Knowledge Graphs

Why, What and How?

Ali Khalili, PhD

May 2022

About me

VU VRIJE UNIVERSITEIT AMSTERDAM

Located in

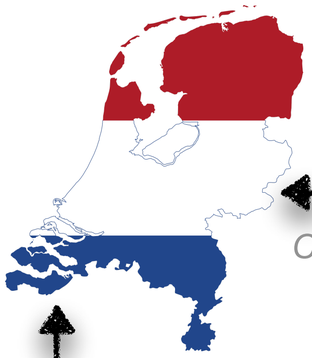
Professor

Frank van Harmelen

Deloitte.
Risk Advisory

knows

Works at



Capital of

Amsterdam

Lives in

Ali Khalili

Likes

JavaScript

Server-side

Likes

node

Neighbour

Located in

Leipzig

Lived in
{PhD}

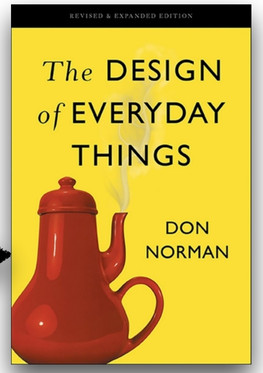
Lived in
{Before 2009}

knows

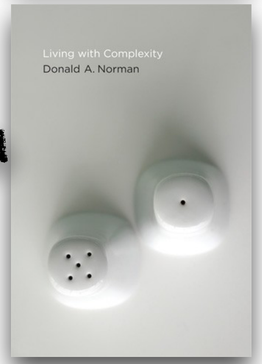
Favourite author

Don Norman

Author of



Author of



Sören Auer

Lives in

Capital of

Tehran

Co-founder

DBpedia

Why Knowledge Graphs?

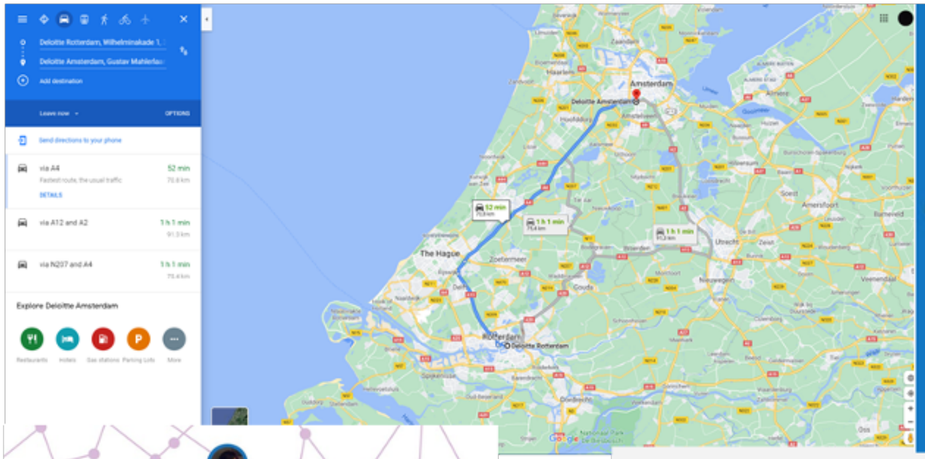
Knowledge Graphs form the foundation of many modern AI & data analytics capabilities

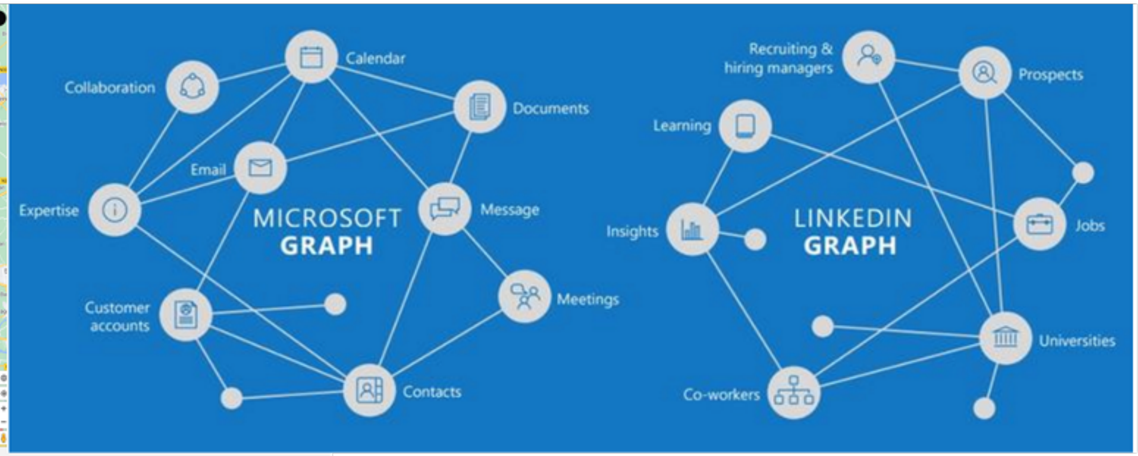
Gartner predicts that by 2025, graph technologies will be used in **80%** of data and analytics innovations, up from 10% in 2021, facilitating rapid decision making across the organization.

Graph Technology Landscape

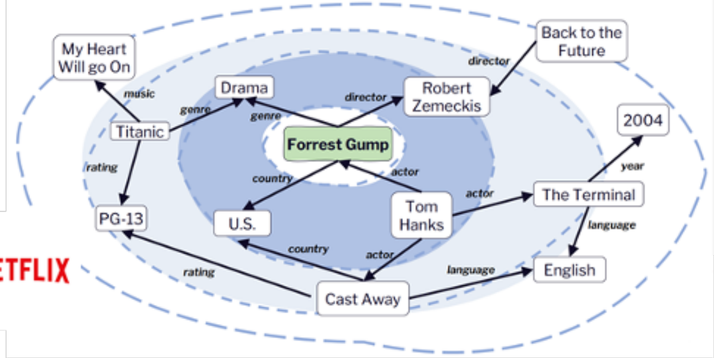
Who is using Knowledge Graphs?

