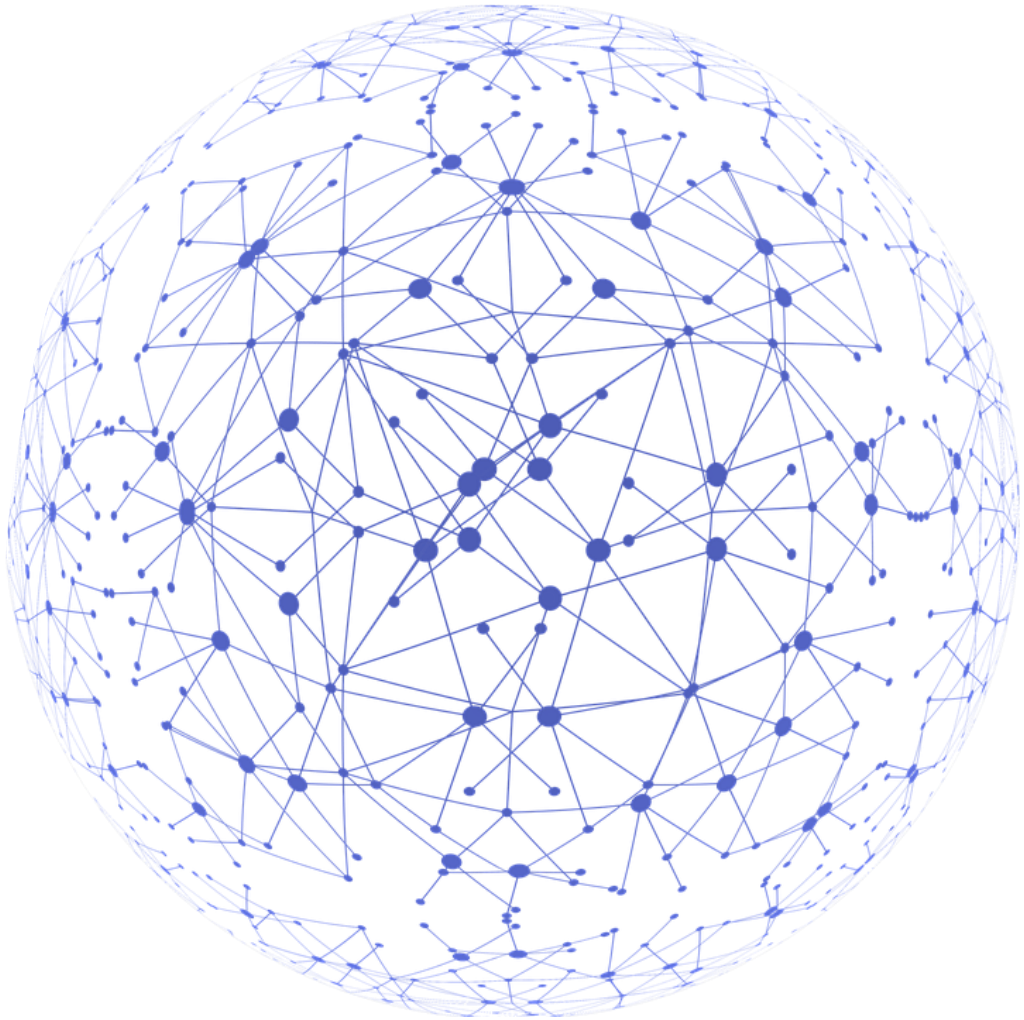
Introduction to Knowledge Technologies: Knowledge Graphs, the Semantic Web and Linked Data

Manolis Koubarakis
Eleni Tsalapati

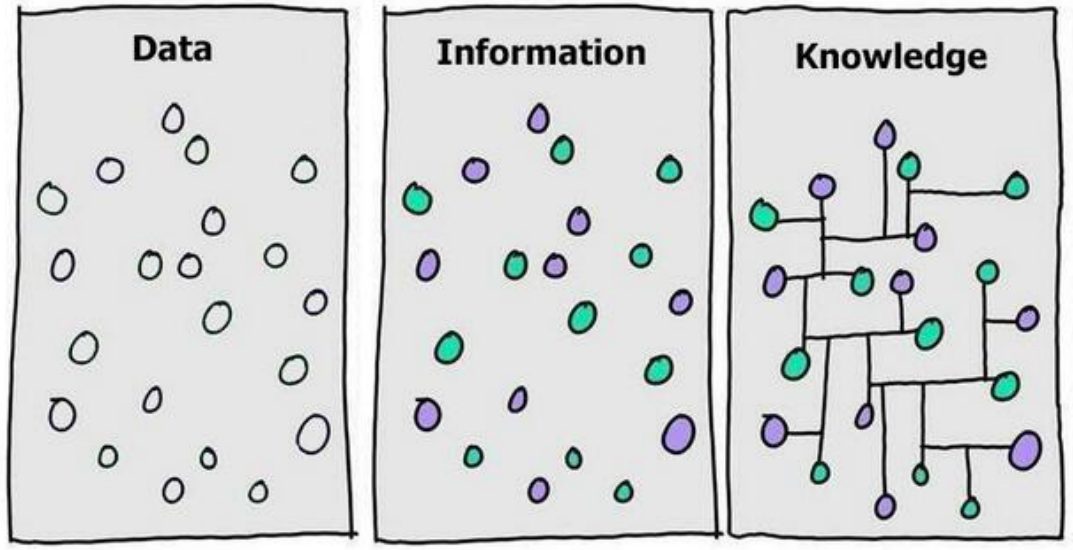
Outline

- **Basic concepts**
  - **knowledge**
- Some History: How did it all start?
- Why do we need them?
- So, what is an ontology?
  - RDFS, OWL, Description logics
- Leading Ontologies and KGs
- How do we query an ontology/KG?
  - SPARQL
- Linked Data
- Linked geospatial and temporal data
- High Value use cases



Basic concepts: Knowledge

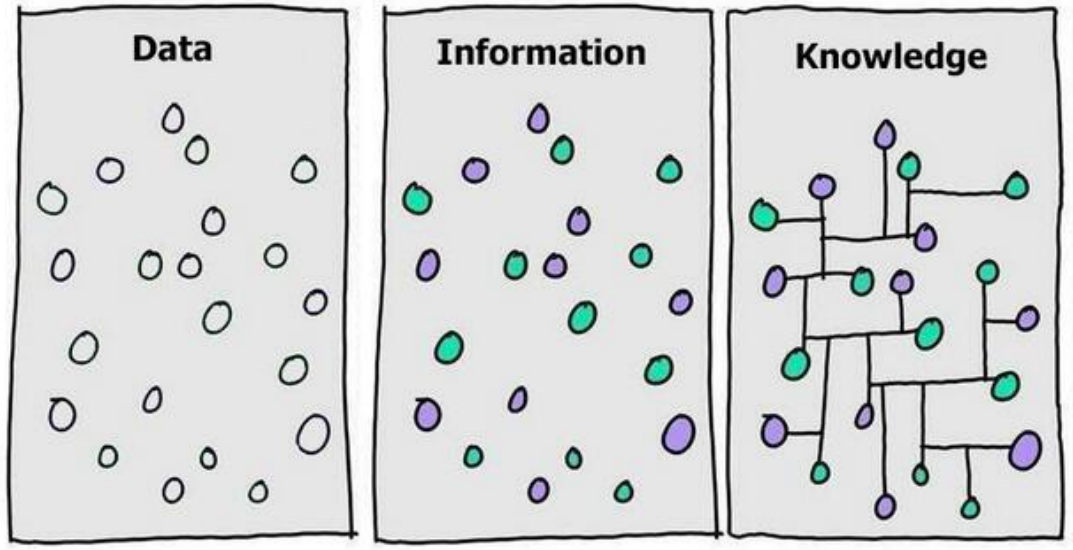
- **Data**
  - **Unorganized**
  - **Unprocessed**,
  - discrete, objective **facts** or
  - **observations**
  - Without a meaning (e.g., sensor measurements)



Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science. 2007;33(2):163-180.

Basic concepts: Knowledge

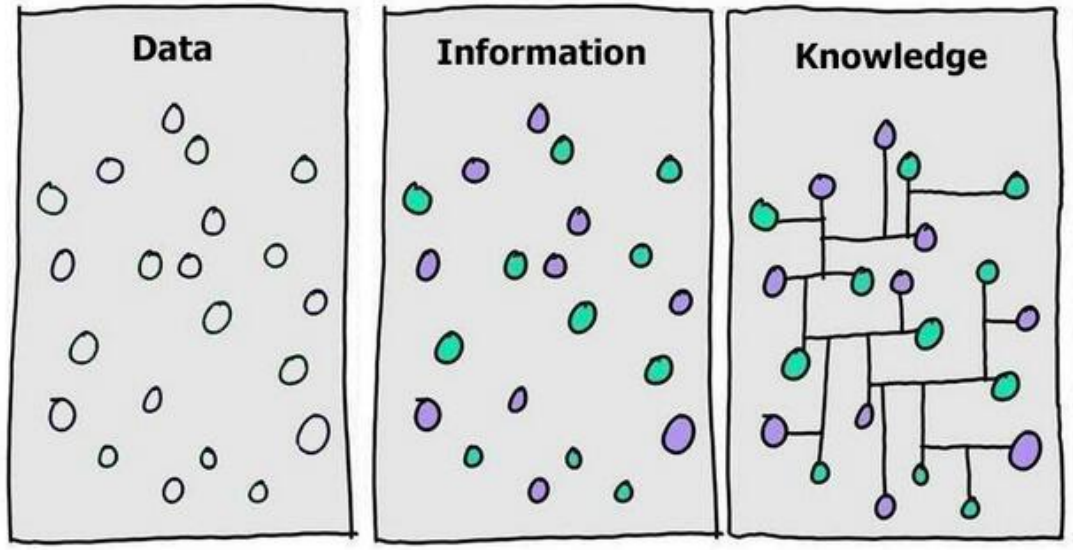
- **Information**:
  - Data processed for a purpose
  - E.g., statistical analysis



Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science. 2007;33(2):163-180.

Basic concepts: Knowledge

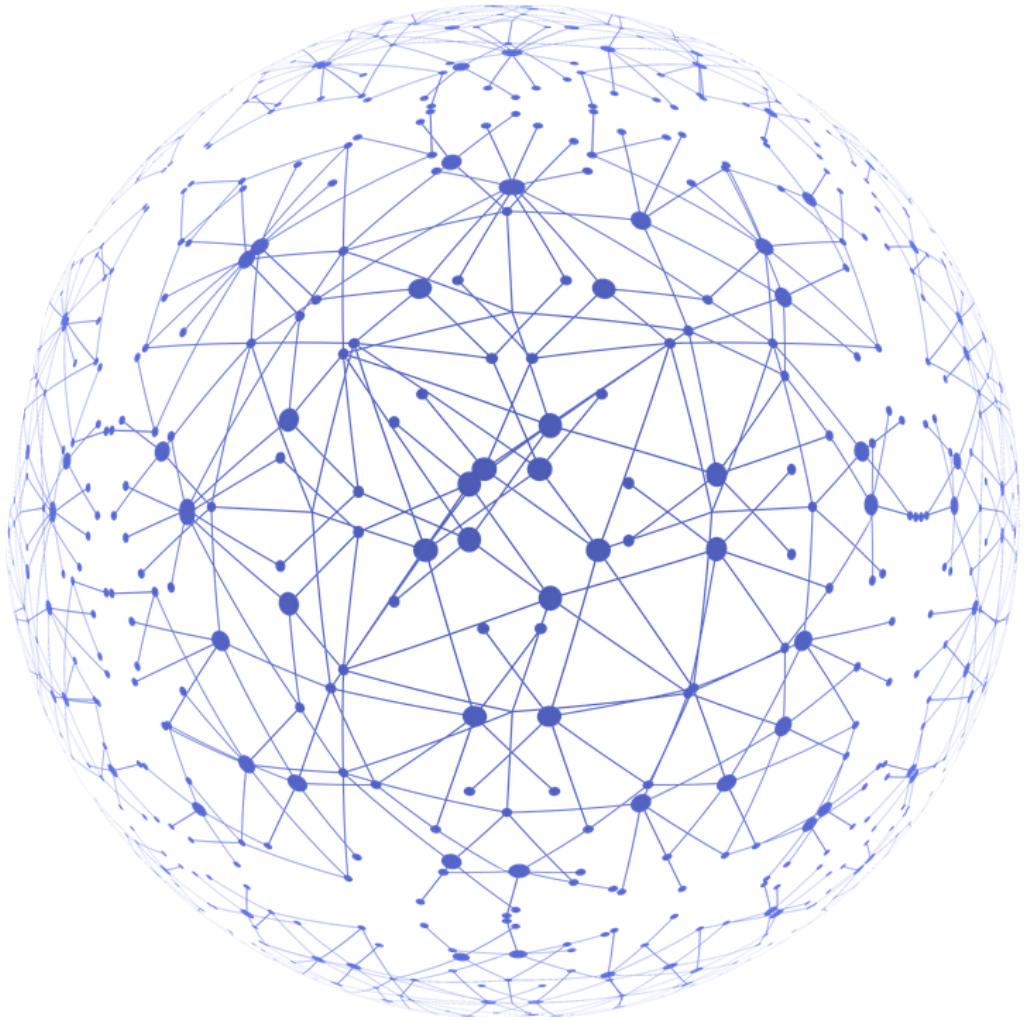
- **Knowledge:**
  - Combination of:
    - Data
    - Information
    - Expert's Knowledge
- Decision Making



Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science. 2007;33(2):163-180.

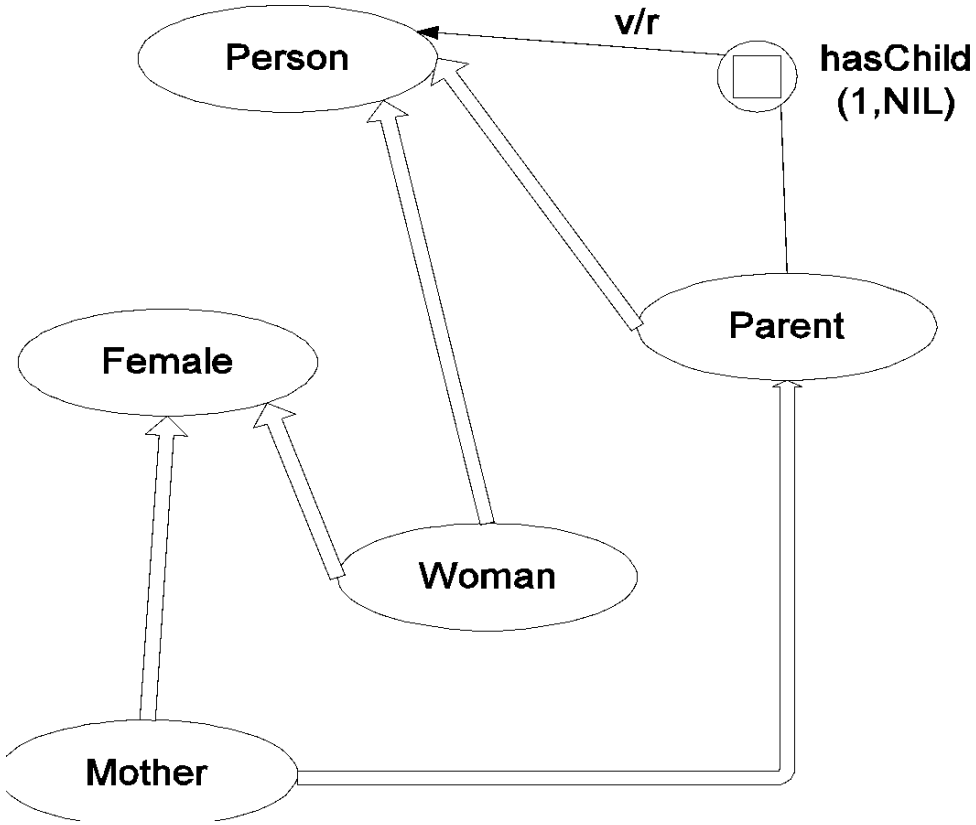
Outline

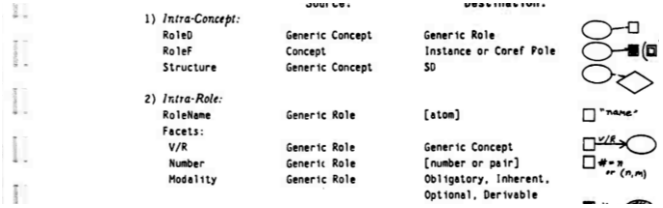
- **Basic concepts**
  - **knowledge,** ontology, knowledge graphs
- **Some History: How did it all start?**
- Why do we need them?
- So, what is an ontology?
  - RDFS, OWL, Description logics,
- How do we query an ontology/KG?
  - SPARQL
- Linked Data
- Linked geospatial and temporal data
- High Value use cases



Some History: Knowledge as Graph - KL-ONE

- Appeared in **1977** by Brachmann
- Idea: **Structured inheritance** networks
- **KL-ONE descriptions** are always formed from other **more general KL-ONE descriptions**
- **Innovation**: use of a deductive classifier
- Can **validate** a frame ontology and **deduce** new information



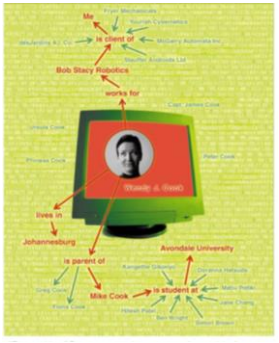


KL-ONE manual: https://apps.dtic.mil/sti/tr/pdf/ADA122437.pdf

Some History: Knowledge as Graph

"The Semantic Web". Scientific American. *Tim Berners-Lee, James Hendler and Ora Lassila (May 2001)*

- "The Semantic Web will bring **structure** to the meaningful content of Web pages, creating an environment where **agents roaming** from page to page readily carry out **sophisticated** tasks for users."



Some History: Knowledge as Graph

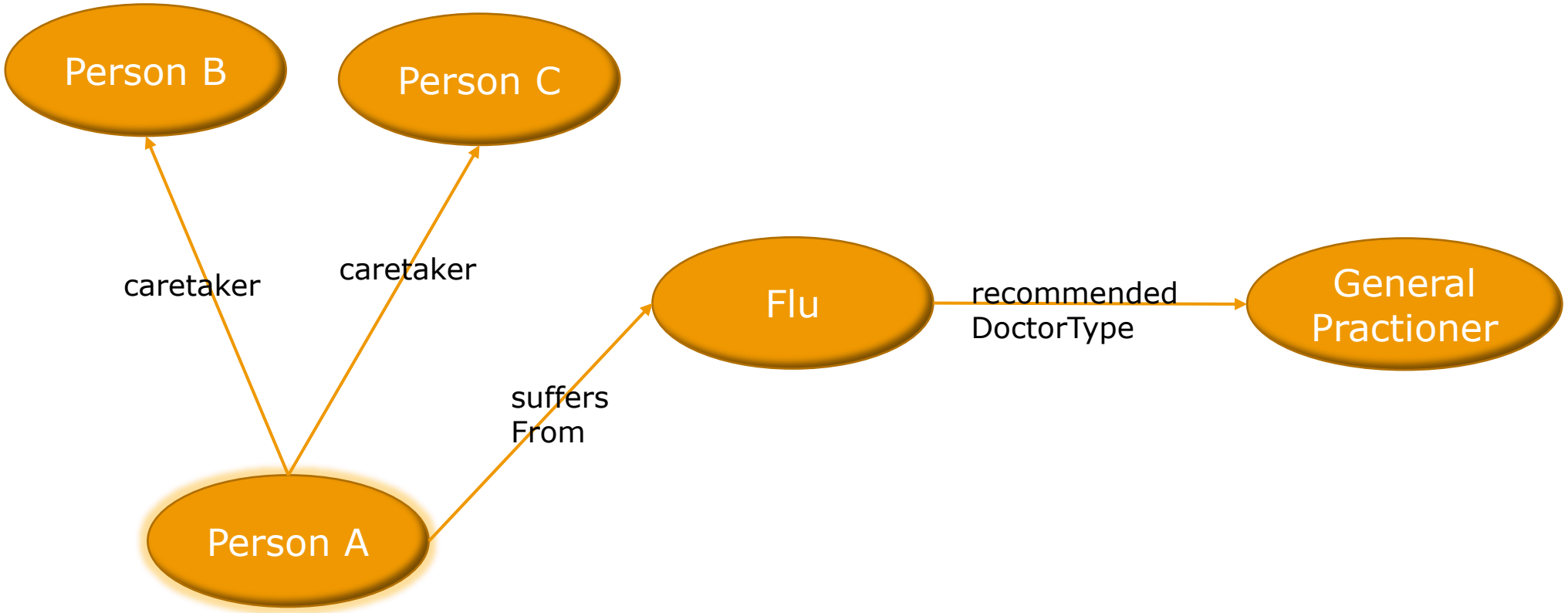
- Scientific American: The Semantic Web, Tim Berners-Lee, James Hendler and Ora Lassila (May 2001)

- "The Semantic Web will bring **structure** to the meaningful content of Web pages, creating an environment where **agents roaming** from page to page readily carry out **sophisticated** tasks for users."

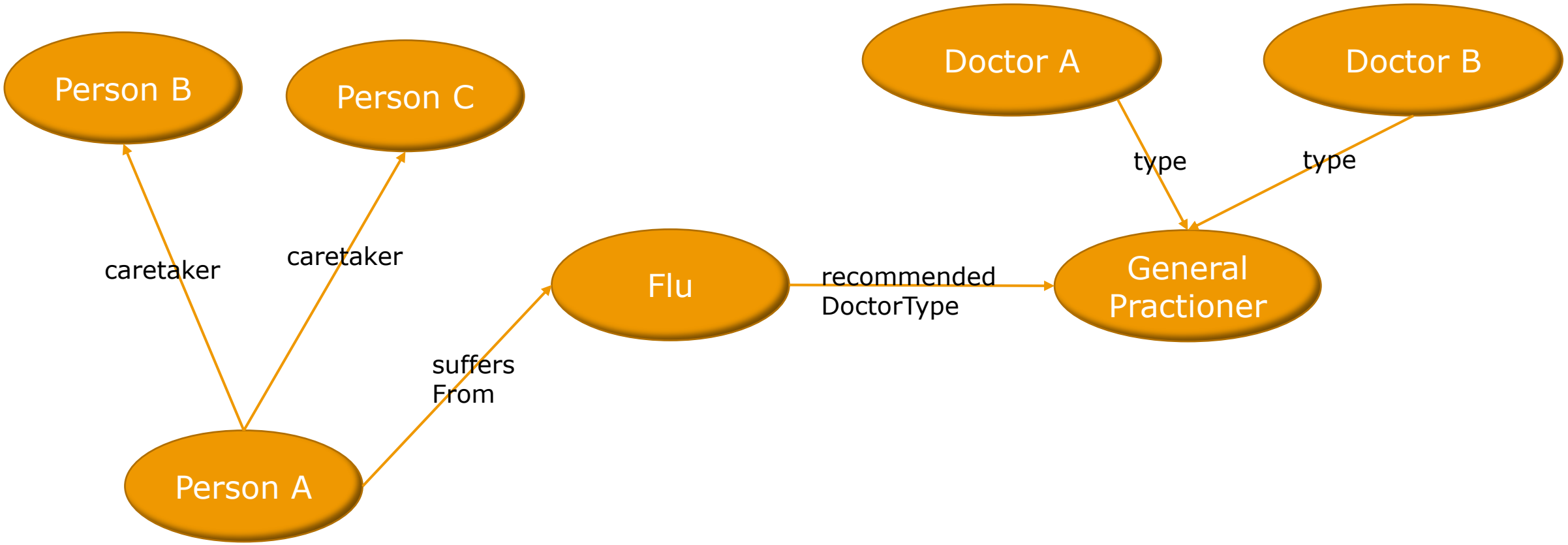
This practically means that the data will be published on the web:

- in a **machine** and **human understandable** way
- the machines will be able to **collect data** from various **distant resources**
- and they will be able to make **inferences** from these data

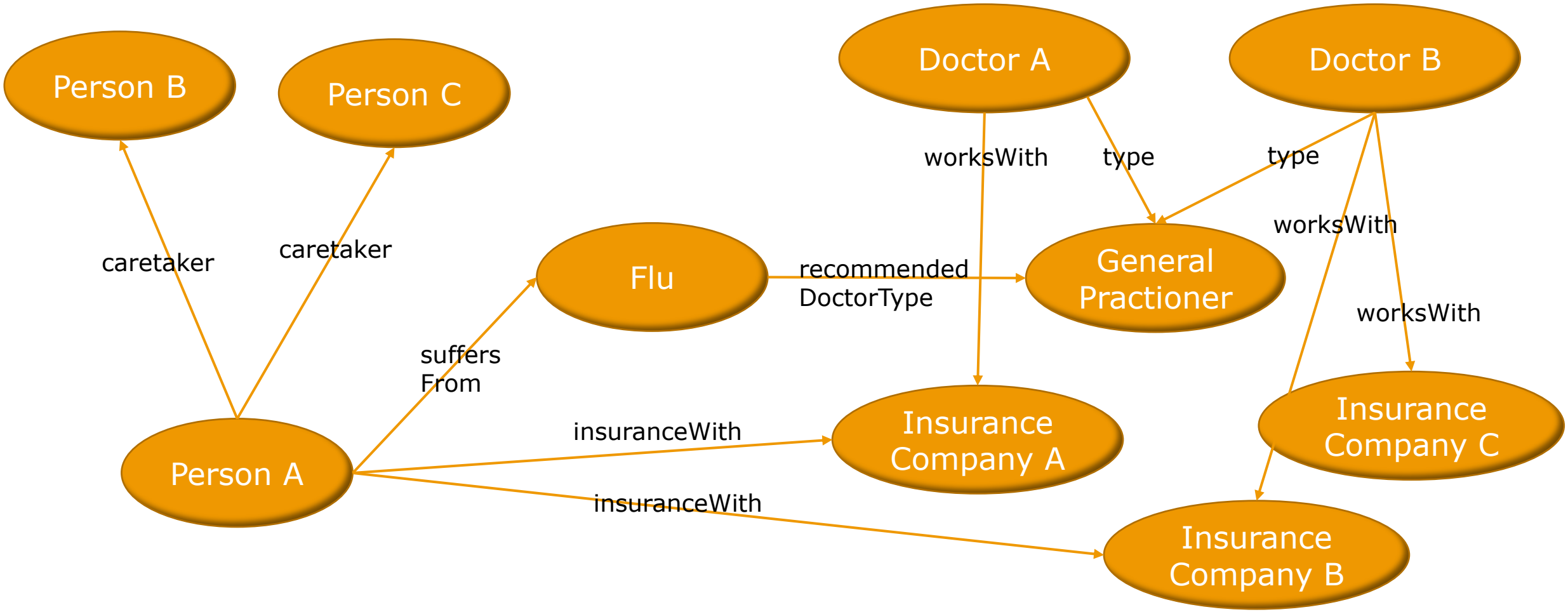
Towards a Better Web!



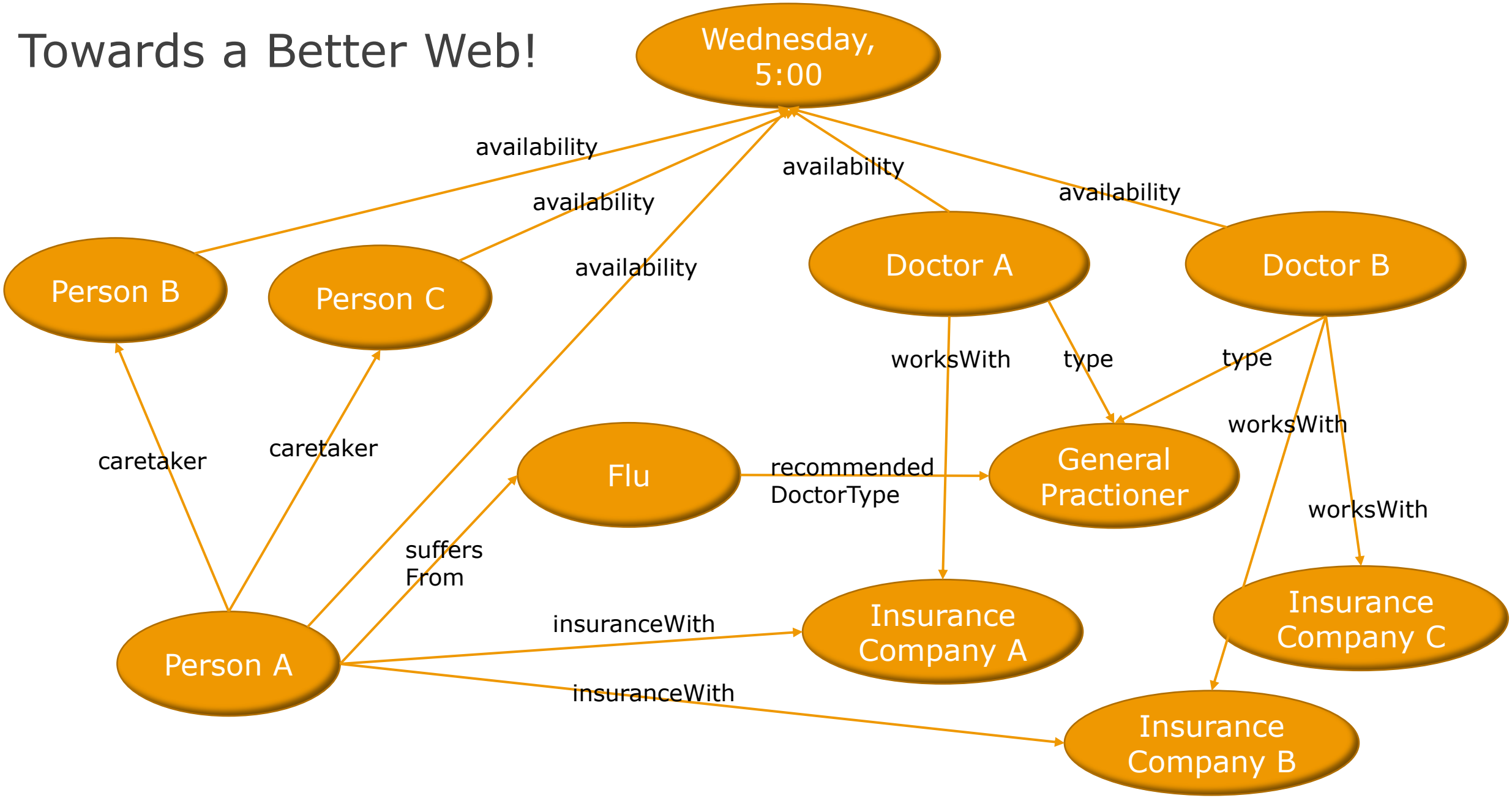
Towards a Better Web!



