Virtual Knowledge Graphs — Applications and Tools

Guohui Xiao

University of Bergen & Ontopic

Huawei Ireland Research Centre

7 July 2022

Outline

1. A Brief History of Semantic Web & Knowledge Graphs

2. The Virtual Knowledge Graph (VKG) Framework

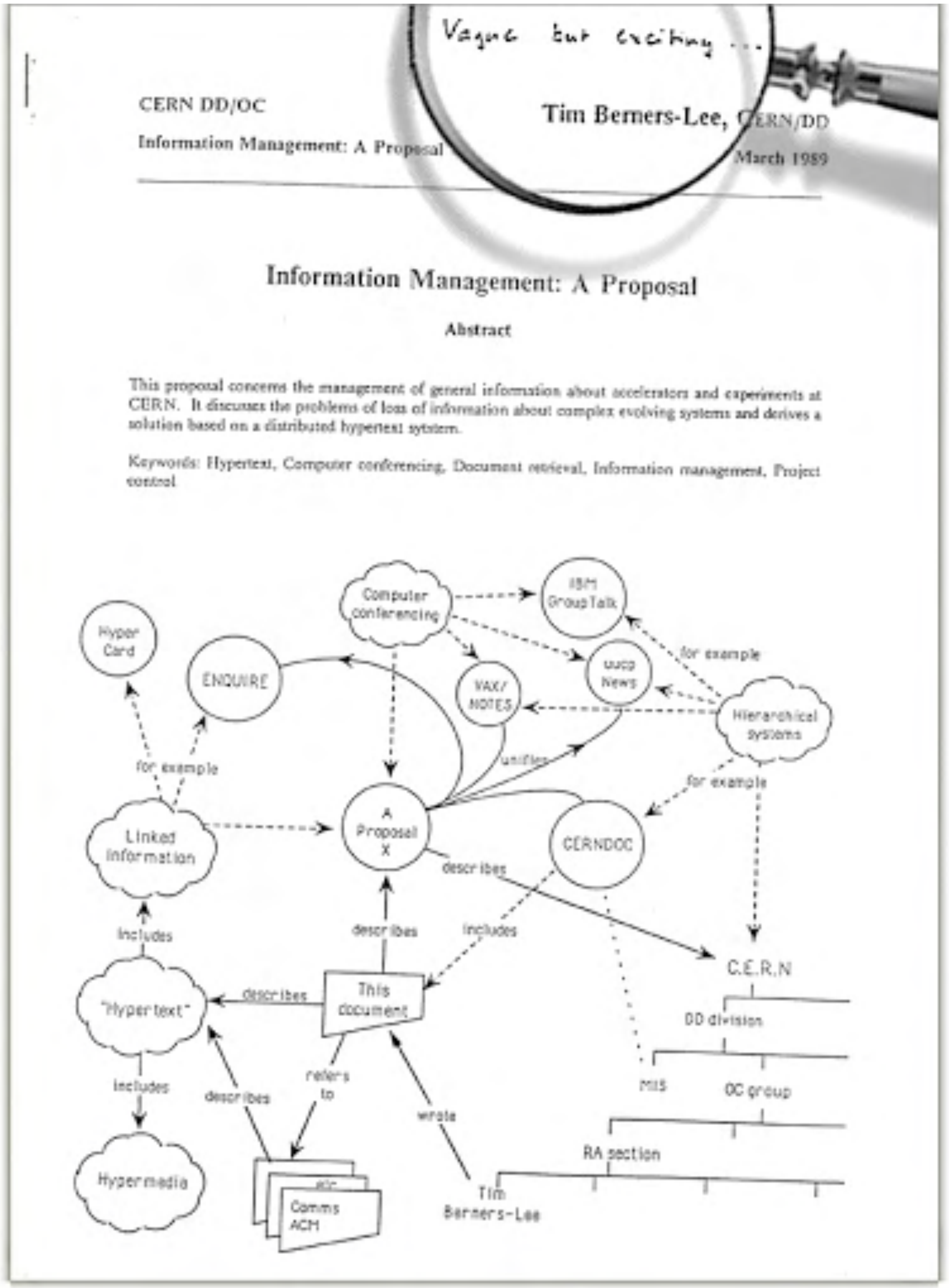
3. VKG Systems and Applications

4. Key Technologies in VKGs

5. A Demo for Data Integration through VKGs

1. A Brief History of Semantic Web & Knowledge Graphs

World Wide Web (1989)



Vague but exciting ...

CERN DD/OC

Information Management: A Proposal

Tim Berners-Lee, CERN/DD

March 1989

Information Management: A Proposal

Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext sytstem.

Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control

... This is why a "web" of notes with links between them is far more ***useful than a fixed hierarchical system****. ... The system we need is like* ***a diagram of circles and arrows****, where circles and arrows can stand for anything.*

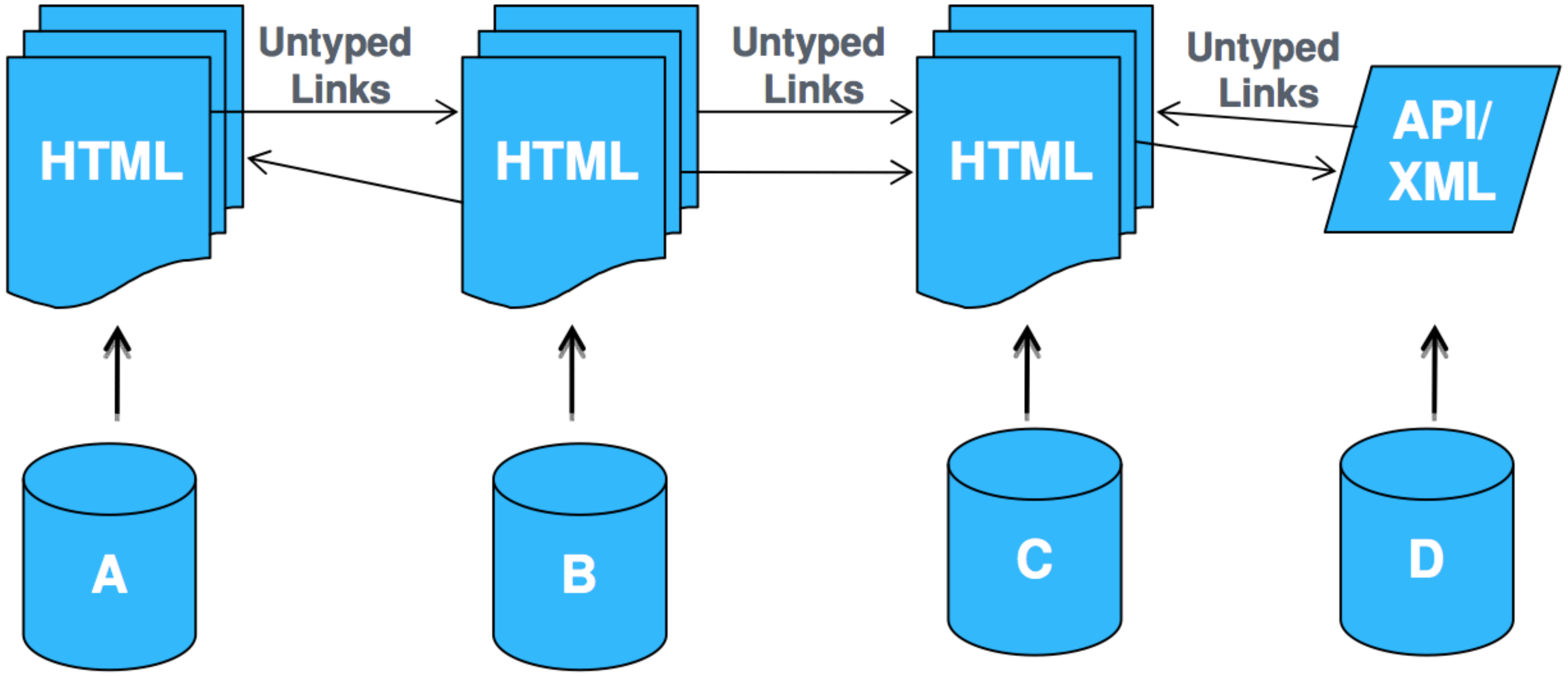
— Information management: A proposal. CERN. 1989

Links are the core of the Web.