Knowledge Graphs are Everywhere…

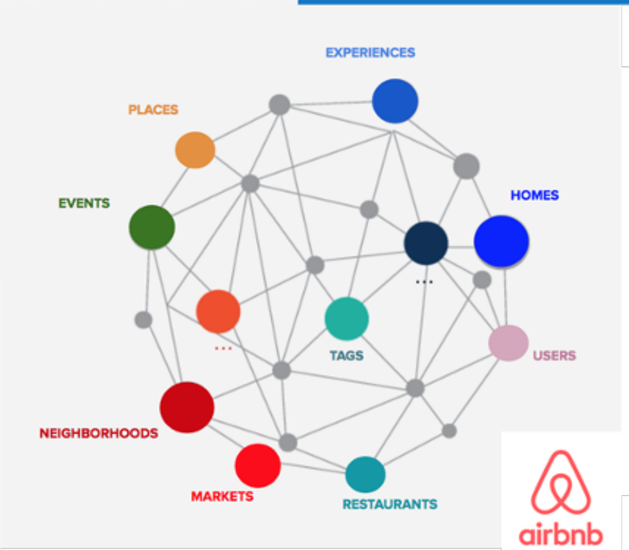




facebook

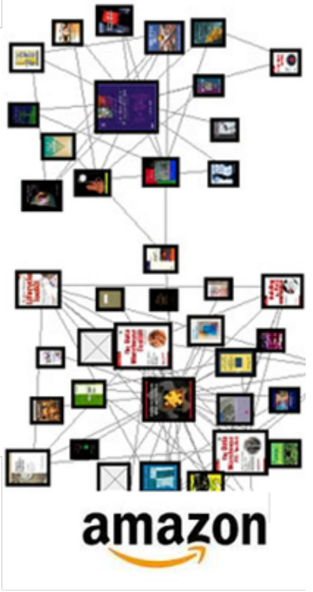


NETFLIX



airbnb

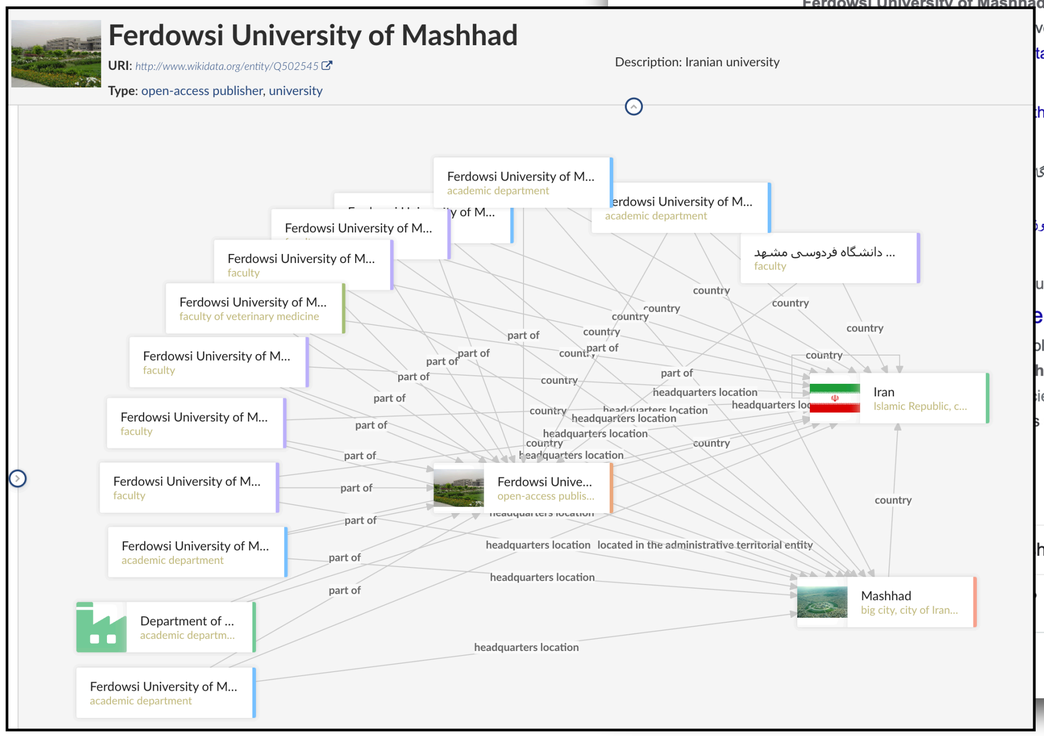
Uber Eats

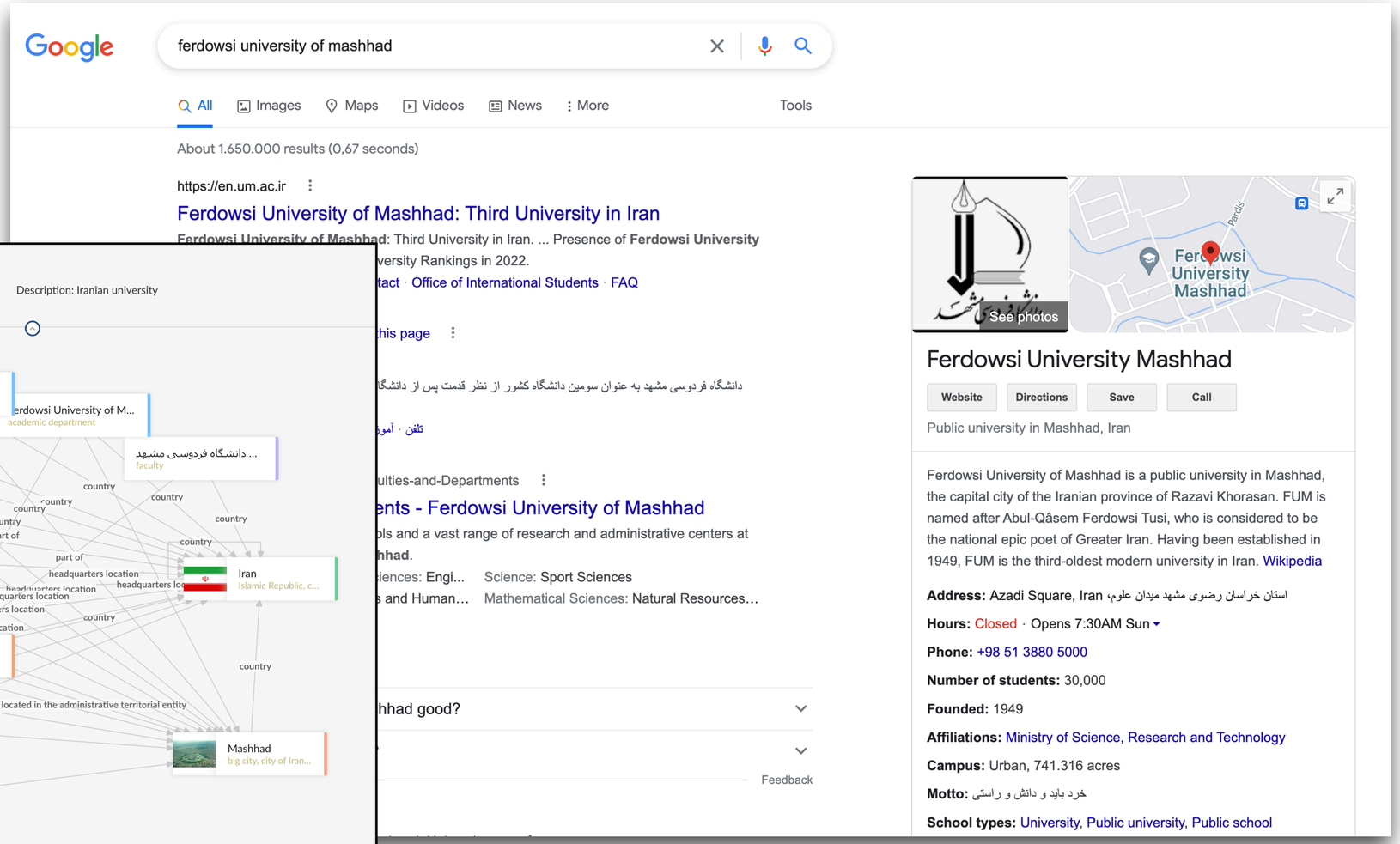


amazon

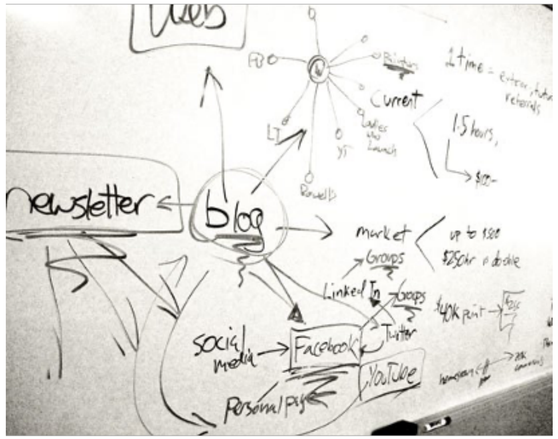
What is a Knowledge Graph?

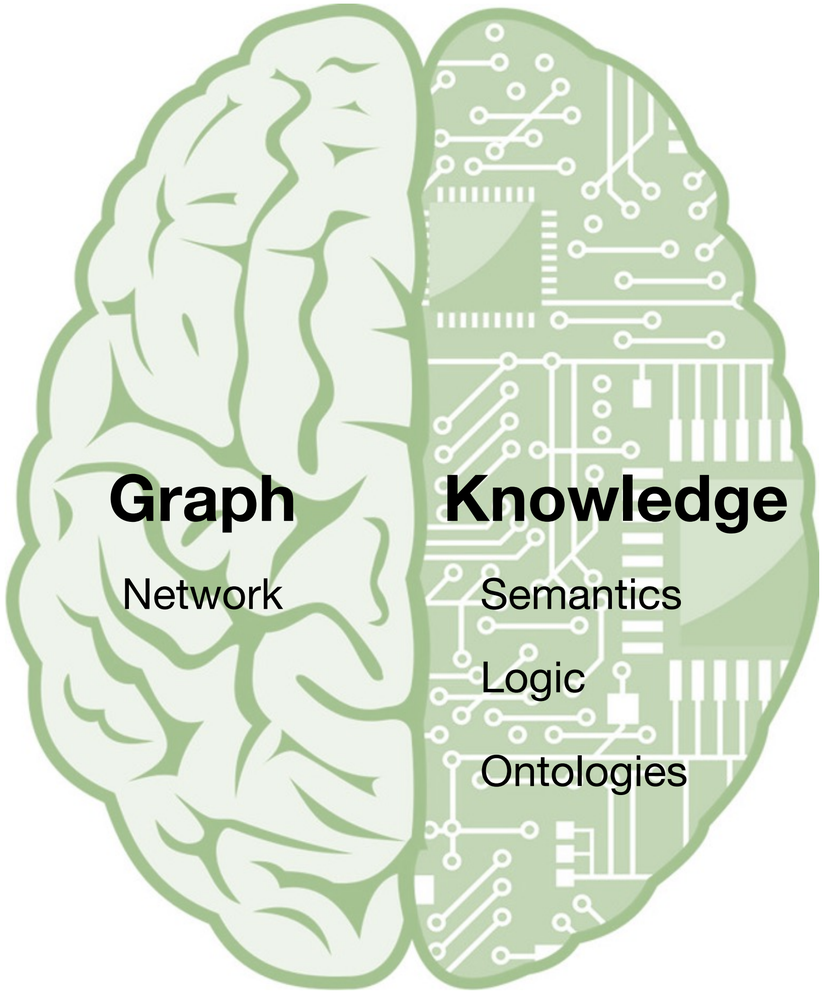
A knowledge graph is a means to represent knowledge in a domain of interest using a graph structure.





A knowledge graph is a means to represent knowledge in a domain of interest using a graph structure.





Human-oriented Machine-oriented

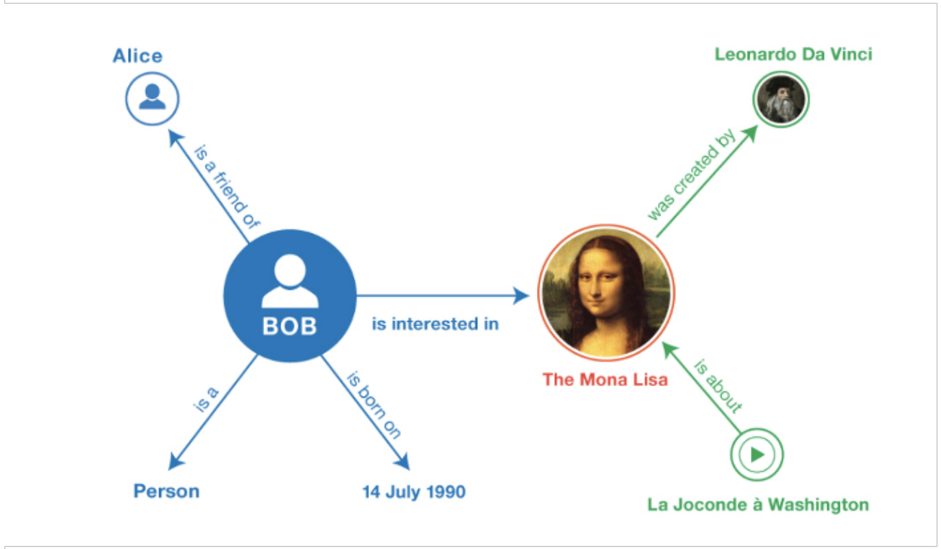
DL	Syntax	Semantics	Name
$\mathcal{EL}$	$\top$	$\Delta^{\mathcal{I}}$	top
	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$	conjunction
	$\exists R.C$	$\{a \mid \exists b : (a,b) \in R^{\mathcal{I}} \wedge b \in C^{\mathcal{I}}\}$	existential restriction
$\mathcal{ALC}$	$\bot$	$\emptyset$	bottom
	$\neg C$	$\Delta^{\mathcal{I}} \backslash C^{\mathcal{I}}$	negation
	$C \sqcup D$	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$	disjunction
	$\forall R.C$	$\{a \mid \forall b : (a,b) \in R^{\mathcal{I}} \rightarrow b \in C^{\mathcal{I}}\}$	universal restriction
$\mathcal{SROIQ}$	$\geq (\leq)\ nR.C$	$\{a \mid \lvert\{b : (a,b) \in R^{\mathcal{I}} \wedge b \in C^{\mathcal{I}}\}\rvert \geq (\leq)\ n\}$	atleast(atmost) restriction
	R^{-}	$\{(b,a) \mid (a,b) \in R^{\mathcal{I}}\}$	role inverse
	$R \circ S$	$\{(a,c) \mid \exists b : (a,b) \in R^{\mathcal{I}} \wedge (b,c) \in S^{\mathcal{I}}\}$	role composition
	U	$\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$	universal role
	...		

