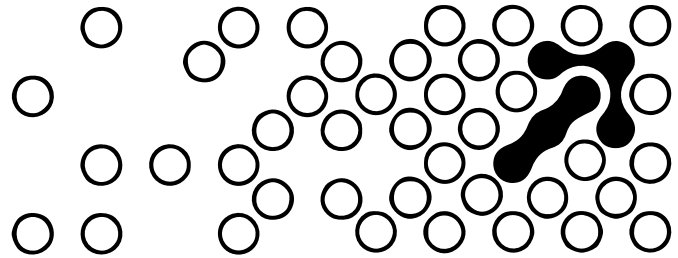


Graph Databases What Enterprise Data People Need to Know

May 12, 2022

DAMA Phoenix

Dave McComb

Bob Muglia

Former head of Microsoft's SQL Server Division

Former CEO of Snowflake

Speaking last week (May 4, 2022) at the Knowledge Graph Conference in NYC had a number of profound things to say about Knowledge Graphs in the Enterprise

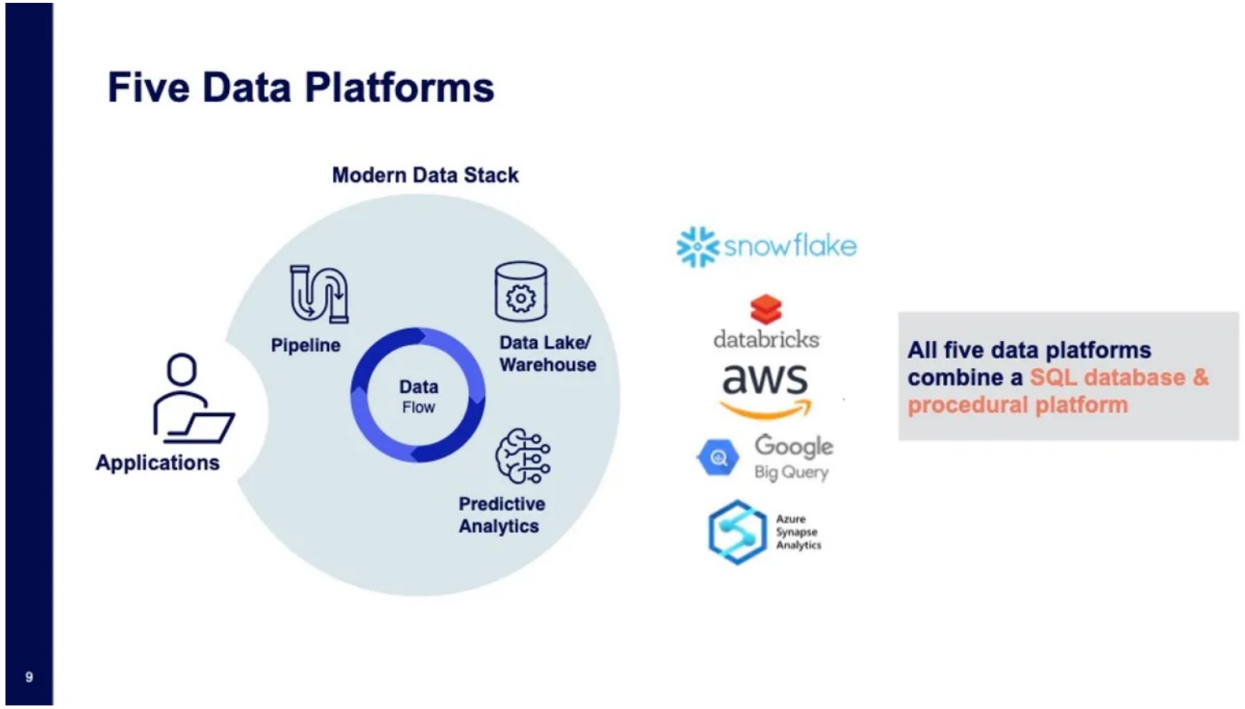
Muglia …

- "So, if you wanted to perform a query to say, 'Hey, tell me all of the resources that Fred Jones has access to in this organization' — that's a hard query to write," he said. "In fact, it's a query that probably can't execute effectively on any modern SQL database if the organization is very large and complex.”

- The problem, said Muglia, was that the algorithms based off of structured query language, or SQL, can't do such complex "recursive" queries.

The “modern data stack”

"We're at the start of a whole new era," said Muglia. "It's like the modern data stack in 2013, 2014 — that's where we are in that lifecycle.

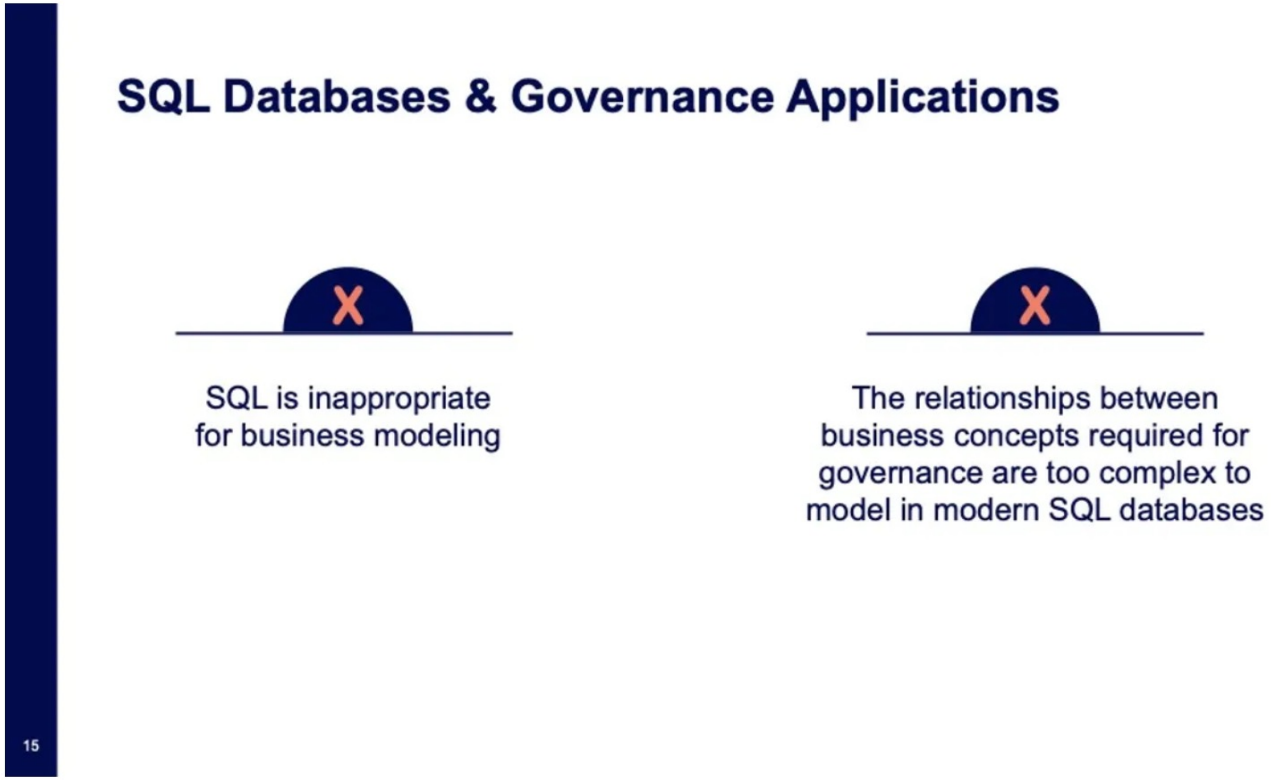


Muglia’s definition of a Knowledge Graph

- "What is a knowledge graph?" asked Muglia, rhetorically. He offered his own definition for what can be a sometimes mysterious concept. "A knowledge graph is a database that models business concepts, the relationships between them, and the associated business rules and constraints."

SQL not up to the task

"So, if you wanted to perform a query to say, 'Hey, tell me all of the resources that Fred Jones has access to in this organization' — that's a hard query to write," he said. "In fact, it's a query that probably can't execute effectively on any modern SQL database if the organization is very large and complex.”



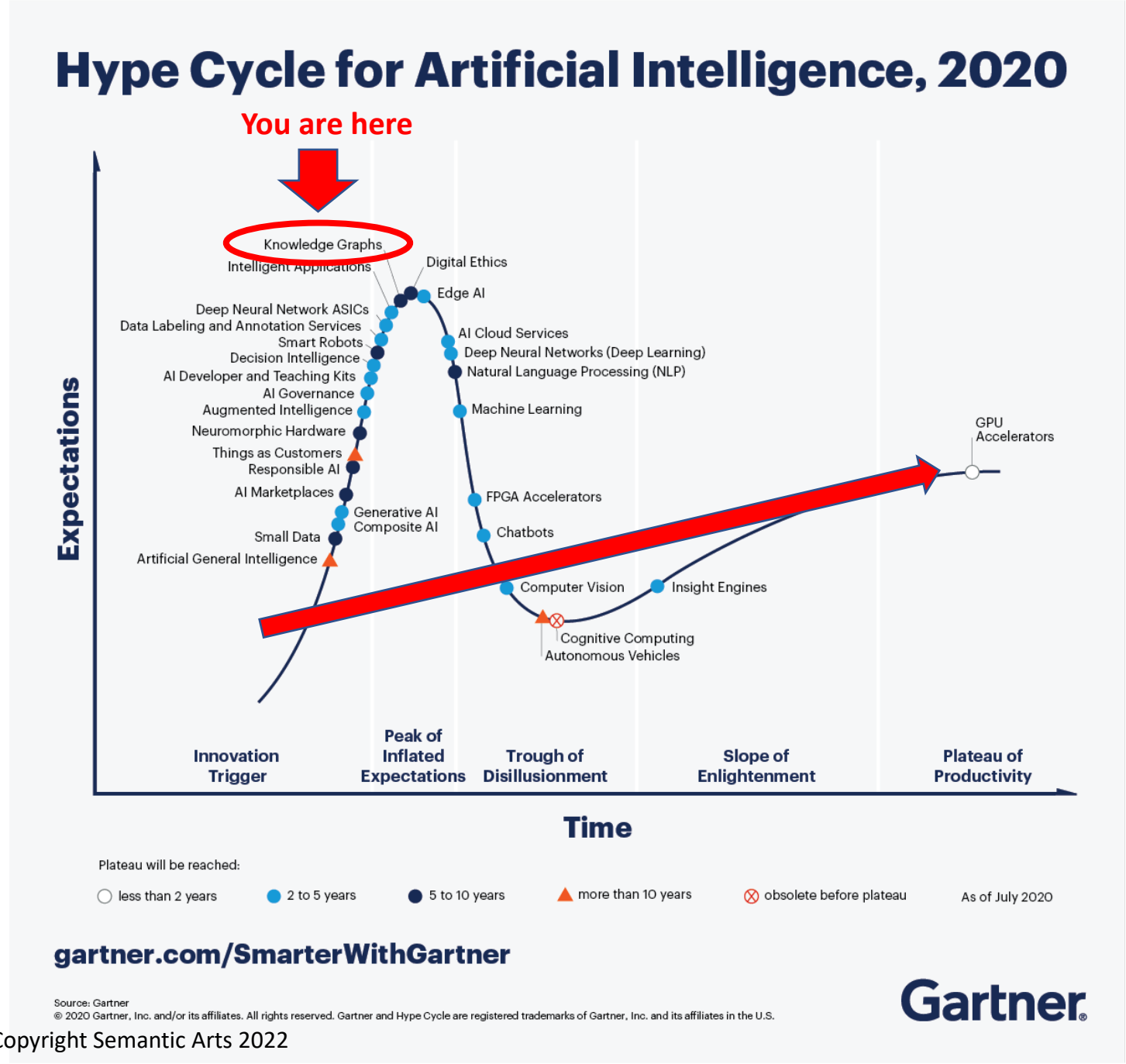
Democratization of Information

- "The model," said Muglia, "becomes the program, and so business analysts can become involved, and make changes to the data structures."
- "Think about thousands of people getting involved who know about the business — think about that!"

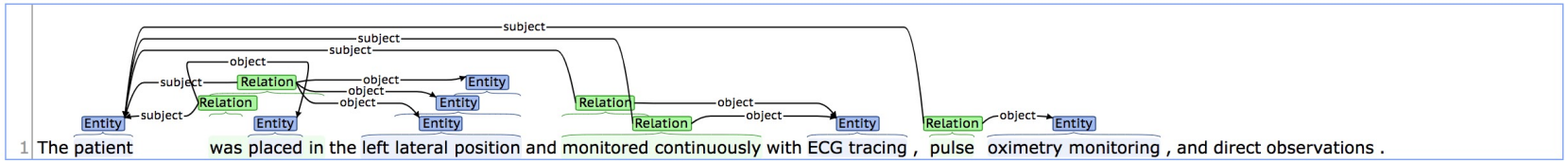
> “A motivated analyst could learn this over the weekend”

One of our clients

We're at the peak of perhaps the most hyped sector



The Knowledge Graph Ecosystem



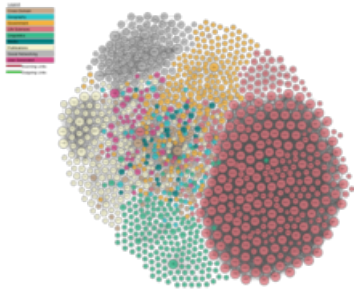
Native Graph Functionality

Ontology

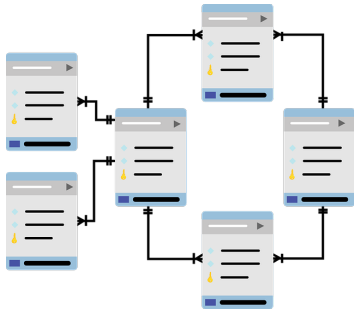
Data as "Triples"