WWW Inventor

Sir Tim Berners-Lee

Web of Documents

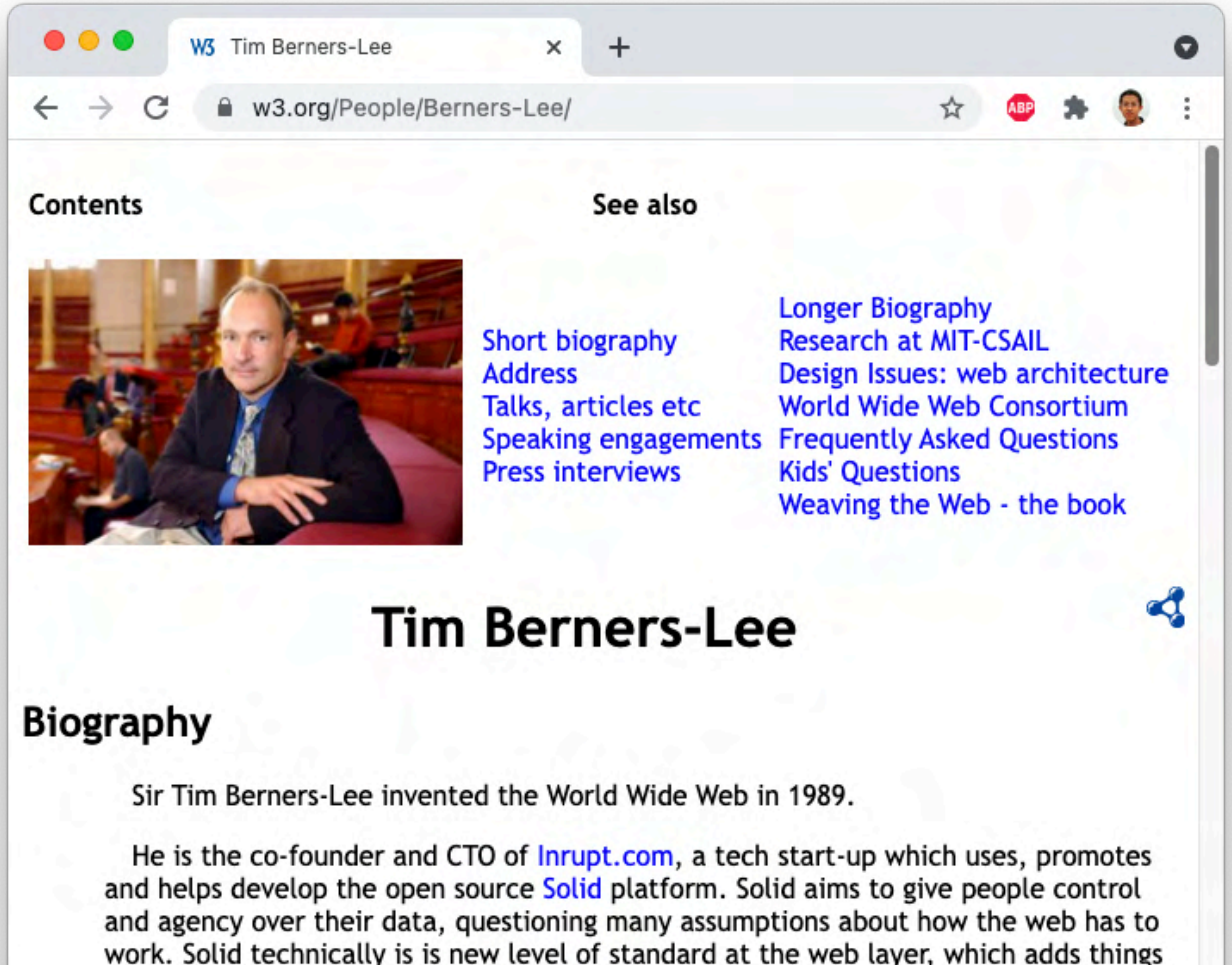


- Primary objects: **documents**
- Degree of structure in data: **low**
- Semantics of content: **implicit**
- Designed for: **human consumption**

Links between **documents**

Problem of the Web of Documents

- wealth of information
- … targeted at human users



Contents | **See also**

Short biography
Address
Talks, articles etc
Speaking engagements
Press interviews

