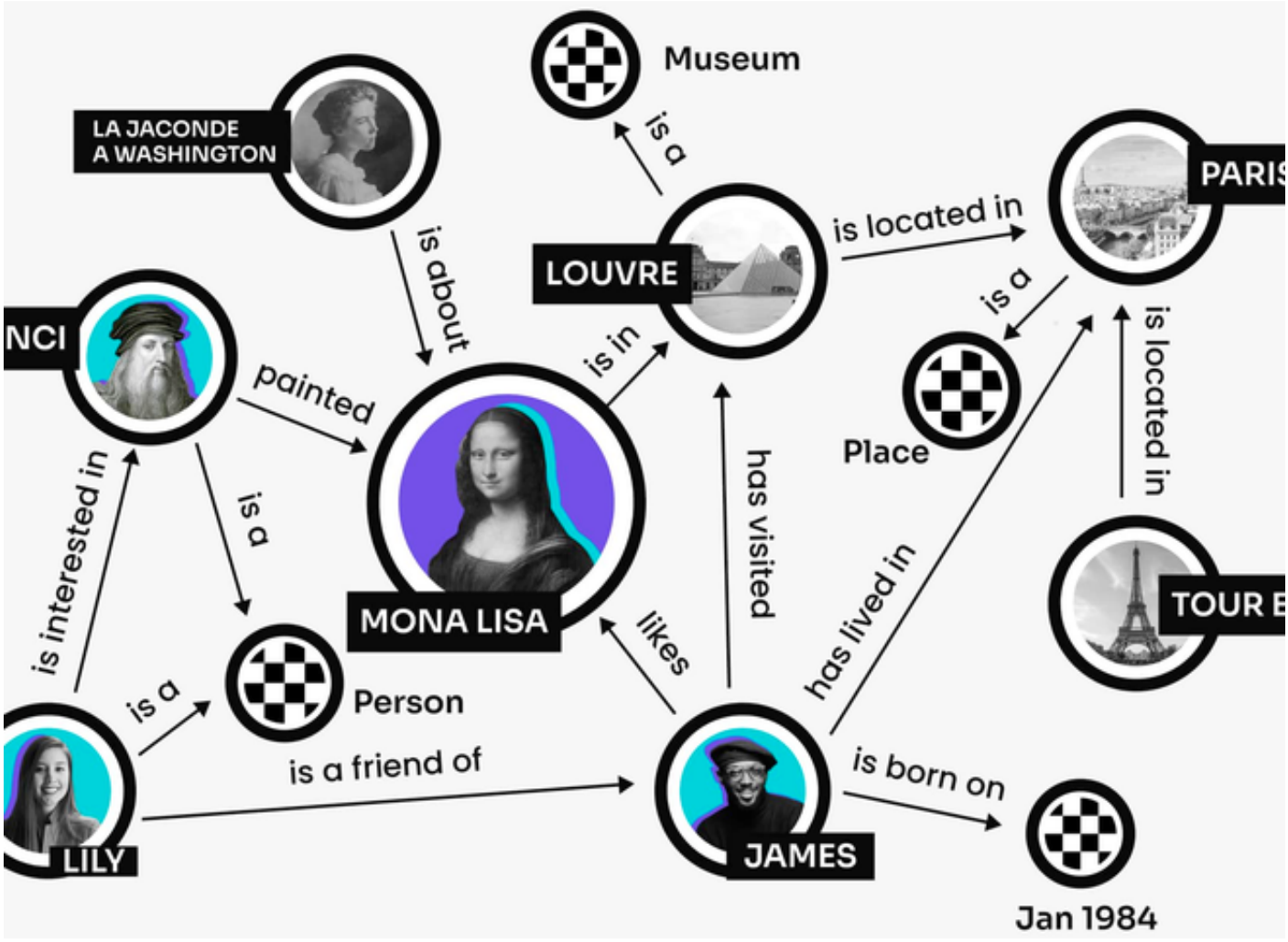
Knowledge Graph

Jennifer Morales (C)
Contingent Worker



What are Knowledge Graphs

A knowledge graph, also known as a semantic network, represents a network of real-world entities—i.e. objects, events, situations, or concepts—and illustrates the relationship between them. This information is usually stored in a graph

database and visualized as a graph structure, prompting the term knowledge "graph." – *IBM*