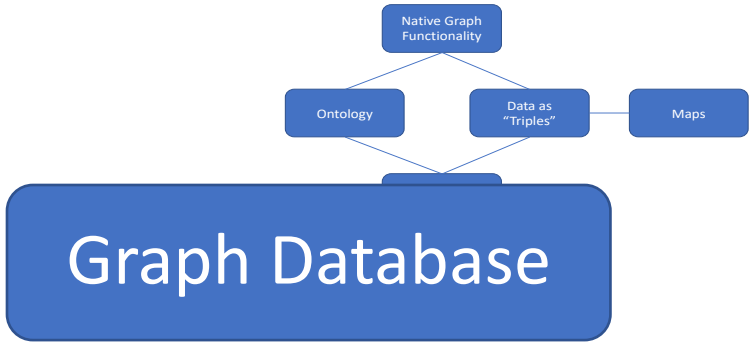
Maps

Graph Database



{JSON}





Two Species of Graph Databases

Labeled Property Graphs

- Developer Friendly
- Proprietary

RDF / Triple Stores

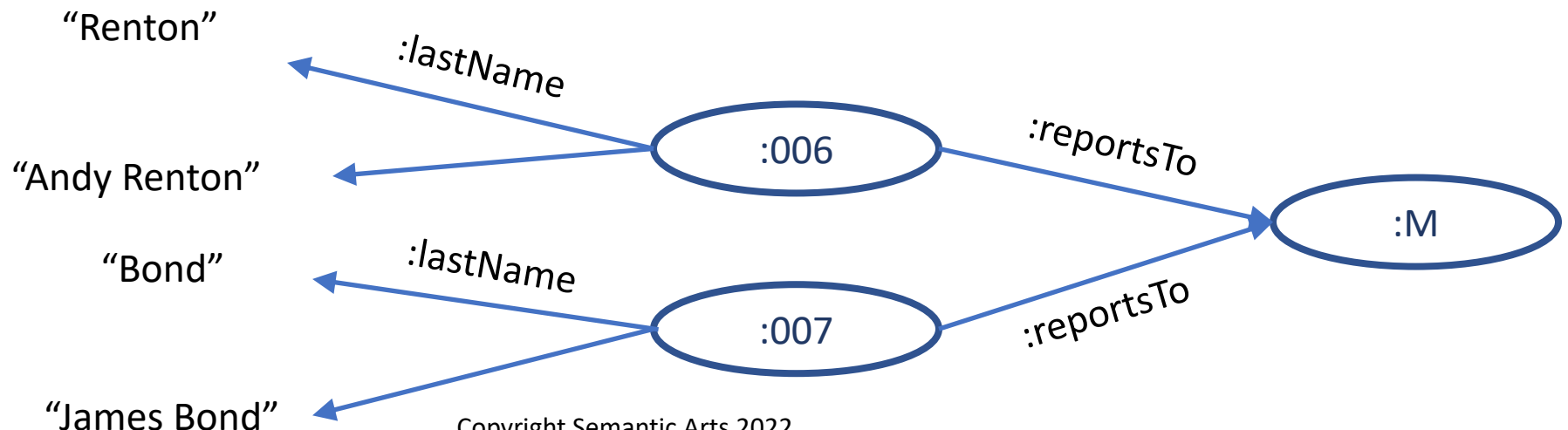
- Solves Integration Problems
- W3C Standards Based

Let’s Dive in and see how this really works

Tables, “Documents” or Graphs

Id	Last Name	Full Name	ReportsTo
006	Renton	Andy Renton	M
007	Bond	James Bond	M

```
{ ‘Agents’:
 [
{‘id’:”006”, ‘LastName’: ”Renton”, ’FullName’:”Andy Renton”,
‘reportsTo’:”M”},
{‘id’:”007”, ‘LastName’: ”Bond”, ’FullName’:”James Bond”,
‘reportsTo’:”M”}}
]
}
```



URIs

- One of the magic bits is URIs (Universal Resource Identifiers) or their superset IRIs (Internationalized Resource Identifiers) which can use Unicode characters

Table 6. Bidi Examples

	Samples
1	http://دائم.سلام.com
2	http://سلام.a.دائم.com

Traditional IDs

Id	Last Name	Full Name	ReportsTo
006	Renton	Andy Renton	M
007	Bond	James Bond	M

{ 'Agents':
[
{"id":"006", 'LastName': "Renton", 'FullName':"Andy Renton",
'reportsTo':"M"},
{"id":"007", 'LastName': "Bond", 'FullName':"James Bond",
'reportsTo':"M"}}
]
}

These are:

-Strings

-Hyperlocal

Hyper local strings, means

- That the same string means something different depending on where it is

Espionage DB

SpyTable

Id	Last Name	Full Name	ReportsTo
006	Renton	Andy Renton	M
007	Bond	James Bond	M

Deli DB

TodaysSpecialTable

Id	Desc	Price
006	BLT	$9.95
007	Ham Sandwich	$8.95

Furthermore,

- Sometimes it is referring to the same thing, but called something different

Espionage DB

SpyTable

Id	Last Name	Full Name	ReportsTo
006	Renton	Andy Renton	M
007	Bond	James Bond	M

Espionage DB

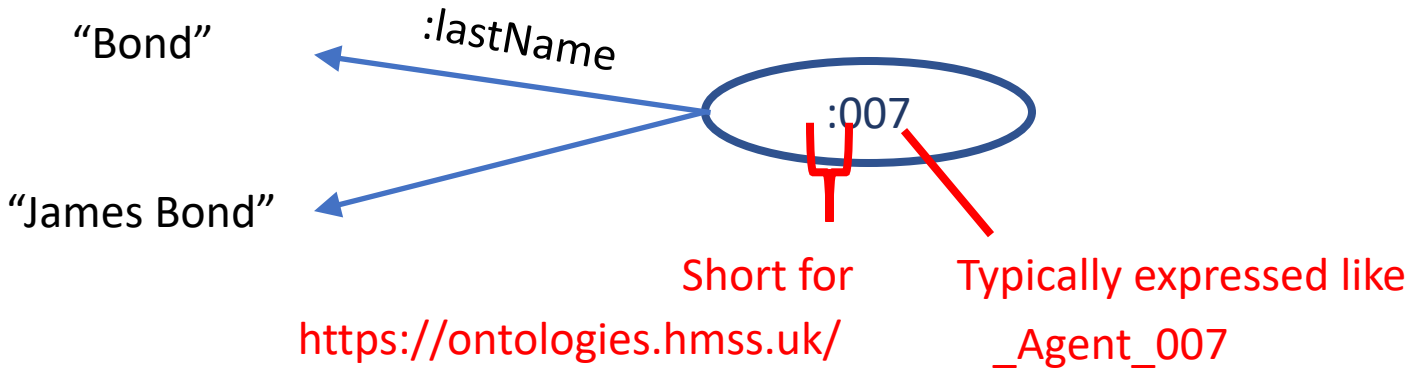
WeaponAssignmentTable

Id	Desc	Agent
001	Watch Garrot	007
002	Dagger Shoes	007

SELECT …. WHERE WeaponAssignmentTable.Agent = SpyTable.Id

URIs/ IRIs are globally unique

- Which means you don't need metadata to do joins

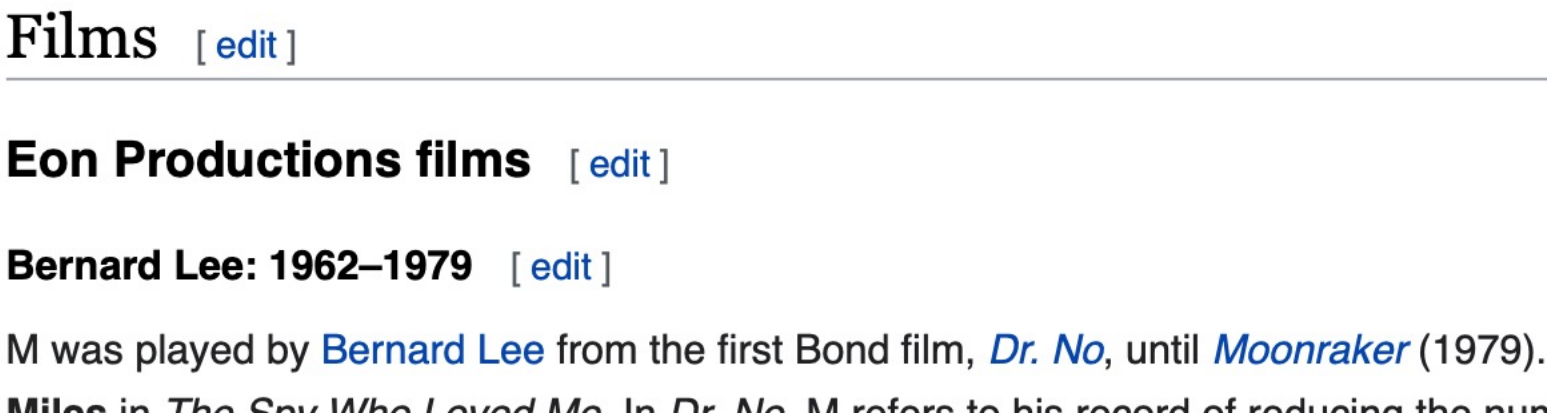


https://ontologies.hmss.uk/_Agent_007

Strings v Things

URIs/ IRIs are globally unique

- Which means you don't need metadata to do joins



Films [edit]

Eon Productions films [edit]

Bernard Lee: 1962–1979 [edit]

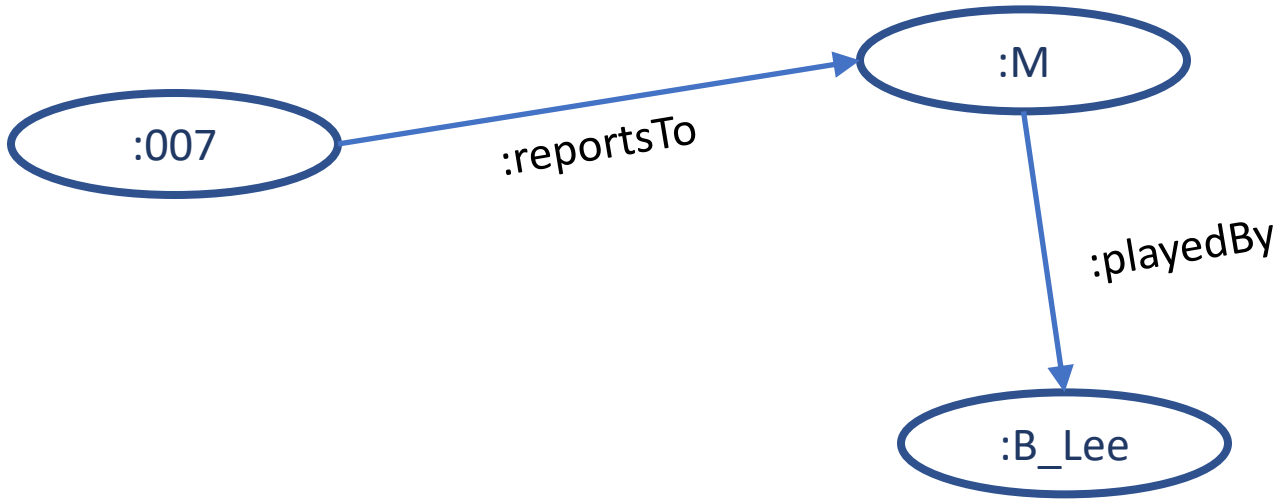
M was played by Bernard Lee from the first Bond film, *Dr. No*, until *Moonraker* (1979).

Globally unique / location independent

https://ontologies.hmss.uk/_Agent_007

Globally unique / location independent

https://ontologies.hmss.uk/_Agent_007



Tinker toys

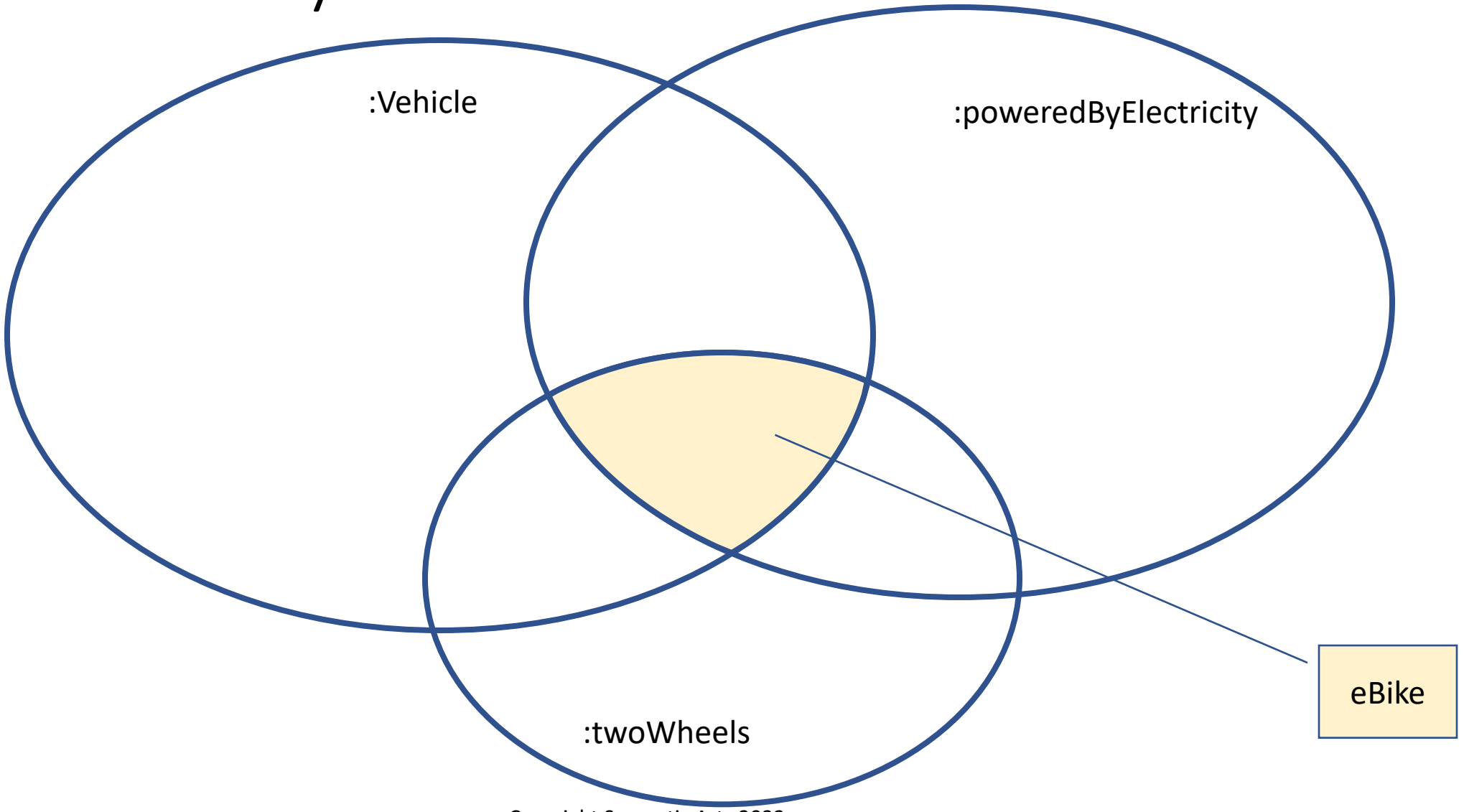
An ontology is a model of the concepts (meaning) of the domain

