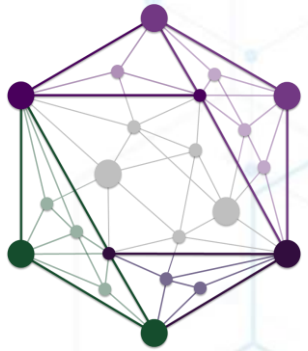
- A specific business/use case, not just curiosity to try new technologies
- Implementation expertise with software tools and guidance from consultants
- Commitment from all stakeholders
- Sufficient time, effort, and expertise to deal with a very complex project
- Data quality

Knowledge Graph Applications

- Semantic search
- Recommendation
- Compliance and risk prediction
- Question answering engines

- Chatbots
- Insight engines
- Expert finder
- Customer 360

An organization typically builds its own web-browser-based knowledge graph application.



Further Reading

Enterprise Knowledge White Papers:

- "How to Optimize Data Governance with Enterprise Knowledge Graphs" August 22, 2019
- "Using Knowledge Graph Data Models to Solve Real Business Problems" June 10, 2019

Enterprise Knowledge Blog Articles:

- "How a Knowledge Graph Supports AI: Technical Considerations" September 26, 2023
- "How a Knowledge Graph Can Accelerate Data Mesh Transformation" July 11, 2023
- "Elevating Your Point Solution to an Enterprise Knowledge Graph" November 16, 2022
- "Digital Twins and Knowledge Graphs" May 5, 2022
- "Where Does a Knowledge Graph Fit Within the Enterprise?" April 21, 2022
- "Integrating Search and Knowledge Graphs" October 19, 2020
- "How to Build a Knowledge Graph in Four Steps: The Roadmap From Metadata to AI" September 9, 2019

Q&A

Thank you for listening.

Questions?

Heather Hedden
Senior Consultant
Enterprise Knowledge, LLC
www.enterprise-knowledge.com
hhedden@enterprise-knowledge.com
www.linkedin.com/in/hedden

ENTERPRISE KNOWLEDGE