How to build a Knowledge Graph?

1. Give all things a name

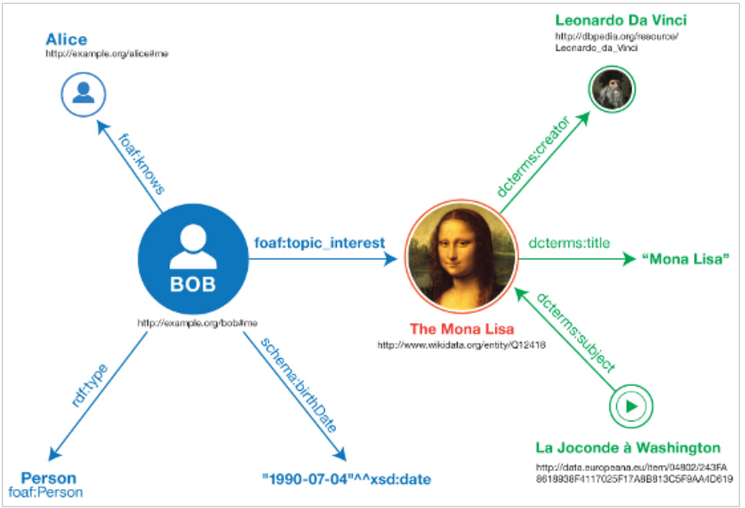
"Now! *That* should clear up a few things around here!"