It is analogous to a schema

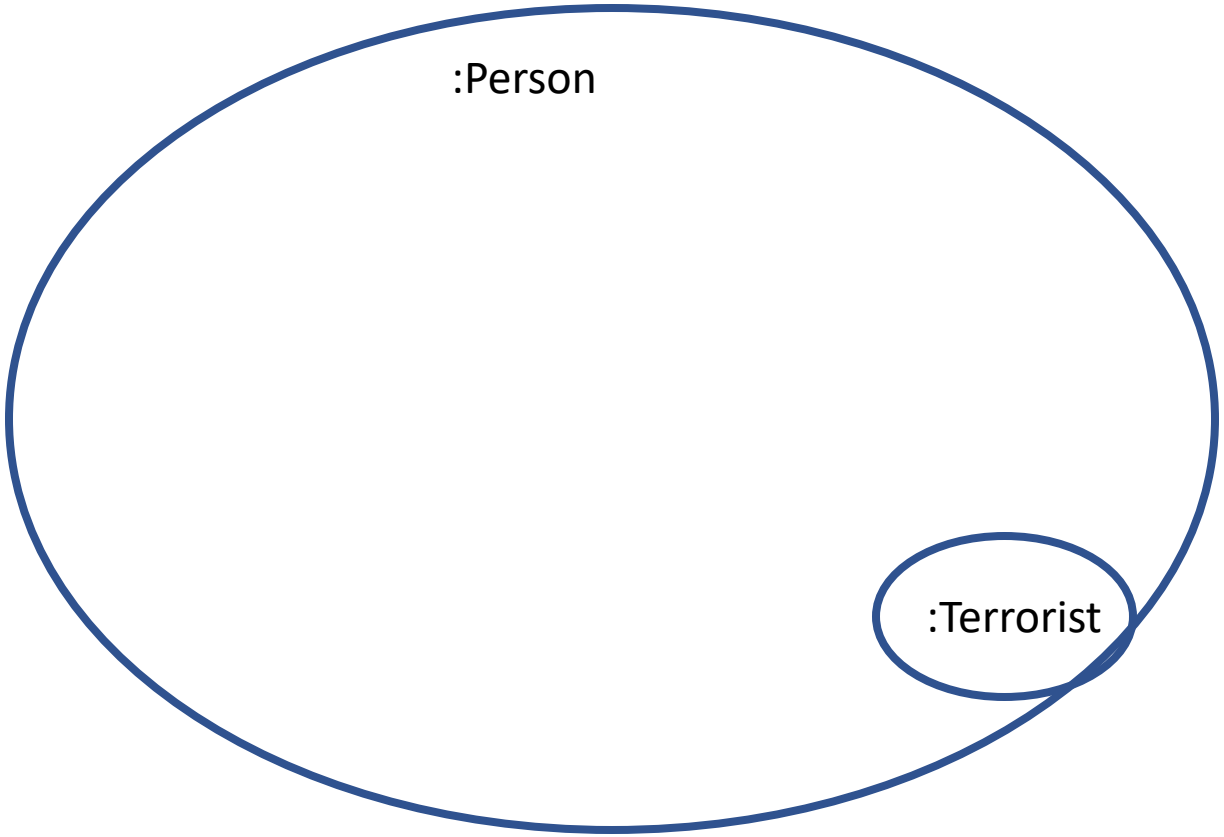
Except the schema is more about structure

The ontology, if well designed, is orders of magnitude simpler than the schemas it subsumes

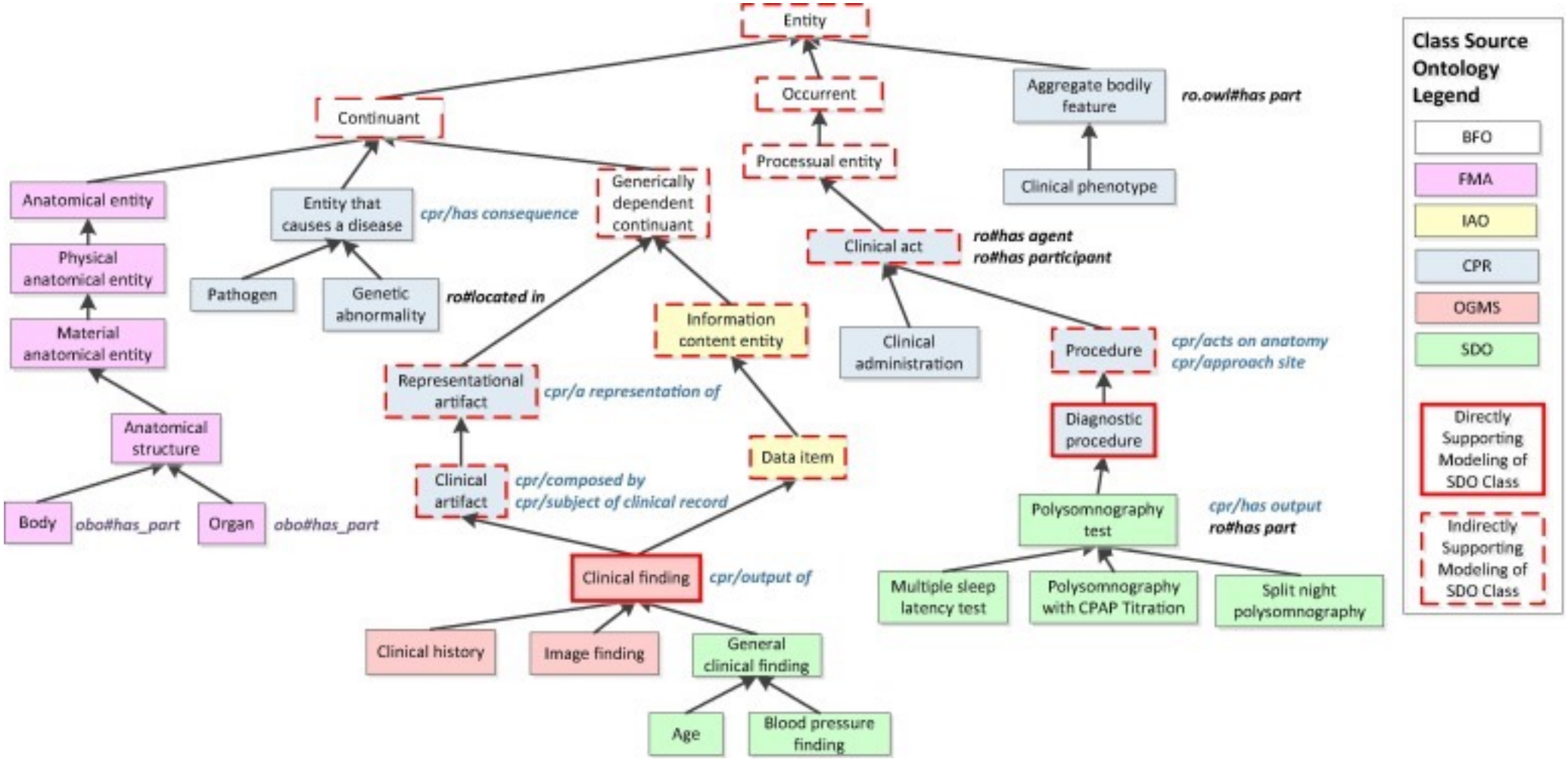
Set Theory



Open World



The Medical Industry is Way Ahead



Accommodates change in place

- **A semantic system can grow and evolve in place**
- **Example: Tasks -> Projects -> Backlogs -> Assignments -> Expenses**

Triples in DBPedia

(Wikipedia)

DBpedia

SPARQL

270 million triples

Endpoint

https://dbpedia.org/sparql

Common prefixes

dbo	DBPedia ontology
dbp	predicates
dbr	Resources (data)

Useful queries

```
SELECT ?s ?name
WHERE {
?s rdf:type dbo:Company .
?s rdfs:label ?name .
FILTER CONTAINS (?name, ”Amgen")
}
LIMIT 100
```

Useful queries

```
SELECT ?s ?p ?o
WHERE {
?s ?p ?o .
VALUES ?s { dbr:Amgen}
}
LIMIT 10000
```

Time card

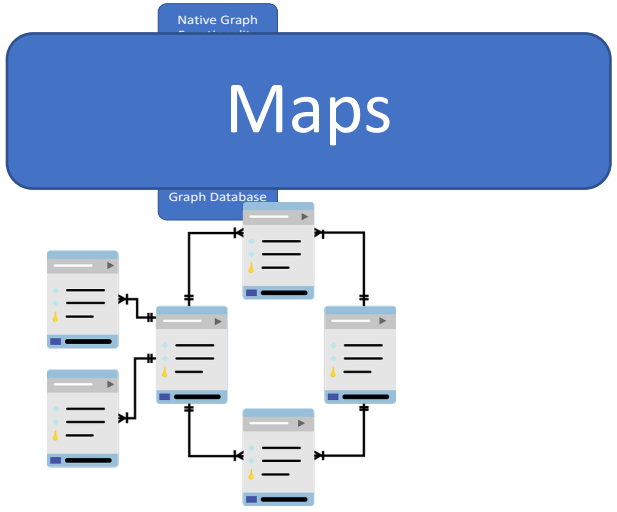
```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX gist: <http://ontologies.semanticarts.com/gist#>
PREFIX sa: <http://ontologies.semanticarts.com/SemArts#>

SELECT ?client ?name WHERE {
  ?client rdf:type sa:Client .
 ?client gist:name ?name .
  FILTER CONTAINS (?name, ”Amgen")
}
```

Federating from our Triplestore to DBpedia

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX gist: <http://ontologies.semanticarts.com/gist#>
PREFIX sa: <http://ontologies.semanticarts.com/SemArts#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
SELECT ?client ?name ?ticker ?numEmp ?projectName
WHERE {
  ?client rdf:type sa:Client .
  ?client gist:name ?name .
  ?project sa:hasSponsor ?client .
  ?project rdfs:label ?projectName .
  FILTER CONTAINS (?name, "Amgen")
  #?client owl:sameAs ?s .
  {
    service <https://dbpedia.org/sparql>
    {
                        SELECT ?s ?numEmp ?ticker
                        WHERE {
                        ?s <http://dbpedia.org/ontology/numberOfEmployees> ?numEmp .
                        ?s <http://dbpedia.org/property/symbol> ?ticker .
                        VALUES ?s { <http://dbpedia.org/resource/Amgen>
                        }
}
  }
}
}
ORDER BY ?projectName
```

Getting Relational data in



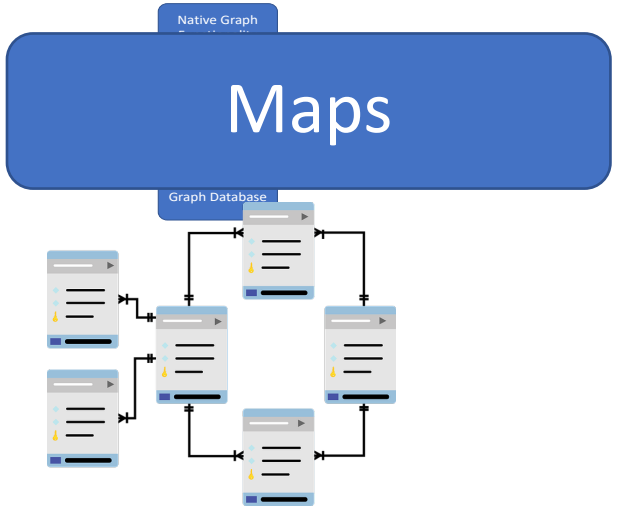

```
@prefix rr: <http://www.w3.org/ns/r2rml#>.

<#TriplesMap1>
    rr:logicalTable [ rr:tableName "EMP" ];
    rr:subjectMap [
        rr:template "http://example.com/employee/{EMPNO}";
        rr:class ex:Employee;
    ];
    rr:predicateObjectMap [
        rr:predicate ex:name;
        rr:objectMap [ rr:column "ENAME" ];
   ];
    rr:predicateObjectMap [
        rr:predicate ex:job;
        rr:objectMap [ rr:column “JOB” ];
   ];
    rr:predicateObjectMap [
        rr:predicate ex:depNr;
        rr:objectMap [ rr:column “DEPTNO ];
   ].
```

ETL Mode

ETL Mode

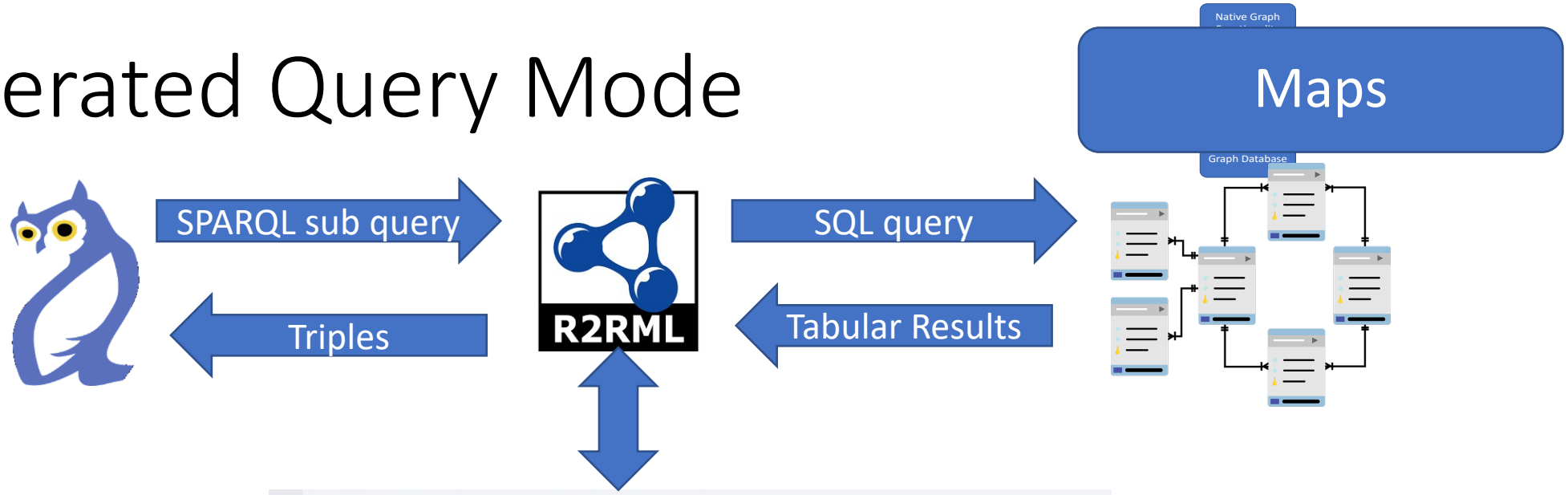
ETL Mode




```
@prefix rr: <http://www.w3.org/ns/r2rml#>.