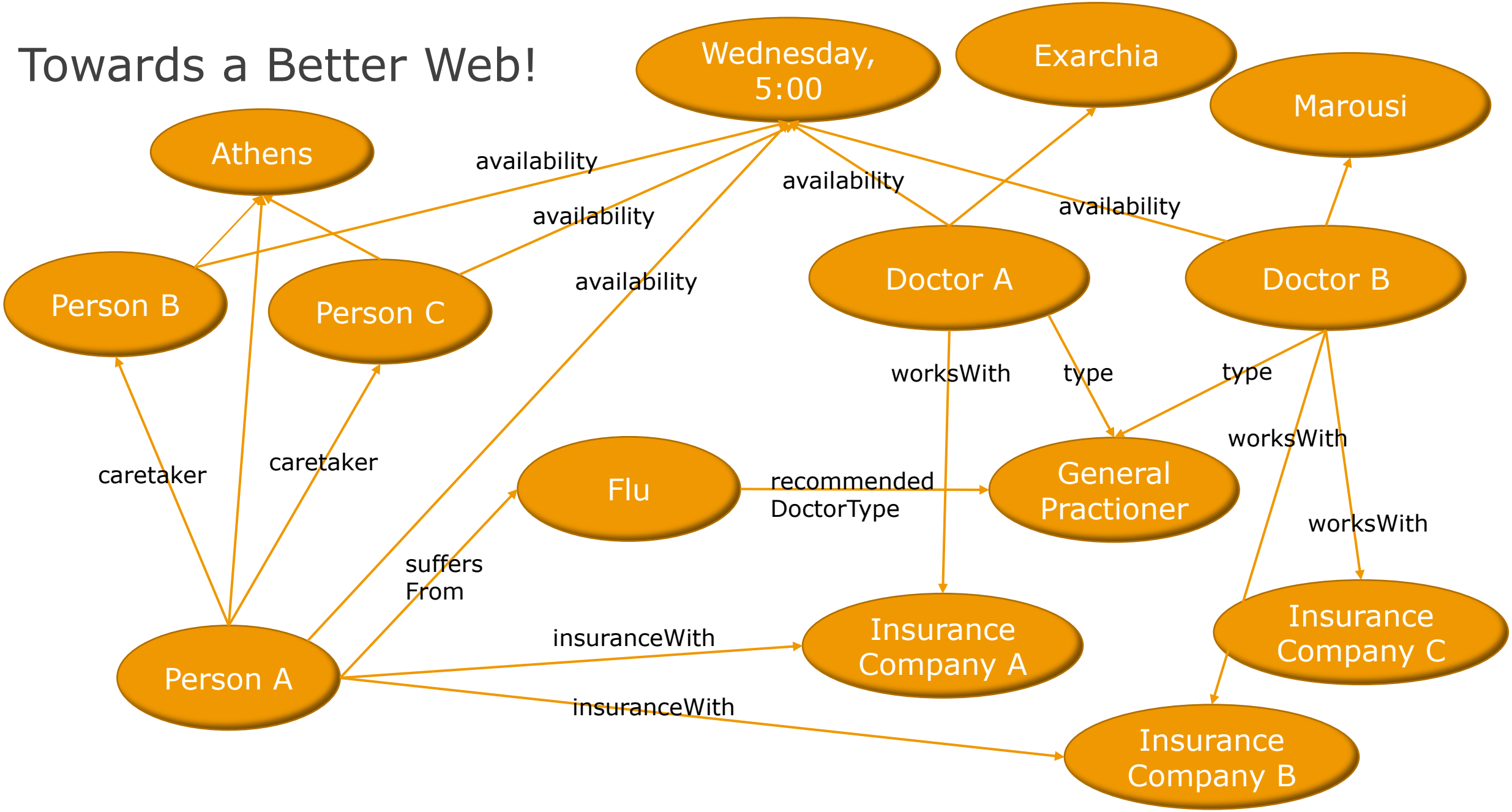
Towards a Better Web!



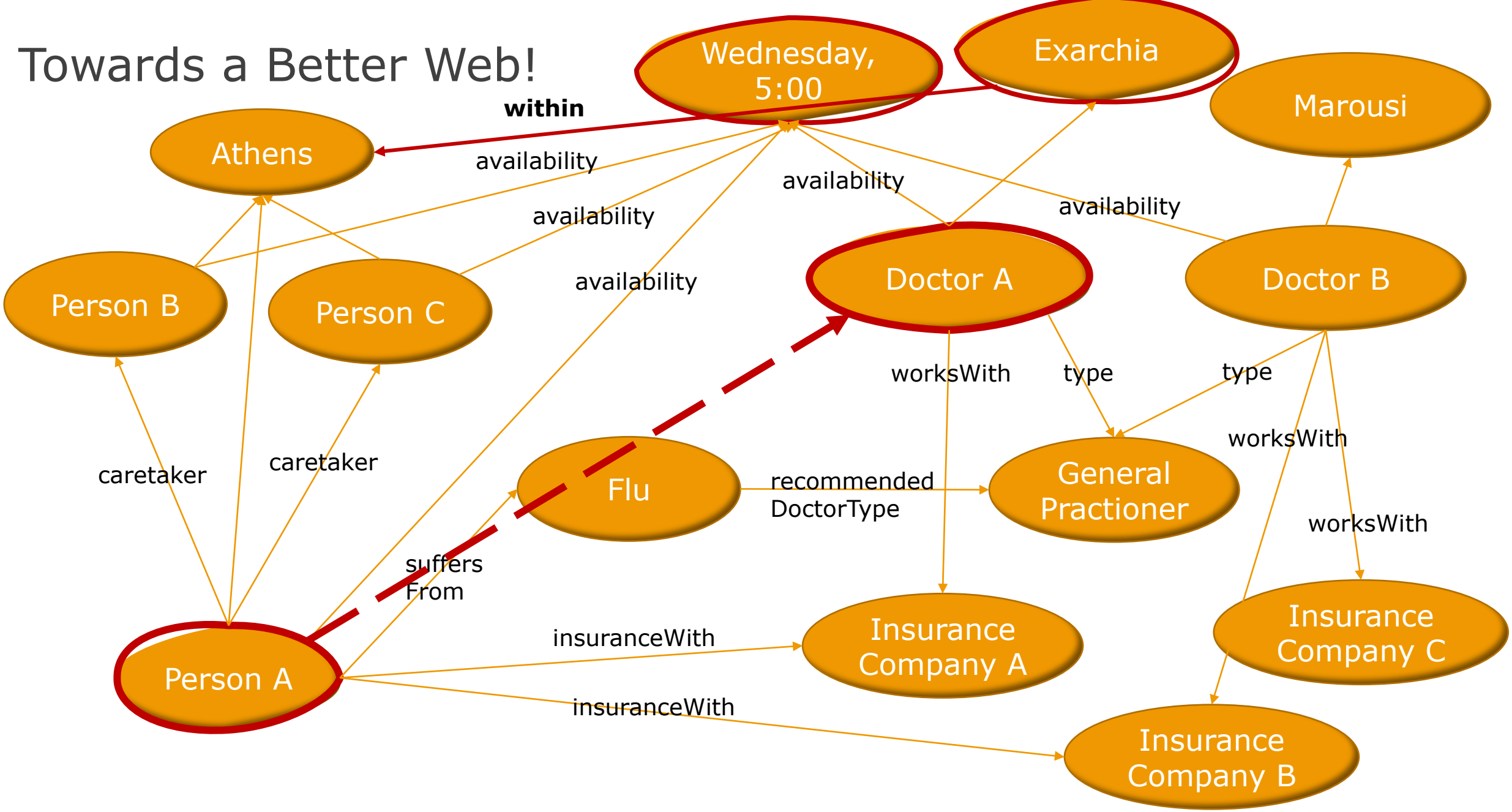
Towards a Better Web!



Towards a Better Web!

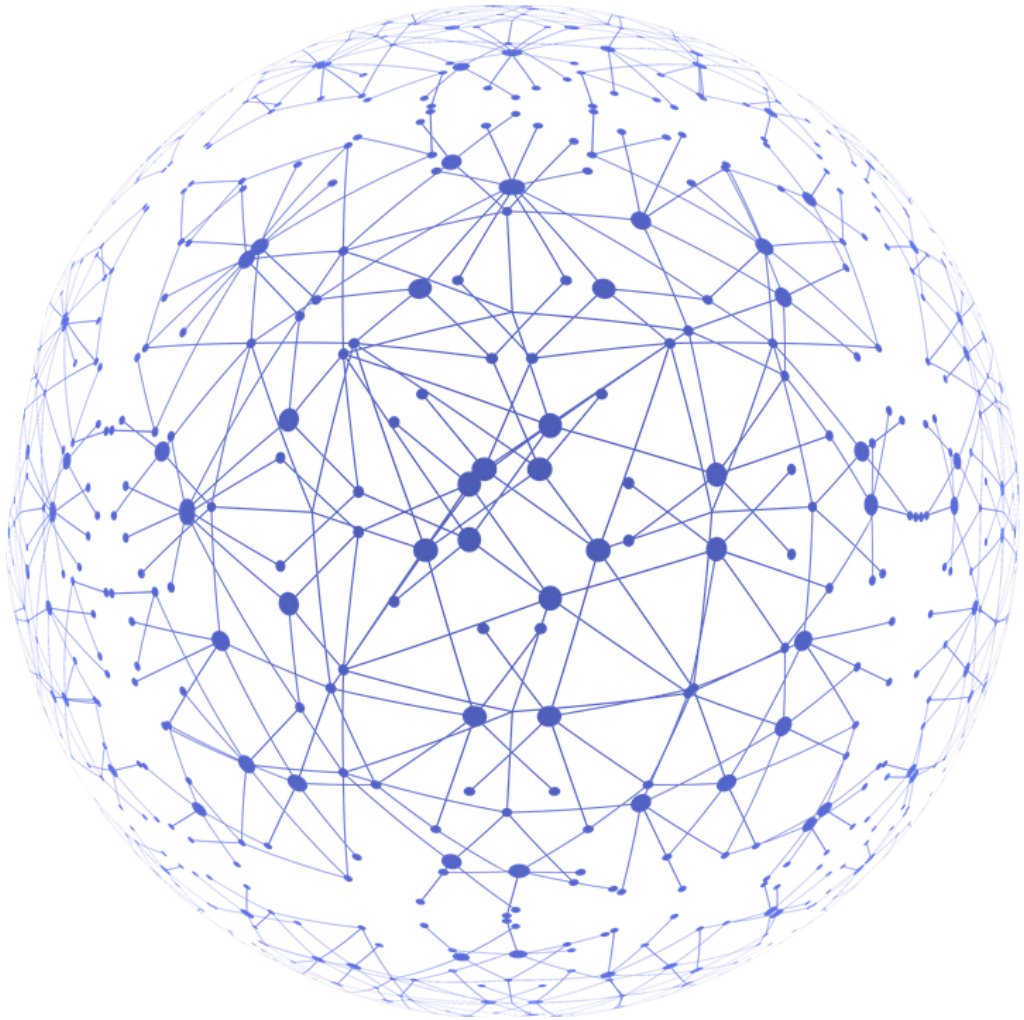


Towards a Better Web!



Outline

- Basic concepts
  - knowledge, ontology, knowledge graphs
- Some History: How did it all start?
- **Why do we need them?**
- Where are we today?
- So, what is an ontology?
  - RDFS, OWL, Description logics,
- How do we query an ontology/KG?
  - SPARQL
- Linked Data
- Linked geospatial and temporal data
- High Value use cases



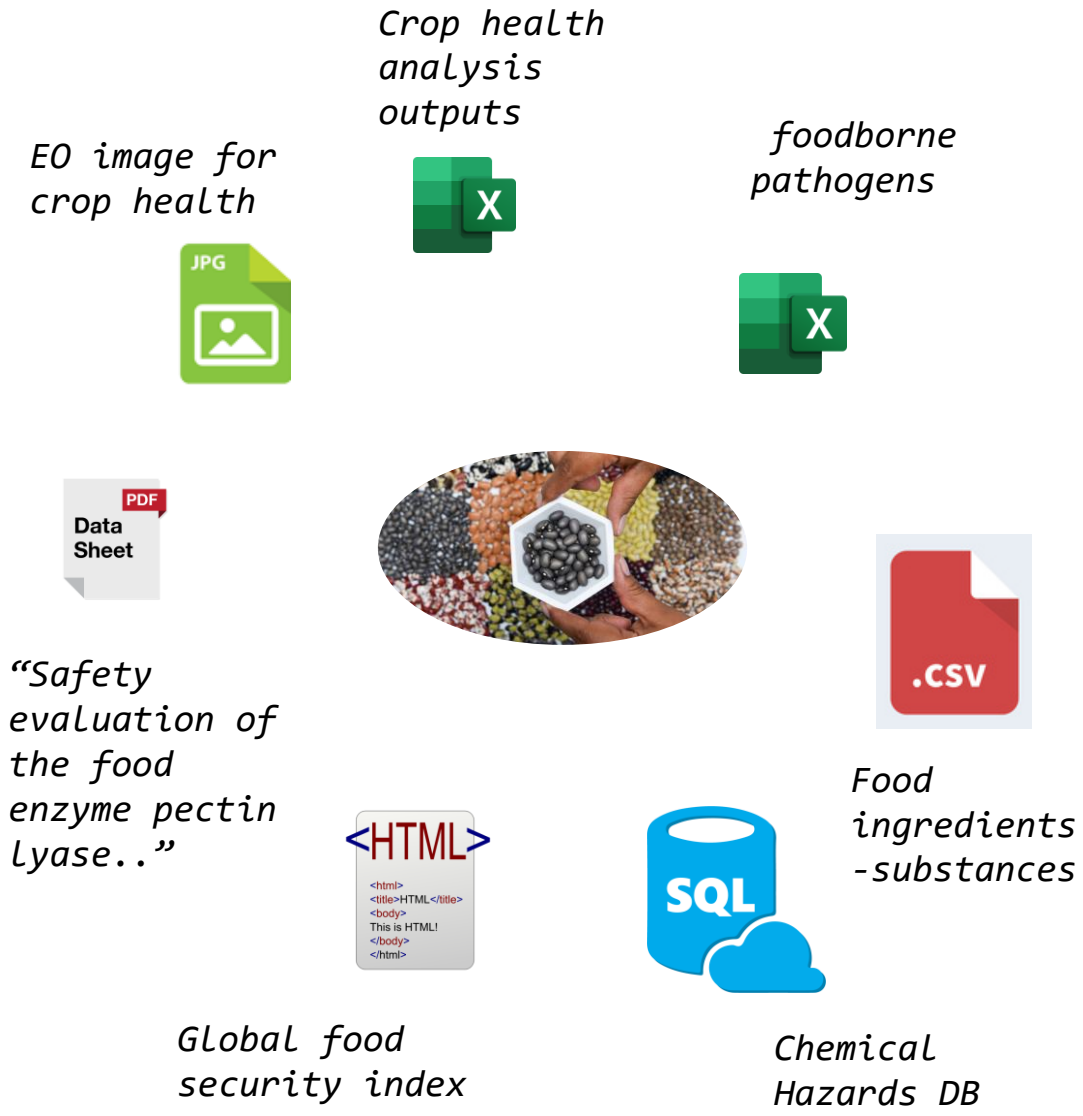
Motivating Example: Food Security

- *Suppose that you work in Food Security and you want to find out if some food is edible.*

- *What piece of information you would have to look for?*

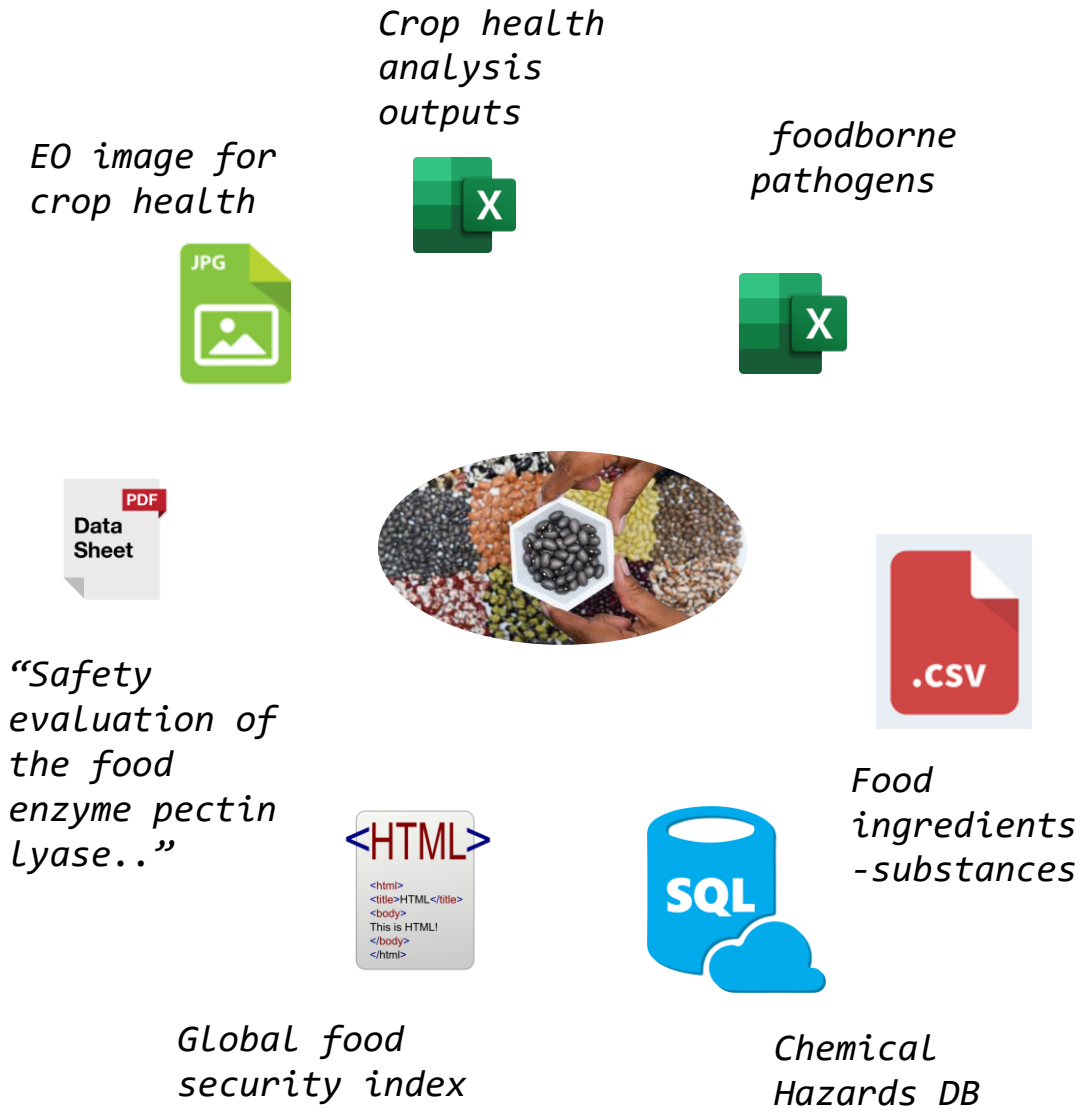
Motivating Example: Food Security

- *Suppose that you work in Food Security and you want to find out if some food is edible.*
- *What piece of information you would have to look for?*



Motivating Example: Food Security

- Various types of **format**
- **Heterogeneous** systems (different dialects)
- **Multiple** resources
- **Different terminology**
  - in North America "biscuit" refers to a softer "quick bread"
  - in Britain it usually means a hard, flat unleavened baked product



Motivating Example: Food Security

- **Aggregate:**
  - Metadata
  - Machine/Human-readable language
- **Integrate:**
  - Interlink metadata, interlink data
  - Shared understanding of the terminology
- **Make sense of these data**
  - Implied information

Format: png
About: EO image
Id: xyz
Area: <3.2444, …>
Resolution: 3.2m
Date:…
Crop type: Sugar crops

Format:xlsx
About: Image analysis outputs
Id:wvy
Input:xyz
Date:…

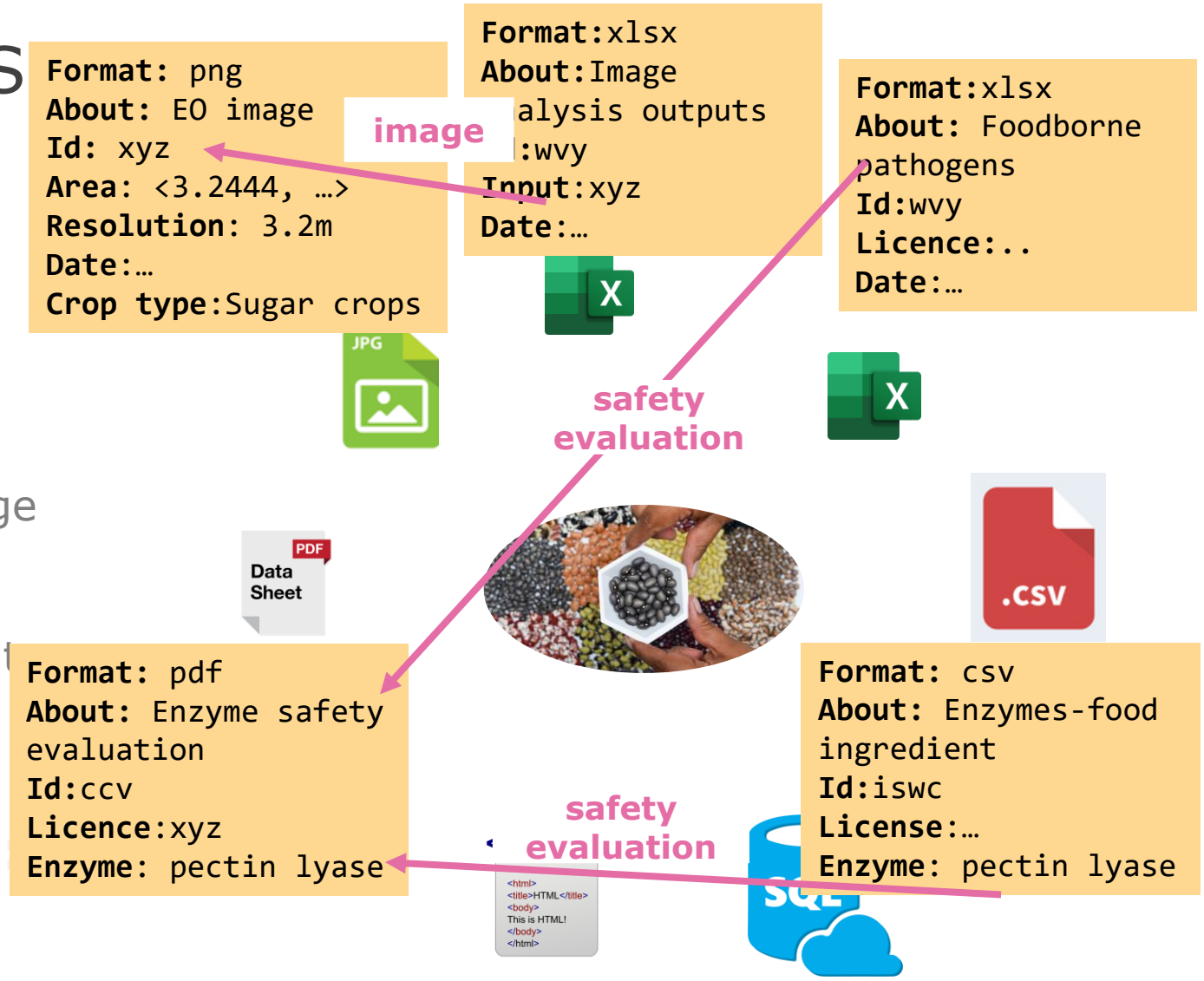
Format:xlsx
About: Foodborne pathogens
Id:wvy
Licence:..
Date:…

Format: pdf
About: Enzyme safety evaluation
Id:ccv
Licence:xyz
Enzyme: pectin lyase

Format: csv
About: Enzymes-food ingredient
Id:iswc
License:…
Enzyme: pectin lyase

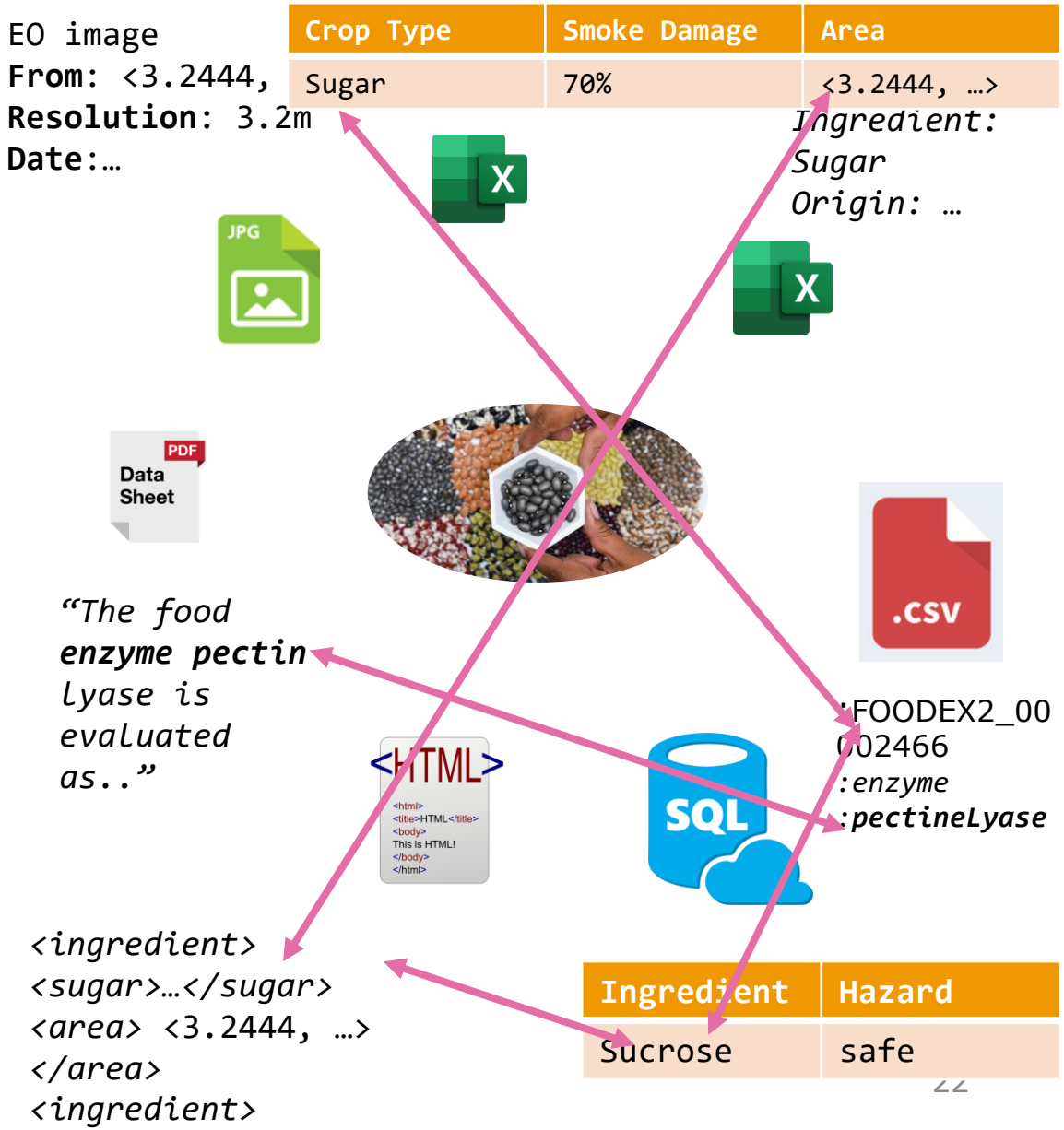
Motivating Example: Food S

- **Aggregate:**
  - Metadata
  - Machine/Human-readable language
- **Integrate:**
  - **Interlink metadata**, interlink dat
  - Shared understanding of the terminology
- **Make sense of these data**
  - Implied information



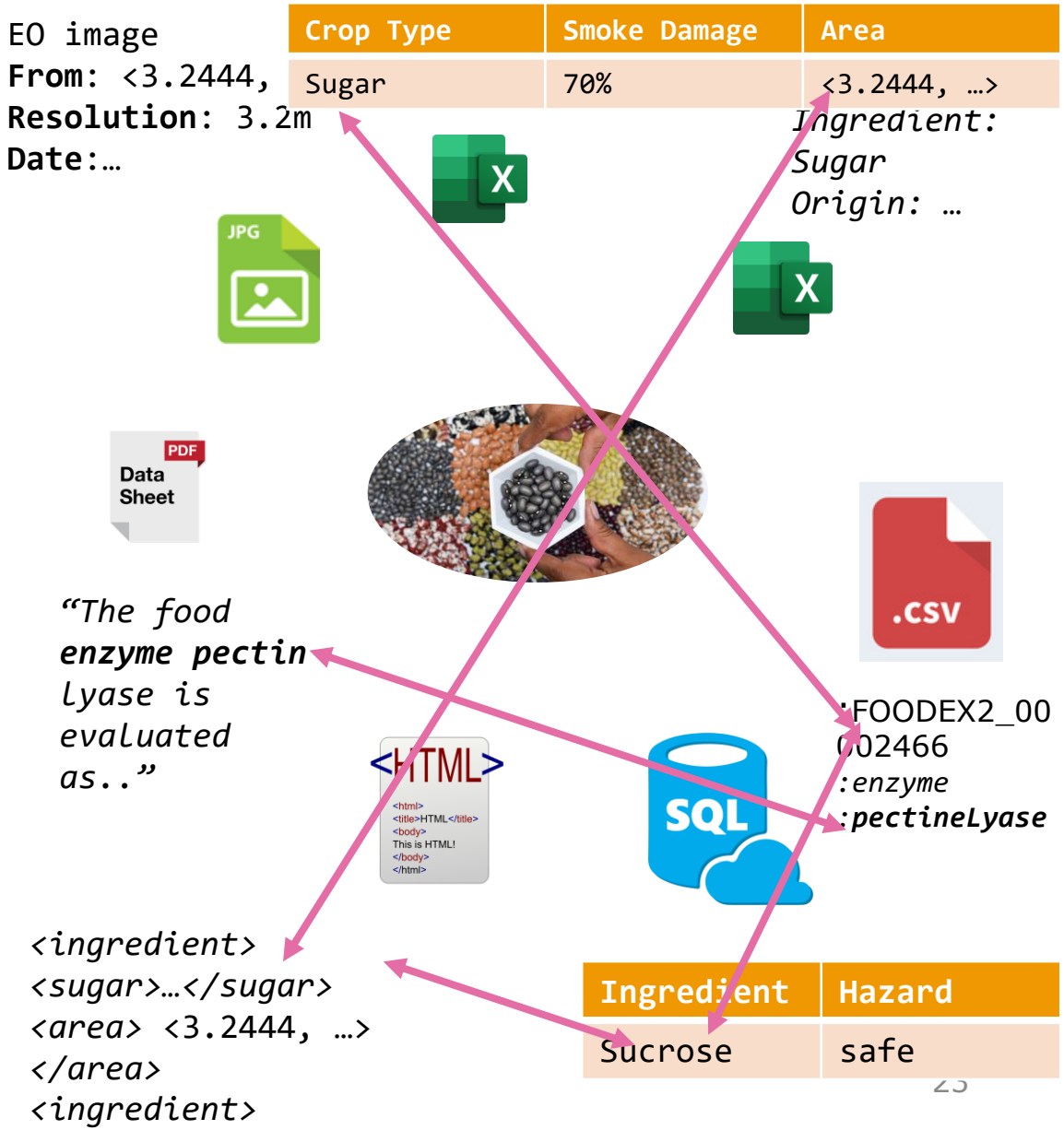
Motivating Example: Food Security

- **Aggregate:**
  - Metadata
  - Machine/Human-readable language
- **Integrate:**
  - Interlink metadata, **interlink data**
  - Shared understanding of the terminology
- **Make sense of these data:**
  - Implied information