2. Make a graph of relations between the things



This makes a *Giant Graph*

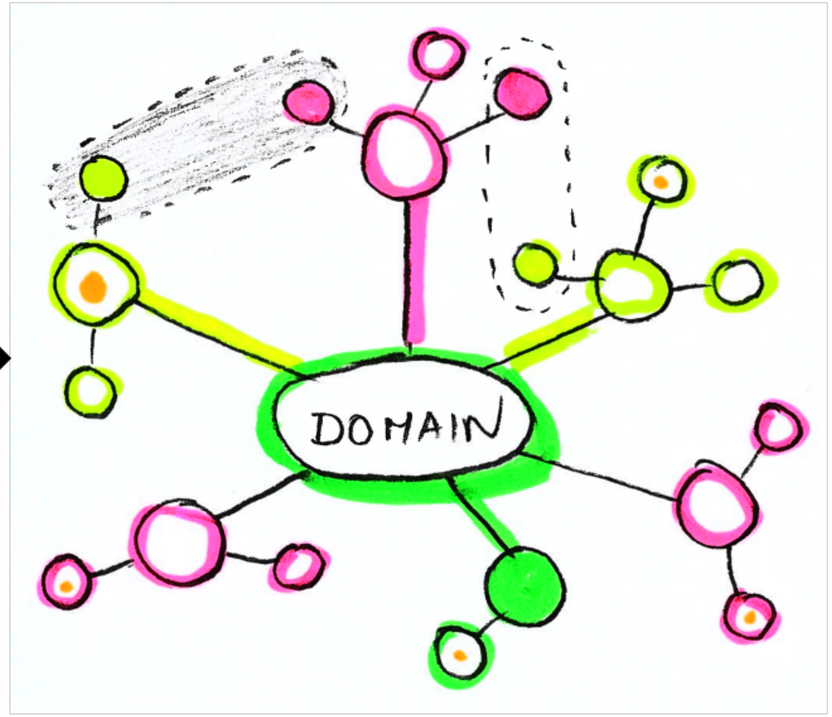
3. Make sure all names are URIs



This makes a *Giant Global Graph*

4. Add semantics (= predictable inference)

Ontology

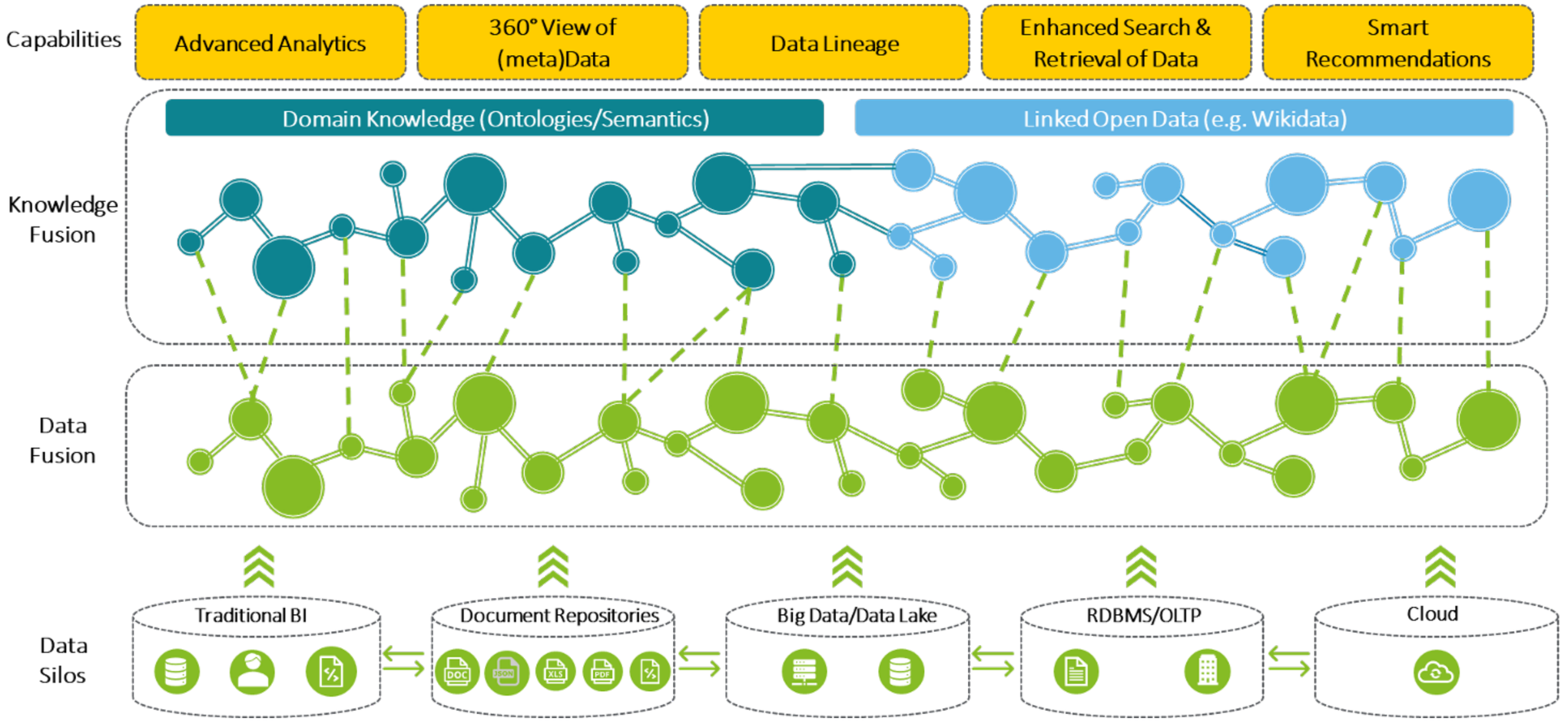


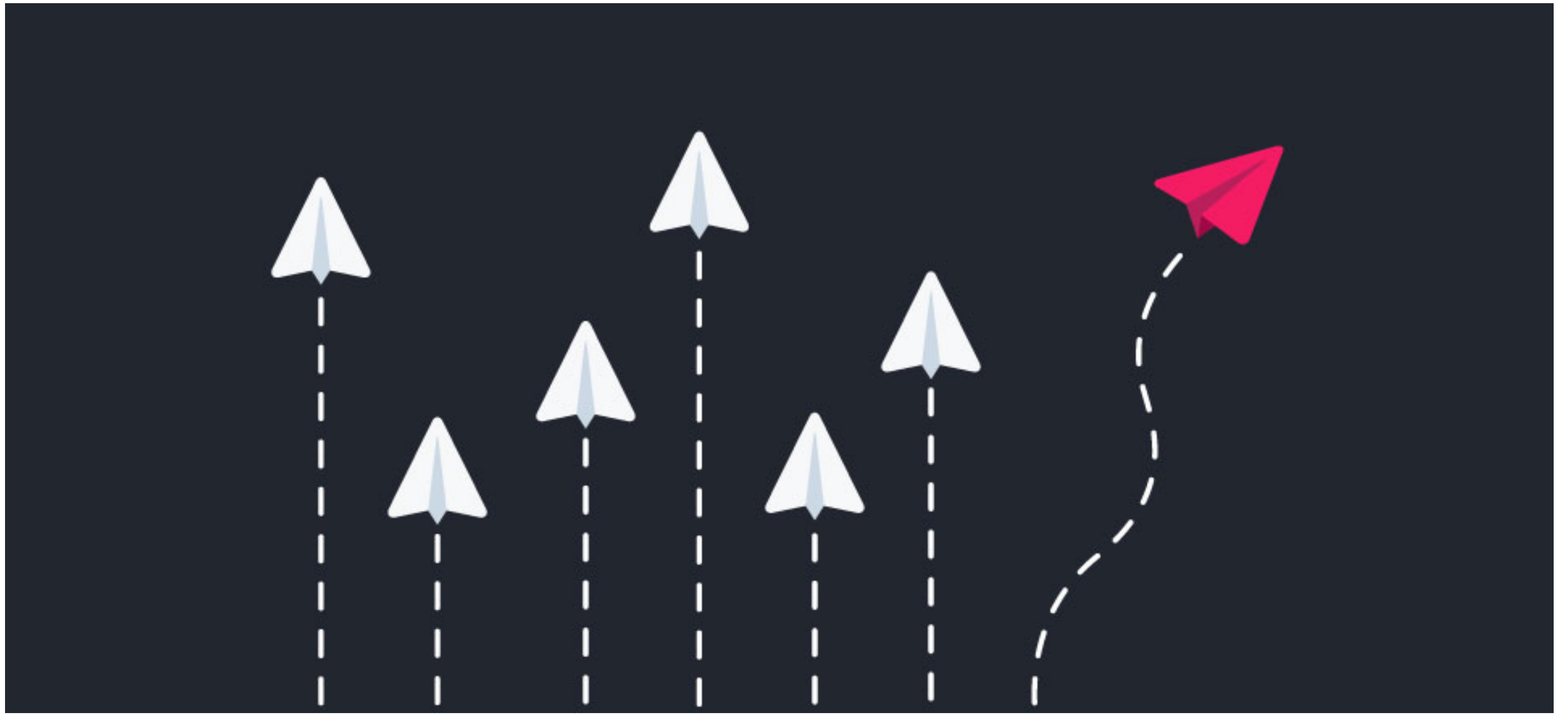
This makes a *Giant Global Knowledge Graph*

What is an Ontology?

- A formal explicit description of
  - **concepts** in a domain of interest (a.k.a classes),
  - **properties** of each concept,
  - and **restrictions** on concepts and properties
- An ontology together with a set of individual instances of classes constitutes a **knowledge base**.
- In reality, there is a fine line where the ontology ends and the knowledge base begins.

Enterprise Knowledge Graphs

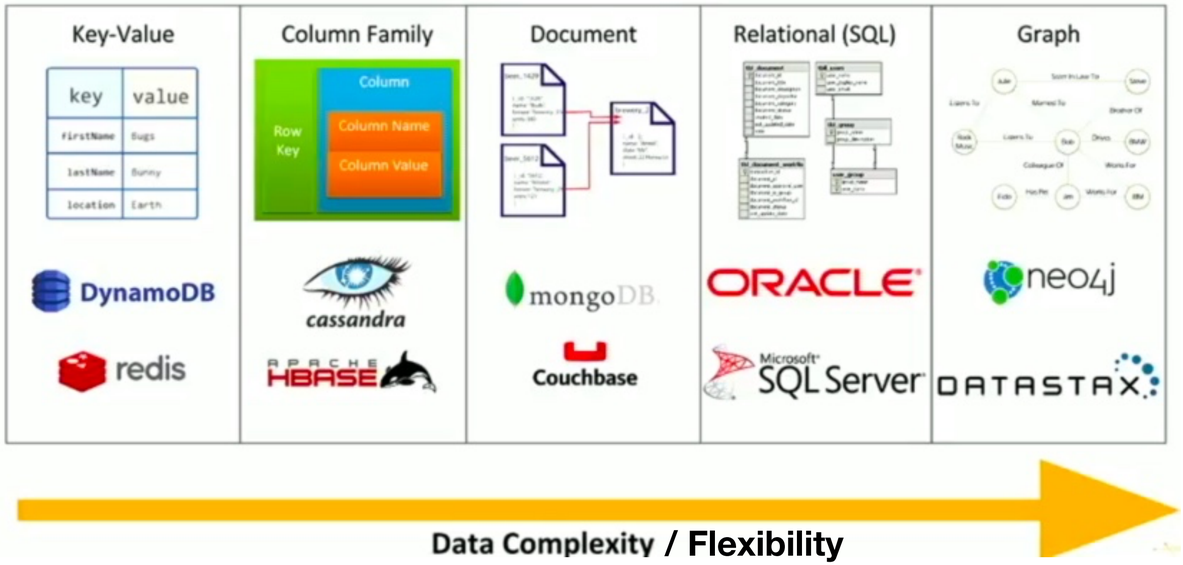


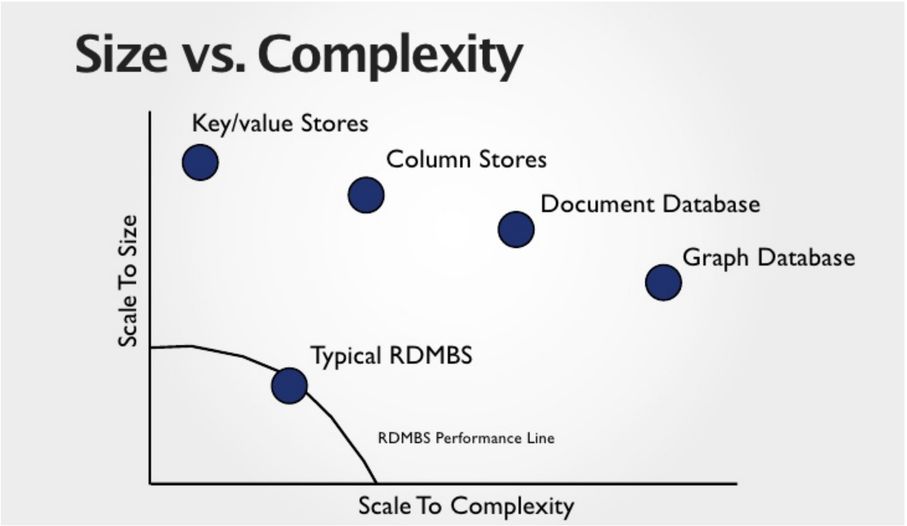


When to use a Knowledge Graph?