<#TriplesMap1>
    rr:logicalTable [ rr:tableName "EMP" ];
    rr:subjectMap [
        rr:template "http://example.com/employee/{EMPNO}";
        rr:class ex:Employee;
    ];
    rr:predicateObjectMap [
        rr:predicate ex:name;
        rr:objectMap [ rr:column "ENAME" ];
   ];
    rr:predicateObjectMap [
        rr:predicate ex:job;
        rr:objectMap [ rr:column “JOB” ];
   ];
    rr:predicateObjectMap [
        rr:predicate ex:depNr;
        rr:objectMap [ rr:column “DEPTNO ];
   ].
```

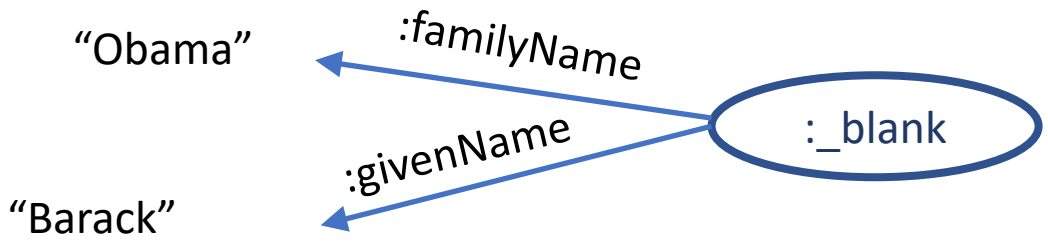
Federated Query Mode

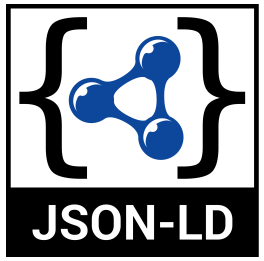



```
@prefix rr: <http://www.w3.org/ns/r2rml#>.

<#TriplesMap1>
    rr:logicalTable [ rr:tableName "EMP" ];
    rr:subjectMap [
        rr:template "http://example.com/employee/{EMPNO}";
        rr:class ex:Employee;
    ];
    rr:predicateObjectMap [
        rr:predicate ex:name;
        rr:objectMap [ rr:column "ENAME" ];
   ];
    rr:predicateObjectMap [
        rr:predicate ex:job;
        rr:objectMap [ rr:column “JOB” ];
   ];
    rr:predicateObjectMap [
        rr:predicate ex:depNr;
        rr:objectMap [ rr:column “DEPTNO ];
   ].
```

Json-LD

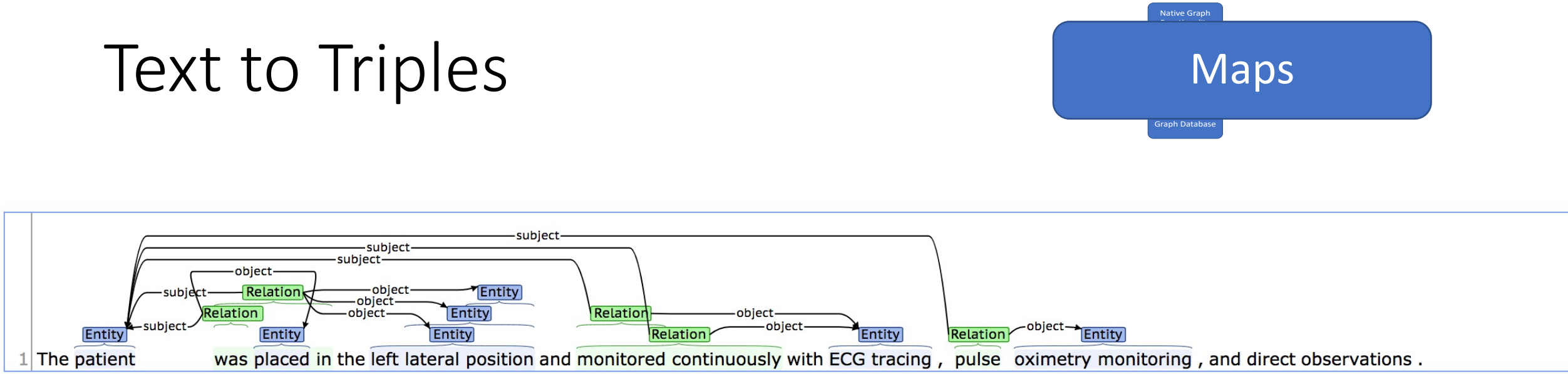





```
{ "@context": "http://schema.org",
 "name": "Barack Obama",
 "givenName": "Barack",
 "familyName": "Obama",
 "jobTitle": "44th President of the United States"
```