Motivating Example: Food Security

- **Aggregate:**
  - Metadata
  - Machine/Human-readable language
- **Integrate:**
  - Interlink metadata, **interlink data**
  - Shared understanding of the terminology
- **Make sense of these data:**
  - Implied information

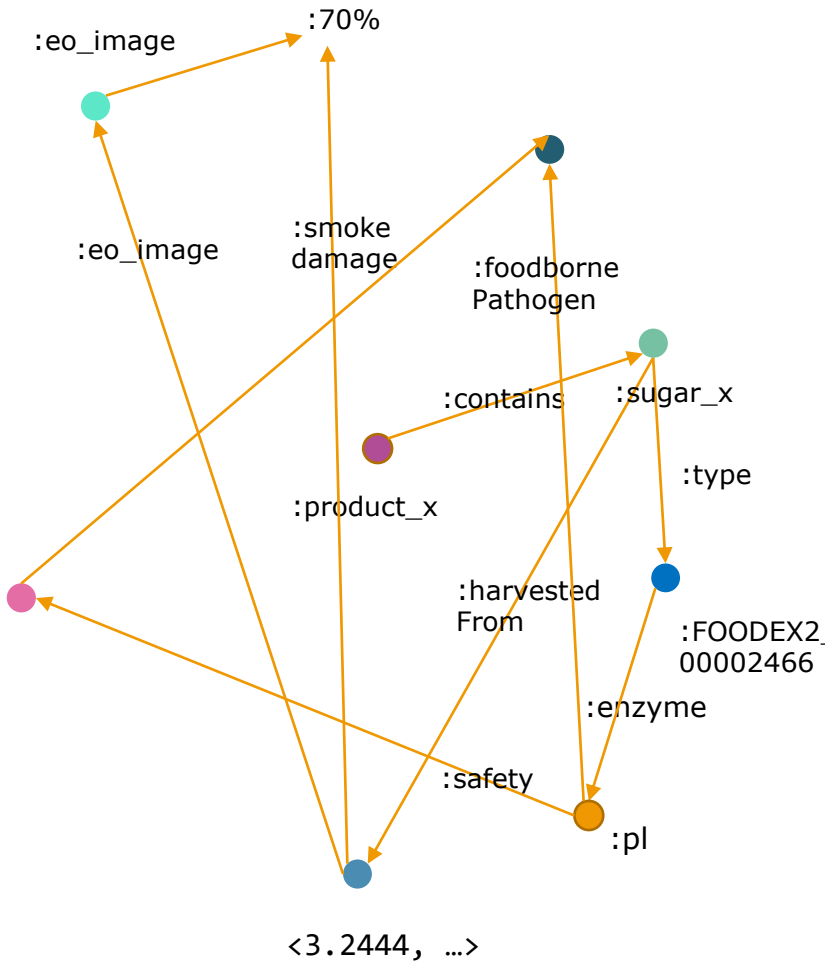


Core Idea

- Information **no longer on sheer sheets of data**

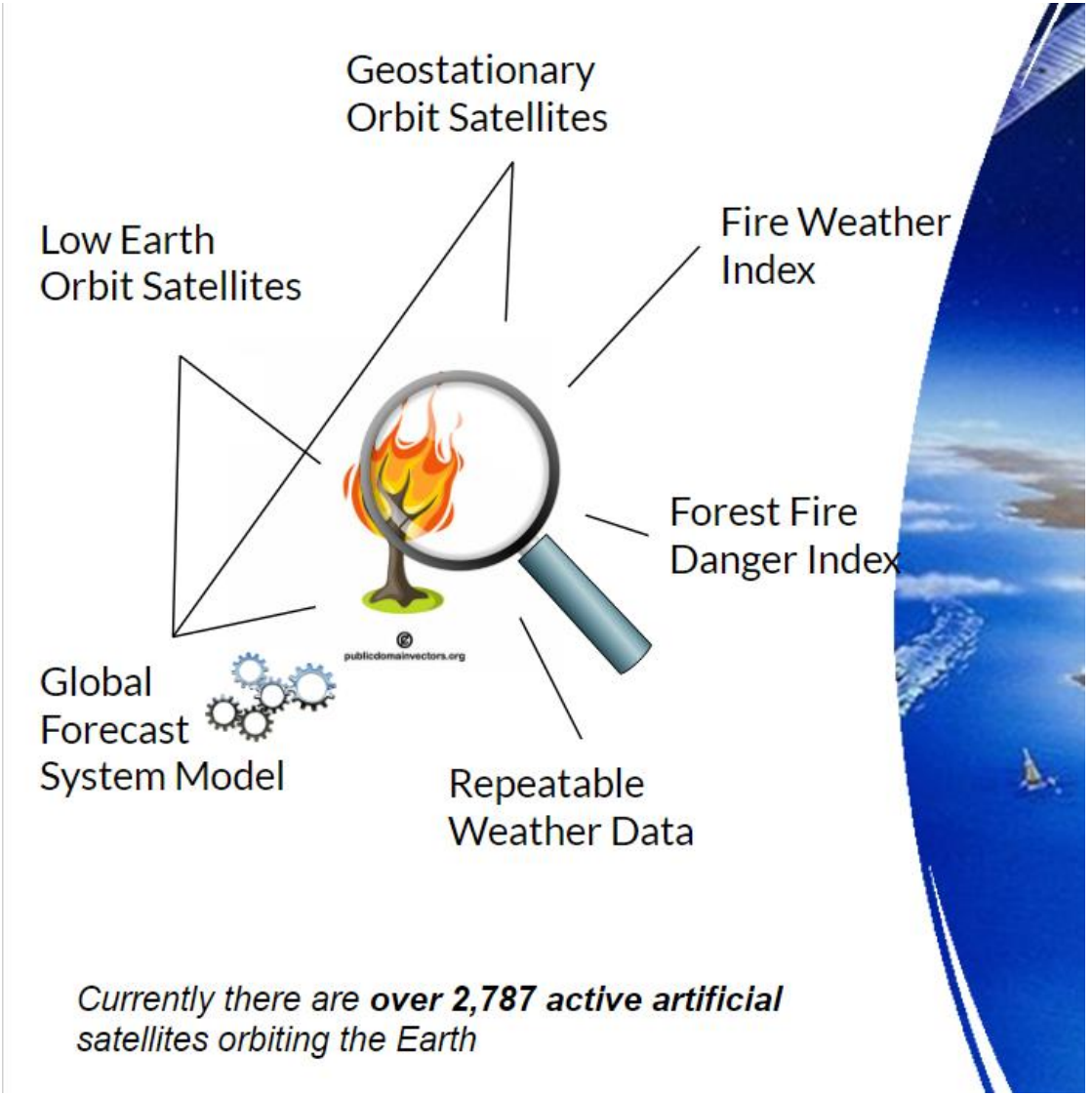
But:

- **Each** piece of information is represented as a **unique node** and the nodes are **interrelated** with **labeled links**
- From human readability to machine **processibility**



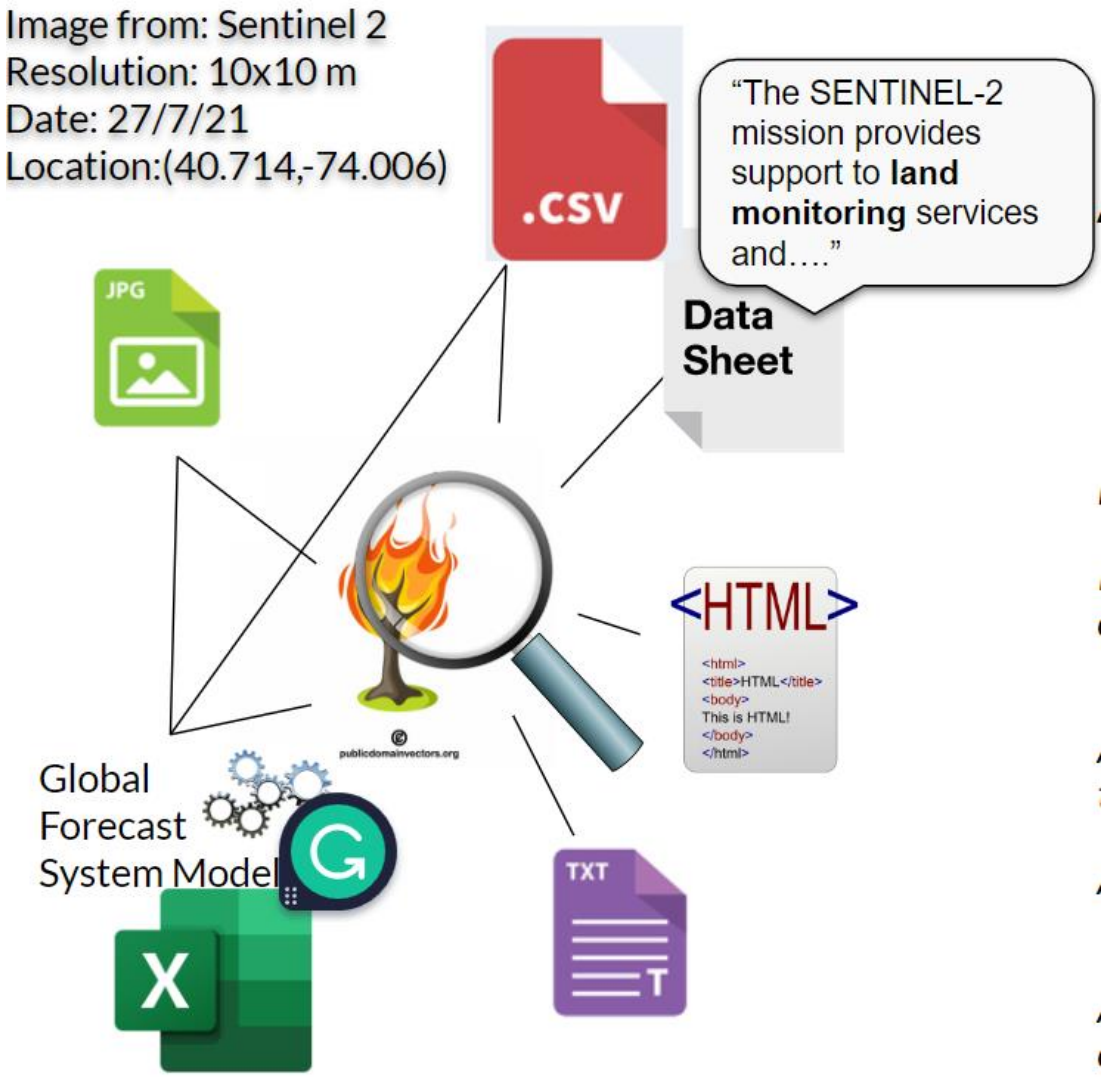
Motivating Example: Earth Observation & Fire monitoring

- **Aggregate:**
  - Metadata
  - Machine/Human-readable language
- **Integrate:**
  - Interlink metadata, **interlink data**
  - Shared understanding of the terminology
- **Make sense of these data:**
  - Implied information



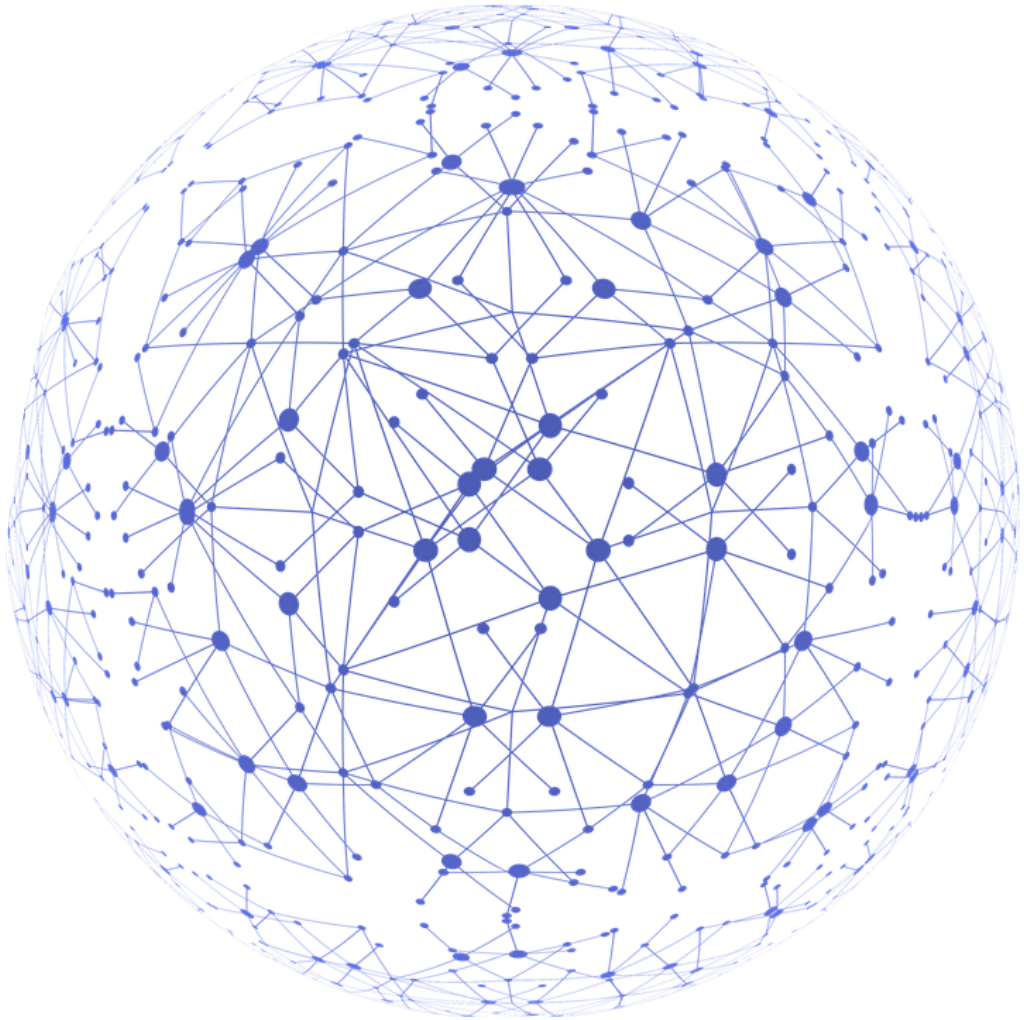
Motivating Example: Earth Observation & Fire monitoring

- How can we find the data that we want?
  - Metadata: data about data.
- Is this enough? How Sentinel 2 data relate to fire detection?
  - Data/metadata must be machine & human readable
- Are you sure the two documents talk about the same thing?
  - Data and metadata must be interlinked
- And what do you mean with "Resolution"?
  - We must have a common understanding of the terminology used
- And why is land monitoring is related to fire detection???
  - Machines must be able to make inferences from data



Outline

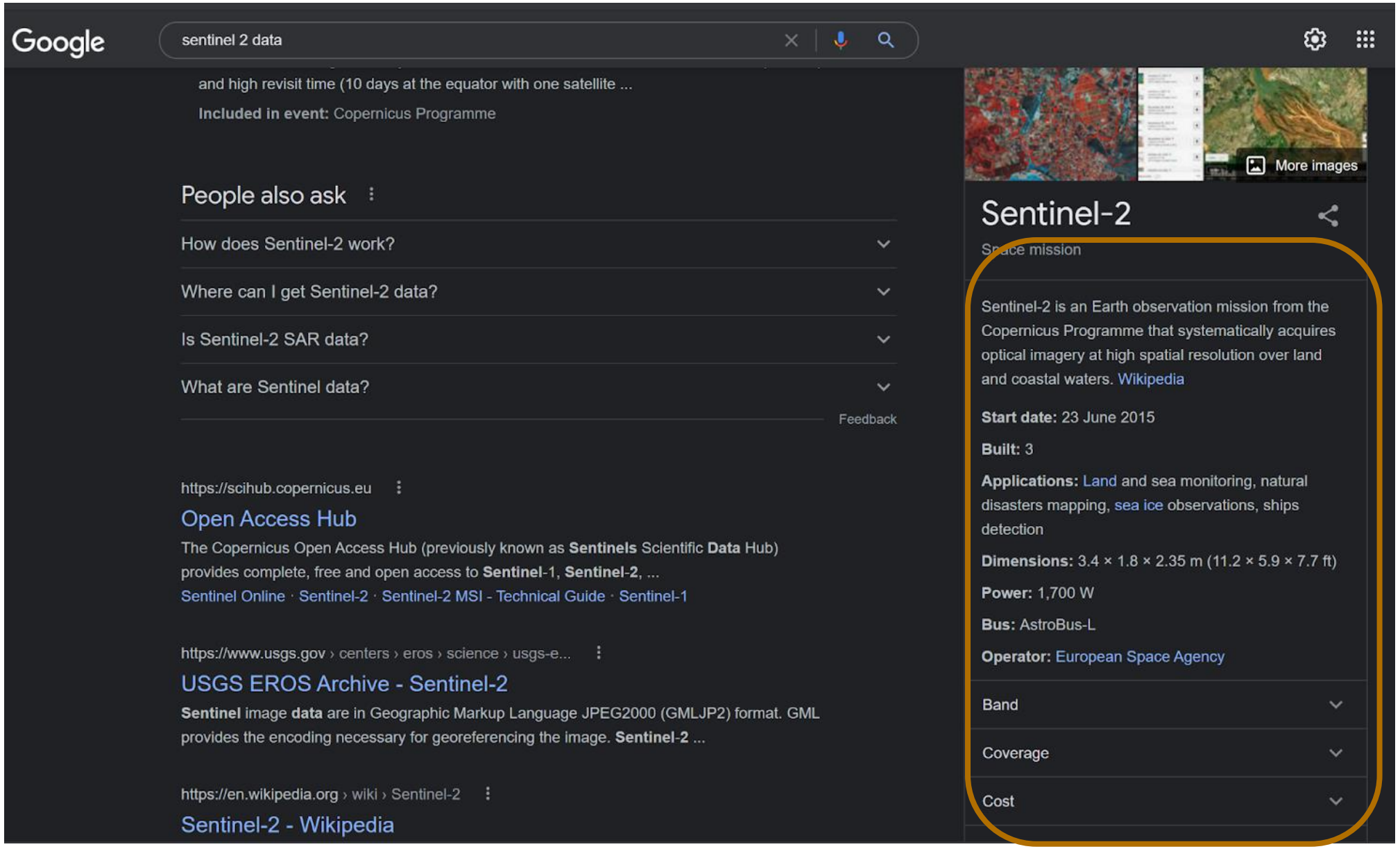
- Basic concepts
  - knowledge, ontology, knowledge graphs
- Some History: How did it all start?
- Why do we need them?
- **Where are we today?**
- So, what is an ontology?
  - RDFS, OWL, Description logics,
- How do we query an ontology/KG?
  - SPARQL
- Linked Data
- Linked geospatial and temporal data
- High Value use cases



The Web Today

- **Universal resource identifiers** (URIs) to identify **documents**.
- The **Hypertext Transfer Protocol** (HTTP) to **exchange documents** between a client and a server.
- **HTML for marking up** information to be presented to human readers through a browser.
- **Search Engines** (Google, duckduckgo, WolframAlpha, Qwant, etc) to discover information.
- Is this information **machine readable?**
- **to some extent..(NLP, KGs)**

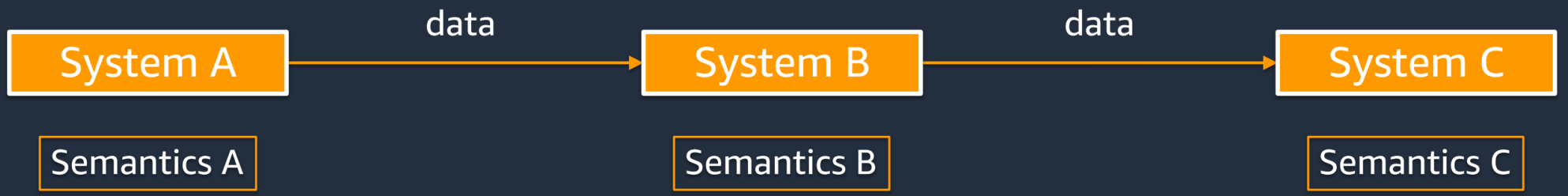
Finding satellite data for fire detection



September 2022, Ora Lassila:

This is what we have

Data moves, but every system has its own idea of semantics



Redefining or “re-articulating” semantics over and over not only is error-prone, it is a lot more work!

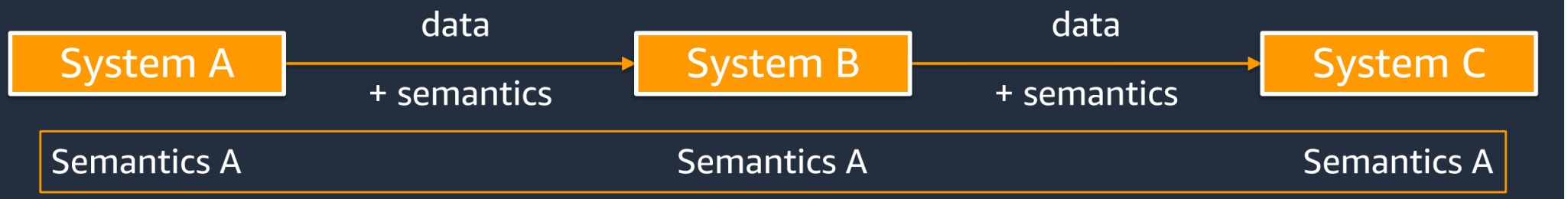
aws

https://lassila.org/publications/2022/US2TS2022-Lassila.pdf

September 2022, Ora Lassila:

This is what we should have!

Data moves with semantics!



What do we need to get there?

The Semantic Web Vision Today (cont'd)

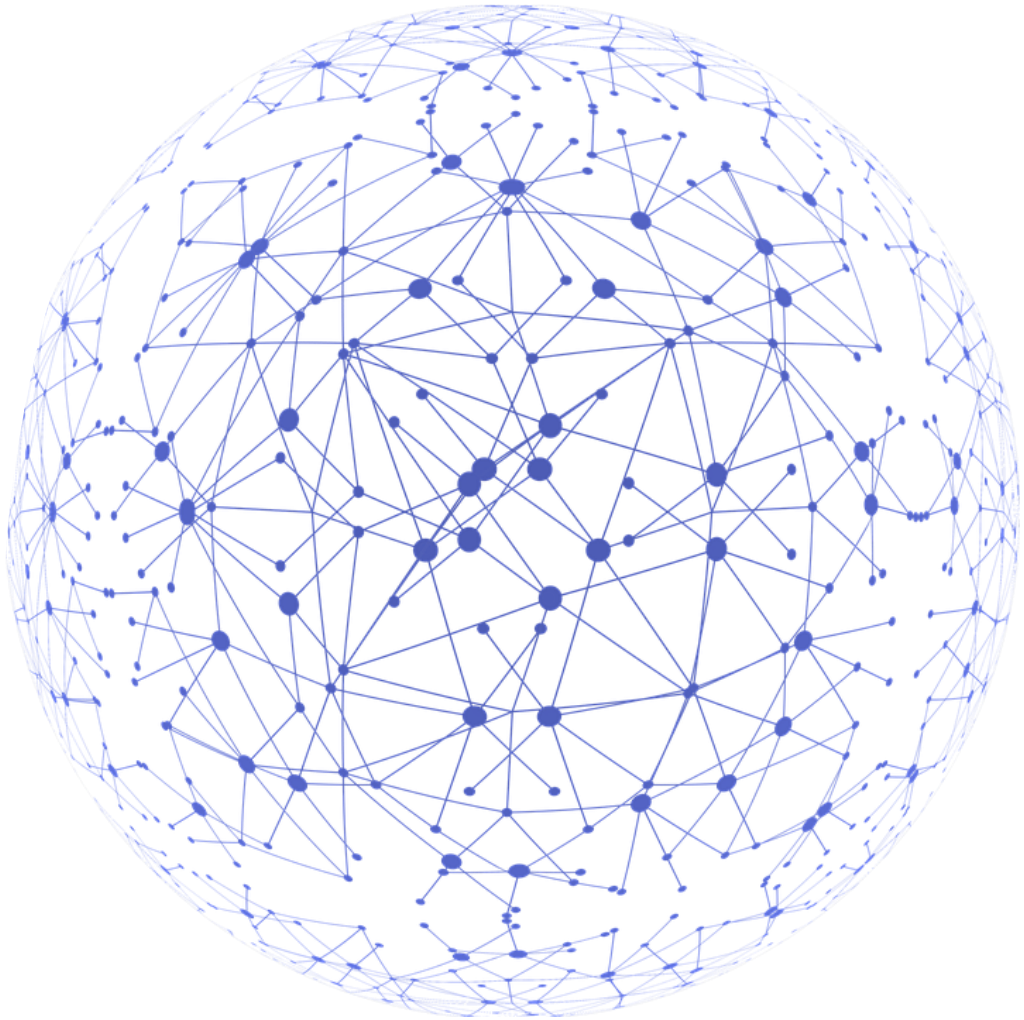
- Why **linked data and Web data integration**?
- Lots of important **Web applications** demand this e.g., e-science and e-government.

- Why Web standards/languages for expressing shared meaning?
- This is important if we want **agents** that are **not handcrafted** only for **particular** tasks to be developed ("roaming from page to page")

- See the paper "The Semantic Web Revisited" by Nigel Shadbolt, Wendy Hall and Tim Berners-Lee at http://eprints.ecs.soton.ac.uk/12614/1/Semantic_Web_Revisted.pdf .

Outline

- Basic concepts
  - knowledge, ontology, knowledge graphs
- Some History: How did it all start?
- Why do we need them?
- Leading Ontologies and KGs
- **So, what is an ontology?**
  - RDFS, OWL, Description logics
- How do we query an ontology/KG?
  - SPARQL
- Linked Data
- Linked geospatial and temporal data
- High Value use cases



Basic Concepts: Ontology

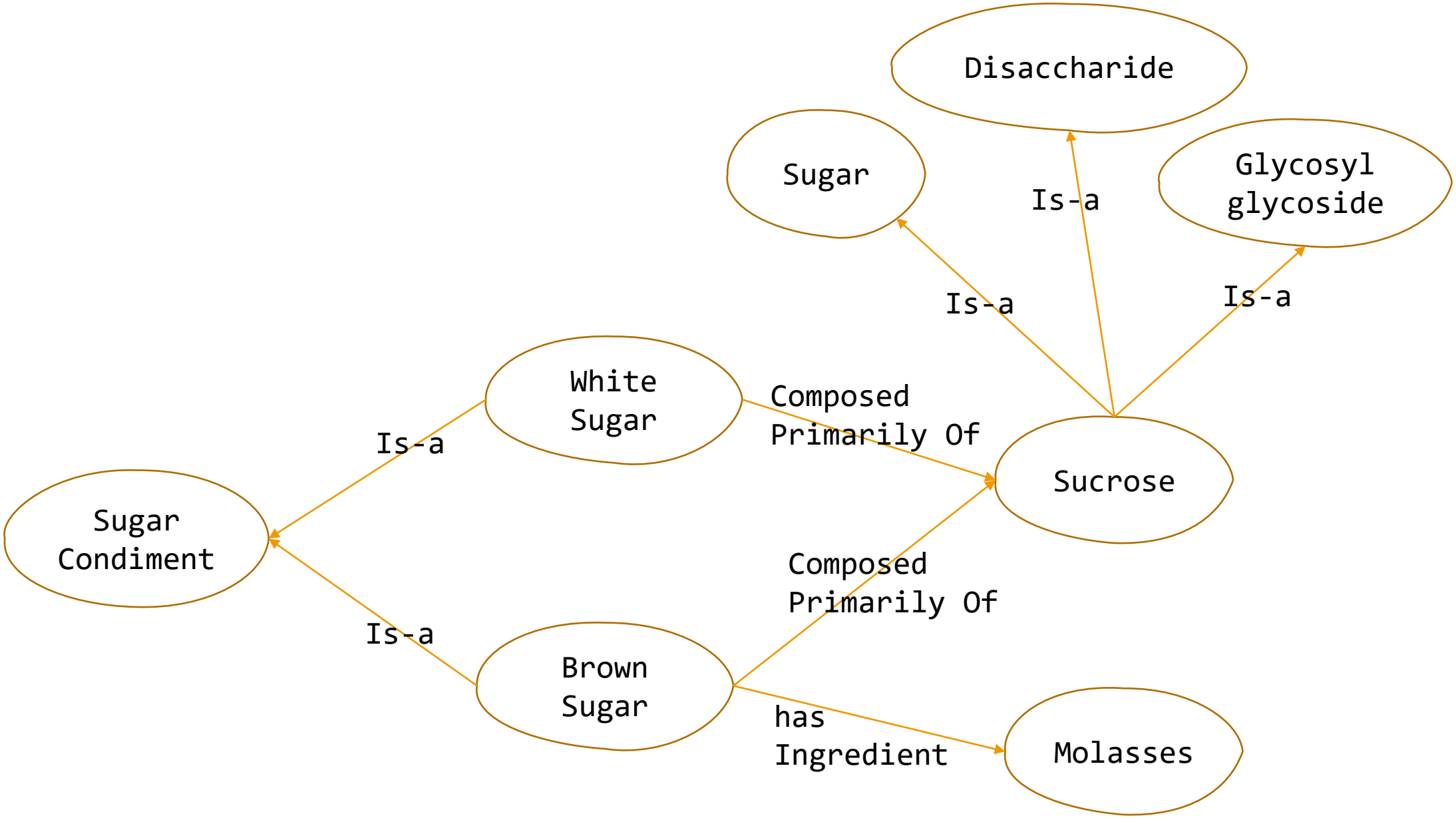
- Philosophy: systematic account of existence
  - what things exist, how they can be differentiated from each other, etc
- An **ontology** is a **formal**, **explicit**, **shared** specification of a **conceptualization** of a domain (Gruber, 1993).
- Intentional (schema-level) knowledge.