Knowledge Graphs are good when dealing with

1. Lots of **relationships** in your data (slow joins issue)
2. Need to **traverse many relationships** quickly
3. **High variability** data that does not fit well in a table
5. Your data model (schema) is constantly **changing**
6. **Complex** rules/patterns that need to be calculated quickly
7. **Integrate** disparate data sources
8. Need to derive **knowledge** from interconnected data





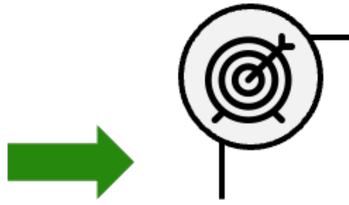
Resource Description Framework

Labelled Property Graphs

What type of Graph DB to use?

Functionality	Labelled Property Graph (LPG)	Resource Description Framework (RDF) graphs	Relevance for knowledge graphs
Theoretical foundations	**Simpler:** no semantics, no inference, basic graph theory graph: nodes + edge	**Formal:** interpretation, entailment, description logic triple: subject-predicate-object	
Associating properties with edges	Easy	Hard. Alternative: **RDF-star**	Important for **versioning/ metadata** addition
Standards	None (yet). **Community driven.**	Numerous W3C and OGC **standards**	Standards e.g. facilitate **mapping data to graph**
Processing multiple graphs	Hard	Very natural to handle **multiple (distributed) graphs** at the same time -> Semantic Web / Linked Data / FAIR Data vision	Ensures **scalability to new sources**
Schema standardization	Has no standard terms, vocabularies	Has many **reusable** curated terms, vocabularies, **ontologies**	Important for linking data
Data validation & reasoning	No standard way for data validation and reasoning	Standard ways such as SHACL, as well as different **reasoning** engines are available	Simplifies **data quality** management
Analytics	A **rich set of graph algorithms**: community detection, pathfinding, similarity detection, centrality, etc.	A limited set of graph algorithms	Important for graph analytics after a KG is created
Flexibility	A property graph can be modelled as a RDF graph	An RDF graph can be modelled as a property graph with a loss of semantics	

- **RDF graphs** present the most appropriate tool for Knowledge Graphs.
- **LPG** provides large advantage for **graph analytics** – for example on top of a structured “knowledge graph”