```
[ { "http://schema.org/familyName": [{"@value": "Obama"}],
"http://schema.org/givenName": [{"@value": "Barack"}],
"http://schema.org/jobTitle": [{"@value": "44th President of the United
States"}],
"http://schema.org/name": [{"@value": "Barack Obama"}] } ]
```

Text to Triples



This example using Stanford NLP Open IE

Another example

Food Tutorials are Infinitely Better When Directed By Wes Anderson. Bruce Lee's biopic, 'Little Dragon', to be directed by Shekhar Kapur. Stallone directed his first short film Vic.

- Wes Anderson **directed** Food Tutorials
- Shekhar Kapur **directed** Little Dragon
- Stallone **directed** Vic

Python spaCy

People v. Hall Sample Case Brief

Style: People (Colorado) v. Nathan Hall

Colorado Supreme Court 2004

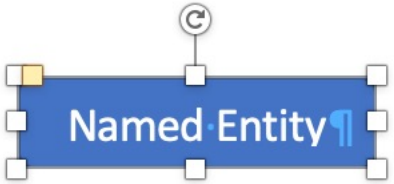
Procedural History:

District court affirmed lack of probable cause (defendant won again)
Appellate court reversed (People won)

At a preliminary hearing, the trial court dismissed case for lack of probable cause (defendant won)

Issue:

People v. Hall Sample Case Brief



Style: People (Colorado) v. Nathan Hall

Colorado Supreme Court 2004

Procedural History:

District court affirmed lack of probable cause (defendant won again)
Appellate court reversed (People won)

At a preliminary hearing, the trial court dismissed case for lack of probable cause (defendant won)

Issue:

Person

People v. Hall Sample Case Brief

Named Entity

Style: People (Colorado) v. Nathan Hall

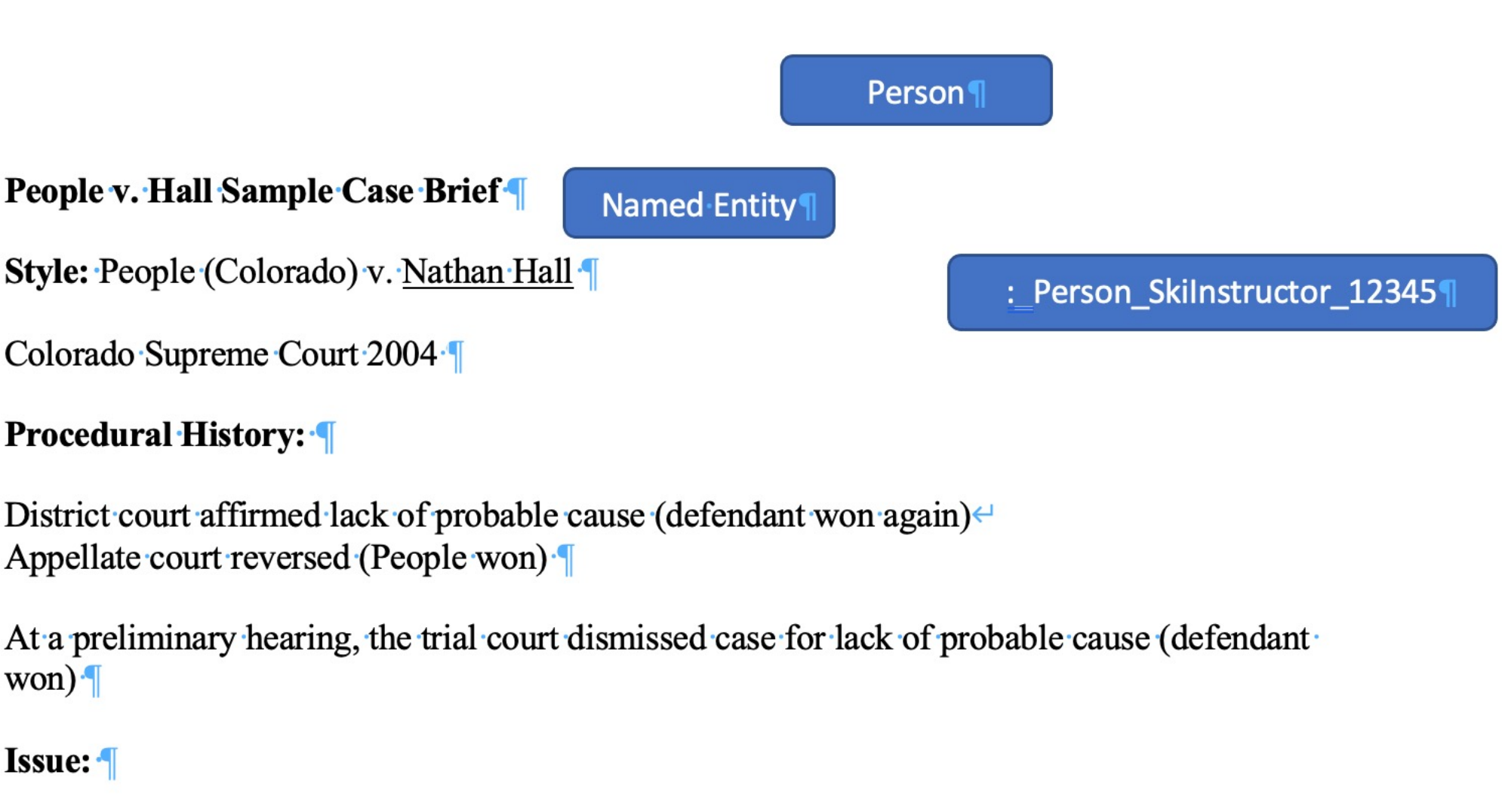
Colorado Supreme Court 2004

Procedural History:

District court affirmed lack of probable cause (defendant won again)
Appellate court reversed (People won)

At a preliminary hearing, the trial court dismissed case for lack of probable cause (defendant won)

Issue:



People v. Hall Sample Case Brief

Style: People (Colorado) v. Nathan Hall

Colorado Supreme Court 2004

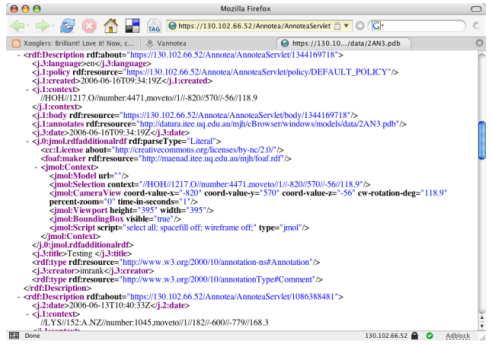
Procedural History:

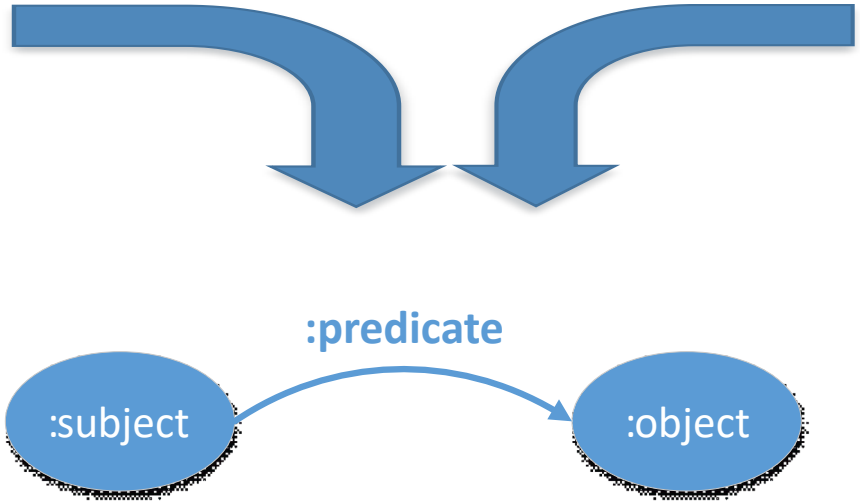
District court affirmed lack of probable cause (defendant won again)