Basic Concepts: Ontology -Conceptualization

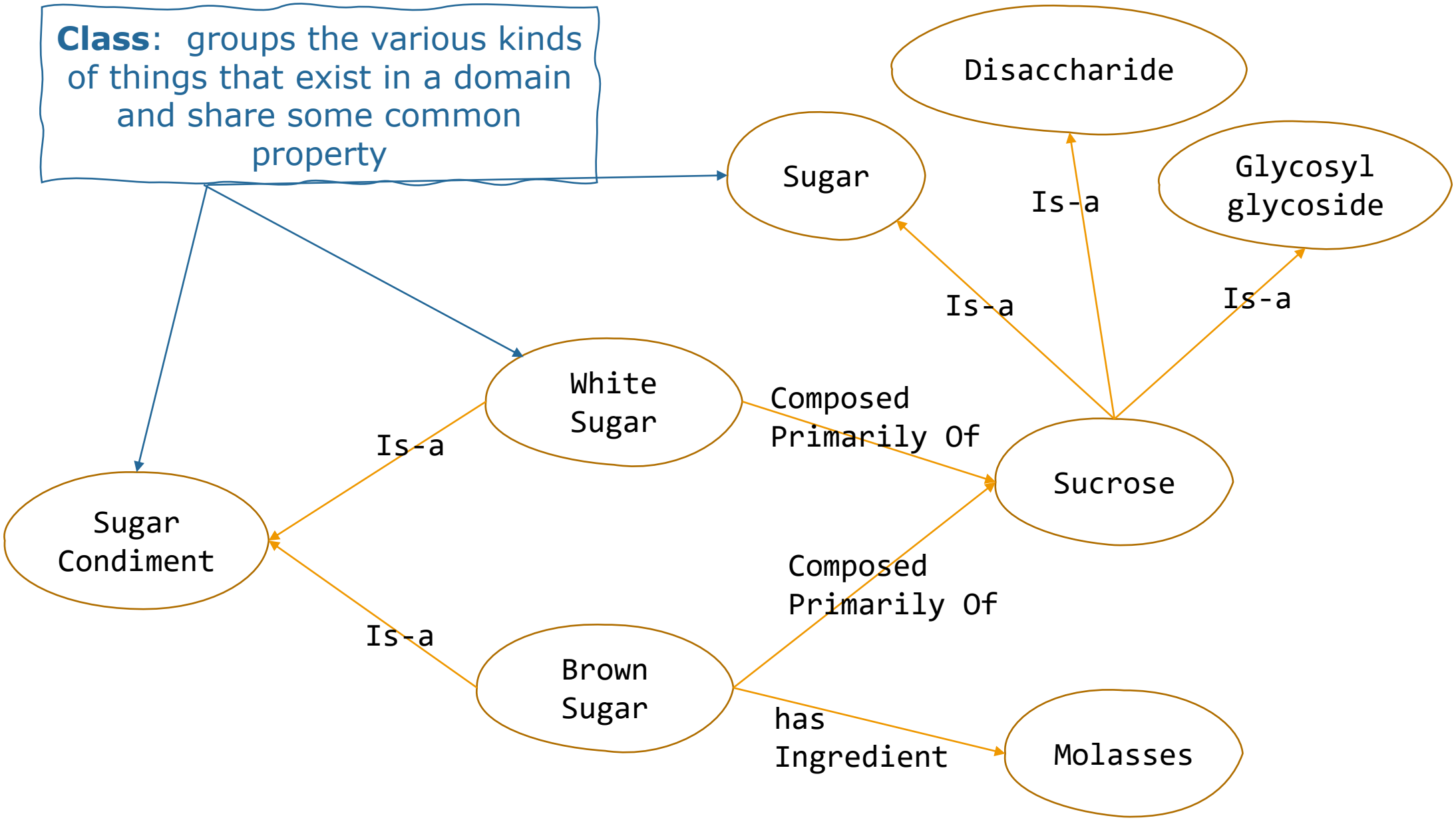
"An **ontology** is a **formal**, **explicit**, **shared** specification of a **conceptualization** of a domain"

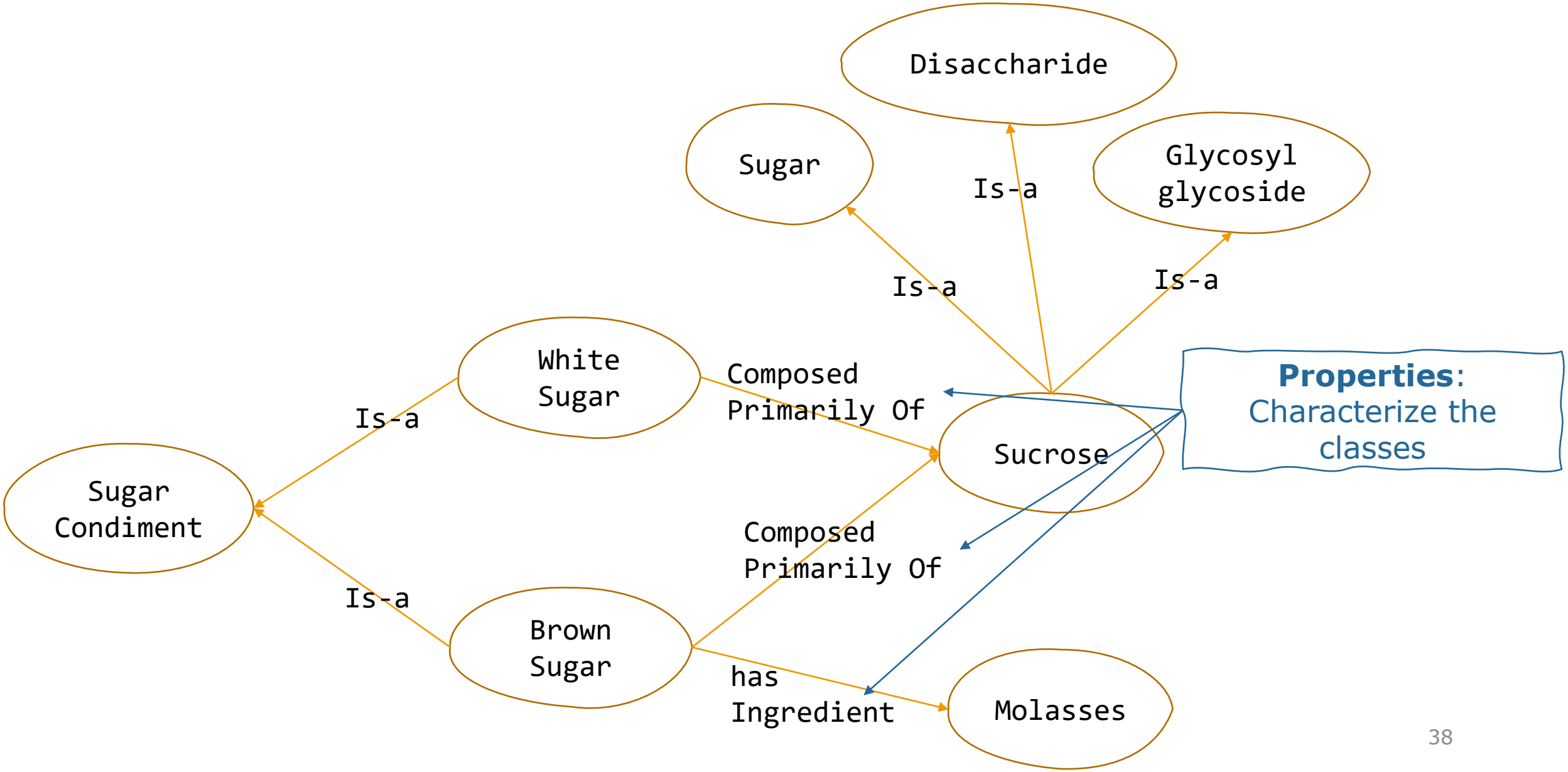
- **Conceptualization**:
  - the **objects**, **concepts**, and other **entities** that are assumed to exist in **some area of interest** and the **relationships** that hold among them.
  - An **abstract, simplified view of the world that we wish to represent for some purpose**.

Example: Foodon

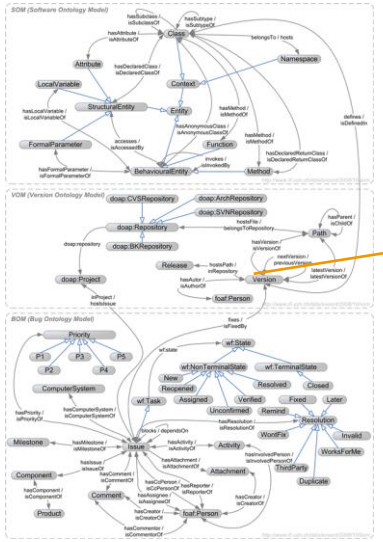


Example: Foodon




Disaccharide
Sugar
Is-a
Glycosyl glycoside
Is-a
Is-a
White Sugar
Composed Primarily Of
Properties: Characterize the classes
Is-a
Sucrose
Sugar Condiment
Composed Primarily Of
Is-a
Brown Sugar
has Ingredient
Molasses

Chemical Hazards Ontology



Foodon Ontology

Disaccharide

Sugar

Glycosyl glycoside

Is-a

Is-a

Is-a

White Sugar

Composed Primarily Of

Sucrose

Is-a

Sugar Condiment

Composed Primarily Of

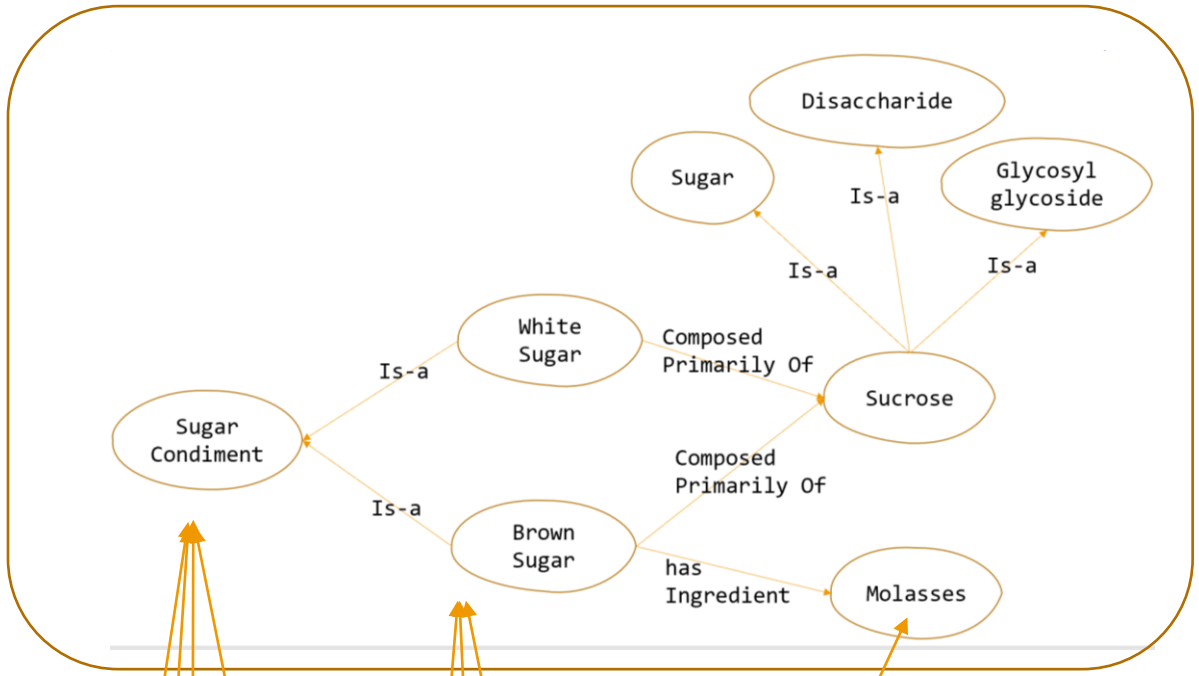
Is-a

Brown Sugar

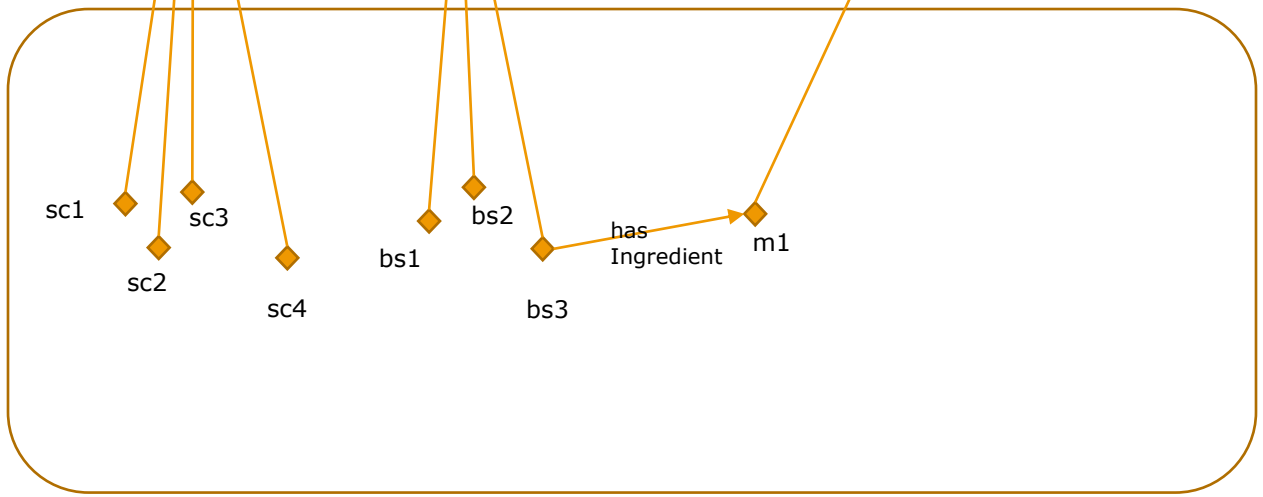
has Ingredient

Molasses

Ontology



Data



Formal

"An **ontology** is a **formal**, **explicit**, **shared** specification of a **conceptualization** of a domain"

- **Eliminates ambiguity**, natural language is slippery!
- Enables **reasoning**
- The literature also offers **special formalisms** for defining ontologies that contain mainly taxonomic knowledge:
  - Semantic networks
  - Frames
  - Description logics
  - **RDF**, **RDFS** and **OWL** (expressive DLs with RDF syntax)

Explicit

"An **ontology** is a **formal**, **explicit**, **shared** specification of a **conceptualization** of a domain"

- Why do you develop your ontology?
- **What questions** do you want your ontology to **answer**?

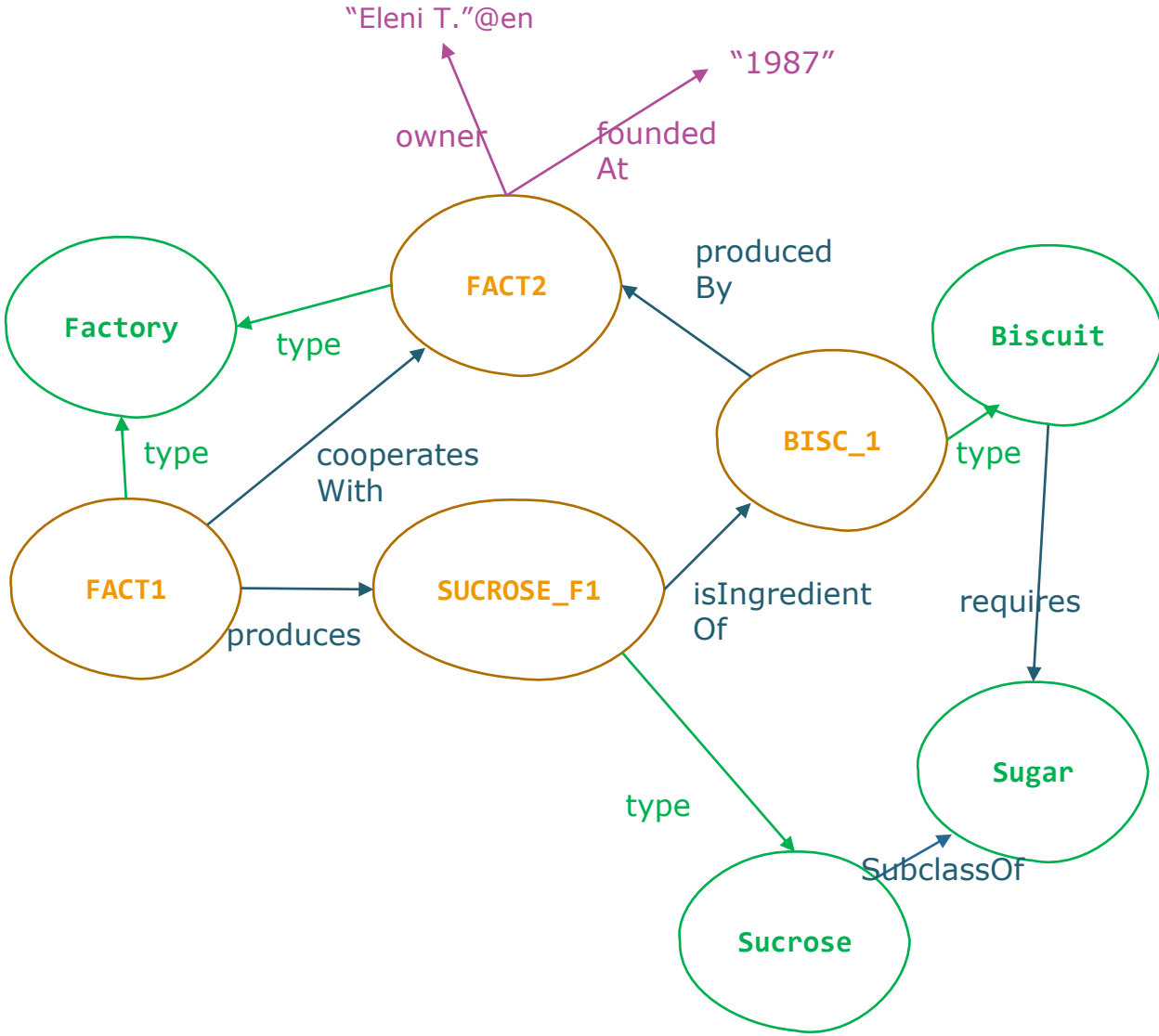
Shared

"An **ontology** is a **formal**, **explicit**, **shared** specification of a **conceptualization** of a domain"

- Research, Industry, Consumer Industrial applications have a **common understanding** of the used terms
  - (e.g., starter, apperizer)
- Domain experts **agree** on the terminology
- Defined in **unambiguous** way (formalism)
- Information **reuse**

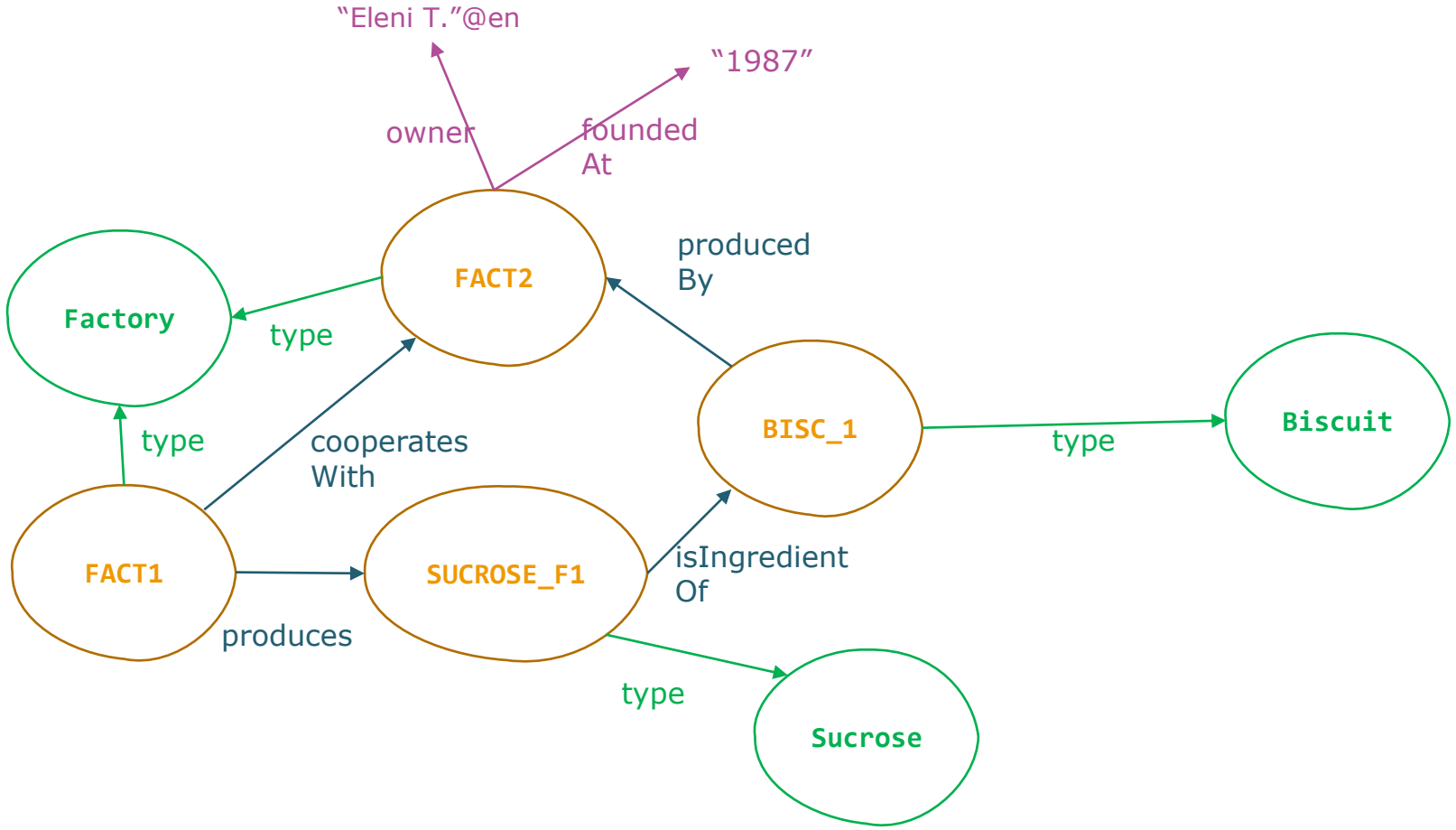
Basic Concepts: Knowledge Graph

- A **knowledge graph** consists of a set of **interconnected typed entities** and their **attributes**
- The term "Knowledge Graph" was introduced by Google in 2012
- Blog post titled: '*Introducing the Knowledge Graph: things, not strings*'
- **but**…



Knowledge Graph vs Graph Databases

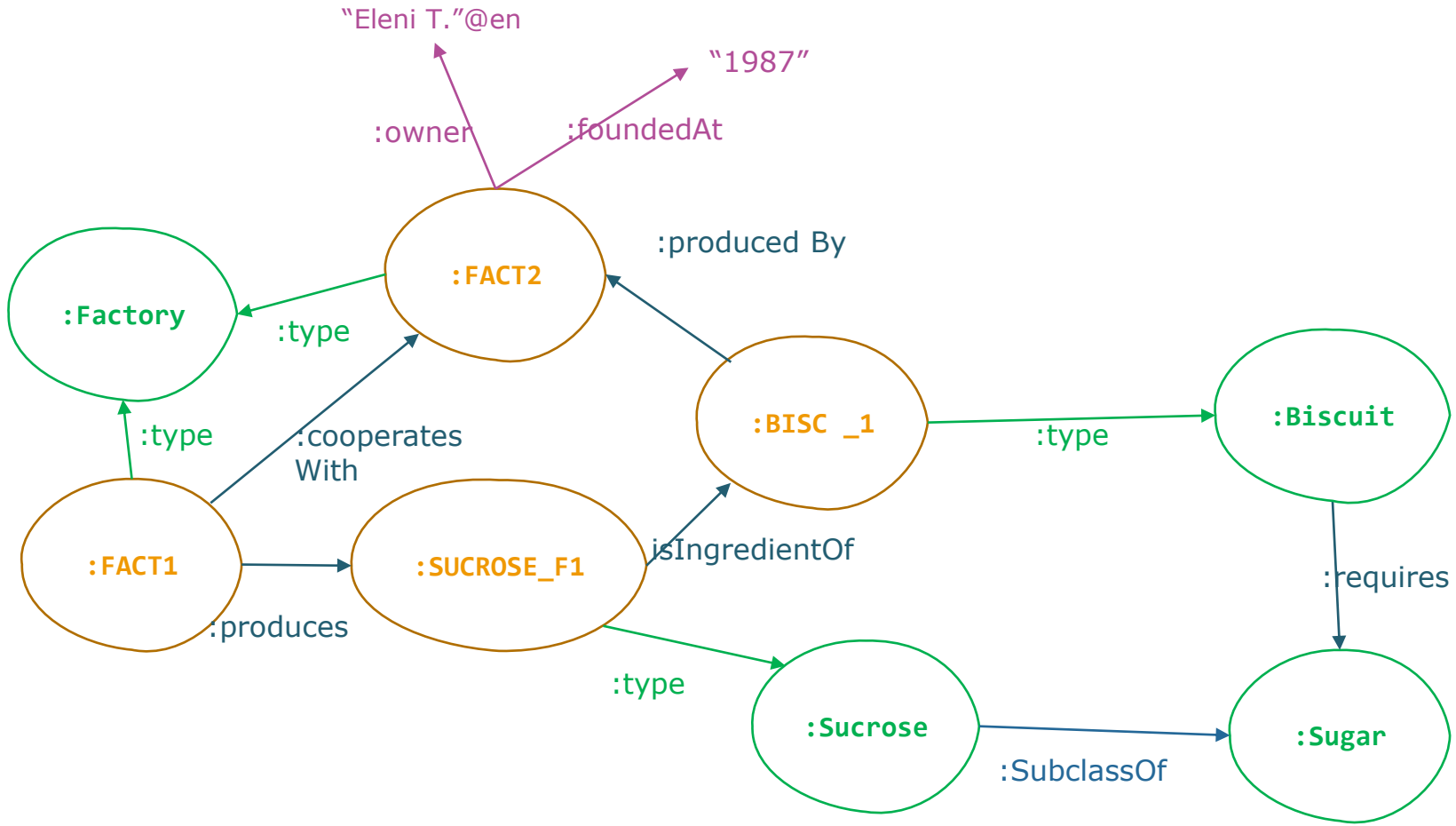
Graph Database



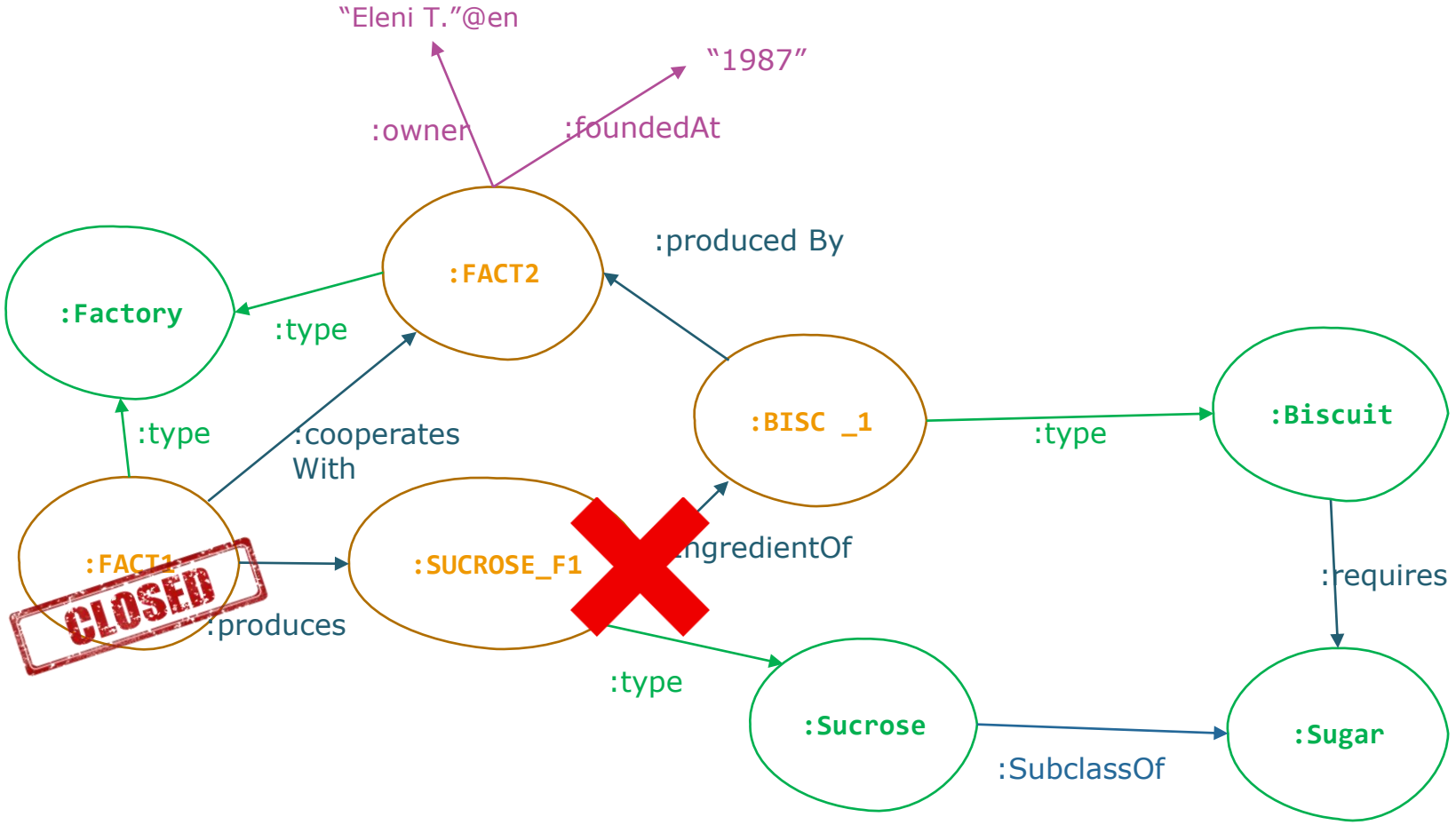
Knowledge Graph vs Graph Databases

Knowledge Graph:

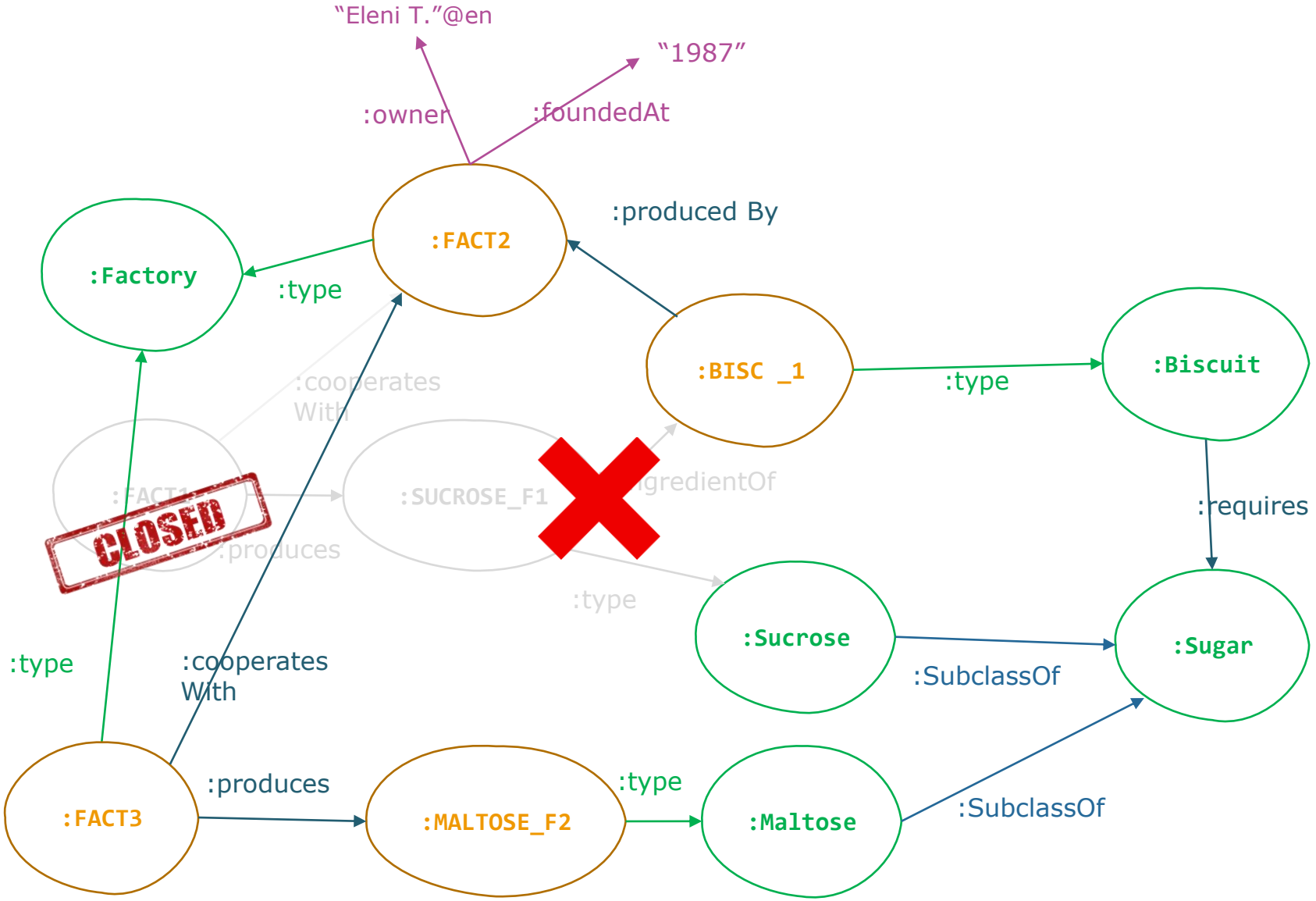
- May contain also **higher-order** organizing principles
- Nodes: **URIs**/literals
- **Standardized** data model (W3C)
- Ontologies freely available for **reuse**



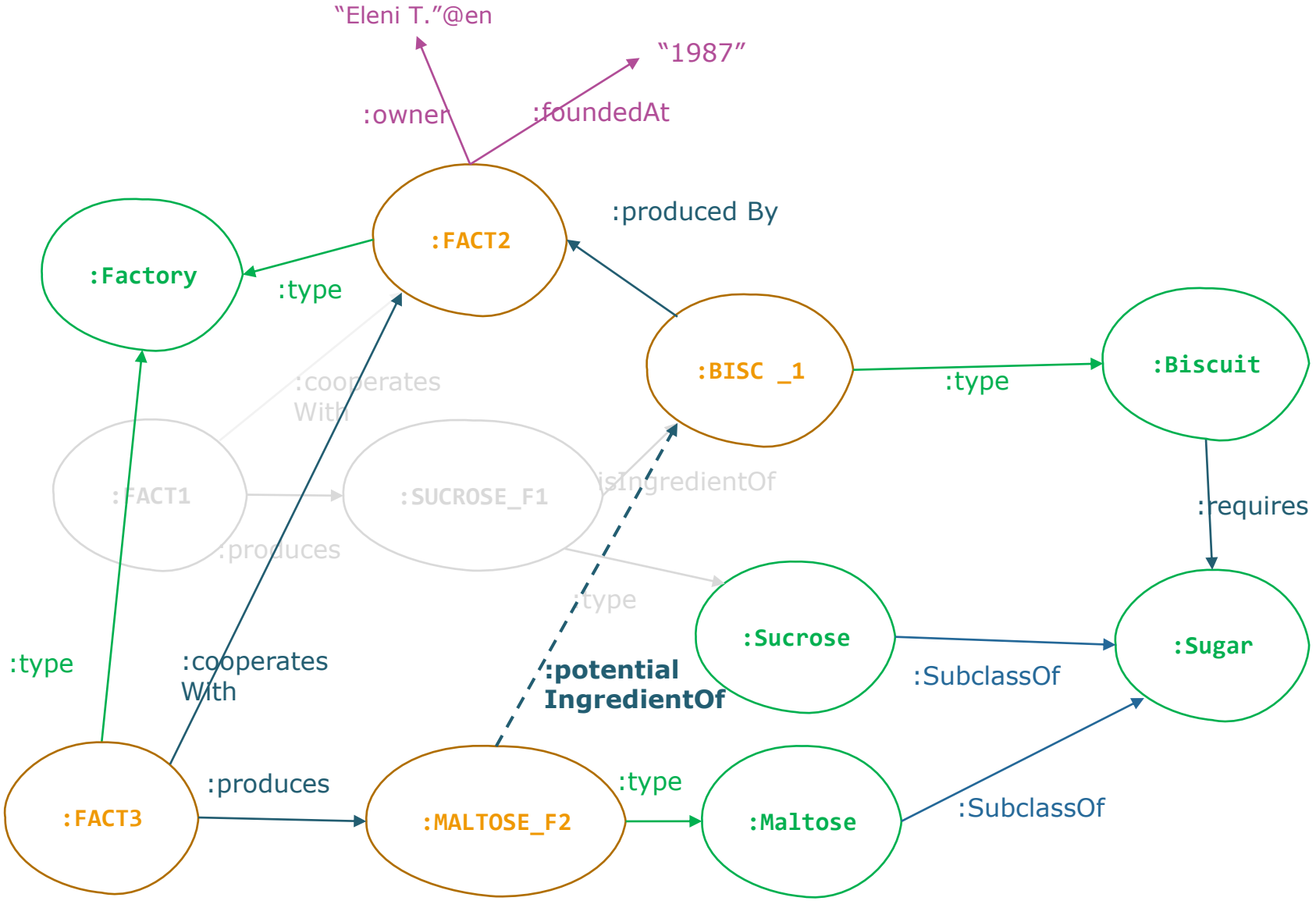
Knowledge Graph vs Graph Databases



Knowledge Graph vs Graph Databases

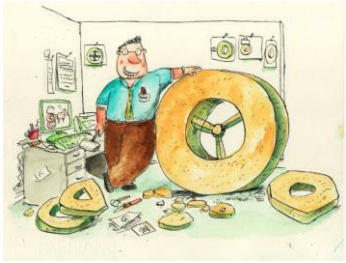


Knowledge Graph vs Graph Databases



Knowledge Graphs vs Graph Databases

- KGs contain also **higher-order** organizing principles (e.g., taxonomy), in a **standardized** way (W3C)
  - **Inferencing**
  - **Knowledge toolkits**
- Nodes: **URIs**/literals
  - **Unified** data
  - Answering queries **across** data silos
- Knowledge models (ontologies) freely available for **reuse**
  - **Shareability**
  - **Interoperability**
  - **Modelling problems are already solved**



- More **flexible**: just adding new nodes, change the ontology

Knowledge Graph vs Ontology

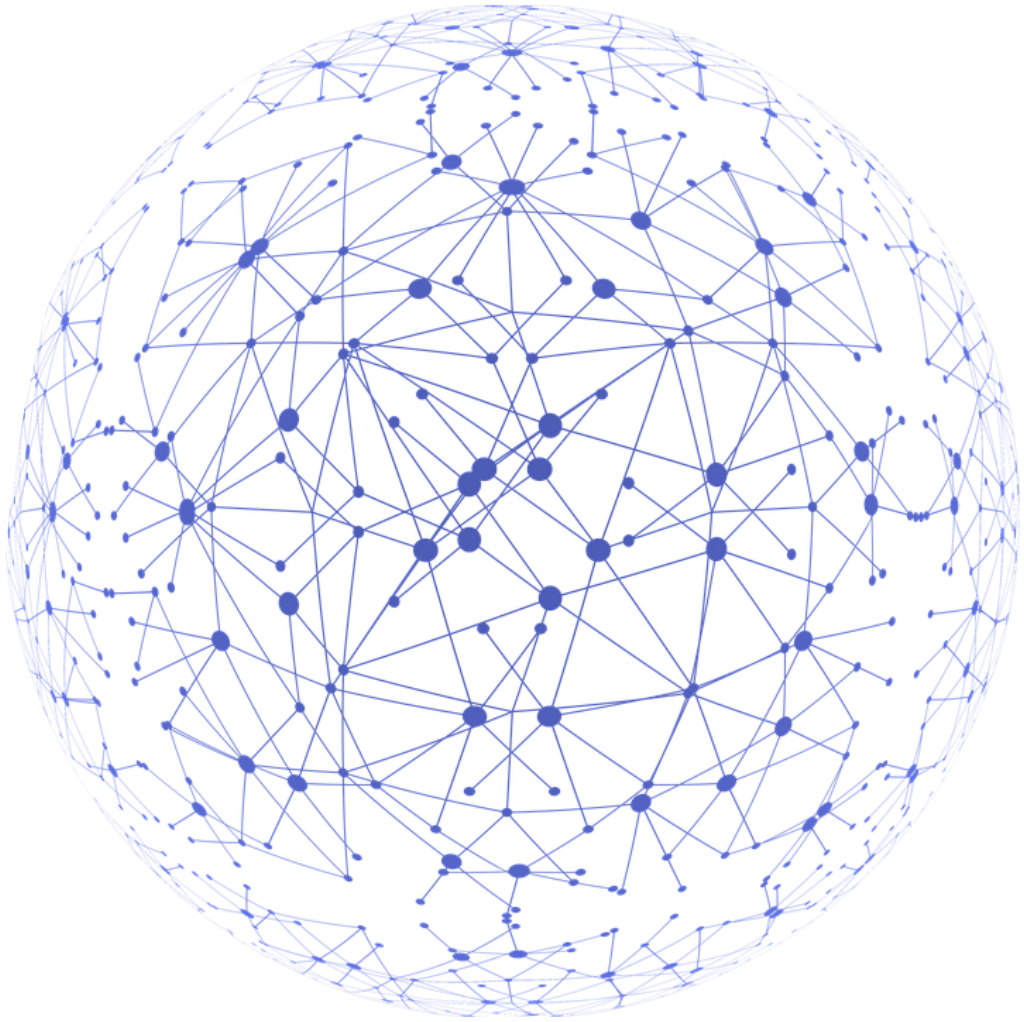
- **Knowledge Graphs** represent formally **interconnected entities** (abstract or not)
- **Ontologies** represent formally the **knowledge** about a domain.
- Ontologies may also include ***complex rules* or *axioms*** e.g.,

"A diet soft drink has always as ingredient something that is of caffeine type"

- These rules are used for ***inferencing***
- **Limitation**: the more complex the ontology is, the harder to get the answers in reasonable time

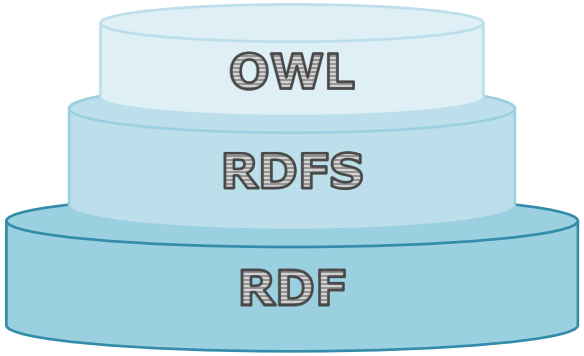
Outline

- Basic concepts
  - knowledge, ontology, knowledge graphs
- Some History: How did it all start?
- Why do we need them?
- Leading Ontologies and KGs
- **So, what is an ontology?**
  - **RDF**, RDFS, OWL, Description logics
- How do we query an ontology/KG?
  - SPARQL
- Linked Data
- Linked geospatial and temporal data
- High Value use cases



Outline

- RDF
- RDFS
- OWL 2.0
- SWRL



Ontology Languages

RDF: represent data as URIs and in graphical form

For describing facts
Data integration from multiple resources
Detachment of data from their schema

RDFS: adds schema to the RDF

Adds constraints on the facts
Hierarchies, domains & ranges of properties
Enables basic inferencing – infer new triples

OWL 2.0: higher expressivity – adds more constraints

Enables more complex inferencing

SWRL: rule language, intuitive, adds expressivity

Resource Description Framework

- The Resource Description Framework (RDF):
  - Data model originally for representing information (especially metadata) about web resources
- Now is used to describe **any data** and not only metadata
- Easy, powerful, expressive W3C **standard**
- For data to be **processed by applications**, rather than being only displayed to people.

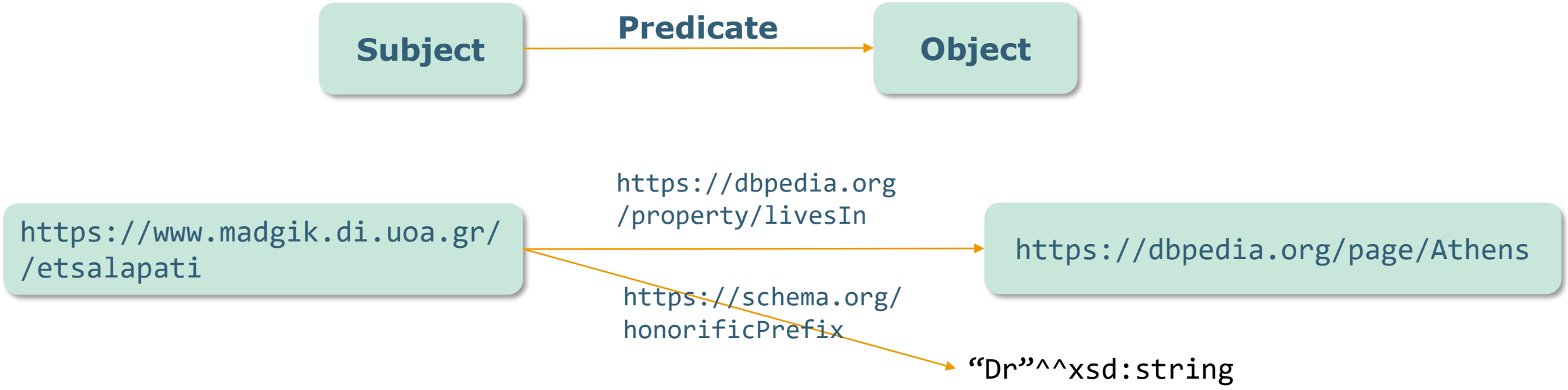
Resource Description Framework

- **Basic idea:**
  - Data objects are identified as web identifiers (**URIs**)
  - Definition of **relationships** between **data objects**

- Unlike URLs, URIs are not limited to identifying things that have network locations, or use other computer access mechanisms.

Resource Description Framework

- **Basic idea:**
  - Data objects are identified as unique resource identifiers (URIs/IRIs)
  - Definition of relationships between data objects



Resource Description Framework

- RDF statements can be written in triple notation:

subject predicate object .

```
<https://www.madgik.di.uoa.gr/etsalapati>
                    <https://dbpedia.org/property/livesIn>
                                          <https://dbpedia.org/page/Athens> .

<https://www.madgik.di.uoa.gr/etsalapati>
                                        <https://schema.org/honorificPrefix> “Dr”.
```

URI reuse

- Reuse existing URIs from well-known vocabularies/ontologies (e.g., DBPedia, Dublin Core)
  - Less effort
  - Shared understanding of the resulted ontology
  - Ambiguity is eliminated
  - Can be looked-up

NY or GR?

```
<https://www.madgik.di.uoa.gr/etsalapati>
                    <https://dbpedia.org/property/livesIn>
                                        <https://dbpedia.org/page/Athens> .
```

What can we do with RDF

- Add machine-readable information using **well-known vocabularies**, e.g. schema.org:
  - Ambiguity is eliminated
  - Shareability is established
- **Enrich** dataset by linking to external datasets.
  - e.g. linking paintings dataset to artists dataset
- Build **aggregations** of data about specific topics
  - e.g., distributed social networks by linking RDF descriptions of people across multiple Web sites
  - e.g., Interlinking various datasets within an organization: cross-dataset QA
- Provide **standard-compliant** way for exchanging data between DBs

Microdata

- Microdata provide a simple way to **annotate HTML elements** with **machine readable tags**.
- Microdata can use **standardized vocabularies** to capture the semantics of HTML items.
- See http://www.w3.org/TR/microdata/ for more details.
- Similar (earlier) approaches are RDFa (RDF in attributes-W3C Recommendation that adds a set of attribute-level extensions to HTML), microformats and JSON-LD.

Example: Microdata – Original HTML

- The following is some HTML code for describing a local business called “Beachwalk Beachwear and Giftware”.

```
<h1>Beachwalk Beachwear & Giftware</h1>
A superb collection of fine gifts and clothing to
accent your stay in Mexico Beach.
3102 Highway 98
Mexico Beach, FL
Phone: 850-648-4200
```

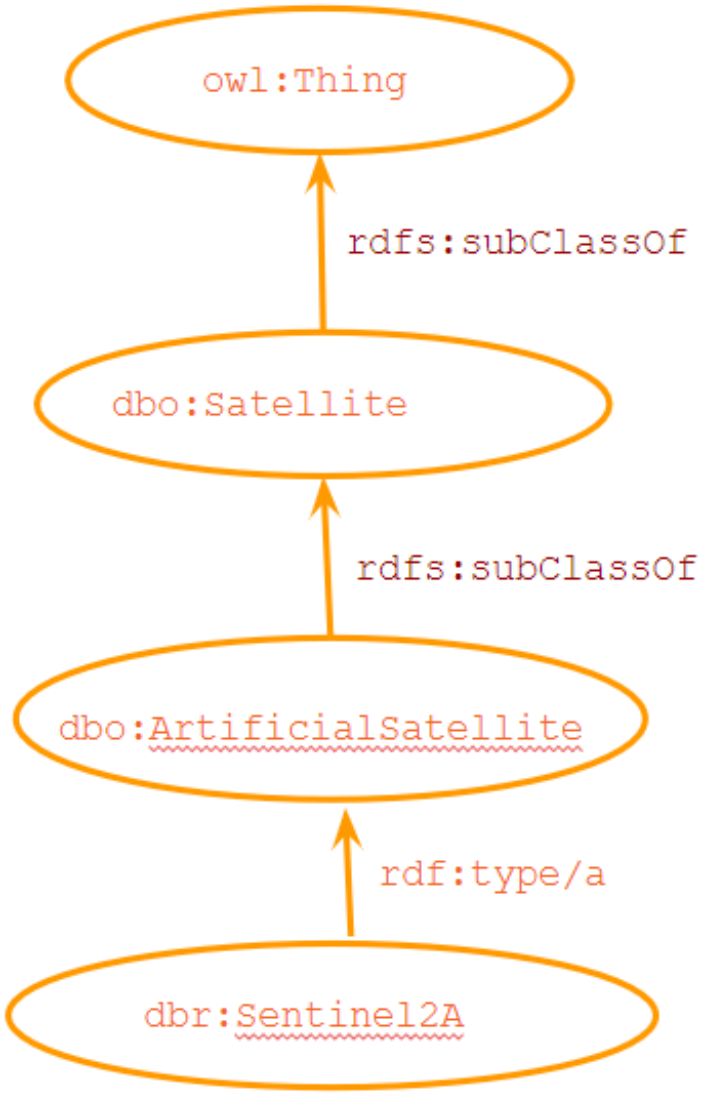
Example: Microdata – Original HTML

```
<div itemscope itemtype="http://schema.org/LocalBusiness">
  <h1><span itemprop="name">Beachwalk Beachwear &
                                    Giftware</span></h1>
  <span itemprop="description"> A superb collection of fine gifts
      and clothing to accent your stay in Mexico Beach.</span>

  <div itemprop="address" itemscope
                  itemtype="http://schema.org/PostalAddress">

      <span itemprop="streetAddress">3102 Highway 98</span>
      <span itemprop="addressLocality">Mexico Beach</span>,
      <span itemprop="addressRegion">FL</span>
  </div>

  Phone: <span itemprop="telephone">850-648-4200</span>
</div>
```

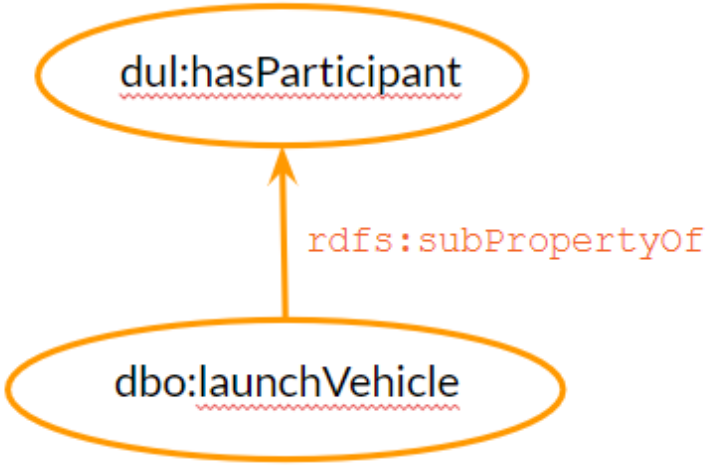
Resource Description Framework Schema (RDFS)

- **RDFS:** a **language** for the definition of the **vocabulary** (i.e., the terms) to be used in an RDF graph.
- RDFS is an **ontology definition language**
- RDFS is used to describe:
  - Specific kinds or **classes** of resources
  - The properties of the resources
  - i.e., to create an **ontology**

Example – Class Hierarchy



Example – Property Hierarchy

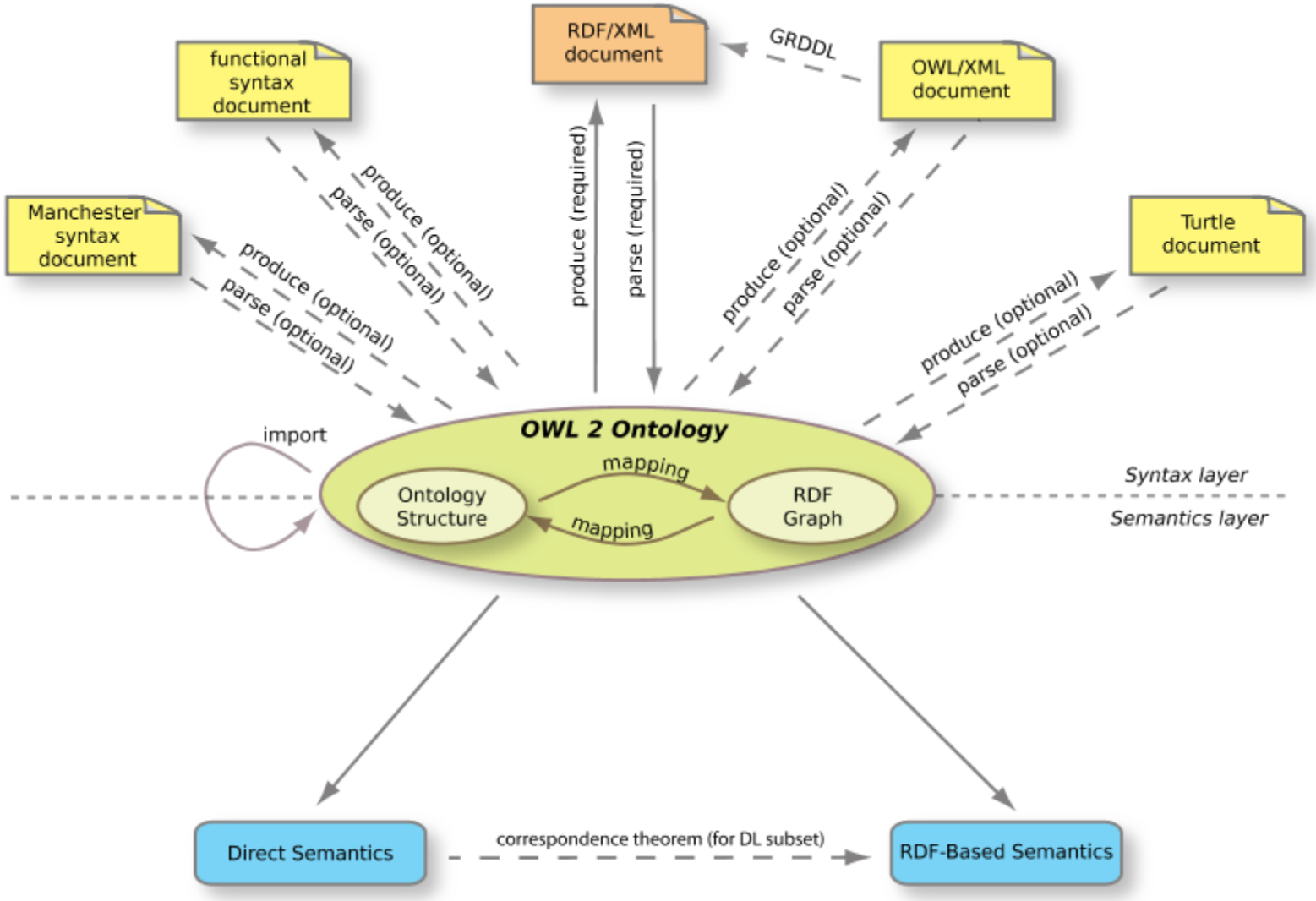


Web Ontology Language (OWL)

- Results from Description Logics
- More expressive than RDFS –allows more powerful ontology modelling

- OWL 2 ontologies consist of the following three different syntactic categories:
  - **Entities: classes**, **properties**, and **individuals**
  - **Expressions:** represent complex notions in the domain being described.
  - **Axioms** are statements that are asserted to be true in the domain being described (e.g., a subclass axiom)

OWL


RDF/XML document
GRDDL
OWL/XML document
functional syntax document
Manchester syntax document
Turtle document
produce (required)
parse (required)
produce (optional)
parse (optional)
import
OWL 2 Ontology
mapping
Ontology Structure
RDF Graph
Syntax layer
Semantics layer
Direct Semantics
correspondence theorem (for DL subset)
RDF-Based Semantics

OWL - Example

```
SubClassOf(:Student  :Person)

EquivalentClasses(:Man
     ObjectIntersectionOf(:Person  :Male))

EquivalentClasses(:Woman
     ObjectIntersectionOf(:Person  :Female))

EquivalentClasses(:Parent
     ObjectSomeValuesFrom(:hasChild  :Person))
```

OWL - Example

```
EquivalentClasses(:Teenager
     ObjectIntersectionOf(Person
          DataSomeValuesFrom(:hasAge
               DatatypeRestriction(xsd:integer
                    xsd:minExclusive "12"^^xsd:integer
                         xsd:maxInclusive "19"^^xsd:integer
))))
```

OWL - Example

```
EquivalentClasses(:ChildlessPerson
            ObjectIntersectionOf(:Person
                  Object Complement Of(ObjectSomeValuesFrom(
                        ObjectInverseOf(:hasParent) owl:Thing))))


SameIndividual(:John :Jack)
SameIndividual(:John otherOnt:JohnBrown)
SameIndividual(:Mary otherOnt:MaryBrown)
DifferentIndividuals(:John :Bill)
```

Description Logics

- DLs are languages for describing the nature and structure of objects.
- Origins: semantic networks and frames.
- Developed in the 80's and 90's in parallel with:
  - pure FOL approaches
  - other languages for structured objects, e.g., Telos and F-logic.
- DLs have been used to provide the foundations for ontology languages for the Web e.g., OWL.
- Many DLs are decidable fragments of first-order logic (FOL)
- Some DLs have features that are not covered in FOL

Description Logics –Complex Class Expressions

- The set of female doctors.

$$\text{Female} \sqcap \text{Doctor}$$

- The set of individuals that have at least 3 children that are male.

$$(\geq 3 \ hasChild \ . \ Male)$$

- The set of individuals such that all their children have graduated from at least one Greek University.

$$(\geq 3 \ hasChild \ . (\forall isAlumniOf \ . \ GreekUniversity))$$

- The set of individuals that have at least three children such that all their degrees are from Greek Universities.

$$(\forall hasChild \ . (\exists isAlumniOf \ . \ GreekUniversity))$$

Description Logics - Assertions

Ann is a female doctor.

$$(Female \sqcap Doctor)(ANN)$$

John is a child of Ann.

$$hasChild(ANN,JOHN)$$

Ann has at least 3 children that are male.

$$(\geq 3\ hasChild\ .\ Male)(ANN)$$

Description Logics - Axioms

A woman is a female person.

$$Woman \equiv Person \sqcap Female$$

A parent is a person who has at least one child.

$$Parent \equiv Person \sqcap (\exists hasChild \; . \; Person)$$

SPARQL

- SPARQL stands for “SPARQL Protocol and RDF Query Language”.
- SPARQL is based on **matching graph patterns against RDF graphs.**

Matching Graph Patterns

“What is the title of ex:book1?”