Most definitions view the topic through a Semantic Web lens and include these features:

- *Flexible relations among knowledge in topical domains:* A knowledge graph (1) defines abstract classes and relations of entities in a schema, (2) mainly describes real world entities and their interrelations, organized in a graph, (3) allows for potentially interrelating arbitrary entities with each other, and (4) covers various topical domains.
- *General structure*: A network of entities, their semantic types, properties, and relationships.
- *Supporting reasoning over inferred ontologies*: A knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge.

There are, however, many knowledge graph representations for which some of these features are not relevant. For those knowledge graphs this simpler definition may be more useful:

- A digital structure that represents knowledge as concepts and the relationships between them (facts). A knowledge graph can include an ontology that allows both humans and machines to understand and reason about its content

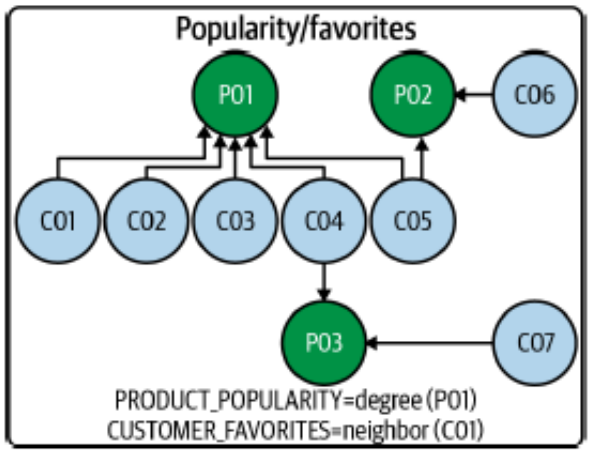
- Wikipedia

Graphs and Graph Databases

a graph database is a database "purpose-built to store and navigate relationships. Relationships are first-class citizens in graph databases, and most of the value of graph databases is derived from these relationships. Graph databases use nodes to store data entities, and edges to store relationships

between entities. An edge always has a start node, end node, type, and direction, and an edge can describe parent-child relationships, actions, ownership, and the like. There is no limit to the number and kind of relationships a node can have". - *AWS*

- Graphs are simple structures that use nodes (or vertices) connected by relationships (or edges) to create high-fidelity models of a domain – *Neo4j*
- Graph data models are able to represent complex, indirect relationships in a way that is both human readable, and machine friendly.



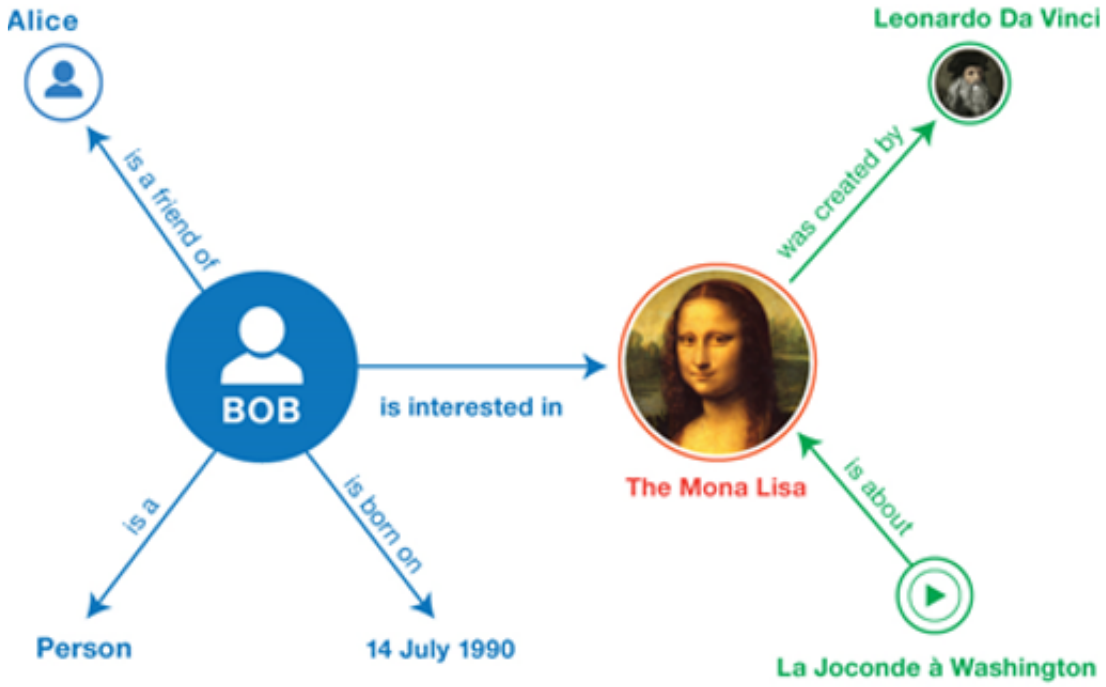
A simple graph showing instances of Customers related to Products bought