Appellate court reversed (People won)

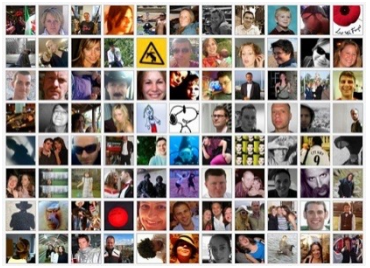
At a preliminary hearing, the trial court dismissed case for lack of probable cause (defendant won)

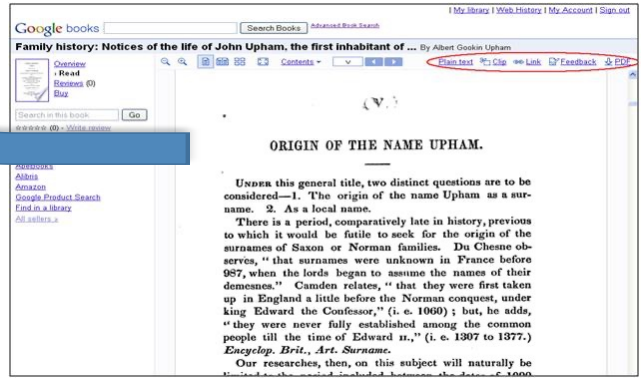
Issue:

Triple as common denominator



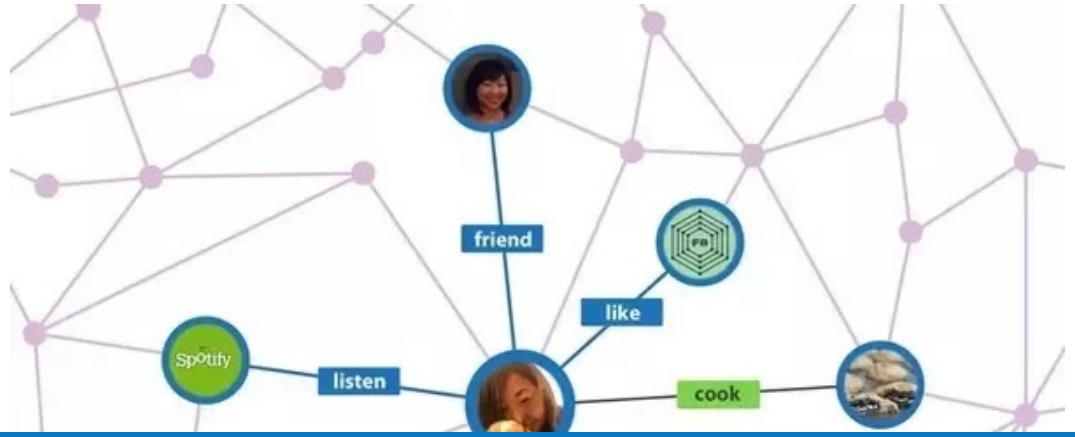


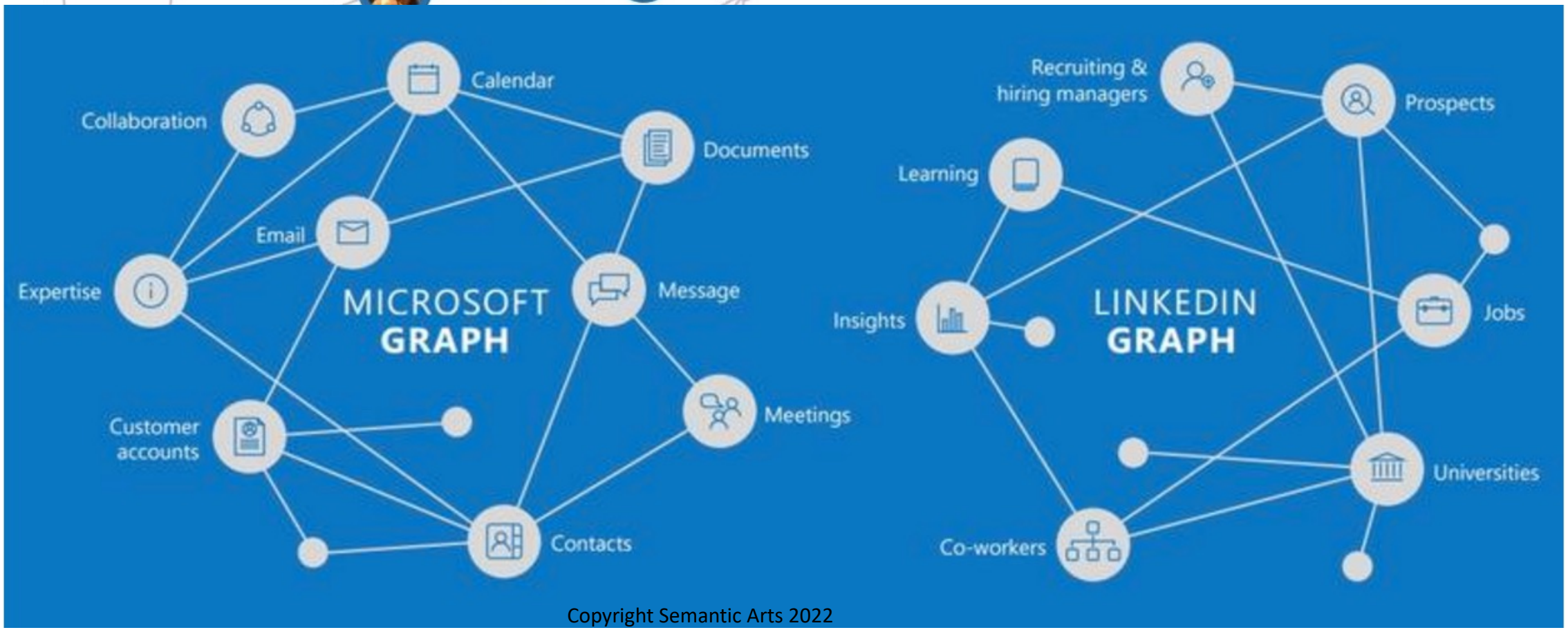


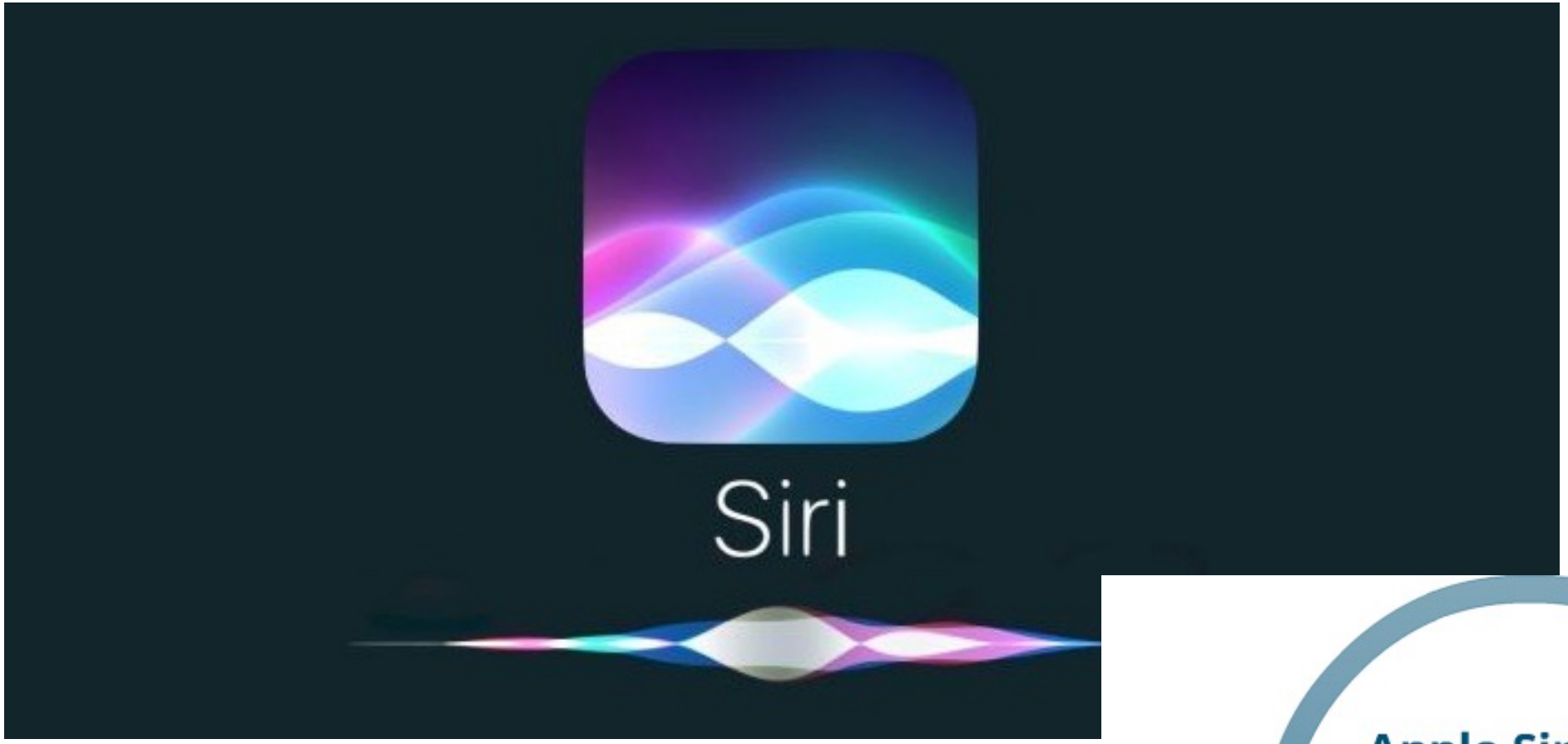


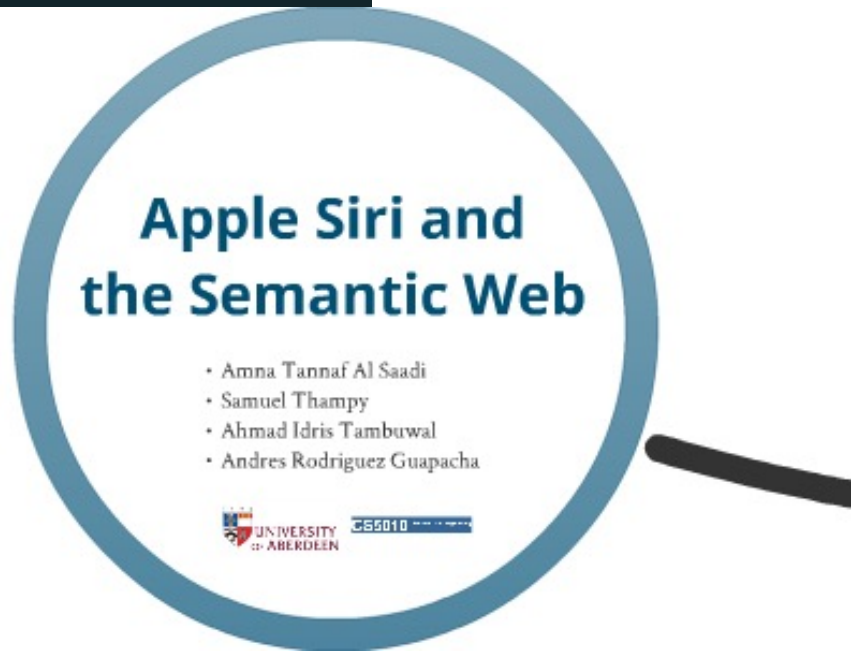
Trend 1: Almost All High Growth Companies use KG

Courtesy Alan Morrison









KG and GDB enable Data-Centric-ness

Very specifically

Starts with the data model (not starting with a business function or process to automate)

And has its' scope as the enterprise (not the application)

A firm is data-centric to the extent:

Application functionality is loosely bound to a single, simple, extensible, federate-able, universal, shared and directly implementable model

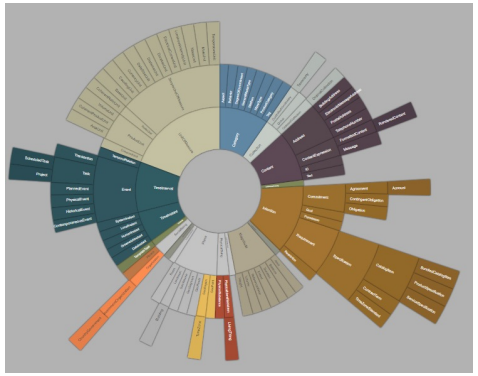
A firm is data-centric to the extent:

Application functionality is loosely bound to a single, simple, extensible, federate-able, universal, shared and directly implementable model

Traditional Enterprise Architecture

Most Enterprises have one data model per application, and therefore thousands of data models

Data-Centric Architecture



A firm is data-centric to the extent:

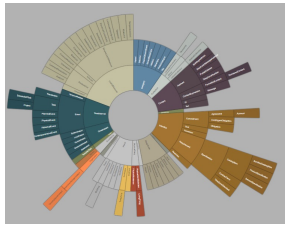
Application functionality is loosely bound to a single, simple, extensible, federate-able, universal, shared and directly implementable model

Traditional Enterprise Architecture

Each application has thousands of concepts.

The enterprise has millions

Data-Centric Architecture



The core ontology has 500-1000 concepts

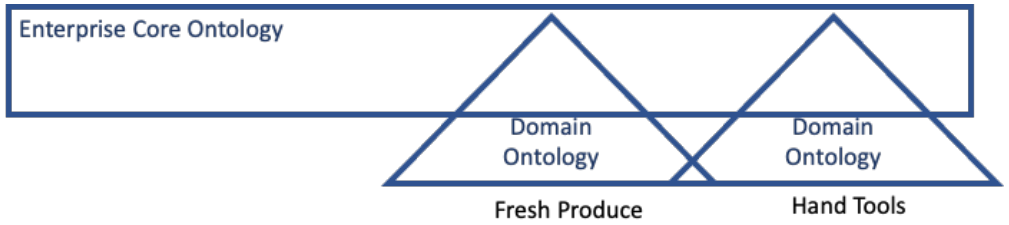
A firm is data-centric to the extent:

Application functionality is loosely bound to a single, simple, extensible, federate-able, universal, shared and directly implementable model

Traditional Enterprise Architecture

There is no mechanism for extending a model for a subset of the enterprise