KG

<http://example.org/book/book1>

<http://purl.org/dc/elements/1.1/title>

"SPARQL Tutorial"

“Query”

<http://example.org/book/book1>

<http://purl.org/dc/elements/1.1/title>

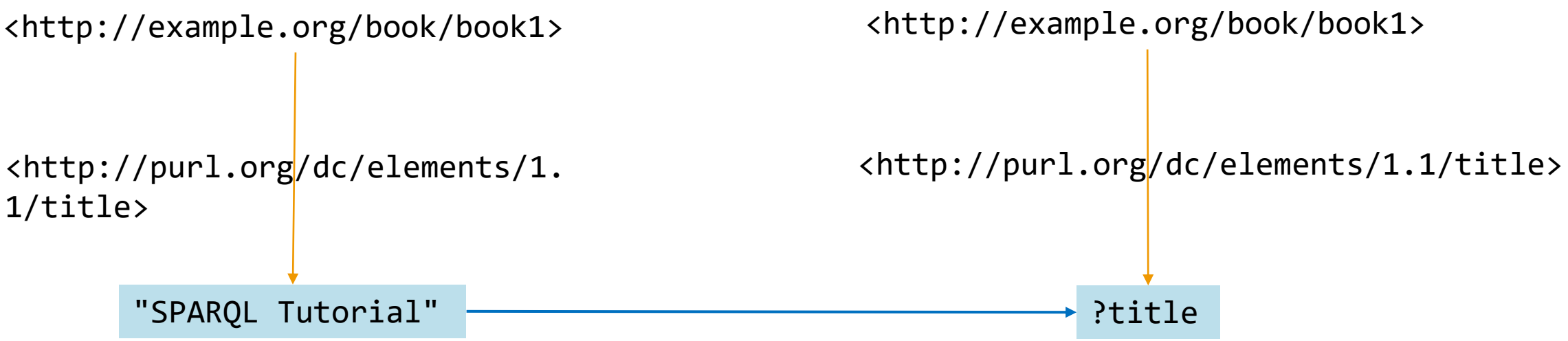
?title

Matching Graph Patterns

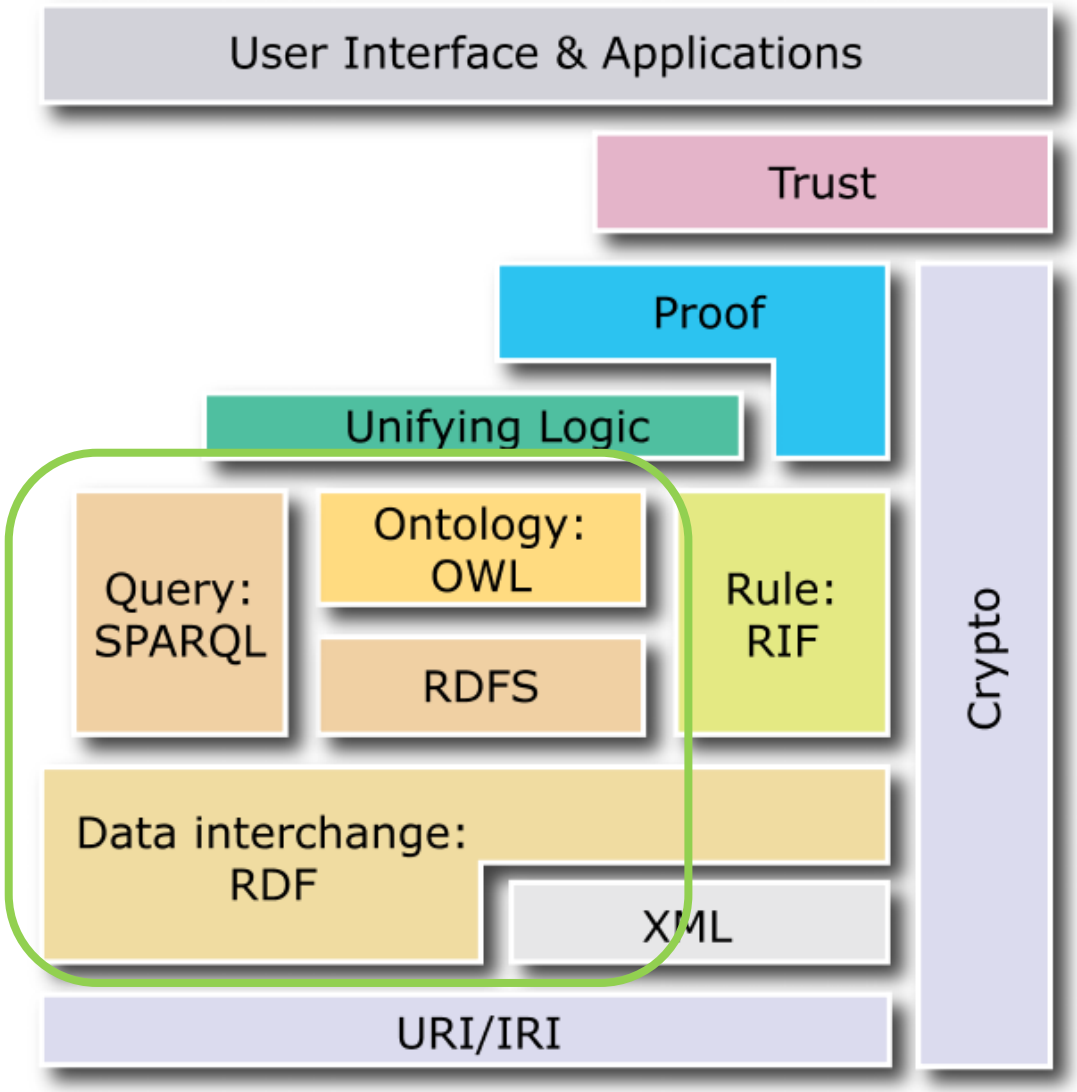
“What is the title of ex:book1?”

KG

“Query”



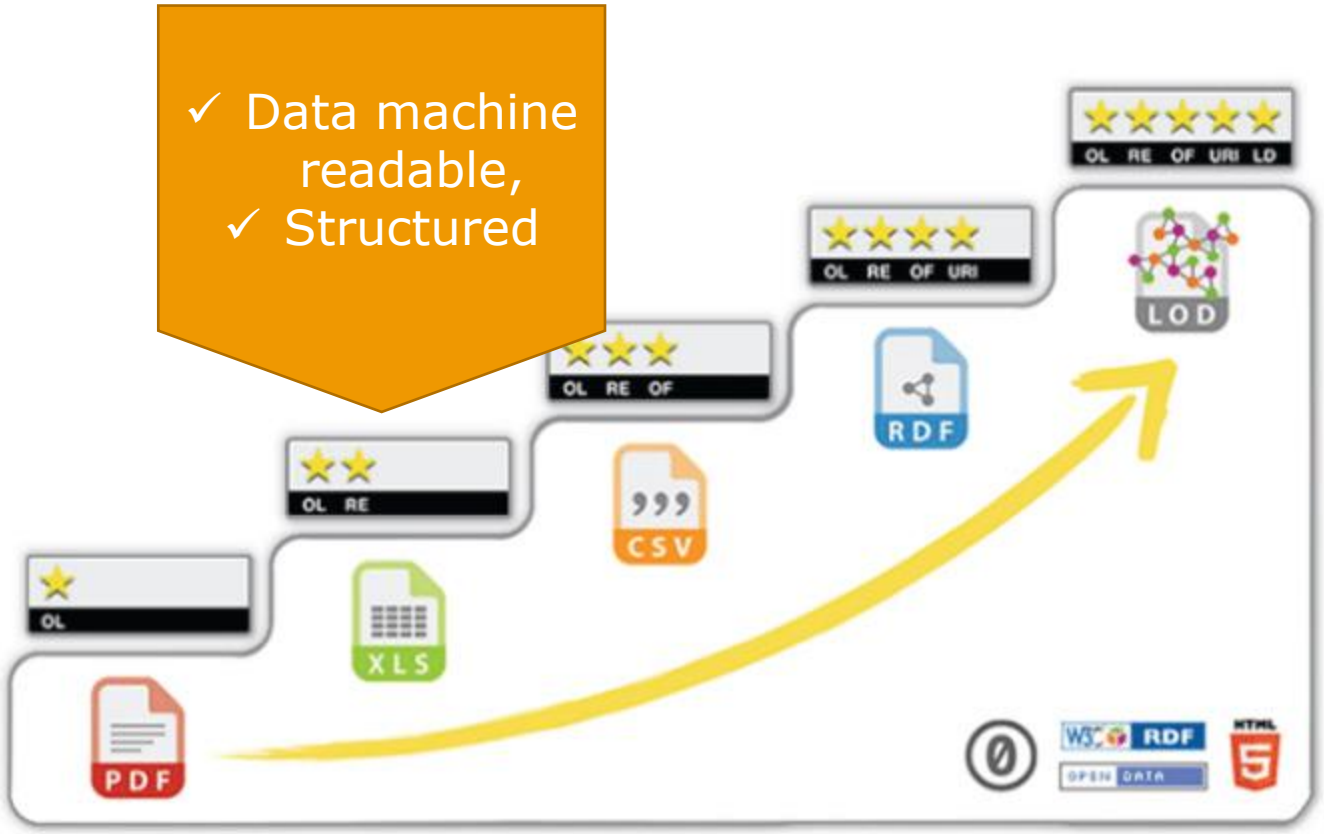
The Semantic Web Layer Cake



5-star deployment scheme for Linked Open Data

- ✓ Data available on the web,
- ✓ In whatever format
- ✓ With an open license, so as to be Open Data

5-star deployment scheme for Linked Open Data



5-star deployment scheme for Linked Open Data

require proprietary software package to access and process

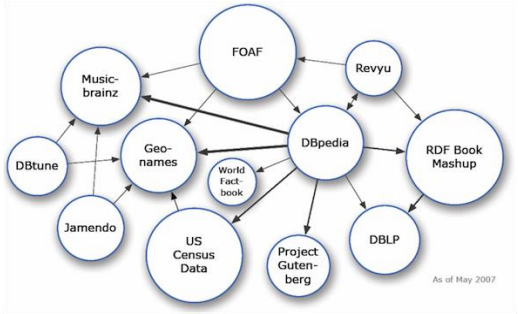
5-star deployment scheme for Data

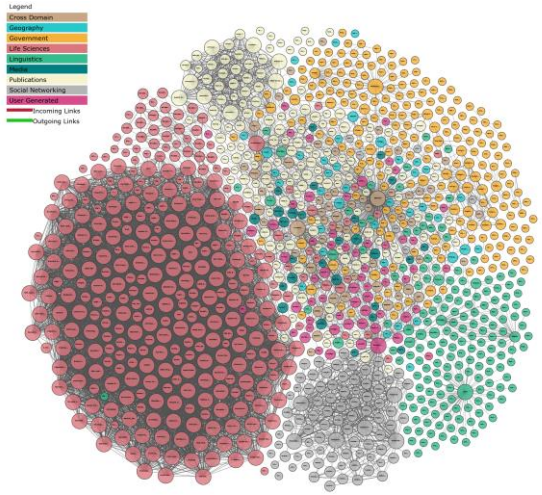
✓ Data that uses open standards from W3C, such as RDF to identify things

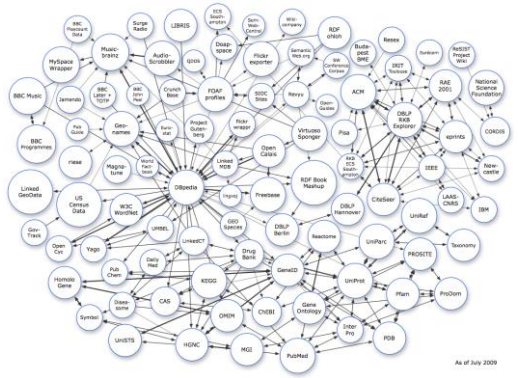
5-star deployment scheme for Linked O

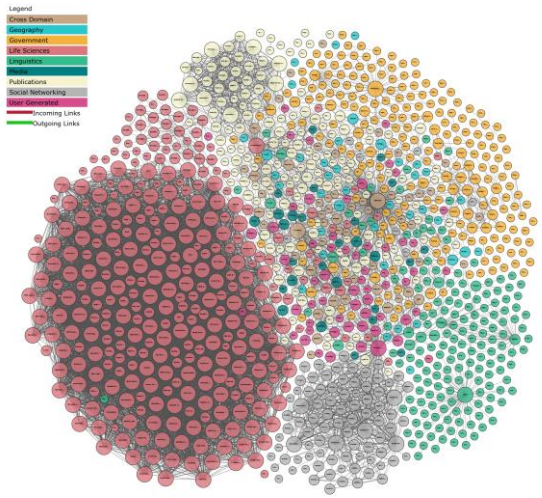
✓ Open data available on the Web linked to other data

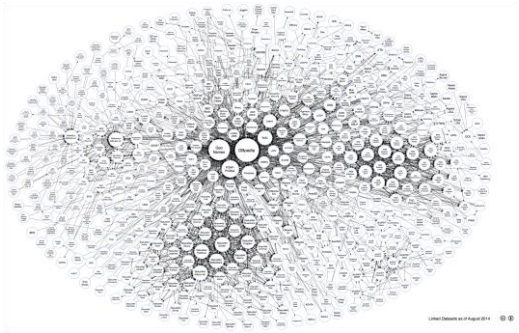
The Linked Open Data Cloud











Linked Open Data

- The goal of this W3C community effort is “to extend the Web with a data commons by publishing various open data sets as RDF on the Web and by setting RDF links between data items from different data sources.”

- The LOD community is developing a set of best practices for achieving this.

Linked Data Principles

- Tim Berners-Lee outlined four principles of linked data in his "Linked Data" note of 2006:
  - Use URIs to name (identify) things.
  - Use HTTP URIs so that these things can be looked up (interpreted, "dereferenced").
  - When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL, etc)
  - Include links to other URIs so that they can discover more things.

Key Linked Data Technologies

- URIs: a generic means to identify entities or concepts in the world
- HTTP: a simple, yet universal, mechanism for retrieving resources, or descriptions of resources
- RDF: a data model for structuring and linking data that describes things in the world
- RDFS: is a general-purpose language for representing simple RDF vocabularies on the Web.
- SPARQL: a query language for querying linked RDF data

Key Linked Data Technologies

- URIs: a generic means to identify entities or concepts in the world
- HTTP: a simple, yet universal, mechanism for retrieving resources, or descriptions of resources
- RDF: a data model for structuring and linking data that describes things in the world
- RDFS: is a general-purpose language for representing simple RDF vocabularies on the Web.
- SPARQL: a query language for querying linked RDF data

Entity Interlinking

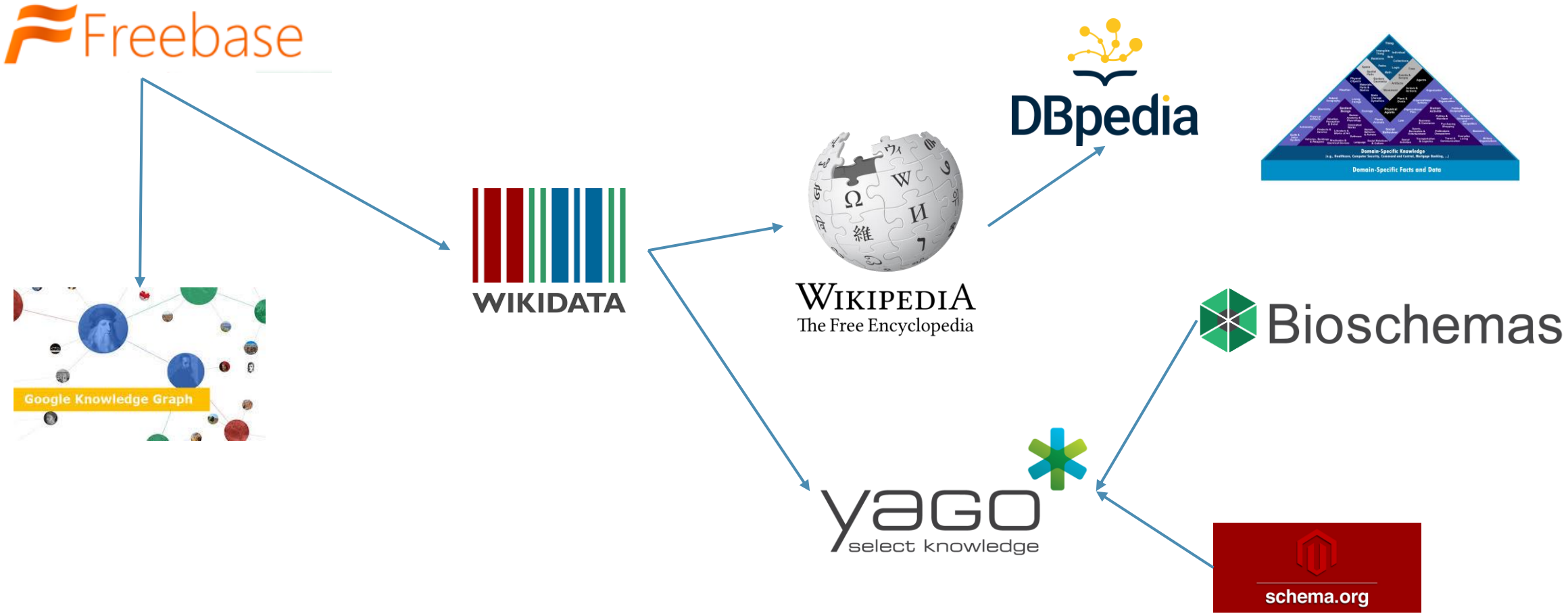
- Data sets will typically talk about the same object using different URIs. For example:

owl:sameAs

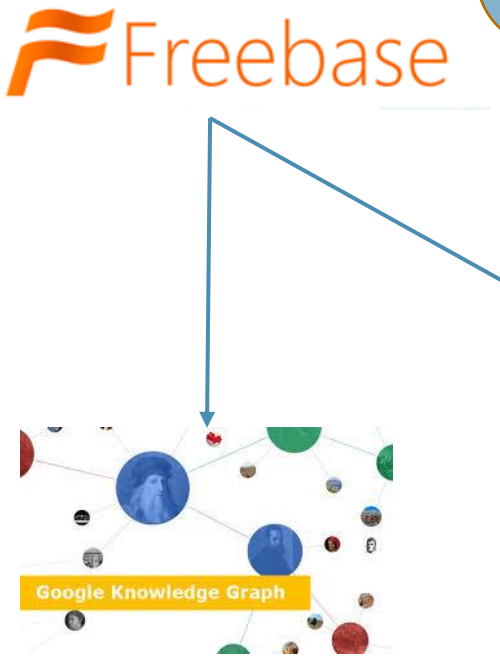
https://www.wikidata.org/wiki/Q20354353

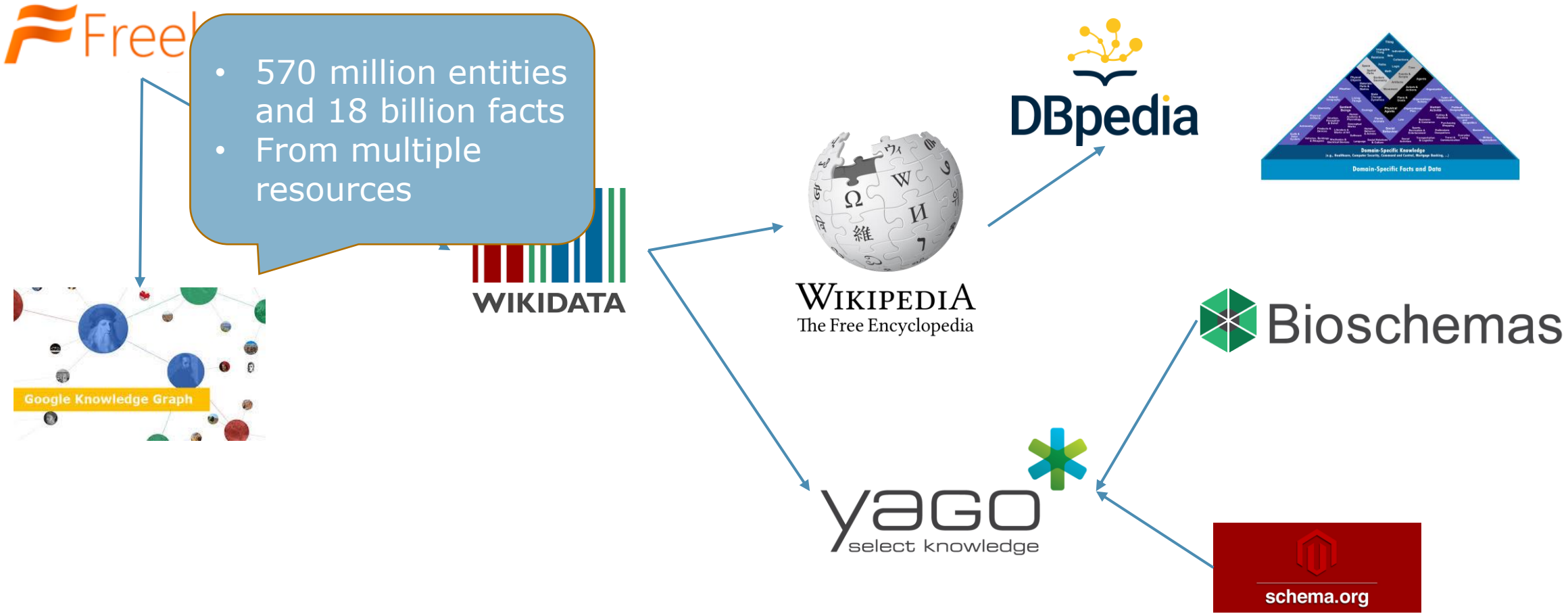
http://dbpedia.org/resource/Sentinel-2A

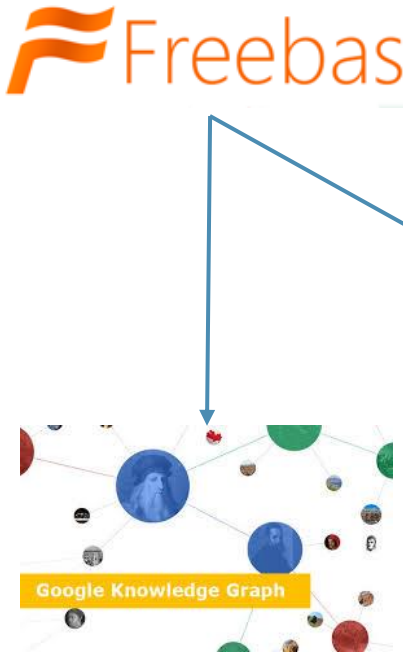
Leading Ontologies and KGs



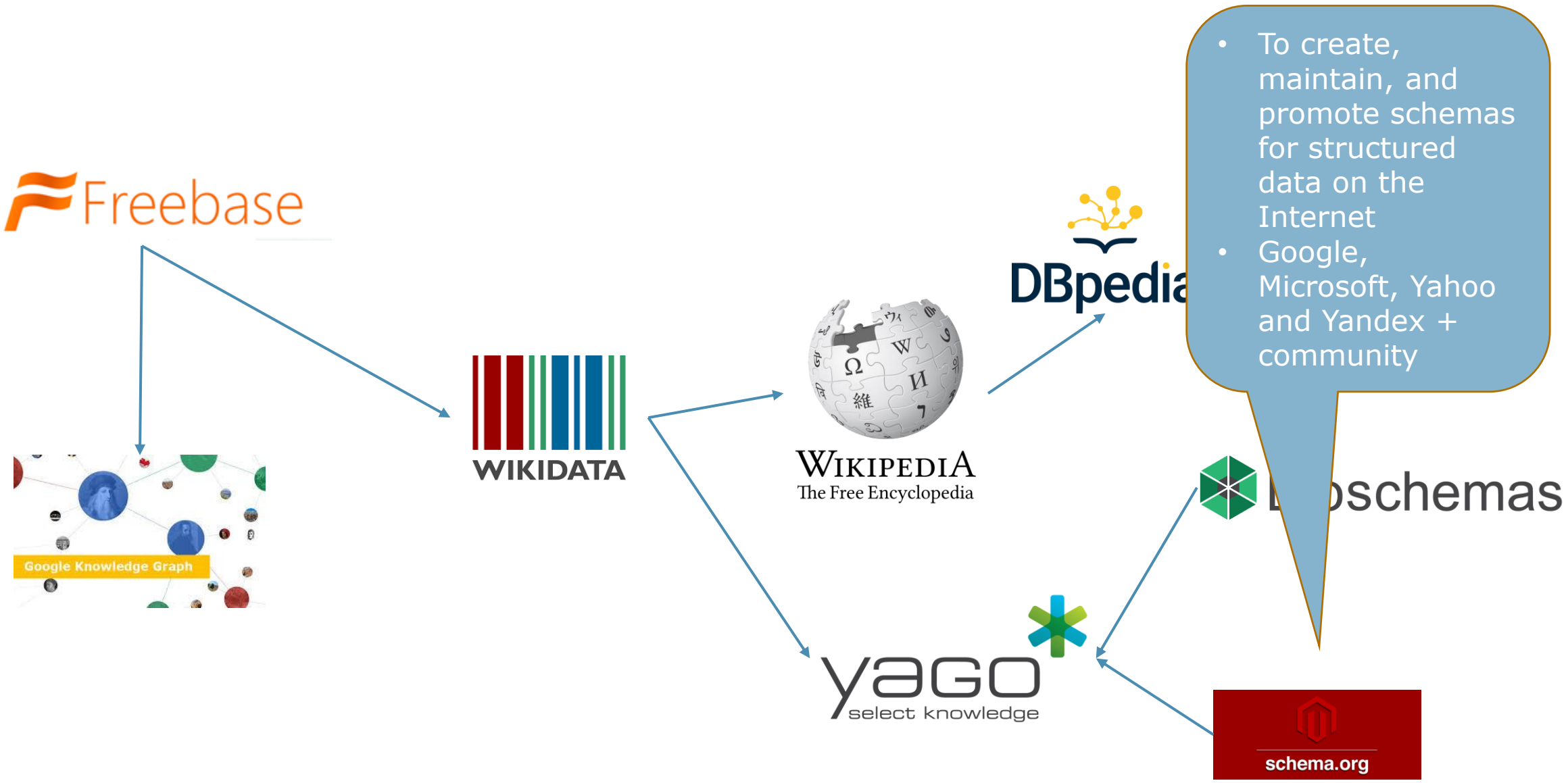
- Large collaborative KB (20 million)
- Community experts
- Started in 2007
- Closed in 2014 (Google)
- 2018 is partially published by Googe



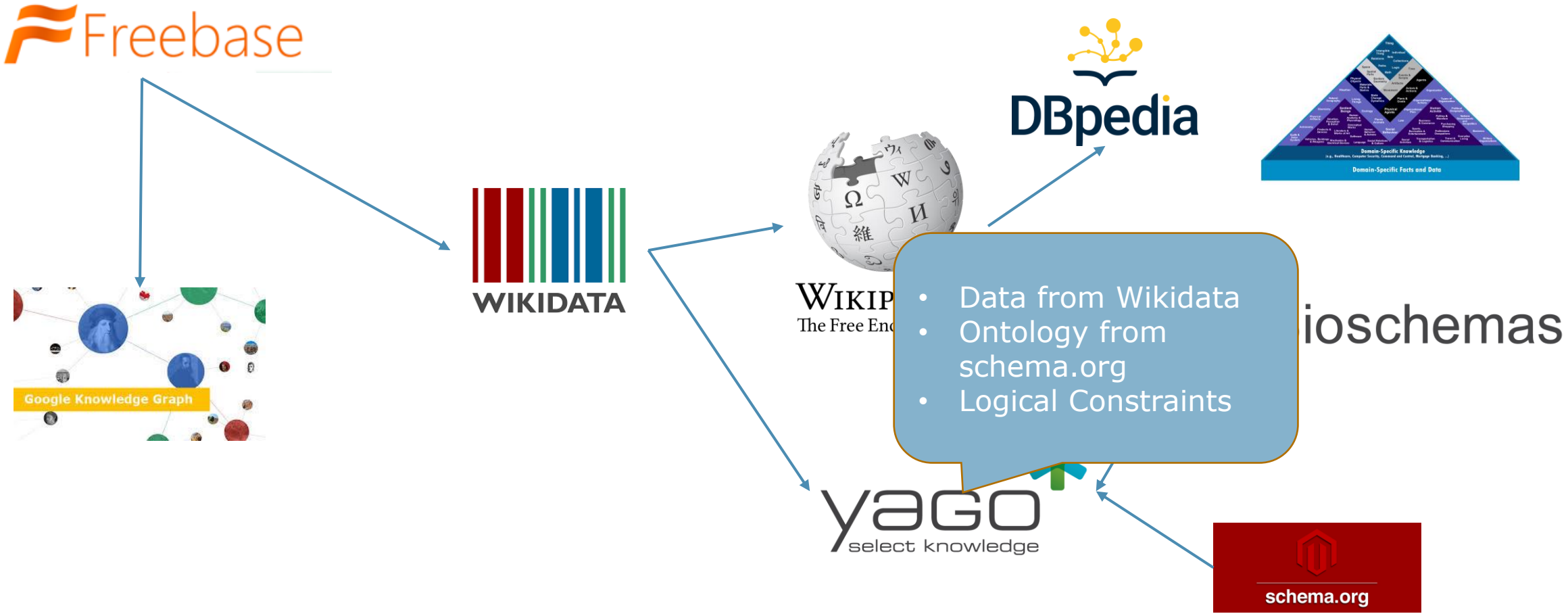

Freeb
570 million entities and 18 billion facts
From multiple resources
WIKIDATA
Google Knowledge Graph
DBpedia
WIKIPEDIA
The Free Encyclopedia
Domain-Specific Knowledge
Domain-Specific Facts and Data
Bioschemas
yago
select knowledge
schema.org