Triples

Triples are the fundamental data structure of the Resource Description Framework (RDF) that is used to define the semantic of the knowledge graph.

A triple is made up of a subject, a predicate and an object.

In the example below, each predicate or edge describes a triple such as Subject: Bob, Predicate: is interested in, Object: Mona Lisa. An Object in one triple can be a Subject in another and visa versa so that the triples can be networked together to form a graph.



Property Graphs

A property graph model allows both nodes and relationships to contain properties.

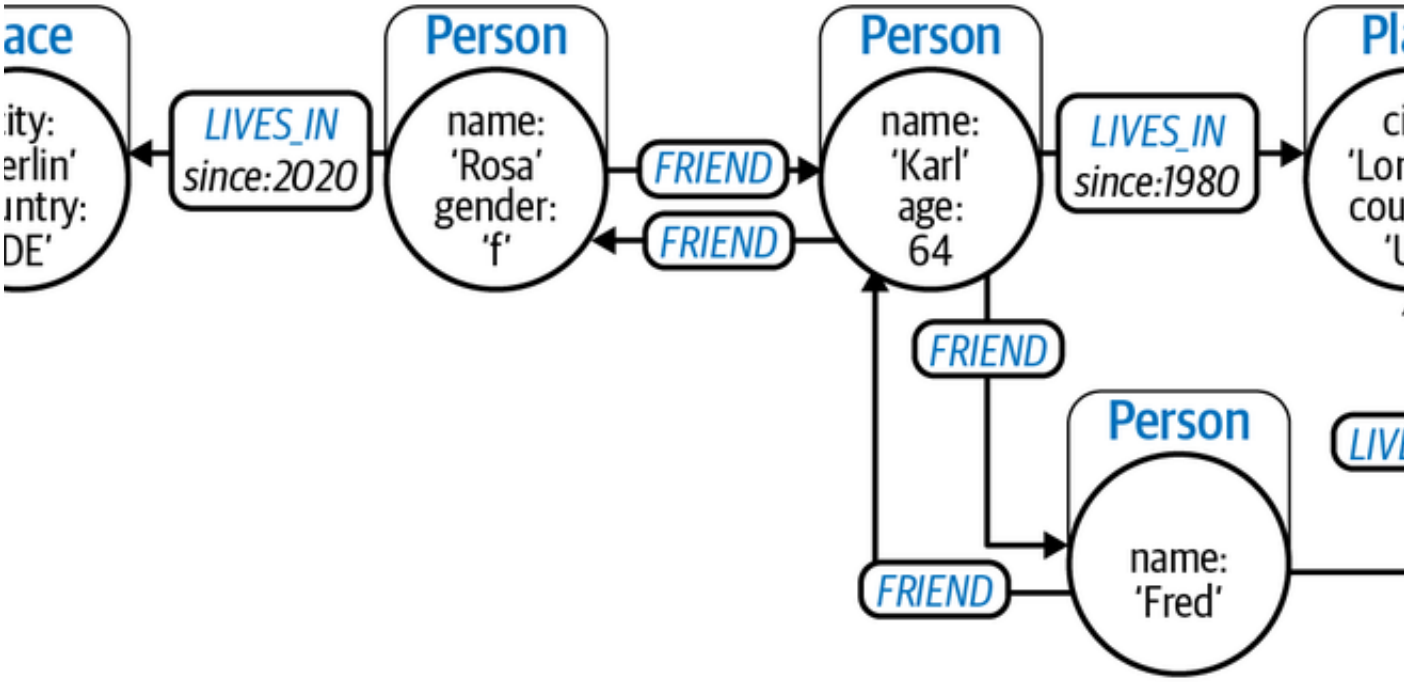
Nodes representing entities in the domain:

- Nodes can contain zero or more properties, which are key-value pairs representing entity data such as price or date of birth
- Nodes can have zero of more labels, which declare the node's purpose in the graph such as representing customers or products

Relationships representing how entities interrelate:

- Relationships have a type, such as bought or liked
- Relationships have a direction from one node to another
- Relationships can contain zero or more properties representing some characteristic of the link such as timestamp or distance
- Relationships always have a start and end node (which can be the same

node)



A Graph showing properties on Nodes such as Age or Gender on a Person node and properties on Relationships such as Since on the Lives In relationship

Ontologies

Well-designed ontologies provide a declarative encoding of the meaning of vocabulary terms that are critical to enabling communication, among people and between machines. They are the starting point for automated extraction of content from documents using natural language processing (NLP) techniques, for seeding learning algorithms, for question answering and recommendation systems, and for understanding and leveraging data gathered from independently developed systems and repositories.

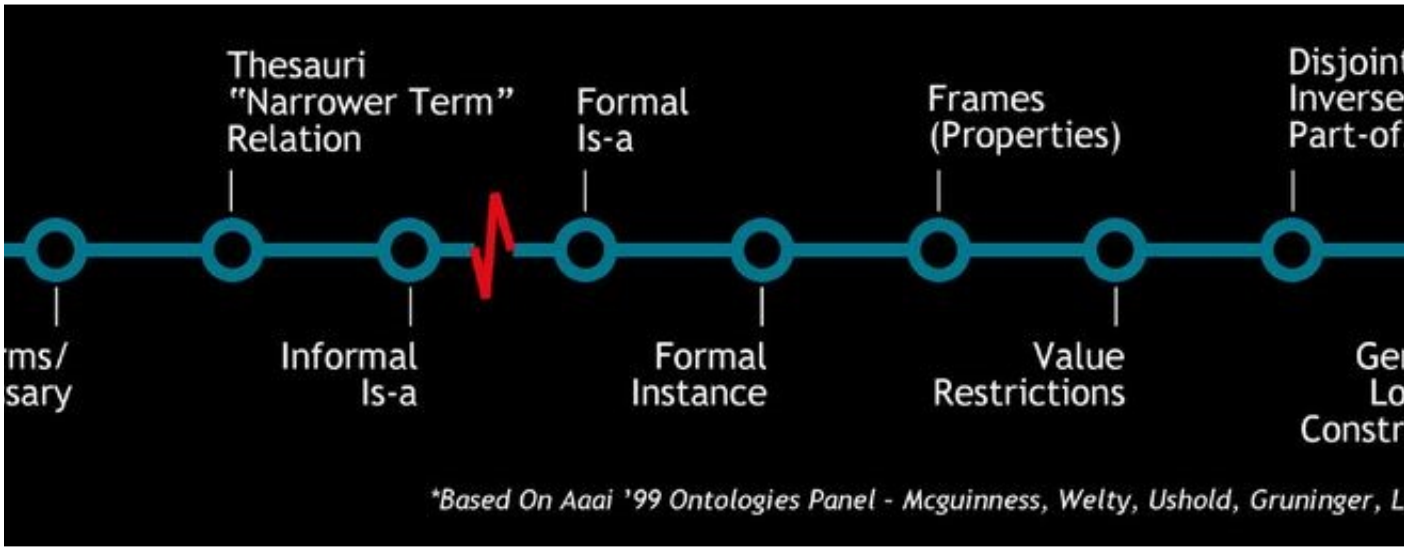
Ontologies play a key role in the vision of the Semantic Web, for example, where they provide the semantic vocabulary used to annotate websites in a way