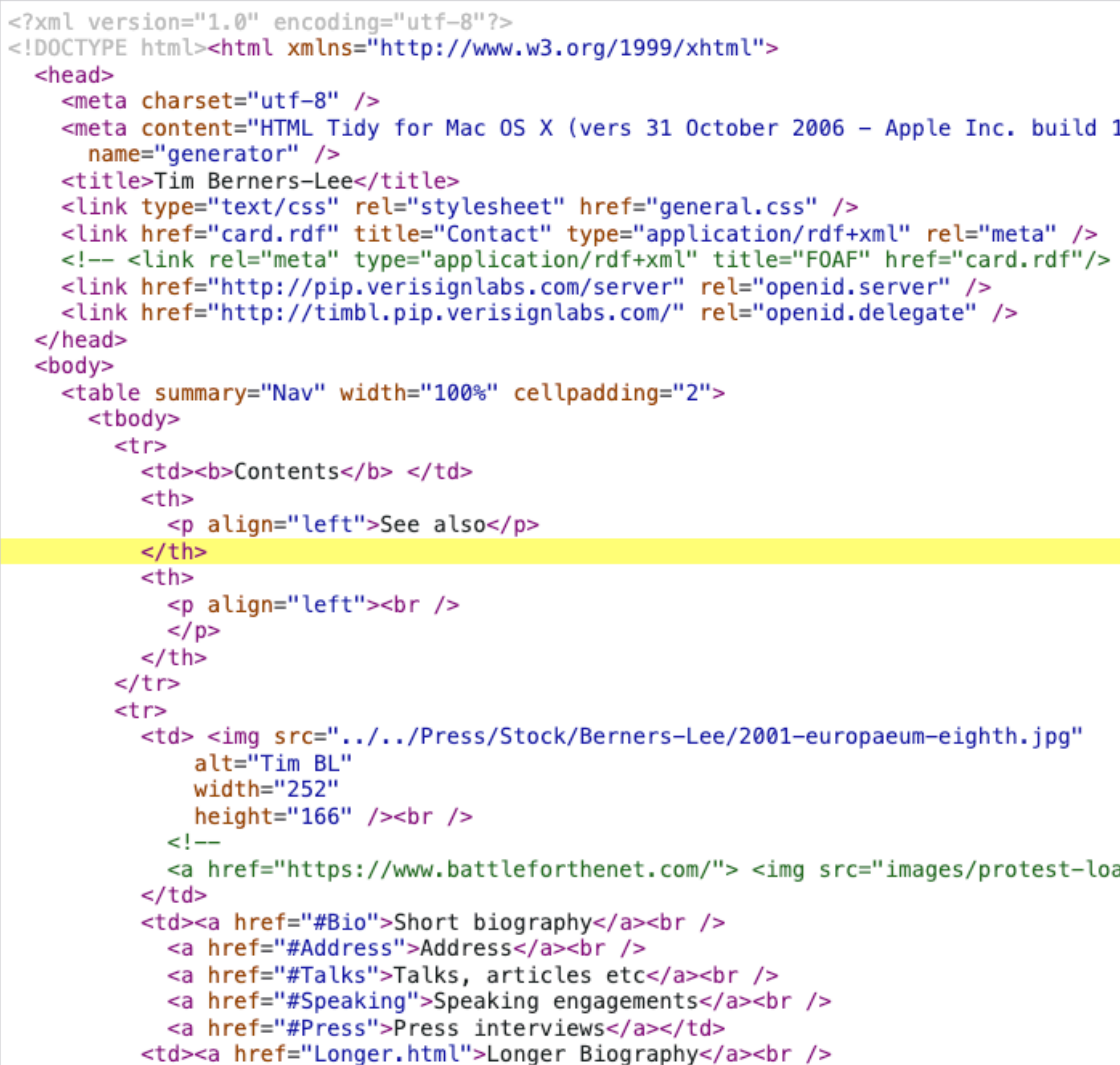
Longer Biography
Research at MIT-CSAIL
Design Issues: web architecture
World Wide Web Consortium
Frequently Asked Questions
Kids' Questions
Weaving the Web - the book

Tim Berners-Lee

Biography

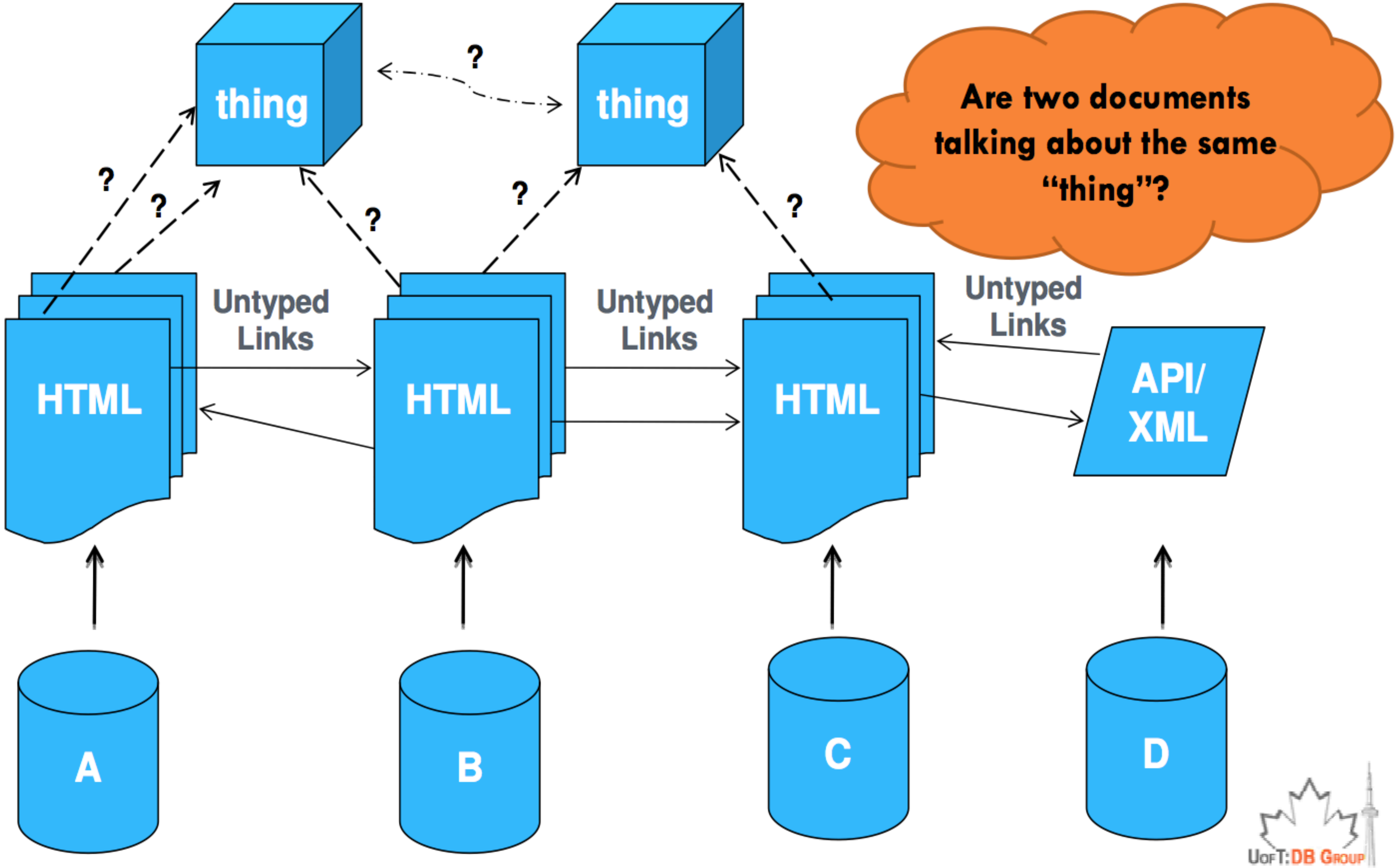
Sir Tim Berners-Lee invented the World Wide Web in 1989.