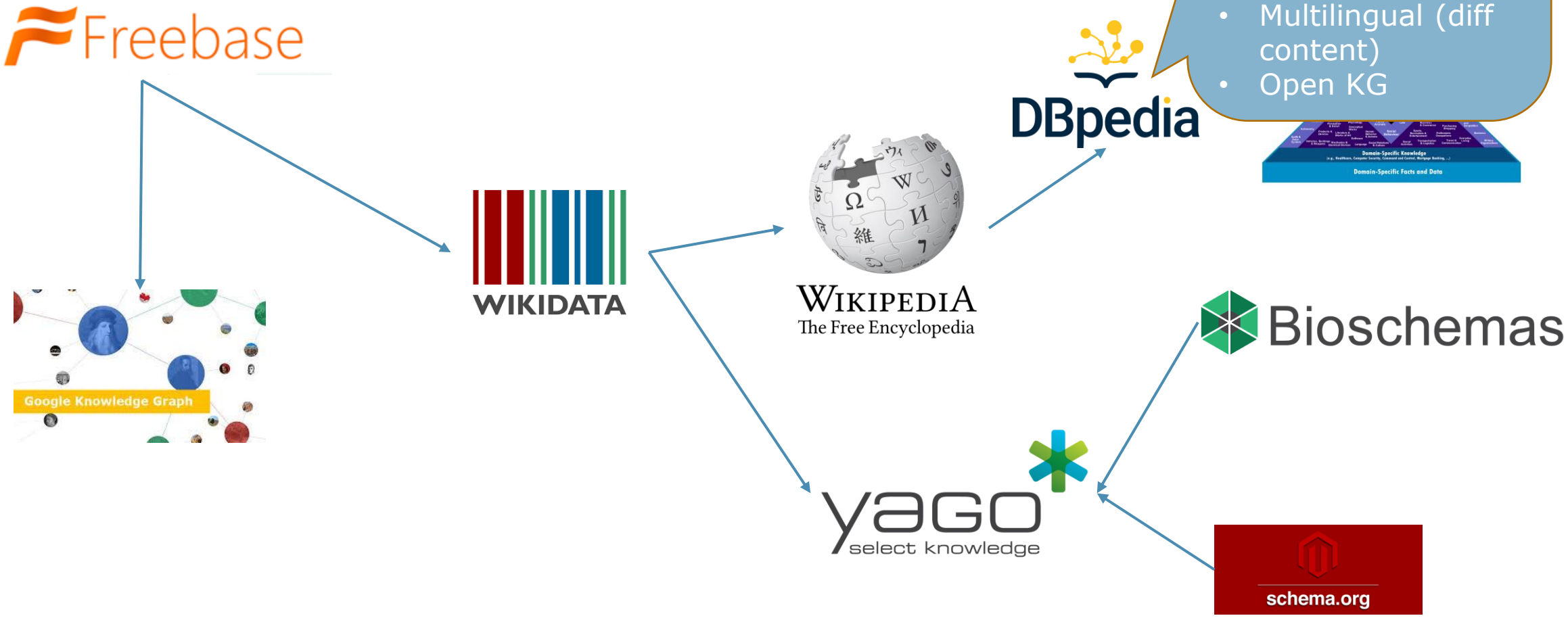


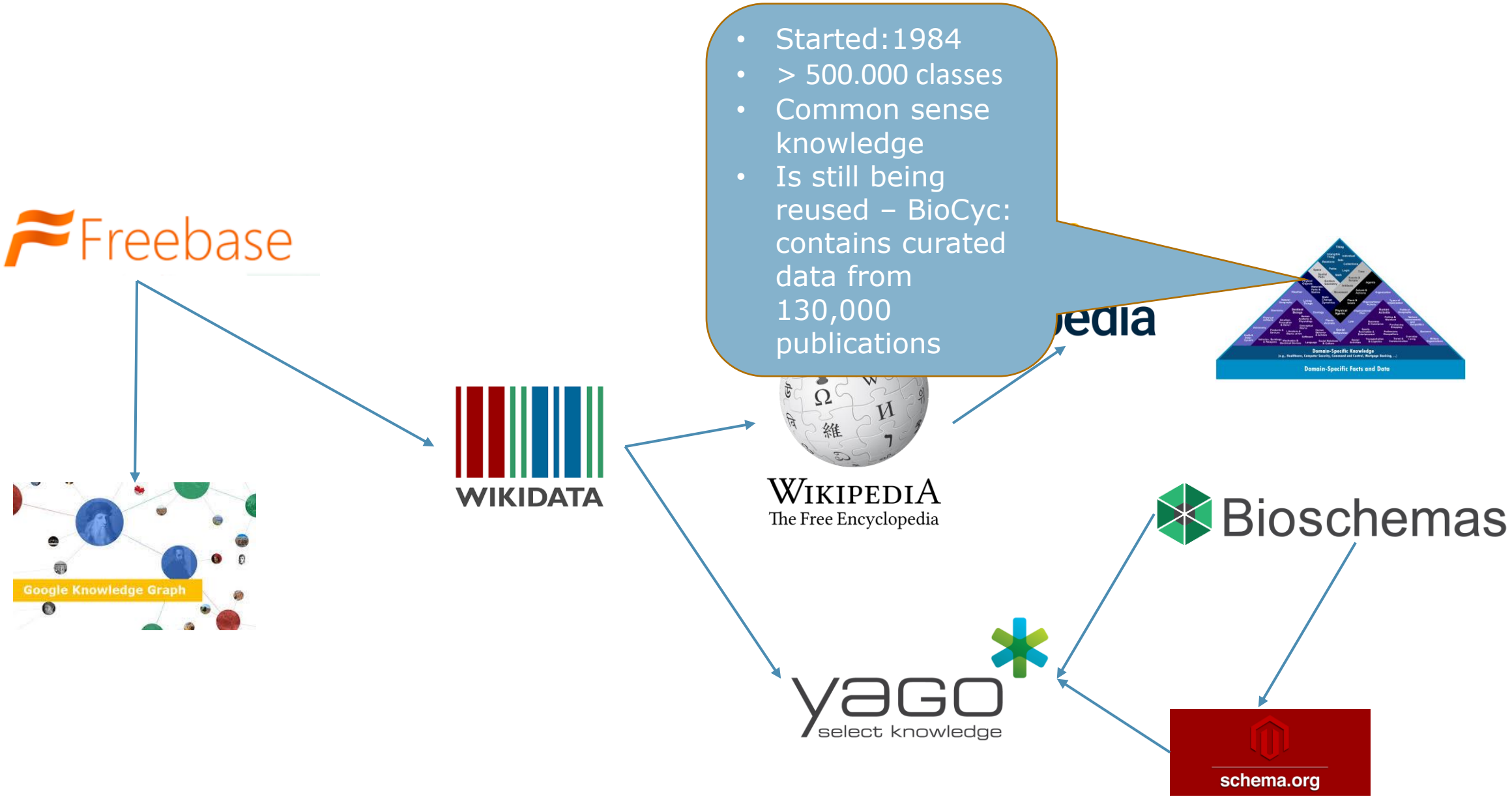
- Crowdsourced (500 edits/min)
- 88 millions entities
- High quality structured data in real time
- Multilingual –same info



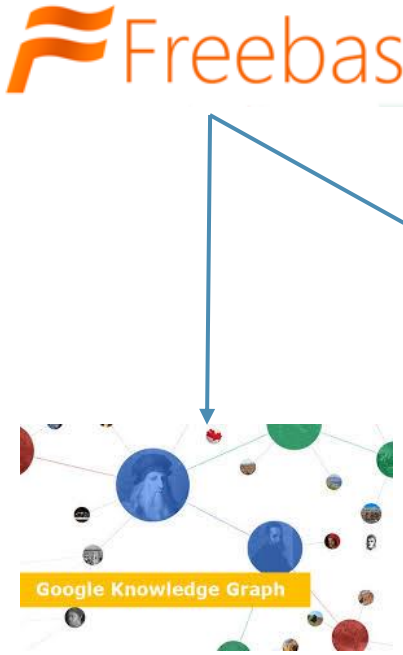
[Schema.org: Evolution of Structured Data on the Web]


Freebase
Google Knowledge Graph
WIKIDATA
DBpedia
Domain-Specific Knowledge
Domain-Specific Facts and Data
WIKIP
The Free Enc
ioschemas
• Data from Wikidata
• Ontology from schema.org
• Logical Constraints
yago
select knowledge
schema.org


Freebase
Google Knowledge Graph
WIKIDATA
WIKIPEDIA
The Free Encyclopedia
DBpedia
• 5 million things
• Extracts automatically content from Wikipedia
• Automatically updated
• Multilingual (diff content)
• Open KG
Domain-Specific Knowledge
Domain-Specific Facts and Data
Bioschemas
yago
select knowledge
schema.org


Started:1984
> 500.000 classes
Common sense knowledge
Is still being reused – BioCyc: contains curated data from 130,000 publications
Freebase
WIKIDATA
WIKIPEDIA
The Free Encyclopedia
Google Knowledge Graph
Bioschemas
yago
select knowledge
schema.org

- Encourages people in the life sciences to use Schema.org
- Proposing new types and properties to Schema.org for life science resources.
- Latest release: 2017
- 500,000 terms, including 15,000 relations, 5M facts (assertions) relating these terms.
- New assertions were added to the KB through a combination of automated and manual means.
- Applications: Glaxo, Cleveland Clinic Foundation, Terrorist KB



WIKIDATA

WIKIPEDIA
The Free Encyclopedia

Bioschemas

yago
select knowledge

schema.org

Standard Vocabularies/Ontologies

- Linked data are written using standard vocabularies i.e., vocabularies that have been **agreed** by certain **communities** for **describing** certain kinds of resources.
  - FOAF: for persons, their activities and their relations to other people and objects.
  - DublinCore: for digital & physical resources (video, images, books, artworks, etch)
  - Schema.org
  - DOLCE: upper level ontology
  - BFO: upper level ontology
  - …

Example: FOAF

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<http://en.wikipedia.org/wiki/Angelina_Jolie#this>
     rdf:type foaf:Person ;
     foaf:name “Angelina_Jolie" ;
     foaf:mbox <mailto:Angelina_Jolie@gmail.com> ;
     foaf:homepage <https://joliestweet.wixsite.com/angelinajolie> ;
     foaf:interest < https://www.unhcr.org/> ;
     foaf:knows [ rdf:type foaf:Person ;
                  foaf:name “Jon Voight" ] .
```

Property: foaf:knows

knows - A person known by this person (indicating some level of reciprocated interaction between the parties).
Status: stable
Domain: having this property implies being a Person
Range: every value of this property is a Person

Dbpedia example

DBpedia Browse using Formats Faceted Browser Sparql Endpoint

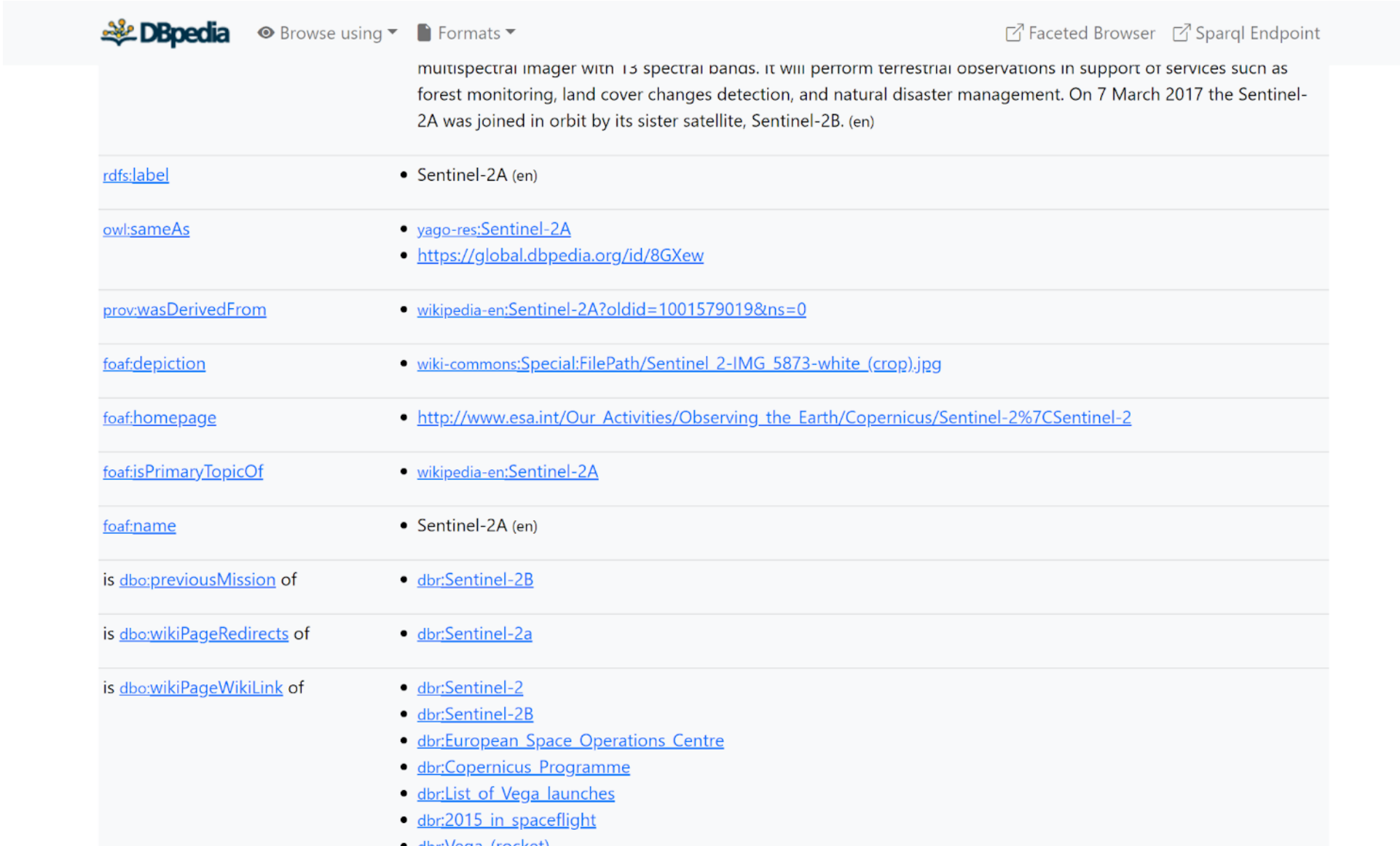
About: Sentinel-2A

An Entity of Type: Earth observation, from Named Graph: http://dbpedia.org, within Data Space: dbpedia.org

Sentinel-2A is a European optical imaging satellite launched in 2015. It is the first Sentinel-2 satellite launched as part of the European Space Agency's Copernicus Programme. The satellite carries a wide swath high-resolution multispectral imager with 13 spectral bands. It will perform terrestrial observations in support of services such as forest monitoring, land cover changes detection, and natural disaster management. On 7 March 2017 the Sentinel-2A was joined in orbit by its sister satellite, Sentinel-2B.

Property	Value
dbo:SpaceMission/missionDuration	• 2556.75
dbo:abstract	• Sentinel-2A is a European optical imaging satellite launched in 2015. It is the first Sentinel-2 satellite launched as part of the European Space Agency's Copernicus Programme. The satellite carries a wide swath high-resolution multispectral imager with 13 spectral bands. It will perform terrestrial observations in support of services such as forest monitoring, land cover changes detection, and natural disaster management. On 7 March 2017 the Sentinel-2A was joined in orbit by its sister satellite, Sentinel-2B. (en)
dbo:cosparId	• 2015-028A
dbo:launchDate	• 2015-06-23 (xsd:date)
dbo:launchVehicle	• dbr:Vega_(rocket)
dbo:manufacturer	• dbr:Airbus_Defence_and_Space
dbo:missionDuration	• 220903200.000000 (xsd:double)

Dbpedia example

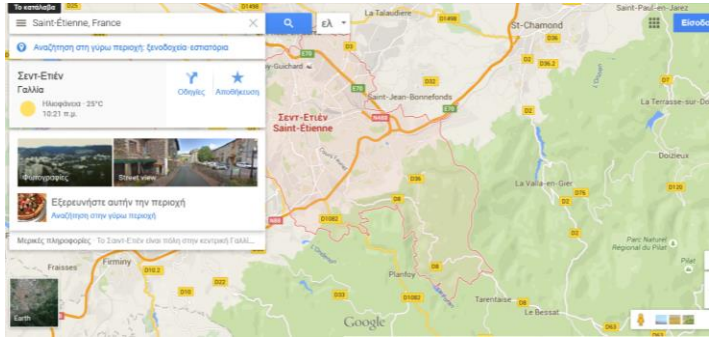


Ontology/KG developement

- Ontologies and knowledge graphs can be created using different means:
- **Cyc, schema.org:** trained ontology engineers
- **WordNet:** trained experts (lexicographers).
- **DBpedia, FreeBase, YAGO4:** automatically importing structured facts from various Web sources possibly with some inference.
- **CaLiGraph, TextRunner, NELL:** parsing textual data and extracting information from them

Geospatial Data on the Web & EO Data

Google Maps



Google earth

LDCM
LANDSAT
Data Continuity Mission

GeoQA

DEEP CUBE

ExtremeEarth

AI4 copernicus

Linked Geospatial and Temporal Data

- This is a research area studied by ai. team since 2010.
- The basic research question is how to manage geospatial and temporal data on the Web using linked data technologies.
- See our team web site for more: http://ai.di.uoa.gr/

Research Contributions

- The model stRDF and stSPARQL
- The systems GeoTriples and Silk
- The spatiotemporal RDF store Strabo2
- The ontology-based data access system Ontop-spatial.
- The visualization tool Sextant.
- Extension of YAGO2 with precise geospatial information for administrative regions: YAGO2geo
- Geospatial Question Answering tool GeoQA
- Many applications, especially with Earth Observation data.
- You will use most of our tools in the course project.

Example of stRDF (Geospatial Dimension)

```
gag:Olympia rdf:type gag:MunicipalCommunity;
            gag:name "Ancient Olympia";
            gag:population "184"^^xsd:int;
            strdf:hasGeometry "POLYGON((21.5 18.5,23.5 18.5, 23.5
                                        21,21.5 21,21.5 18.5));
      http://www.opengis.net/def/crs/OGC/1.3/CRS84"^^strdf:WKT.
```

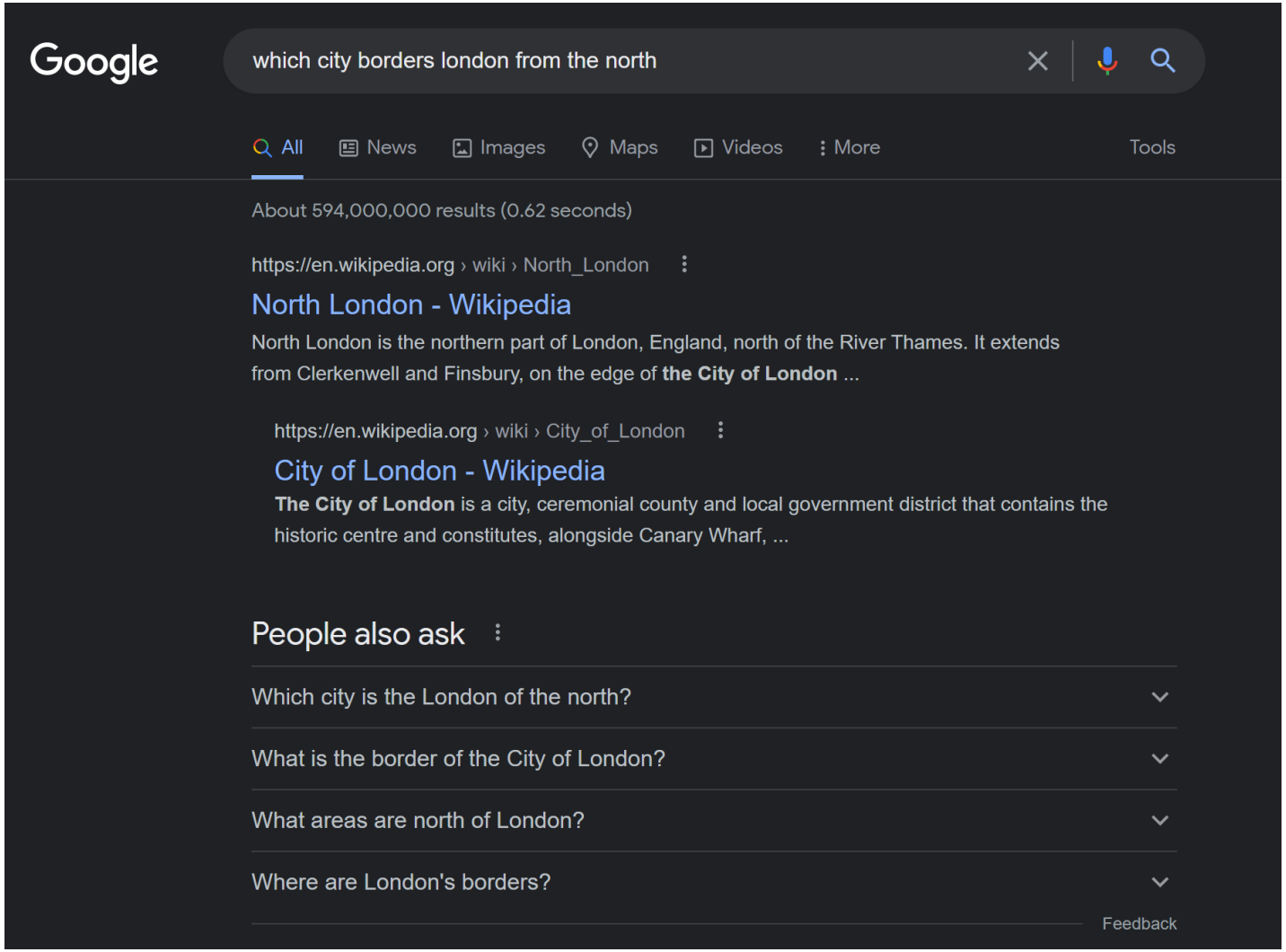
Ancient
Olympia

Example of stSPARQL (Geospatial Dimension)

Query: Compute the parts of burnt areas that lie in coniferous forests

```
SELECT ?burntArea (strdf:intersection(?baGeom, strdf:union(?fGeom)) AS
                                                          ?burntForest)
WHERE {  ?burntArea  rdf:type noa:BurntArea;
                     strdf:hasGeometry ?baGeom.
          ?forest rdf:type clc:Region;
                  clc:hasLandCover clc:ConiferousForest;
                  strdf:hasGeometry ?fGeom.