meaningful for machine interpretation. The Semantic Web is an extension of the World Wide Web through standards set by the World Wide Web Consortium (W3C). To enable the encoding of semantics with the data, technologies RDF and Web Ontology Language (OWL) are used to formally represent an ontology that can describe concepts, relationships between entities, and categories of things. The RDF is the basic framework that the rest of the Semantic Web is based on. RDF provides a mechanism for allowing anyone to make a basic statement about anything and layering these statements into a single model.

An Ontology specifies a rich description of the:

- Terminology, Concepts, Nomenclature
- Relationships among concepts and individuals
- Sentences distinguishing concepts, refining definitions and relationships (constraints, restrictions, regular expressions)

The scope of an ontology is relevant to a particular domain or area of interest.



An ontology is a more formal semantic understanding of terms beyond a simple catalog or glossary which enables humans and machines to reason and gain insight on data

Organizing Principles

What transforms a graph into a knowledge graph is the application of an organizing principle that helps humans and software users to understand it. Sometimes this is called semantics, or it can just be thought of as making data smarter.

- To integrate data from multiple repositories/sources, it's important to know what terms mean – even if just as simple as a "car" is synonymous to "auto"
- Rather than having to repeatedly encode smart behavior into applications, we encode it once, directly into the data. Smarter data benefits knowledge reuse and reduces duplication and discrepancies.
- To collaborate with other people or machines, it is important to agree on terminology and on the context for the meanings of terms

Standards

- There are several standard or widely used ontologies that service a variety of domains, such as Financial Industry Business Ontology (FIBO) for finance and business, Languages, Countries and Codes (LCC), and Schema.org and Dublin Core for general purpose web resource annotation.
- For interoperability or to reuse an existing public model, if working in a domain for which such a standard exists, it may worth considering adopting that model.

Custom

- To build or adapt an ontology to a business domain it can be defined formally using one of the standard languages available. The most widely used ones are Resource Description Framework (RDF) and Web Ontology Language (OWL).

Hybrid

- A mix of standards and adaptive customization is often most aligned with the reality of modern business

Knowledge Graph Capabilities and Uses

Knowledge graphs provide numerous capabilities that enable interoperability, decision support, and insights

- Data linkage and integration despite silos
- Open global reusable data standards
- Alignment based on meaning
- Highly expressive data schemas with built in rules that reflect concepts
- Flexible changeable schemas
- Rich multi-level taxonomies

Some common use cases for knowledge graphs include:

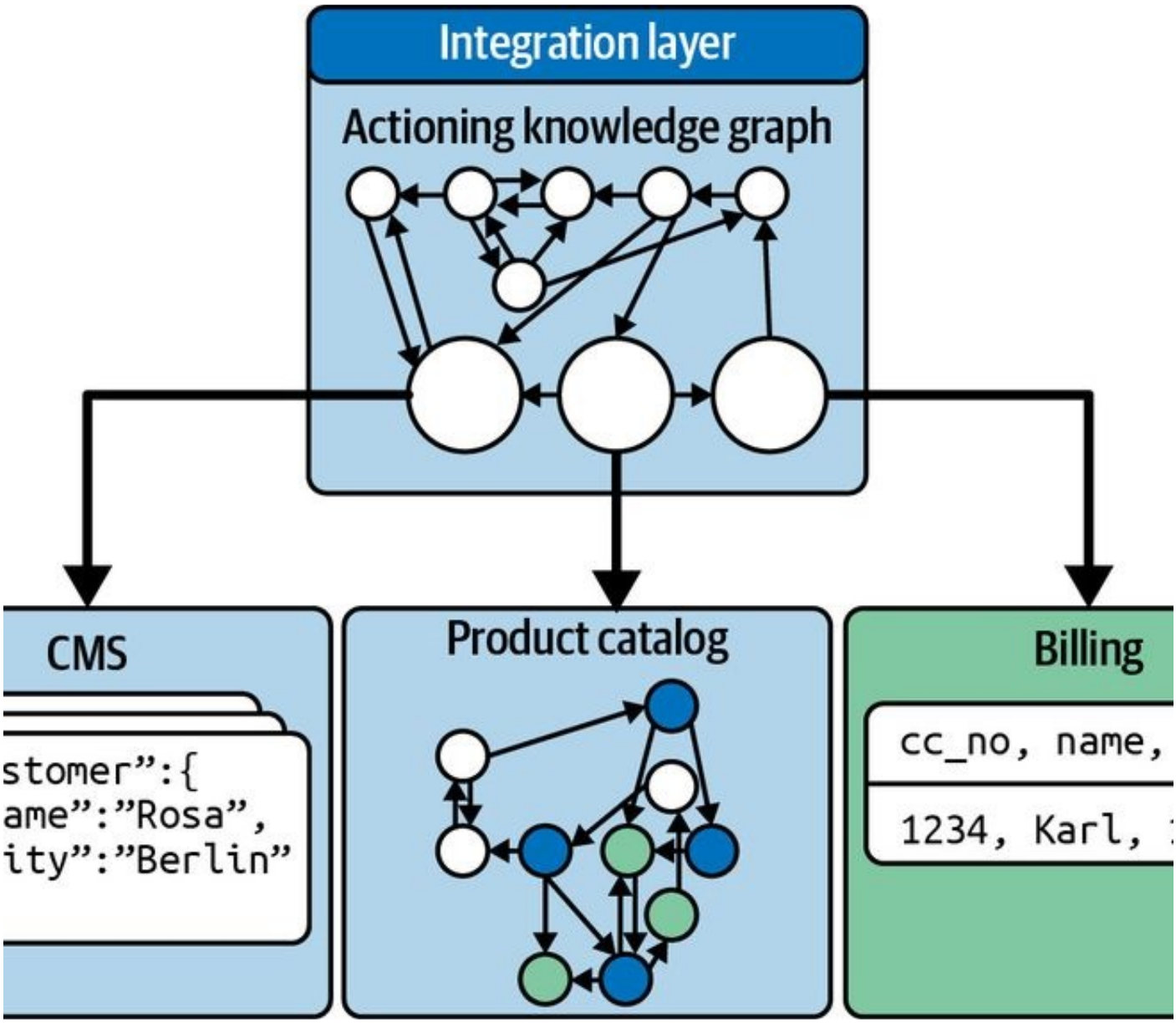
- Data Lineage - trace all steps in data pipelines from source to consumer with high-fidelity provenance information
- Impact analysis and root cause analysis - assess the direct or indirect effects of an event or change in an ecosystem, or map symptoms or consequences to originating causes.
- Information Search - documents, lessons learned, and knowledge in general can be indexed in a knowledge graph to search *things instead of string*s.
- Single view of an entity (customer, product, etc.) - provides a trusted and contextualized aggregation of all information relevant to an entity.
- Fraud detection - based on detecting communities of like behavior, unusual transactions, or suspicious commonalities.
- Customer Experience - surfacing complex sequences of activities for journey analysis
- Recommendations - based on customer history, seasonality, inventory, and product influence on sales of other items.
- Machine learning and Artificial Intelligence - provide contextual knowledge to include in training data and to maintain context for explainability of AI decisions.
- Digital Twin - create a rich virtual view of the business close to reality. The elements in the real world and how they interrelate are captured in a property graph, and the constraints and rules that govern the real world a

captured in the graph.