The best of both words can be achieved by a **hybrid approach**

Resource Description Framework

Labelled Property Graphs

ANZO

STARDOG

blazegraph

OrientDB

neo4j

DataStax Enterprise Graph

GraphDB

OPENLINK VIRTUOSO UNIVERSAL SERVER

ArangoDB

JanusGraph

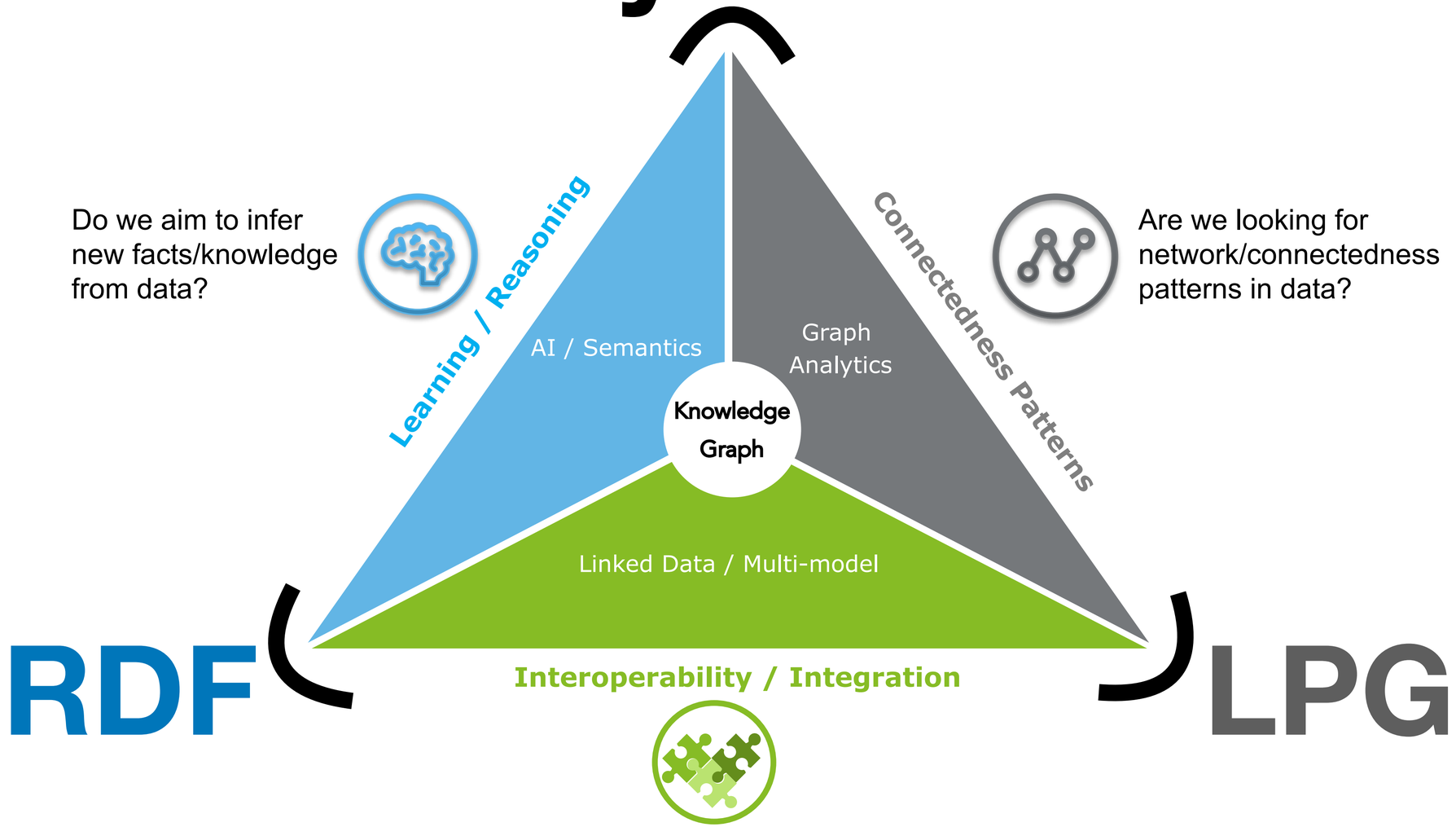
TigerGraph

AllegroGraph

MarkLogic

Amazon Neptune

Hybrid

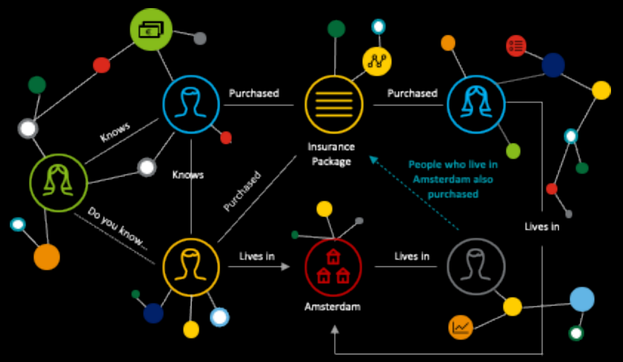

Do we aim to infer new facts/knowledge from data?
Learning / Reasoning
AI / Semantics
Graph Analytics
Connectedness Patterns
Are we looking for network/connectedness patterns in data?
Knowledge Graph
Linked Data / Multi-model
RDF
LPG
Interoperability / Integration
Do we need to integrate multiple varying data sources?

Example Use Cases of Knowledge Graphs

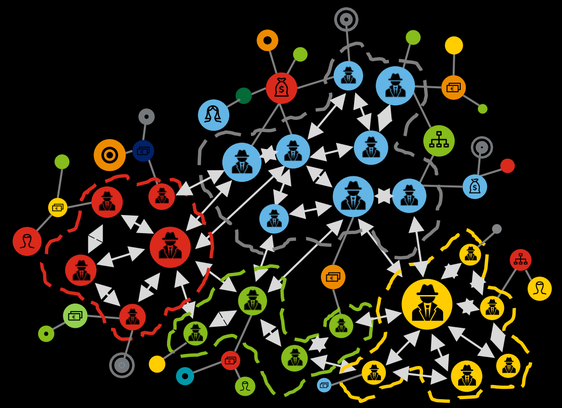
Use Cases of GraphTech in Financial Services [PDF]

Deloitte.

Using graphs, complex business analytics can be accelerated and data integration can be simplified with increased agility and cost efficiency



Semantic Search, Recommender Systems & Conversational AI



Fraud Detection & Financial Crime Analytics

- Spotting Fraud
- Anti Money Laundering
- Anti Terrorist Financing

Data Lineage & Metadata Management

- Risk Data Aggregation & Reporting
- Master Data Management
- Data Migration
- Impact Analysis