He is the co-founder and CTO of Inrupt.com, a tech start-up which uses, promotes and helps develop the open source Solid platform. Solid aims to give people control and agency over their data, questioning many assumptions about how the web has to work. Solid technically is is new level of standard at the web layer, which adds things



```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE html><html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <meta charset="utf-8" />
    <meta content="HTML Tidy for Mac OS X (vers 31 October 2006 - Apple Inc. build 1
      name="generator" />
    <title>Tim Berners-Lee</title>
    <link type="text/css" rel="stylesheet" href="general.css" />
    <link href="card.rdf" title="Contact" type="application/rdf+xml" rel="meta" />
    <!-- <link rel="meta" type="application/rdf+xml" title="FOAF" href="card.rdf"/>
    <link href="http://pip.verisignlabs.com/server" rel="openid.server" />
    <link href="http://timbl.pip.verisignlabs.com/" rel="openid.delegate" />
  </head>
  <body>
    <table summary="Nav" width="100%" cellpadding="2">
      <tbody>
        <tr>
          <td><b>Contents</b> </td>
          <th>
            <p align="left">See also</p>
          </th>
          <th>
            <p align="left"><br />
            </p>
          </th>
        </tr>
        <tr>
          <td> <img src="../../Press/Stock/Berners-Lee/2001-europaeum-eighth.jpg"
              alt="Tim BL"
              width="252"
              height="166" /><br />
            <!--
            <a href="https://www.battleforthenet.com/"> <img src="images/protest-loa
          </td>
          <td><a href="#Bio">Short biography</a><br />
            <a href="#Address">Address</a><br />
            <a href="#Talks">Talks, articles etc</a><br />
            <a href="#Speaking">Speaking engagements</a><br />
            <a href="#Press">Press interviews</a></td>
          <td><a href="Longer.html">Longer Biography</a><br />
```

Web of Documents



Problems of the Web of Documents

Approaches toward a solution:

① Ad hoc: Deployment of AI methods
(most notably NLP techniques) to evaluate
existing unstructured information on the Web

② A priori: Structure information on the Web
at authoring time
in a way facilitating later automated deployment

➔ Semantic Web

The Semantic Web (2001)

The Semantic Web — A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities

By Tim Berners-Lee, James Hendler and Ora Lassila

EXCLUSIVE: WARP DRIVE UNDERWATER • ARCTIC OIL VS. WILDLIFE
SCIENTIFIC AMERICAN
MAY 2001 WWW.SCIAM.COM $4.95
Get the Idea?
(TOMORROW'S WEB WILL)
i know what you mean ...
PLUS:
Antibiotics' Dim Future
Rorschach: A Waste of Ink
The Oldest Stars
Copyright 2001 Scientific American, Inc.