FILTER  (strdf:intersects(?baGeom,?fGeom)) }
GROUP BY ?burntArea ?baGeom
```

Geospatial Question Answering



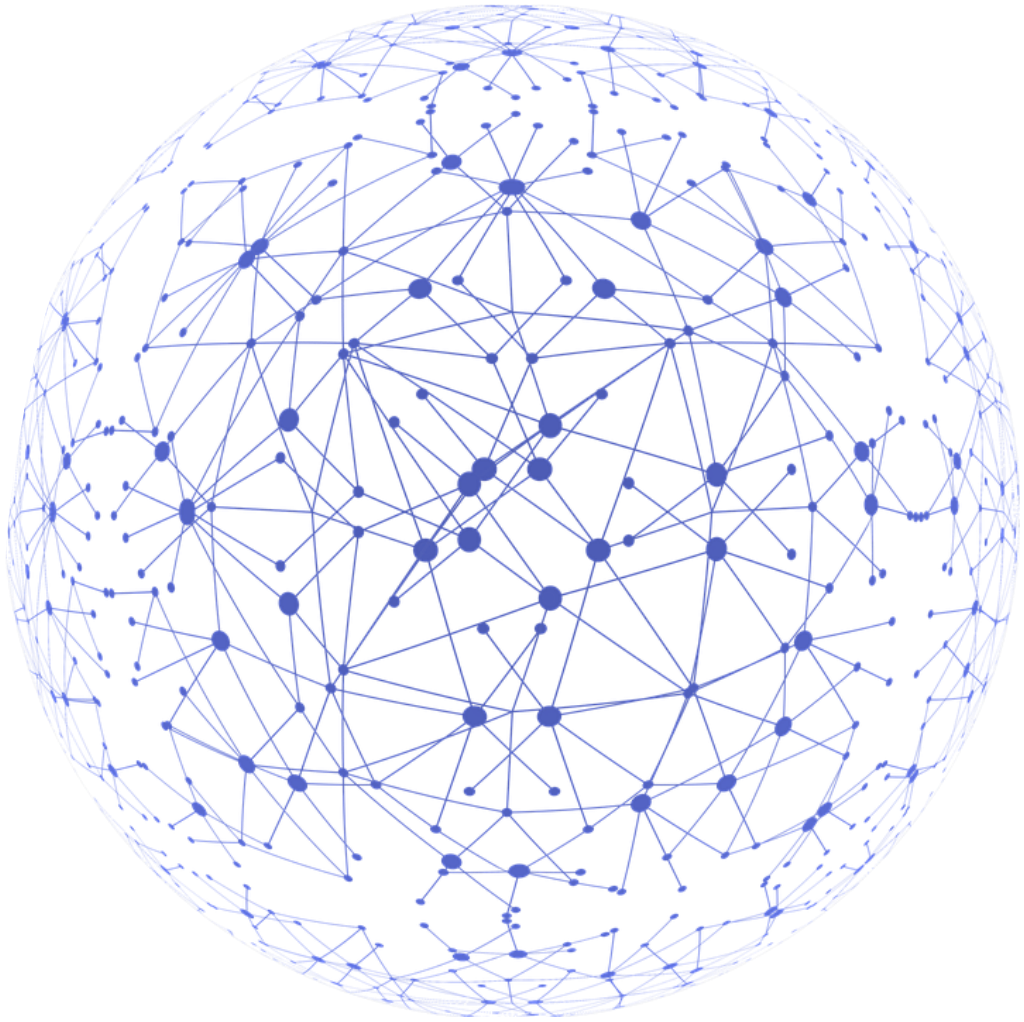
Geospatial Question Answering

- It is now clear that we do not get a definite answer like we got for Sentinel-2 and its start date.
- The reason is that the Google knowledge graph currently does not support geospatial question answering (similarly, temporal or with quantities etc.)
- See our paper on this topic: http://cgi.di.uoa.gr/~koubarak/publications/2023/ISWC_2023_GeoQuestions_paper-3.pdf

- Can we get even more complex??
- Sentinel-2 data from cities north of London during Christmas 2020

Outline

- Basic concepts
  - knowledge, ontology, knowledge graphs
- Some History: How did it all start?
- Why do we need them?
- Where are we today?
- So, what is an ontology?
  - RDFS, OWL, Description logics,
- How do we query an ontology/KG?
  - SPARQL
- Linked Data
- Linked geospatial and temporal data
- **High Value use cases**



High Value Use Cases:

- Public sector:
  - Open Data Directive (EU)
  - Openbudgets.brussels
  - The Zaragoza's Knowledge Graph: Open Data to Harness the City Knowledge
- Law:
  - Legislative data
- Medicine:
  - eHealth Network
  - Biopharma industry
- Industry
- Banking sector

Open Data Directive

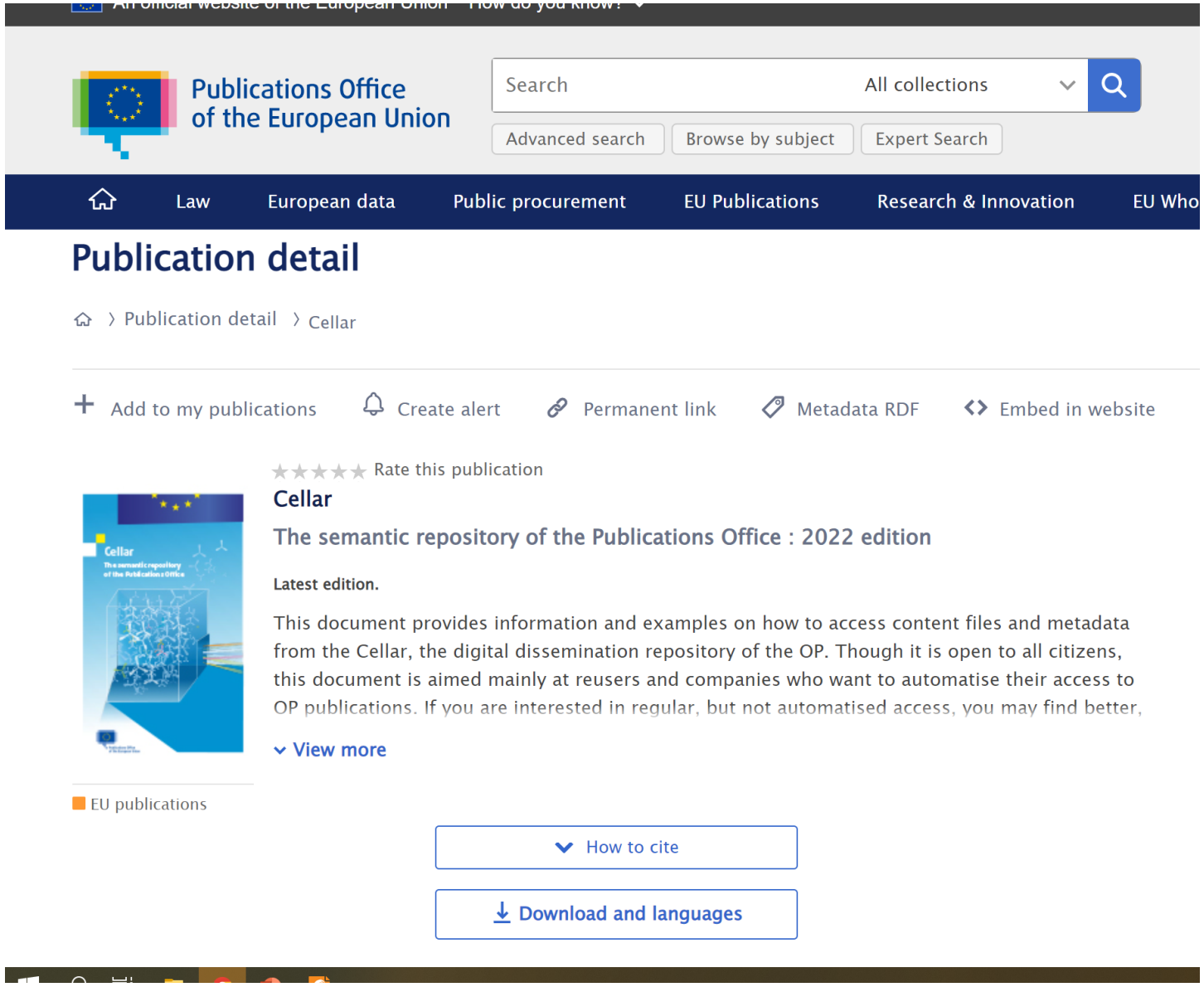
- Entered into force on 16 July 2019 (replacing the PSI Directive of 2013)
- The goal is to make public sector and **publicly-funded data re-usable** and **transparent**.
- Focuses on the **economic aspects of the reuse of information**.
- Strengthens the **transparency** requirements for public–private agreements
- Encourages the Member States to make as much information available for re-use as possible
- High value datasets: beneficial for the society and economy:
  - Geospatial, earth observation and environment, meteorological, statistics, companies and company ownership, mobility

Example: Brussels

- Openbudgets.brussels[1]
  - Better understanding who spent how much and why, makes political decisions more transparent.

[1]https://openbudgets.be.brussels/en/about/about-us

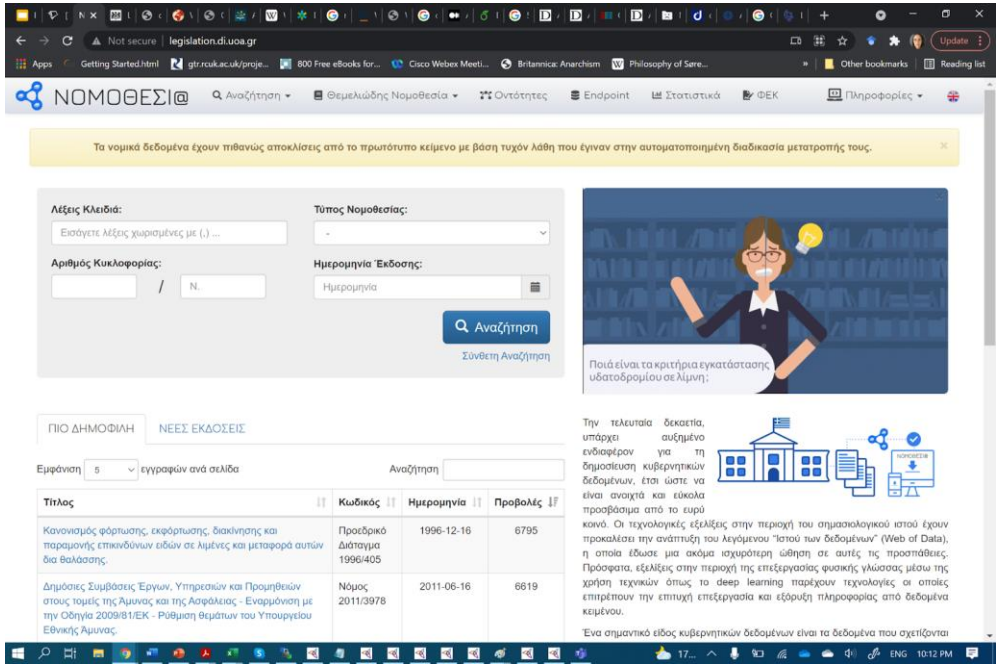
Eur-lex: Access to European Law



Cellar's EU publications can be retrieved directly from Cellar using the HTTP RESTful web services, dedicated to publications.

Greek linked data

- A legal document might refer to **another** legal document
- A legal document might **modify** the content of other legal documents
- Nomothesia:
  - presents the textual **content** of legal documents
  - search for legislative documents based on their **metadata** or **textual content**



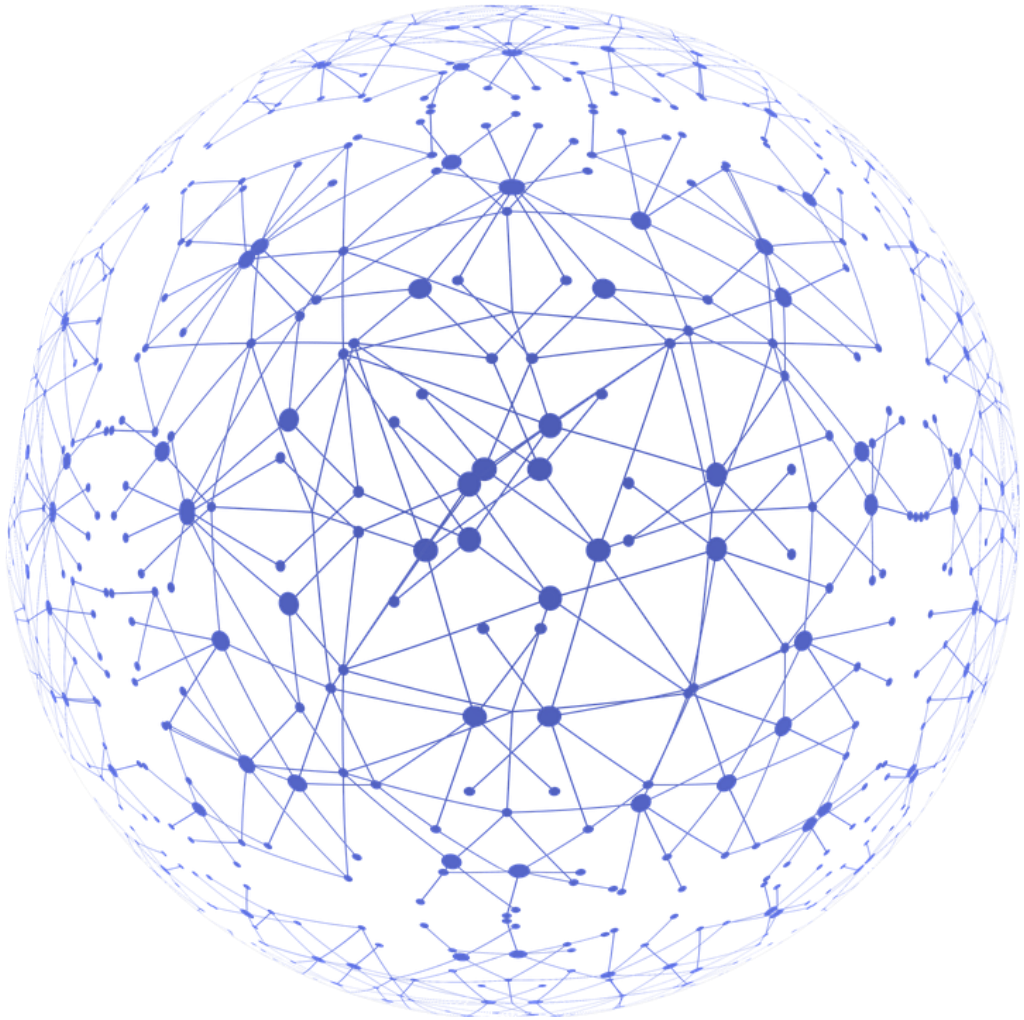
- answers complex analytics
- e.g., "Which are the 5 most frequently modified legal documents during 2008-2013?"
- "Who are the 3 past government members that have signed the most legal documents during their service in 2008-2015?".

ELISE - European Location Interoperability Solutions fc e-Government

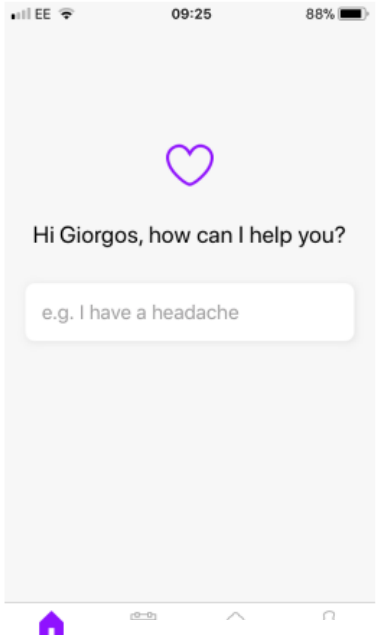
- Goal: Management of knowledge resources from the various organizations:
  - technical resources, training documents, skillsets, insights, and frequently asked questions

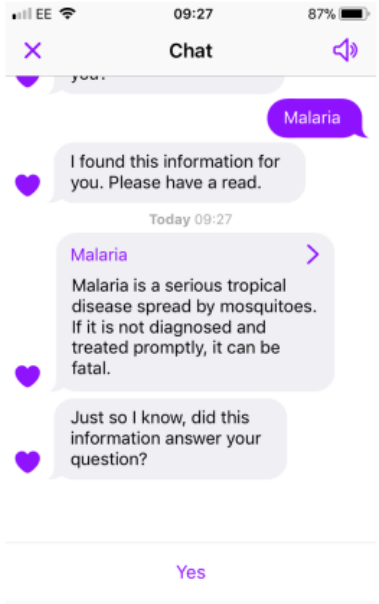
Outline

- Basic concepts
  - knowledge, ontology, knowledge graphs
- Some History: How did it all start?
- Why do we need them?
- Where are we today?
- So, what is an ontology?
  - RDFS, OWL, Description logics,
- How do we query an ontology/KG?
  - SPARQL
- Linked Data
- **Linked geospatial and temporal data**
- High Value use cases



High Value Use Cases: Healthcare

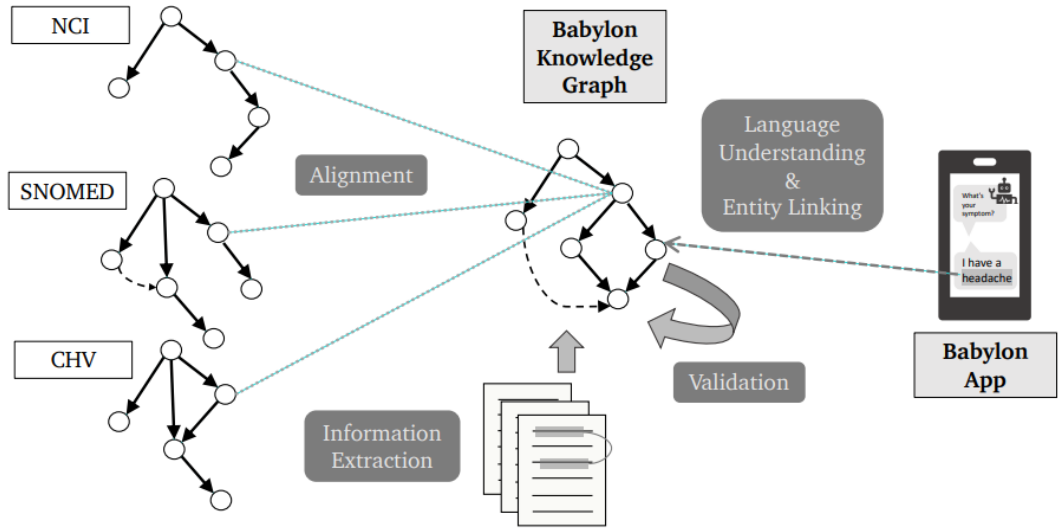




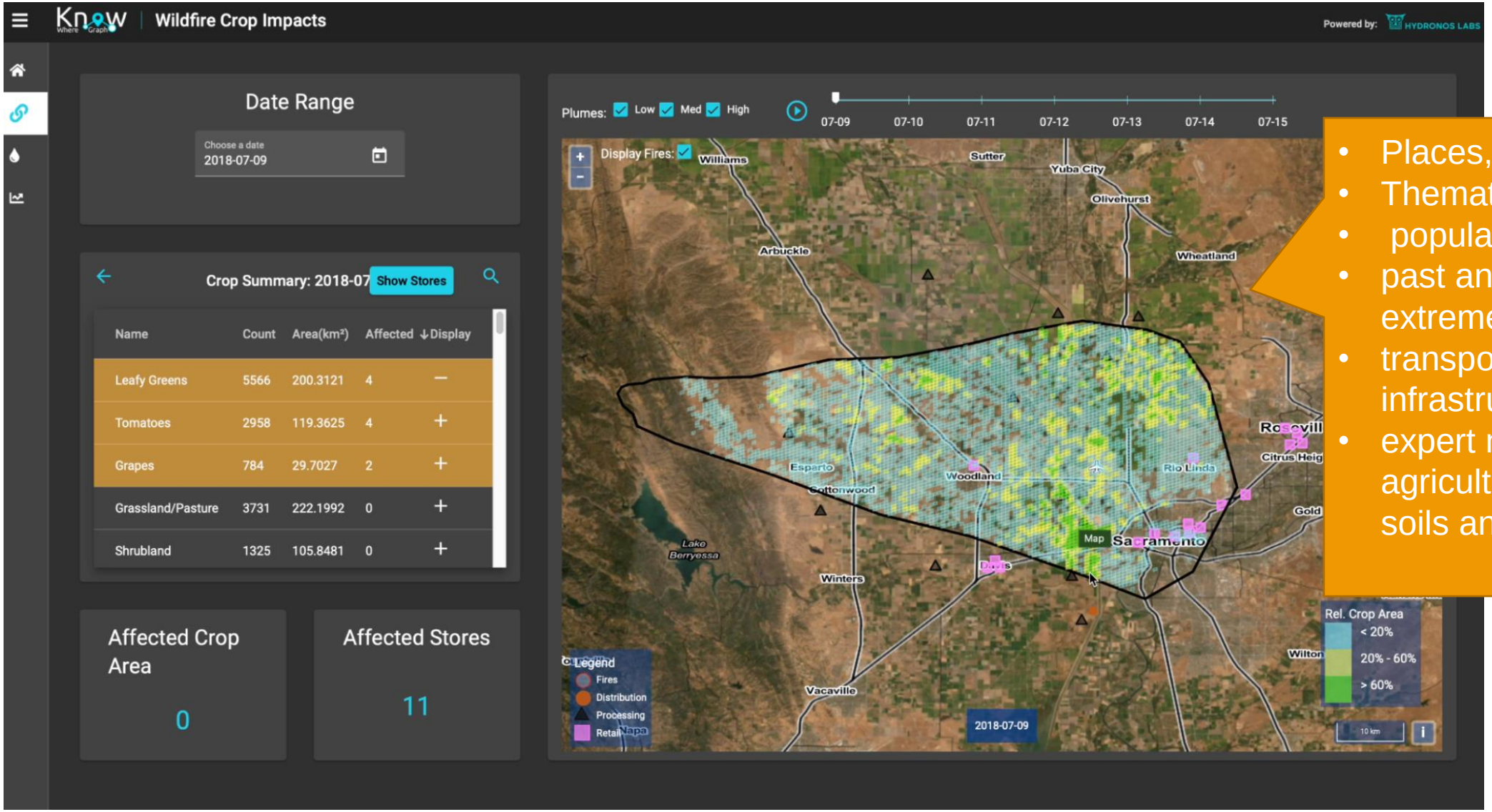
Cancer domain, including cancer related diseases, findings and abnormalities.

Nursing Ontology Diagnosis

Open Access, Collaborative Consumer Health Vocabulary Initiative



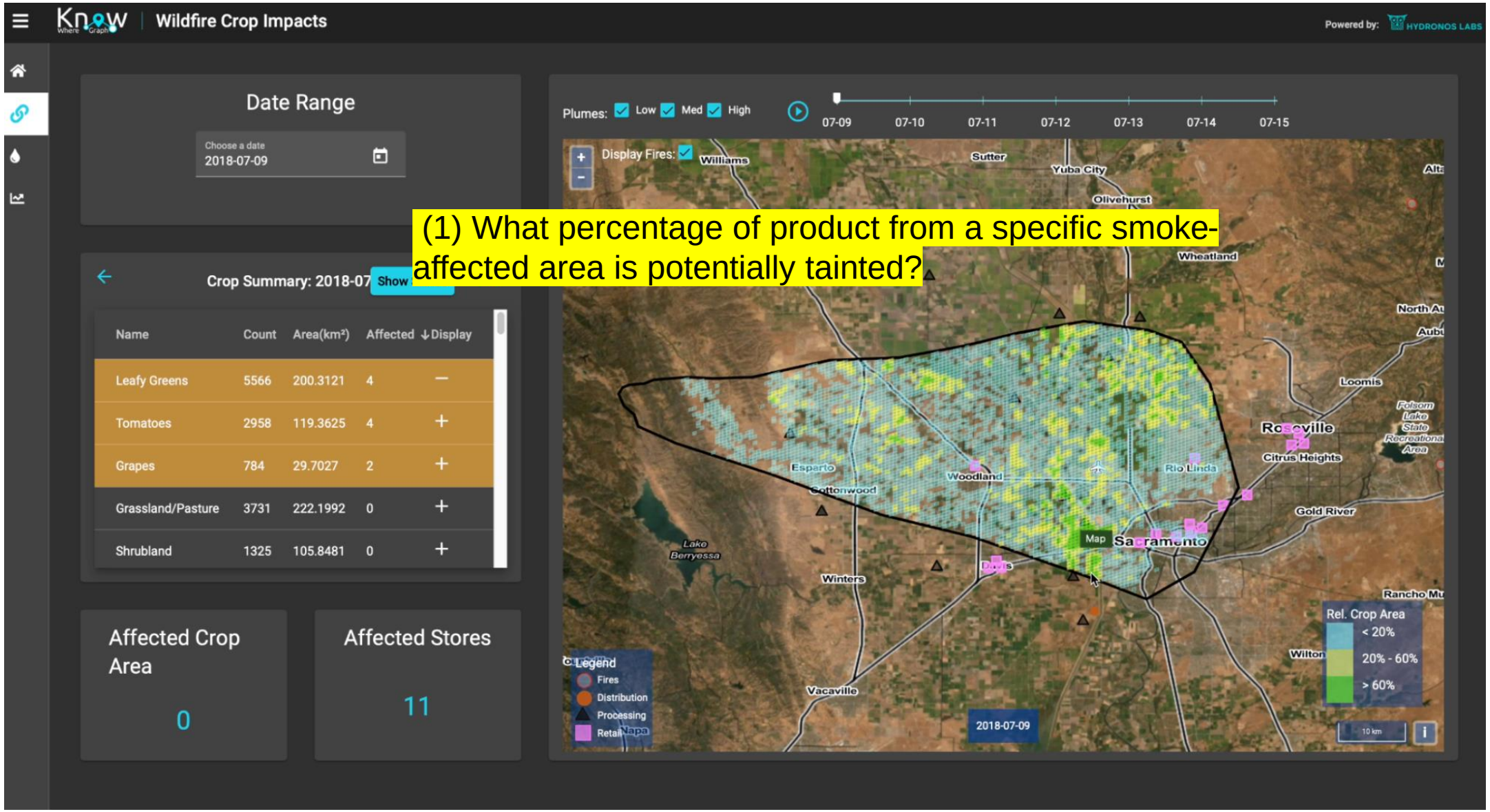
KnowWhereGraph -Food Supply Chain Resilience



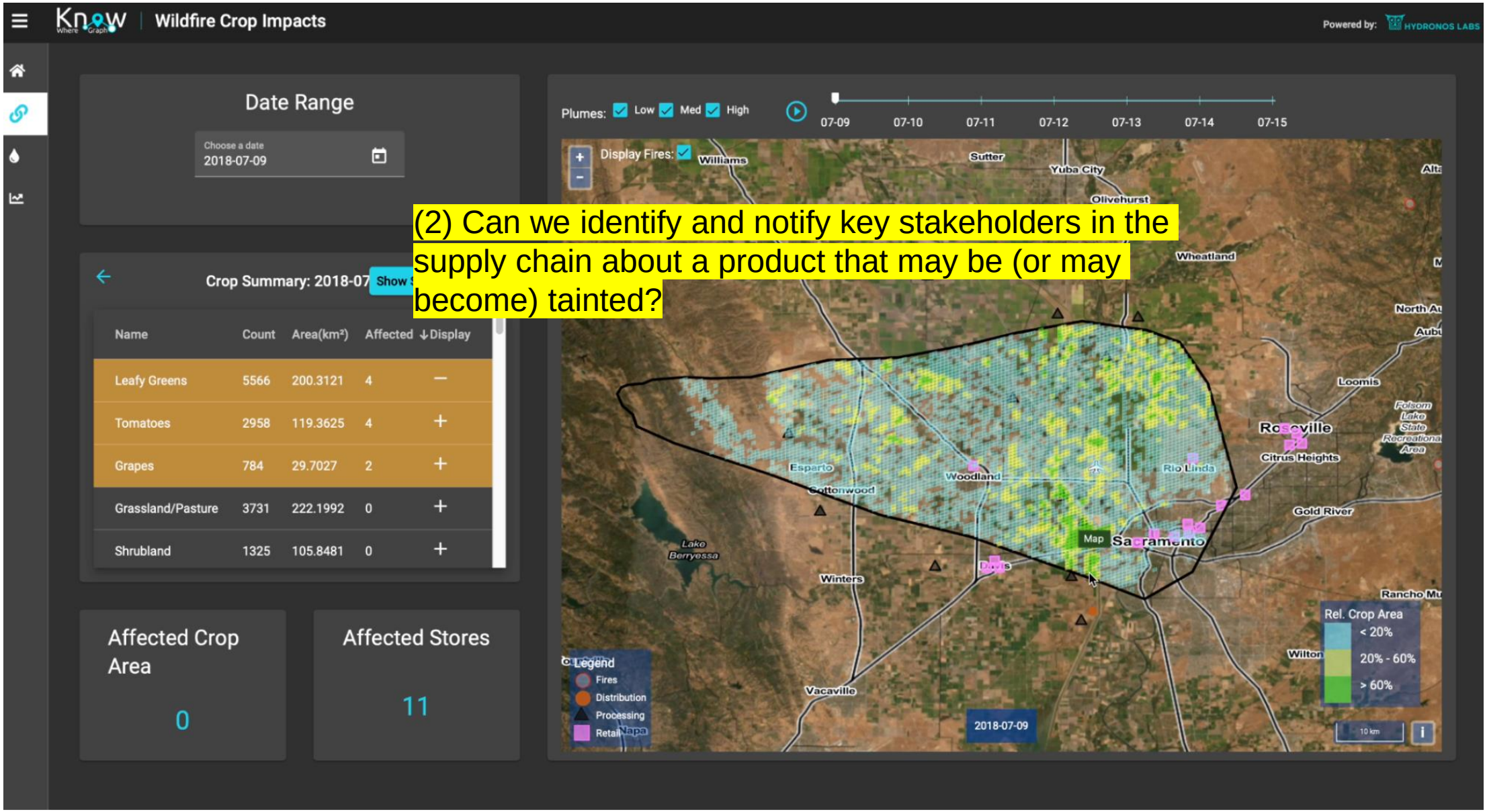
- Places, regions
- Thematic layers on top:
- population health
- past and present extreme events,
- transportation infrastructure,
- expert networks agricultural data about soils and crops

A screenshot of the KWG application developed for the Food Industry Association allowing users to visualize and explore the impact of wildfire events on the full food supply value chain for select crops.

KnowWhereGraph -Food Supply Chain Resilience



KnowWhereGraph -Food Supply Chain Resilience



(2) Can we identify and notify key stakeholders in the supply chain about a product that may be (or may become) tainted?

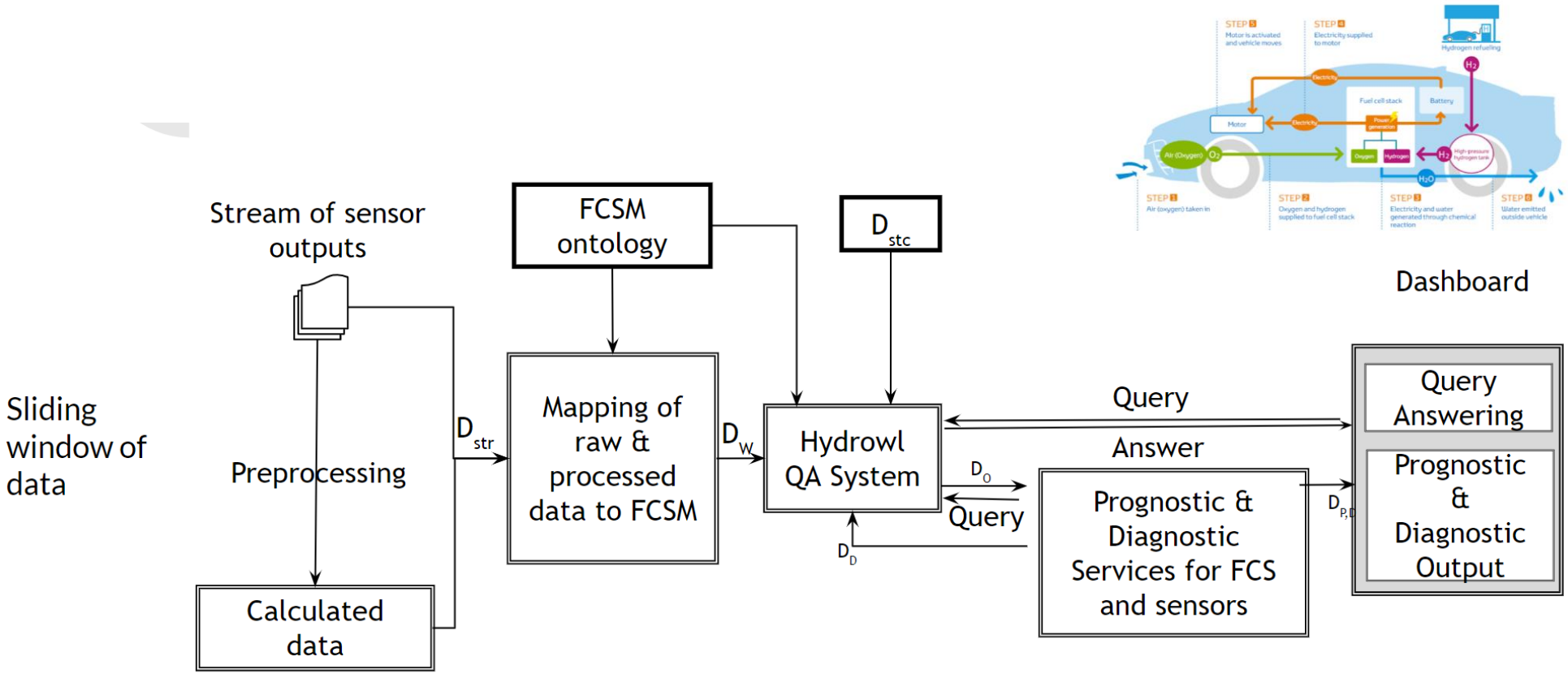
AI on Demand Platform: Working Group for Ontology

- AIoD facilitates **knowledge sharing**, **research experimentation** and development of state-of-the art **solutions and technologies related with AI and AI-based robotics**.

- To avoid replication of work across projects on ontologies
- To seek a "common ontology" to support the future AI on demand platform and related projects
- To reach an agreement on technologies to be used
- To make the knowledge discoverable across platforms
- To incorporate trustworthy AI topics related to the knowledge classification

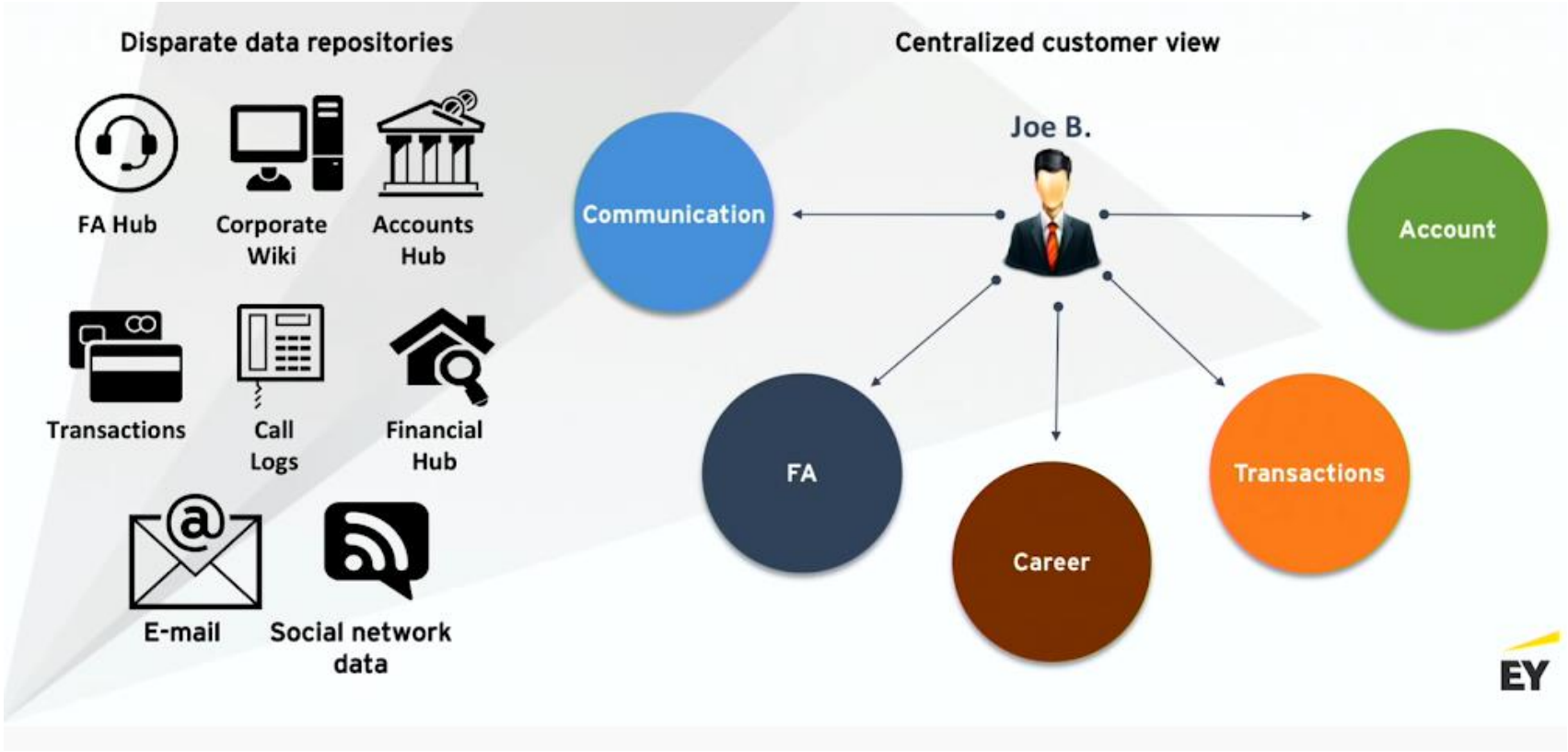
https://www.ai4europe.eu/ai-community/working-groups/working-group-ontology

High Value Use Cases: System Diagnosis



This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101032307

High Value Use Cases: Enterprise



[Knowledge Graphs: The Path to Enterprise — Michael Moore and Al Omar Azhar, EY]

Where are we today? –Enterprise KGs

What have we seen?

- Knowledge vs Information
- History
- Ontology
- Knowledge Graph
- Knowledge Graphs vs Graph DBs
- Linked Data
- Leading Ontologies/KGs
- Use cases

- A **knowledge graph** can be used for:
  a. Store data, just a fancy name for a Graph DB!
  b. For representing knowledge in a flexible, reusable way that allows the derivation of *implied* knowledge
  c. Automated data classification

- A **knowledge graph** can be used for:
  a. Store data, just a fancy name for a Graph DB!
  b. For representing knowledge in a flexible, standardized, reusable way that allows the derivation of *implied* knowledge
  c. Automated data classification

- An **ontology** can be used for:
  a. Store data in structured, reusable way, just a sophisticated name for KGs!
  b. Automated data classification
  c. For representing complex generic knowledge in a flexible, standardized reusable way that allows the derivation of *implied* knowledge

- An ontology can be used for:
  a. Store data, just a fancy name for KGs!
  b. Automated data classification
  c. For representing **complex generic knowledge** in a **flexible**, **standardized, reusable** way that allows the derivation of ***implied*** knowledge

Emphasis of this Course

- The Semantic Web topics that we will cover in this course are:
  - Linked data
  - RDF, RDFS, RDF*, SHACL, SPARQL and RDF stores (Sesame).
  - Ontologies
  - Description logics, OWL 2, tools for developing ontologies (Protégé) and reasoners (Pellet, Hermit).
  - Ontology engineering
  - Rules
  - SWRL and others
  - Ontology engineering
  - Linked geospatial data (main focus of AI.Team)
  - stSPARQL, GeoSPARQL and Strabo

Logistics

- 3 Homeworks
- 1 Project (usually part of the 3rd HW)

That's it!

- **Knowledge Graphs: Synthesis Lectures on Data, Semantics, and Knowledge**, November 2021.
- **Knowledge Graphs: Data in Context for Responsive Businesses**, Specs, By Jesús Barrasa, Amy E. Hodler, and Jim Webber. O'Reilly Media, July 2021.
- Dooley et al. FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration, npj Science of Food (2018) 2:23, https://www.nature.com/articles/s41538-018-0032-6.pdf
- https://www.cdbb.cam.ac.uk/files/a_survey_of_top-level_ontologies_lowres.pdf
- Description Logics: http://www.cs.ox.ac.uk/people/ian.horrocks/Publications/download/2007/BaHS07a.pdf