Knowledge Graphs in Data Fabric

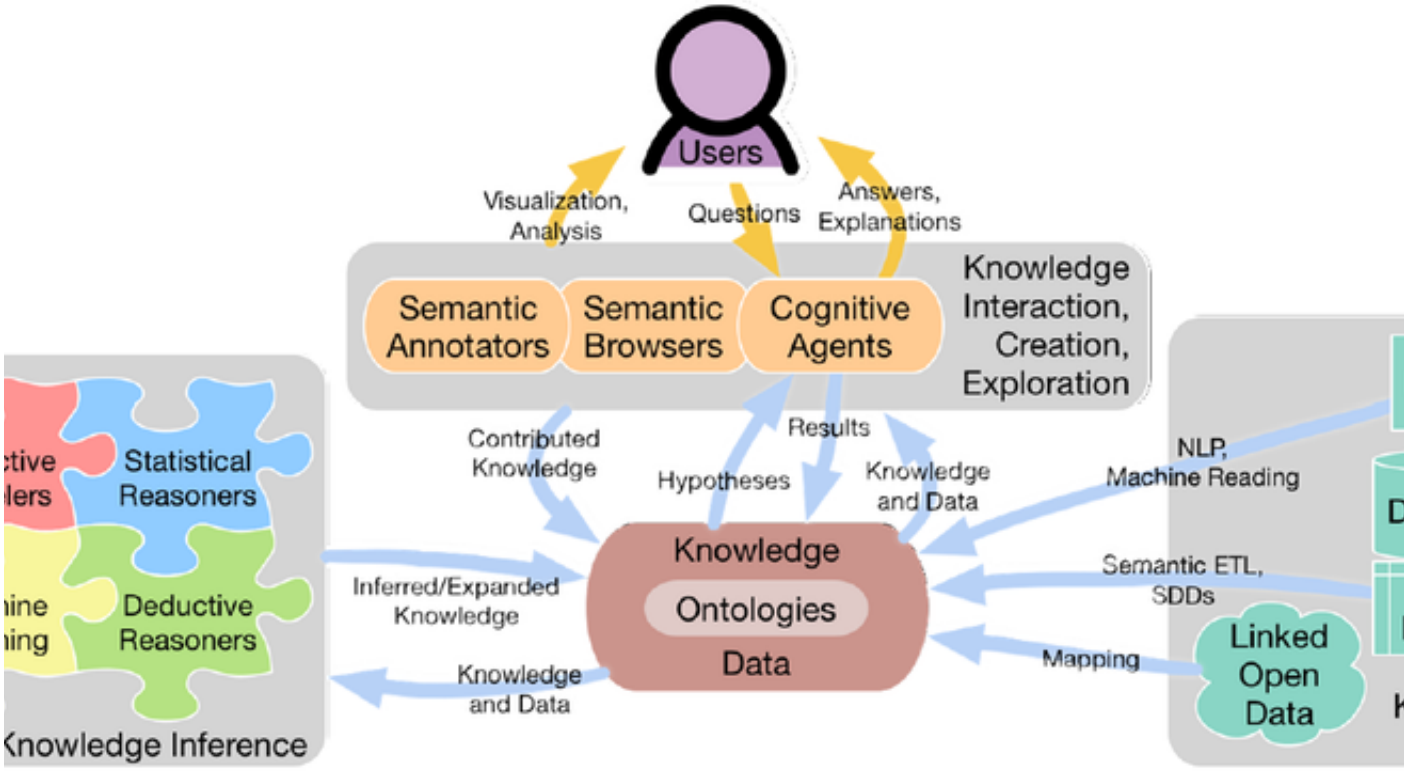
Knowledge Graphs can be used to provide integrated access to multiple data domains in a data fabric.

The knowledge graph provides an index and perhaps some metadata management across data systems and can act as an entry point for application or user requests that can follow the relationships in the graph to leaf nodes, which redirect to records in other systems.



Knowledge Framework

The knowledge graph and its ontologies are important pieces, but are best implemented as part of a larger integrated framework



Ideation & Emerging Technology

Edit **Contact Us**