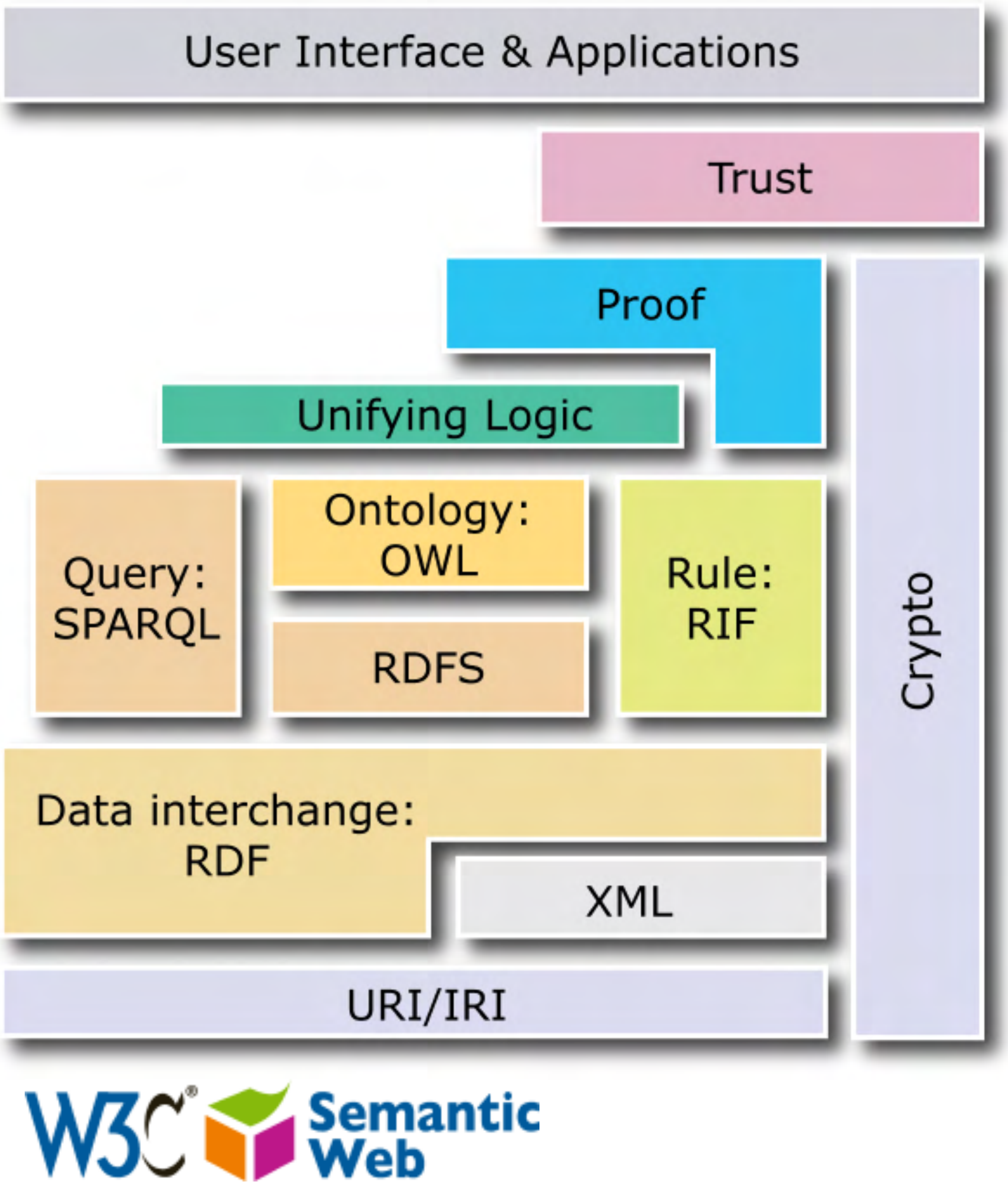
Key ideas:

- Expressing Meaning: things, but not strings
- Knowledge Representation: logic based, triple format
- Ontologies: expressive knowledge
- Agents: Machines should understand each other

Semantic Technologies

- **Not Only About the Web**
- The Semantic Web vision has generated technologies that are applied outside the web context including:
  - Retailing, supermarkets (see web pages of Best Buy)
  - Health care, medicine (see SNOMED ontology)
  - Enterprise intelligence (see solutions by Ontotext)
  - Research: Bio, Geo, Cultural heritage, etc.
  - Government (“smart cities”)
  - Software development (“ontology-driven software develoment”)