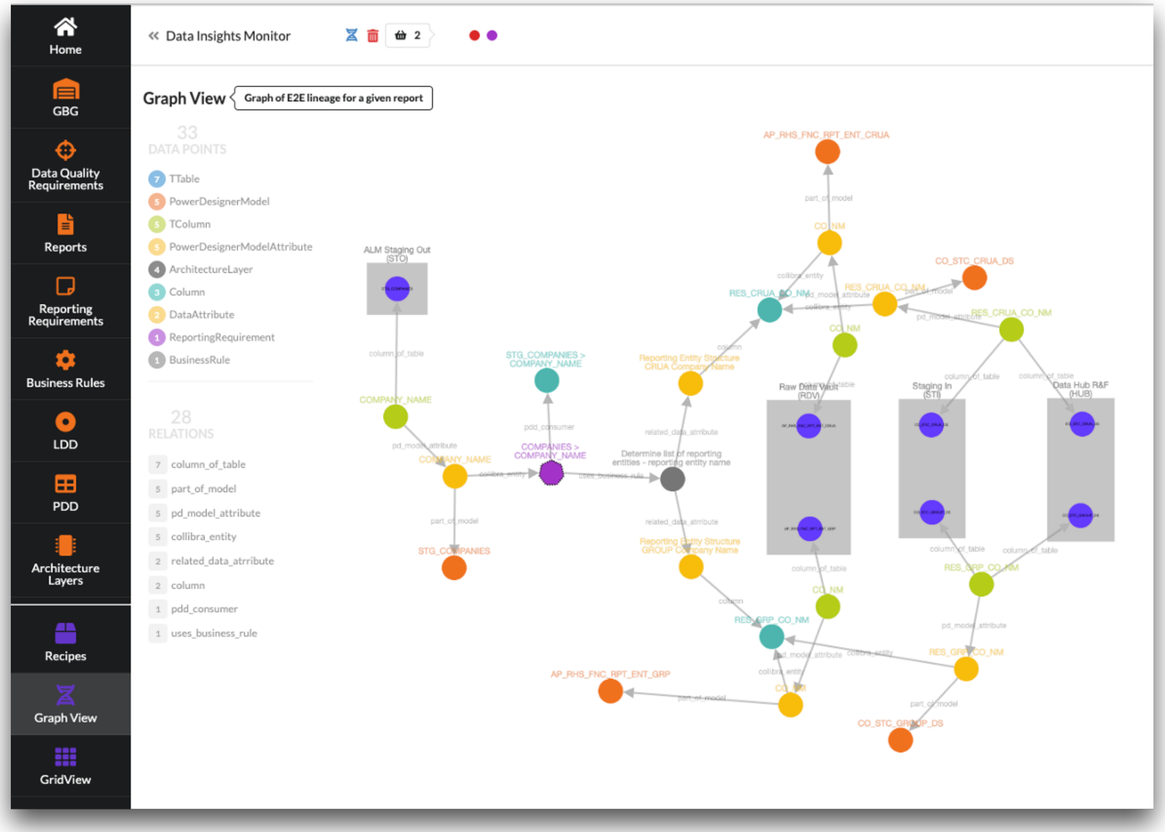
Compliance Management

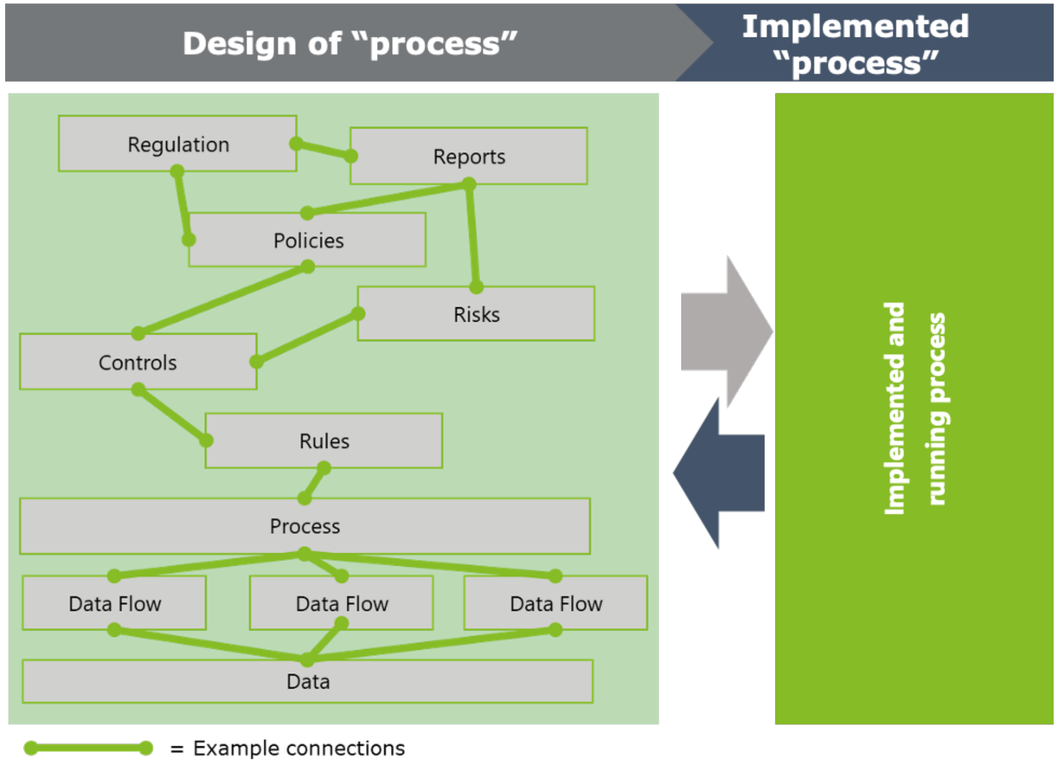
360° View of Risk & Value

- Know Your Customer (KYC)
- Due Diligence
- Investment Research
- Insurance Underwriting & Claim
- Commercial Real Estate

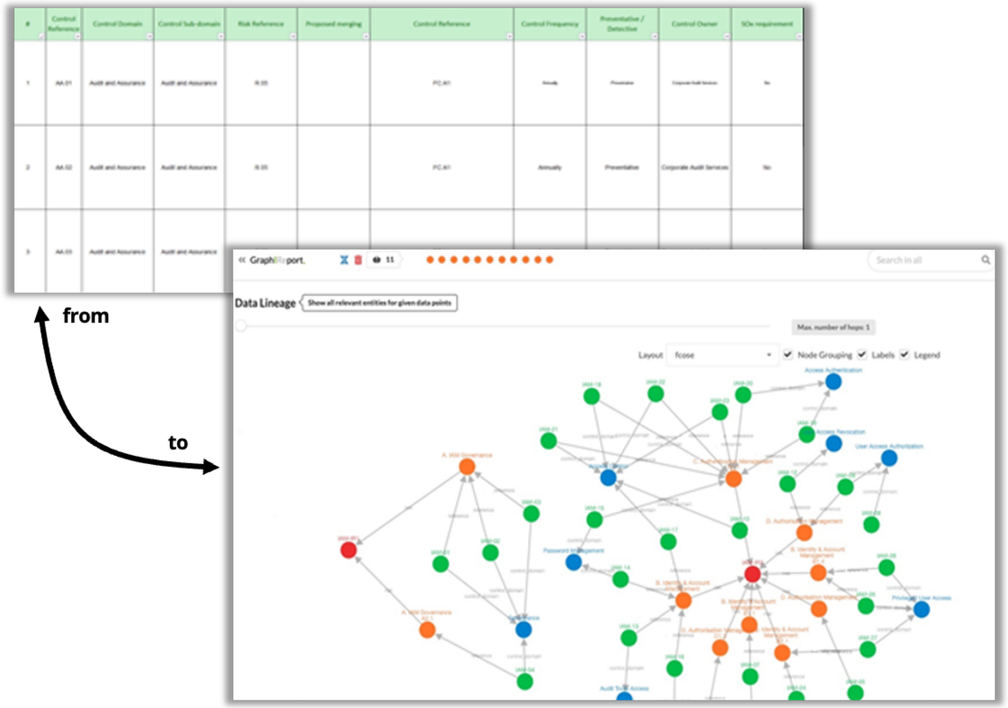
Data Lineage & Metadata Management

Data-driven Regulatory Reporting





Risk Control Frameworks



Insight in risk & control landscape and **connections**

- How does our risk profile look like?
- What are the most common risks?
- How many controls are connected to a risk?
- Do I have any risks that have no control mapped to it?
- How does my control landscape look like (i.e., domains, sub-domains, controls)

Visualization of control **interdependencies** and analysis of **control importance**

- Which dependencies exist between controls?
- What is the impact of a control deficiency?
- Which are the critical controls within the framework (controls with most dependent controls)?
- Do I have any overlapping controls?
- Are there opportunities for rationalization?

Visualization of **ownership** and **workload** of key players

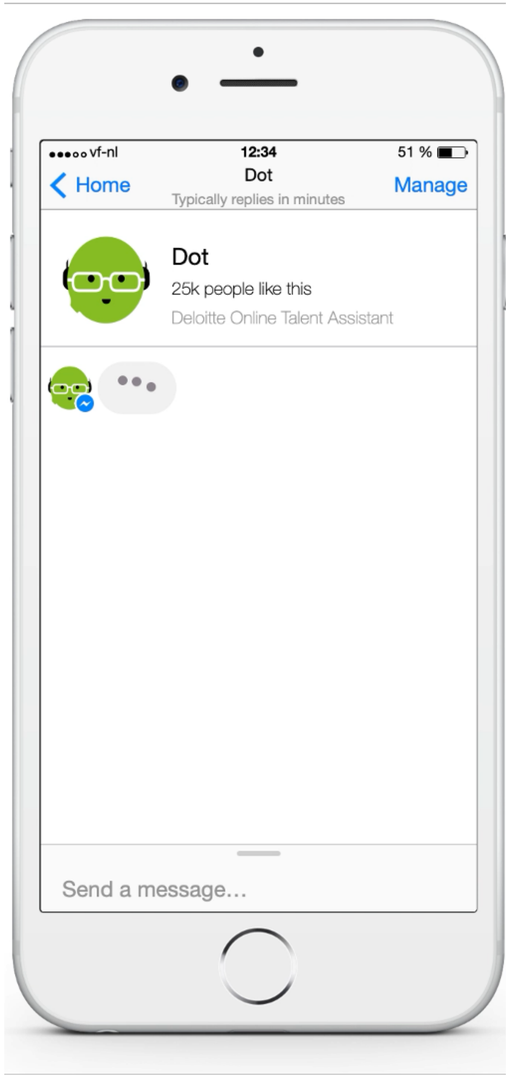
- Who are key players (and key dependencies) in the execution of controls?
- Is there a need to better distribute the workload related to control execution?
- Are controls being executed by the right departments?

Comparison of the number of **automated** controls and **non-automated** controls

- In which processes do we have a lot of manual tasks or controls and which have we automated?
- What are the points where controls need to be manual due to systems restrictions and lack of integration?
- What are the inefficiencies generated by systems limitations or lack of integration?
- Where are the data quality issues that create inneffiencies or manual turnarounds?

Semantic Search & Chatbots

Recommendation Engines



Any

Questions?