Data-Centric Architecture



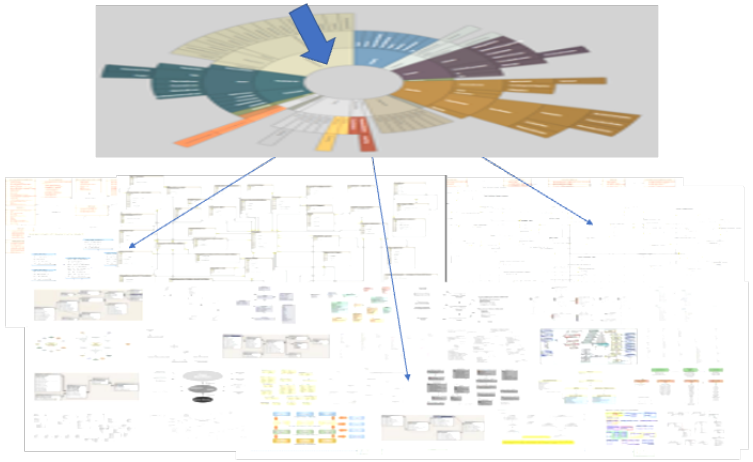
A firm is data-centric to the extent:

Application functionality is loosely bound to a single, simple, extensible, federate-able, universal, shared and directly implementable model

Traditional Enterprise Architecture

Traditional systems co-locate data (data warehouses) instead of querying many databases

Data-Centric Architecture



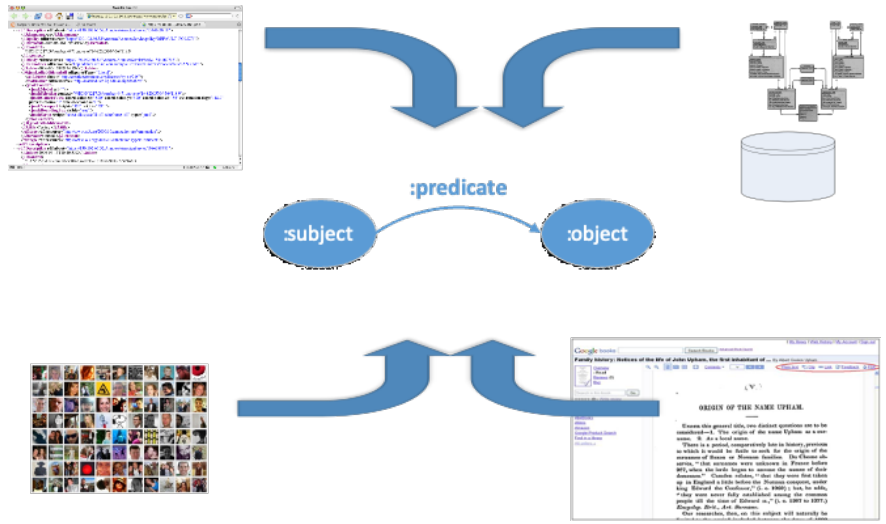
A firm is data-centric to the extent:

Application functionality is loosely bound to a single, simple, extensible, federate-able, universal, shared and directly implementable model

Traditional Enterprise Architecture

There isn't a good way to unite disparate types of data

Data-Centric Architecture



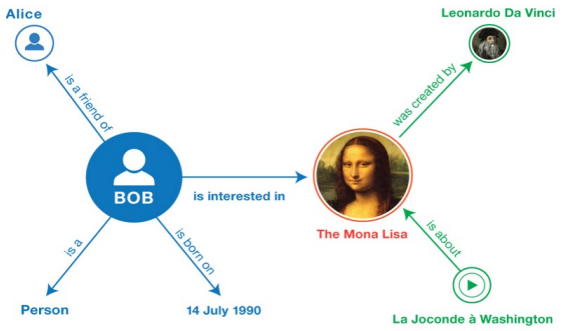
A firm is data-centric to the extent:

Application functionality is loosely bound to a single, simple, extensible, federate-able, universal, shared and directly implementable model

Traditional Enterprise Architecture

Systems share data by copying and transforming

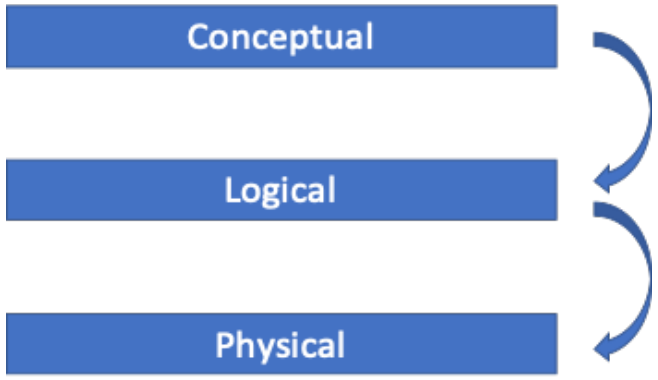
Data-Centric Architecture



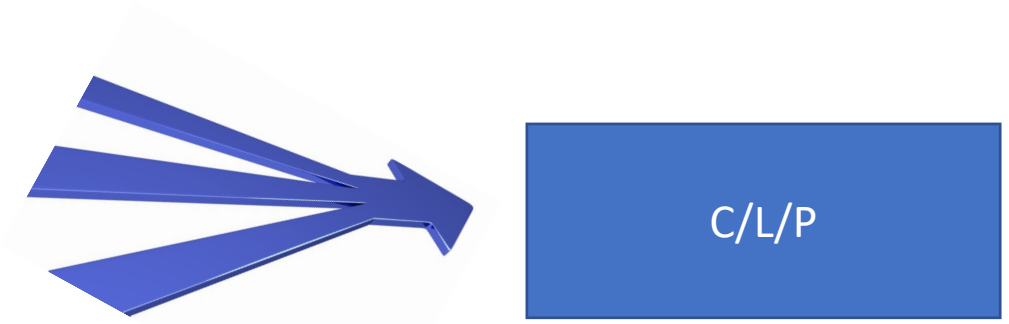
A firm is data-centric to the extent:

Application functionality is loosely bound to a single, simple, extensible, federate-able, universal, shared and directly implementable model

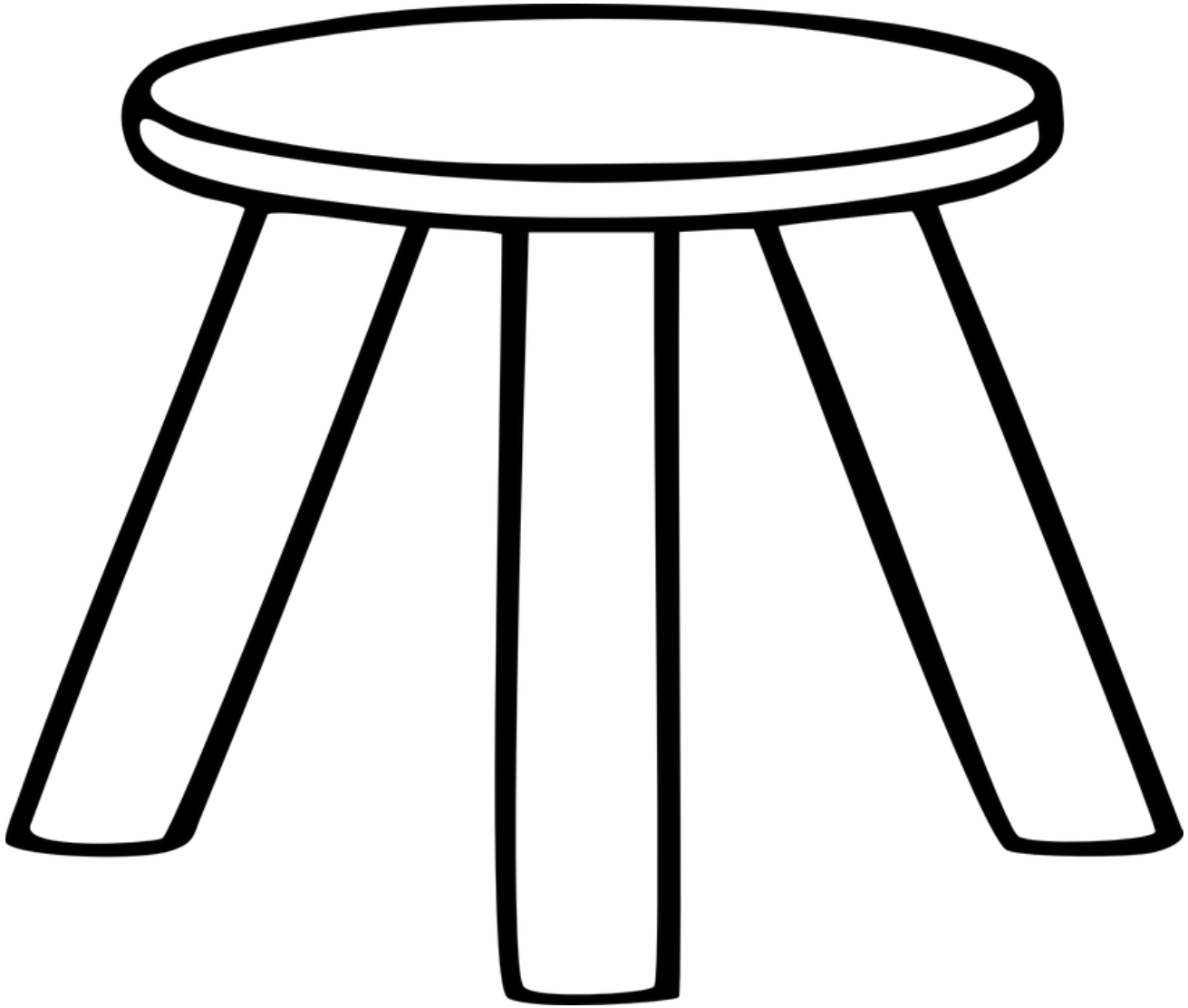
Traditional Enterprise Architecture



Data-Centric Architecture

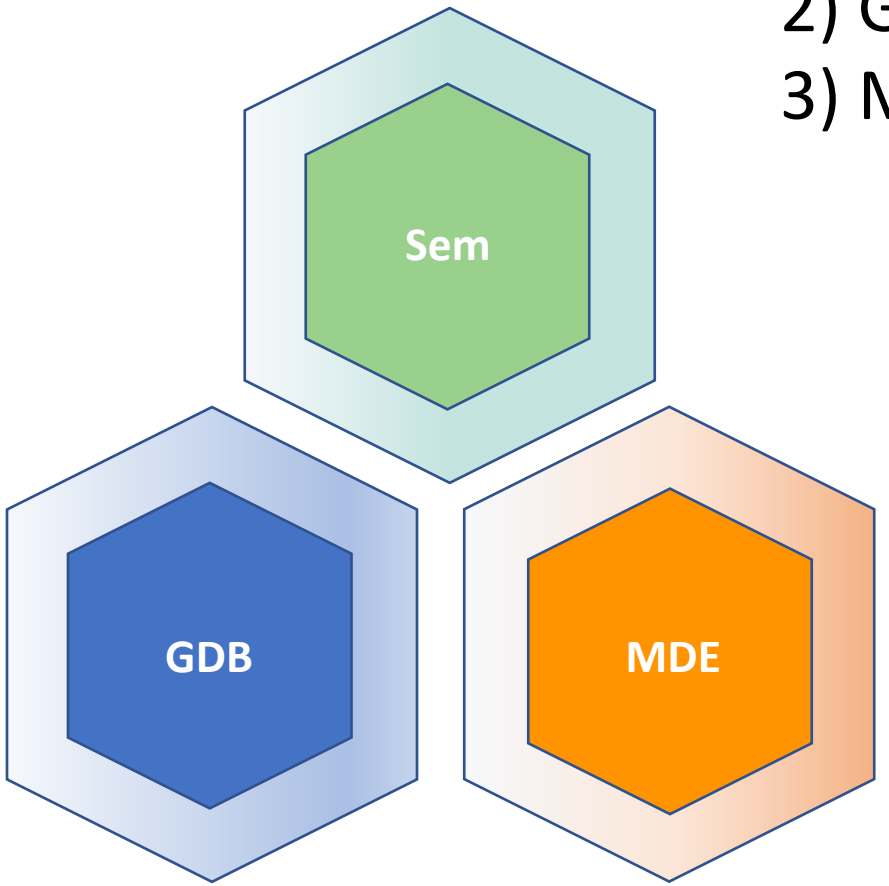



HOW DO
WE GET
THERE?



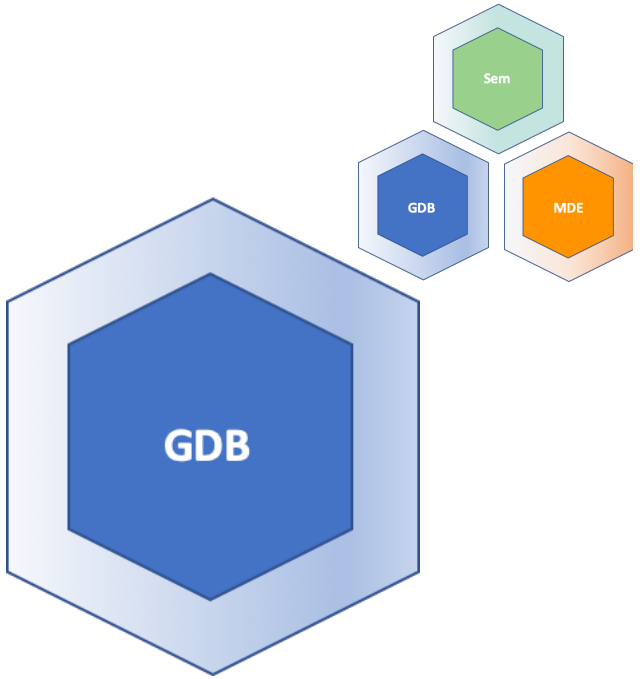
Three Keys to Data-Centric

1) Semantics
2) Graph Databases
3) Model Driven Everything

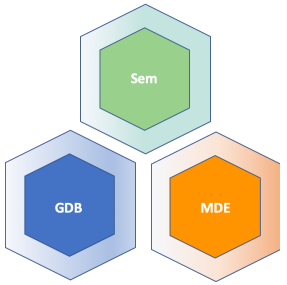


Knowledge Graph Databases

- Flexibility
- Integration almost for free
- Removes Data Silos



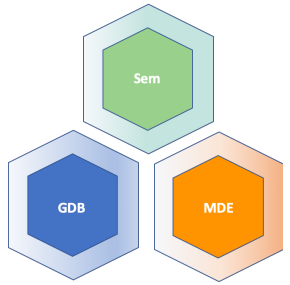
Graph DB and Flexibility



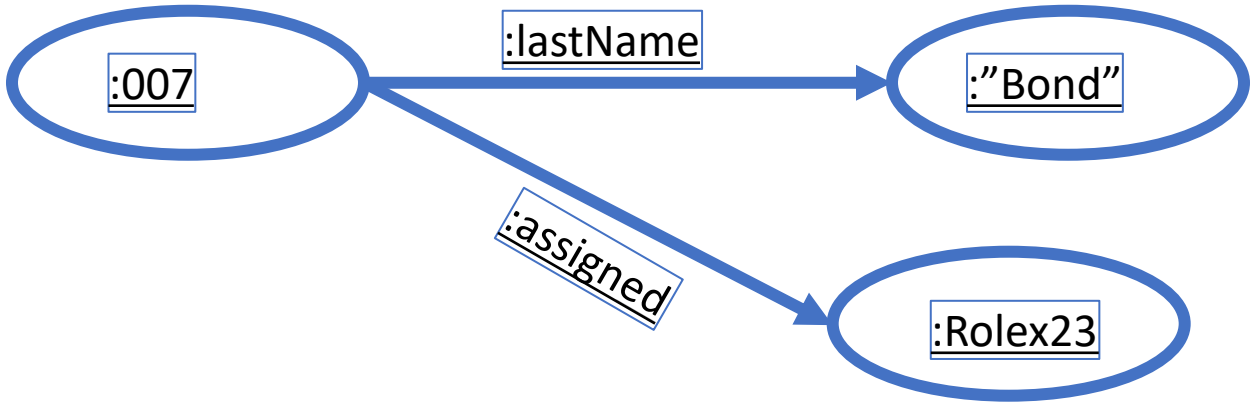
- Instead of storing documents in tables or documents, a graph database stores data in “di-graphs” – a “directed” (has an arrow) graph

agent		