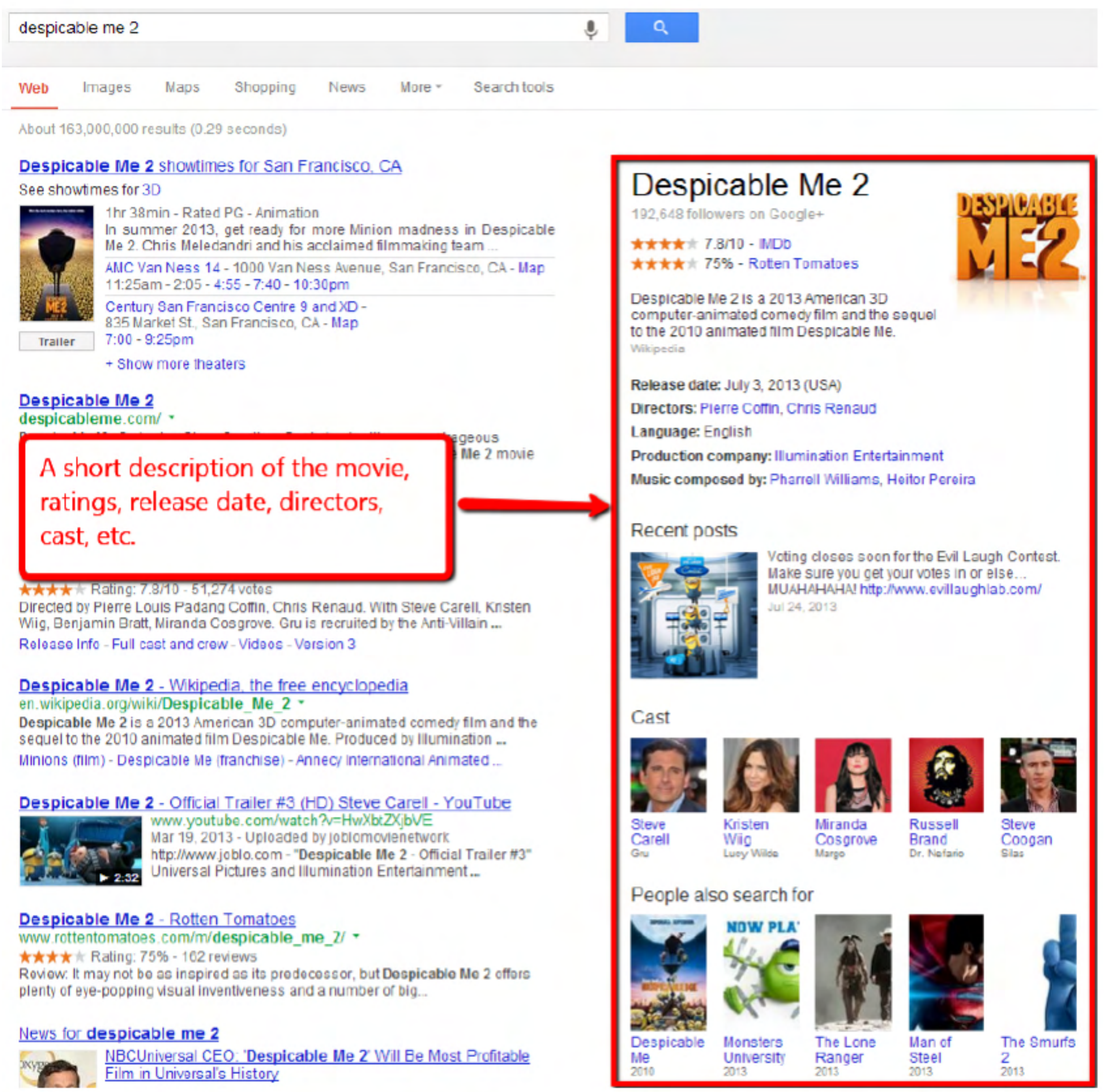
Semantic Web: W3C Standards



- 1994 First public presentation of the Semantic Web idea
- 1998 Start of standardization of data model (RDF) and a first ontology languages (RDFS) at W3C
- 2000 Start of large research projects about ontologies in the US and Europe
- 2002 Start of standardization of a new ontology language (OWL) based on research results
- 2004 Finalization of the standard for data (RDF) and ontology (OWL)
- 2008 Standardization of a query language (SPARQL)
- 2009 Extension of OWL to OWL 2.0
- 2010 Standard Rule Interchange Format (RIF)
- 2012 RDB2RDF Mapping Language (R2RML and Direct Mapping)
- 2013 SPARQL 1.1
- 2014 RDF 1.1

Google Knowledge Graph

“Introducing the Knowledge Graph: things, not strings” (2012)



About 163,000,000 results (0.29 seconds)

Despicable Me 2 showtimes for San Francisco, CA
See showtimes for 3D
1hr 38min - Rated PG - Animation
In summer 2013, get ready for more Minion madness in Despicable Me 2. Chris Meledandri and his acclaimed filmmaking team ...
AMC Van Ness 14 - 1000 Van Ness Avenue, San Francisco, CA - Map
11:25am - 2:05 - 4:55 - 7:40 - 10:30pm
Century San Francisco Centre 9 and XD - 835 Market St., San Francisco, CA - Map
7:00 - 9:25pm
+ Show more theaters

Despicable Me 2
despicableme.com/
ageous
Me 2 movie

A short description of the movie, ratings, release date, directors, cast, etc.