Id	Last Name	Full Name
006	Renton	Andy Renton
007	Bond	James Bond

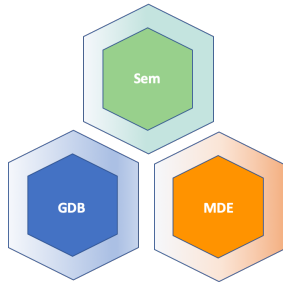
Graph DB and Flexibility



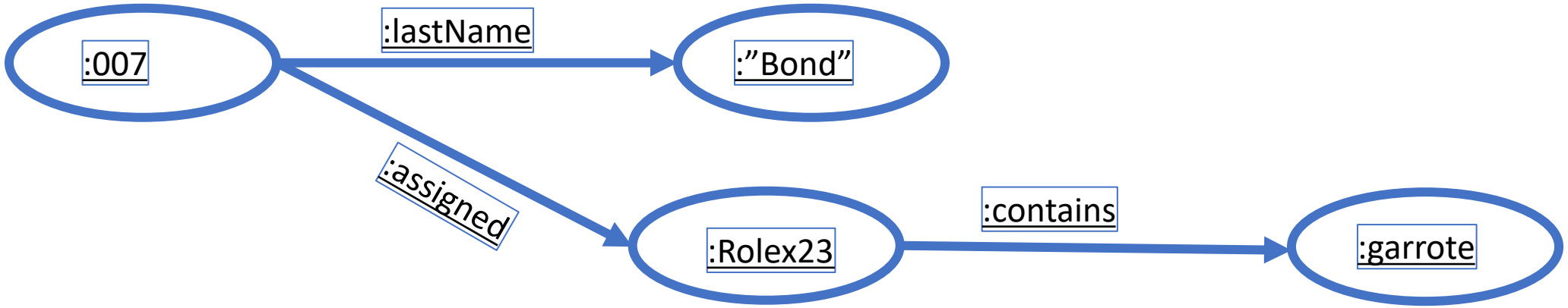
- New information is just accreted, there doesn't have to be a pre-existing table structure



Graph DB and Flexibility

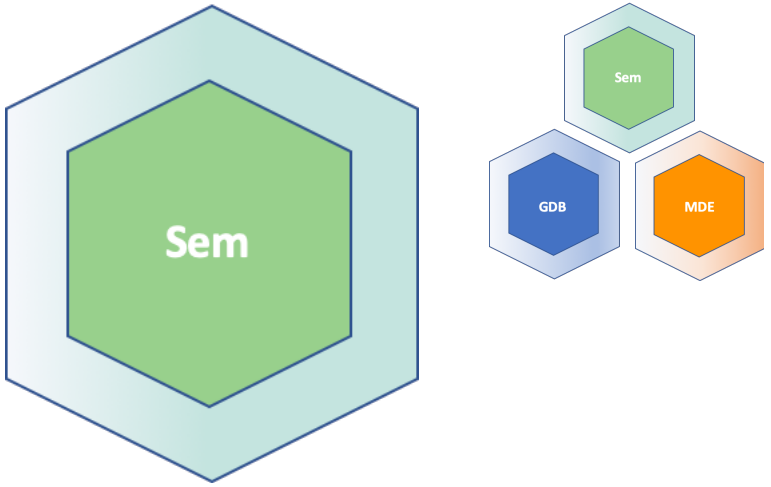


- The graph can be extended indefinitely



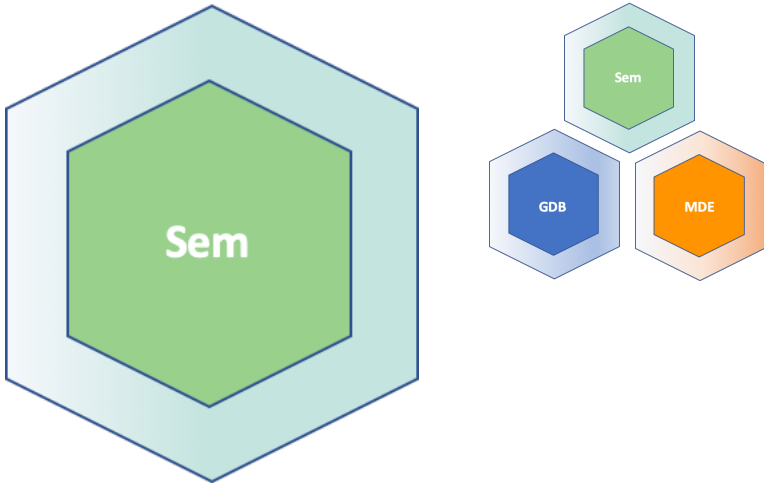
Semantic Technology

- **Semantics** is the study of meaning
- OWL is a W3C standard for formally expressing meaning in a way that humans and computers have a shared definition of what things mean
- A model built in OWL is called an **ontology**



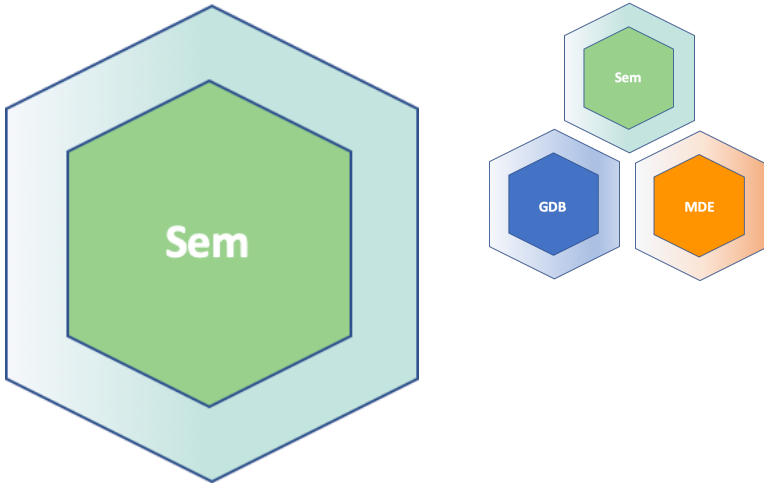
Semantic Technology

- Deep meaning is where simplification comes from



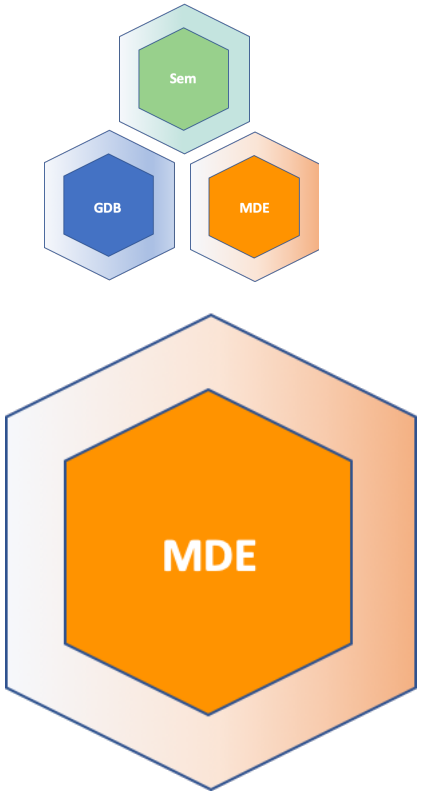
Inference

- Create new information from data + knowledge



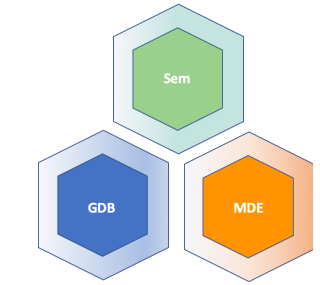
Model Driven Everything

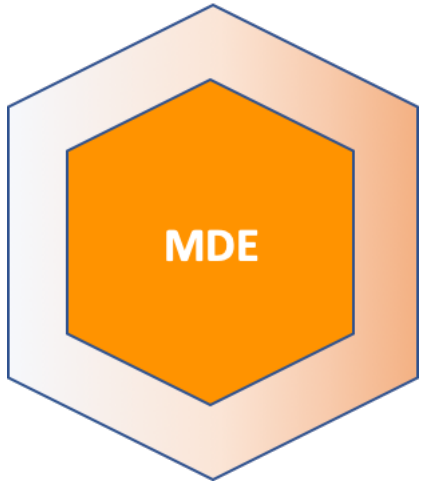
- 90% of 90% of all applications can be implemented without any application code

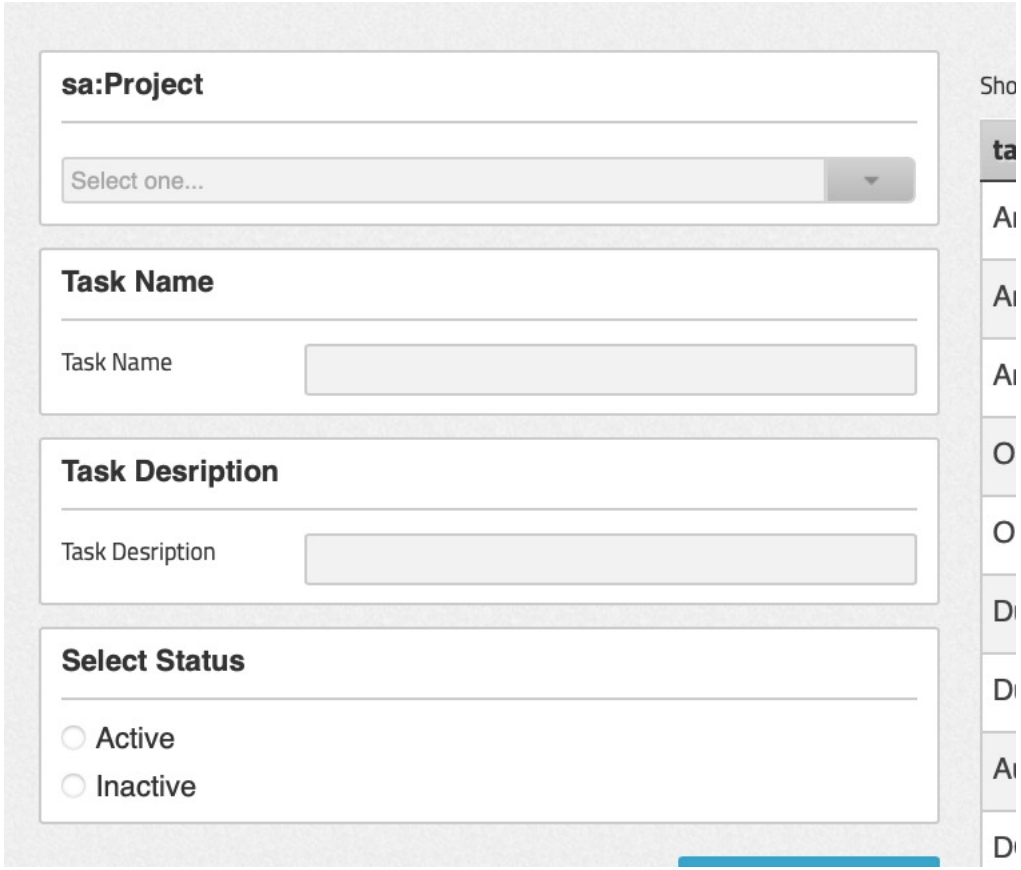


Model Driven Everything

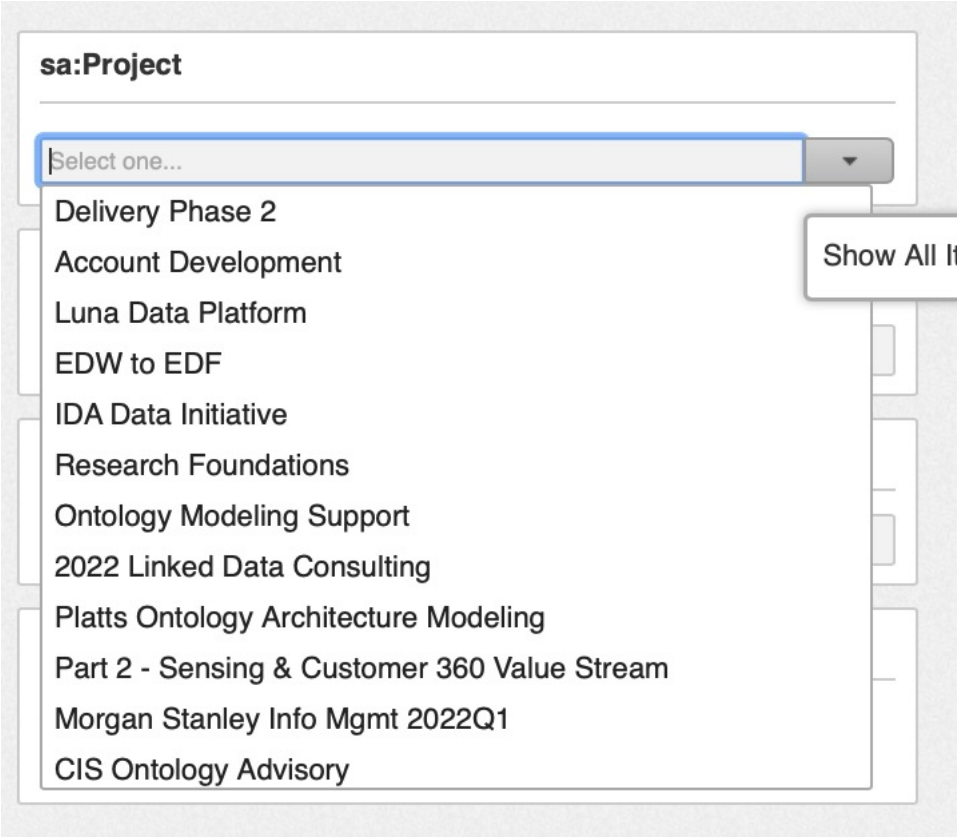
- How Data-Centric Model Driven differs from low code / no code

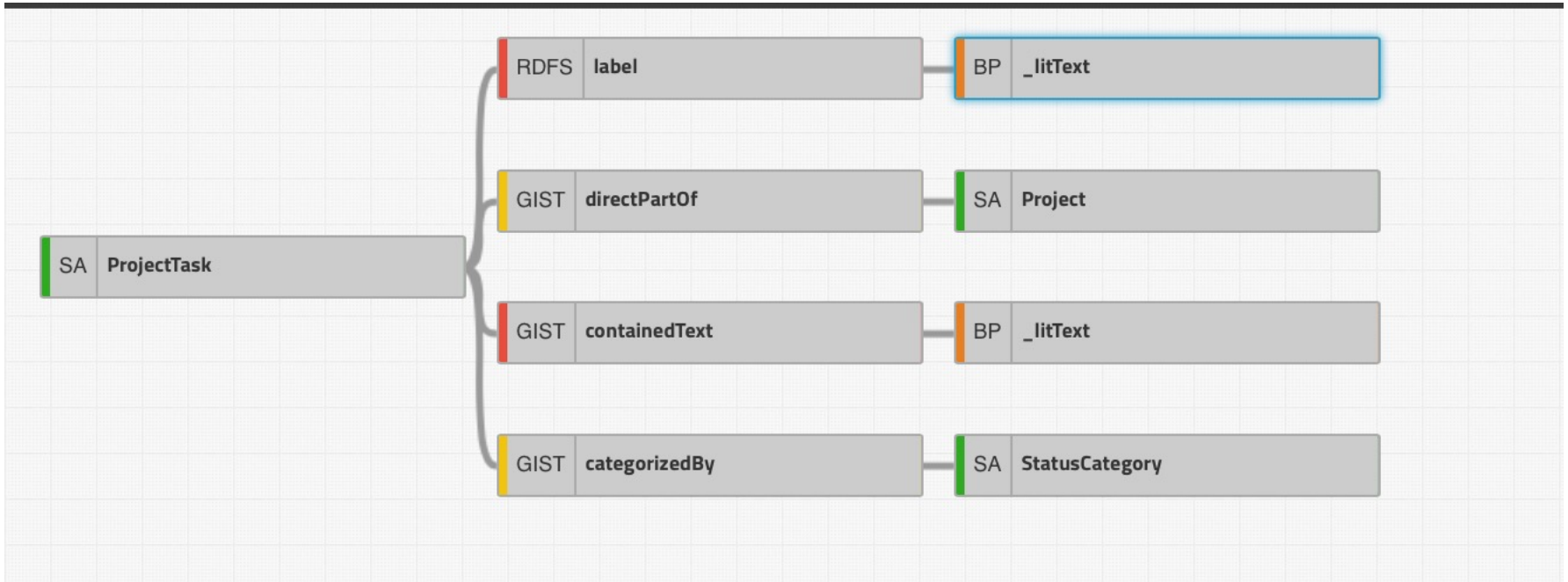


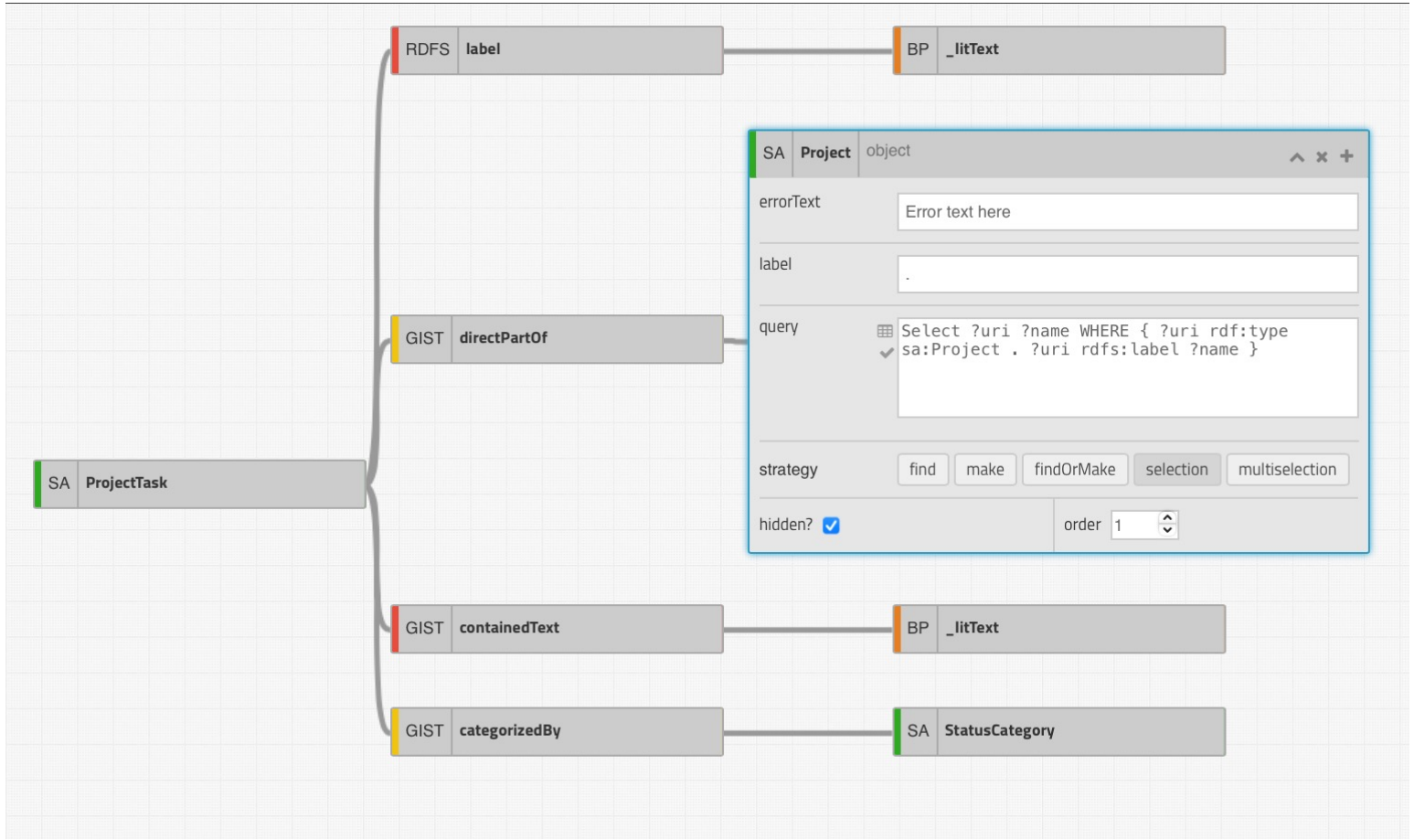




Demo of Model Driven

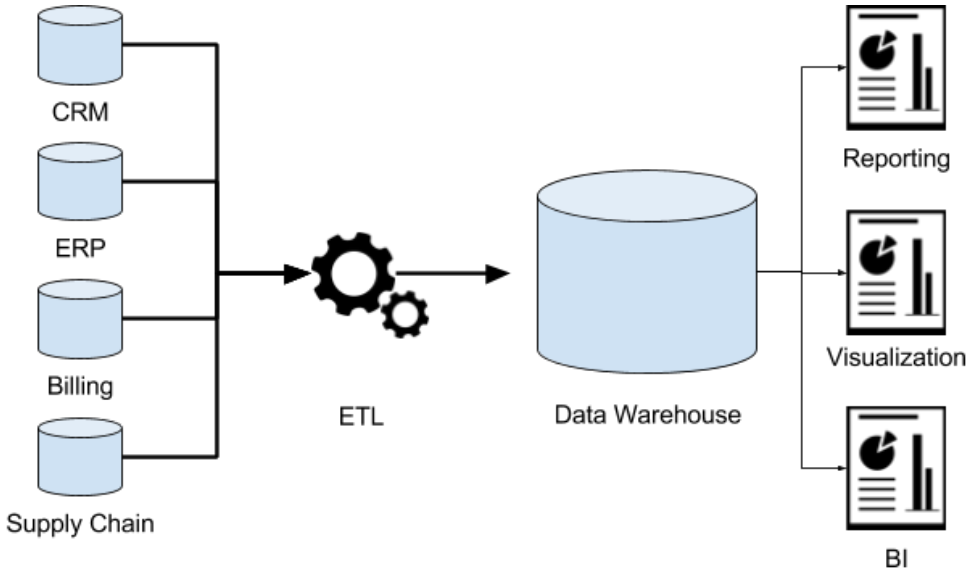






How is this different from …

Data warehouse



or data lake

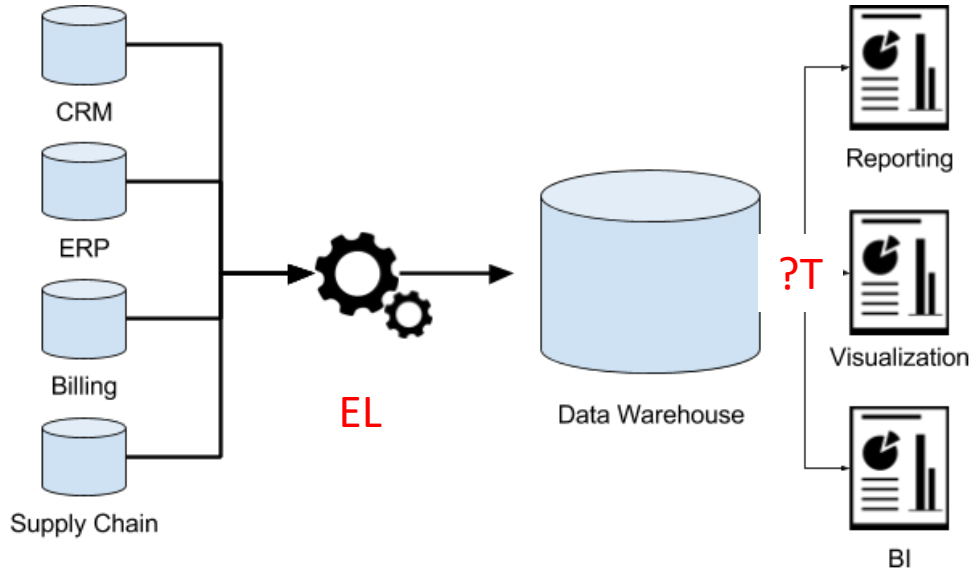
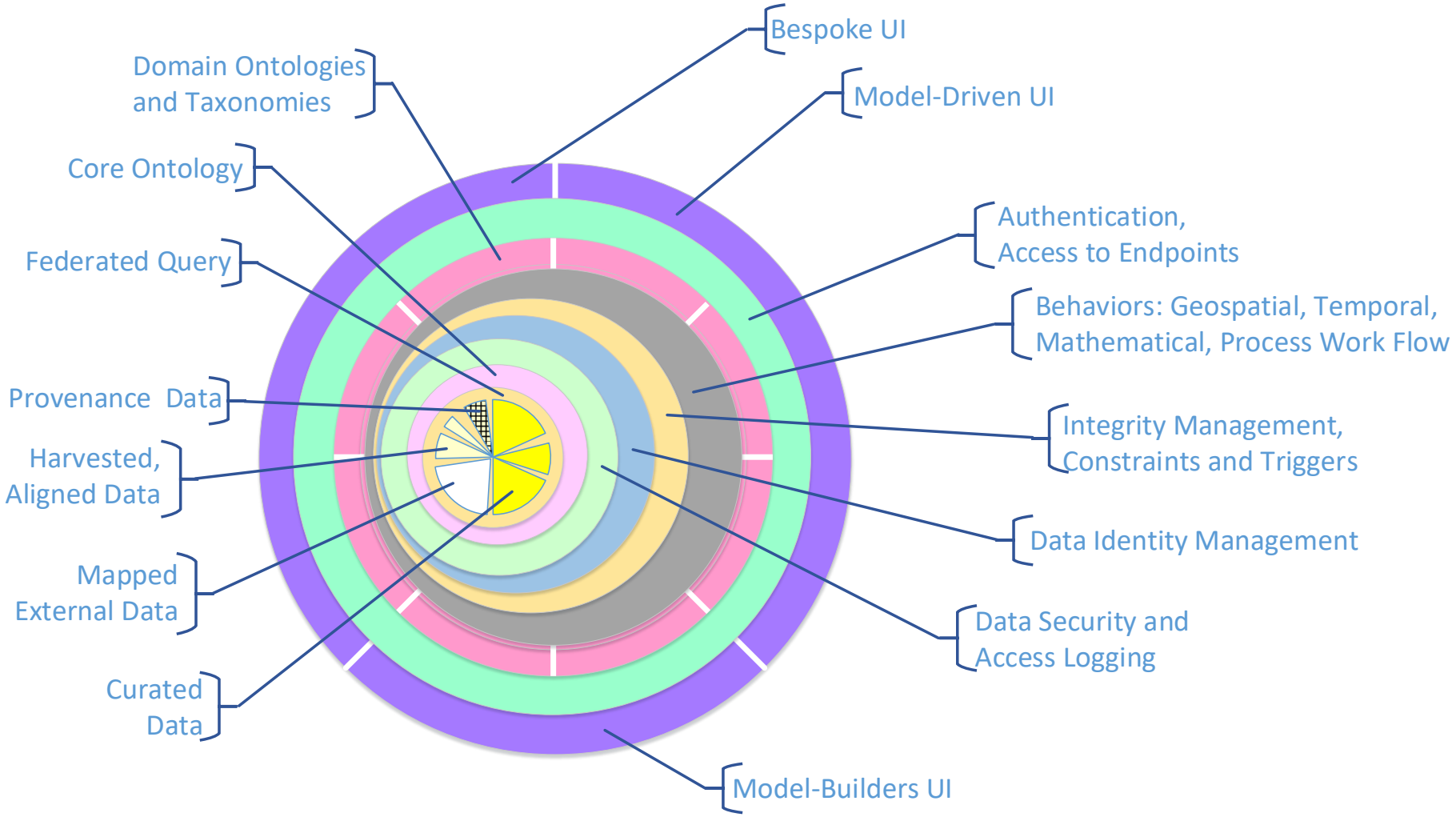


Photo by John Bell on Unsplash

The Data-Centric Architecture



To Pursue this Further

https://www.semanticarts.com

gist

https://www.dcaforum.com/dcaf-2022/

June 6-8 Fort Collins, CO

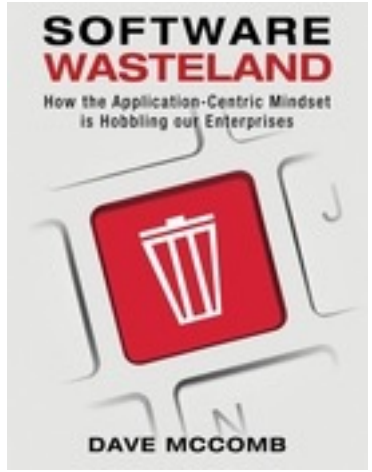
http://www.datacentricmanifesto.org

Join 1300 other professionals, sign the manifesto

Book Raffle

For those who didn't win, these links allow you to get 10 copies for the price of 4

https://technicspub.com/software_wasteland/



https://technicspub.com/data-centric/