★★★★☆ Rating: 7.8/10 - 51,274 votes
Directed by Pierre Louis Padang Coffin, Chris Renaud. With Steve Carell, Kristen Wiig, Benjamin Bratt, Miranda Cosgrove. Gru is recruited by the Anti-Villain ...
Release Info - Full cast and crew - Videos - Version 3

Despicable Me 2 - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Despicable_Me_2**
Despicable Me 2 is a 2013 American 3D computer-animated comedy film and the sequel to the 2010 animated film Despicable Me. Produced by Illumination ...
Minions (film) - Despicable Me (franchise) - Annecy International Animated ...

Despicable Me 2 - Official Trailer #3 (HD) Steve Carell - YouTube
www.youtube.com/watch?v=HwXbtZXjbVE
Mar 19, 2013 - Uploaded by joblomovienetwork
http://www.joblo.com - "**Despicable Me 2** - Official Trailer #3" Universal Pictures and Illumination Entertainment ...
2:32

Despicable Me 2 - Rotten Tomatoes
www.rottentomatoes.com/m/**despicable_me_2**/
★★★★☆ Rating: 75% - 102 reviews
Review: It may not be as inspired as its predecessor, but **Despicable Me 2** offers plenty of eye-popping visual inventiveness and a number of big...

News for **despicable me 2**
NBCUniversal CEO: '**Despicable Me 2**' Will Be Most Profitable Film in Universal's History

Despicable Me 2
192,648 followers on Google+
★★★★☆ 7.8/10 - IMDb
★★★★☆ 75% - Rotten Tomatoes

Despicable Me 2 is a 2013 American 3D computer-animated comedy film and the sequel to the 2010 animated film Despicable Me.
Wikipedia

Release date: July 3, 2013 (USA)
Directors: Pierre Coffin, Chris Renaud
Language: English
Production company: Illumination Entertainment
Music composed by: Pharrell Williams, Heitor Pereira

Recent posts

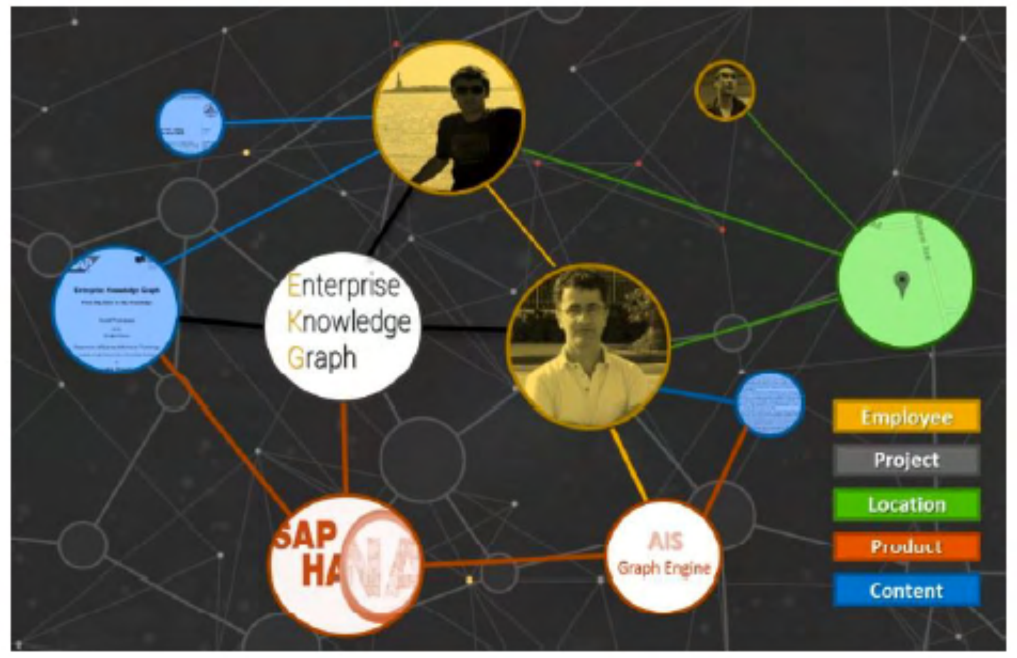
Voting closes soon for the Evil Laugh Contest. Make sure you get your votes in or else... MUAHAHAHA! http://www.evillaughlab.com/
Jul 24, 2013

Cast

Steve Carell – Gru
Kristen Wiig – Lucy Wilde
Miranda Cosgrove – Margo
Russell Brand – Dr. Nefario
Steve Coogan – Silas

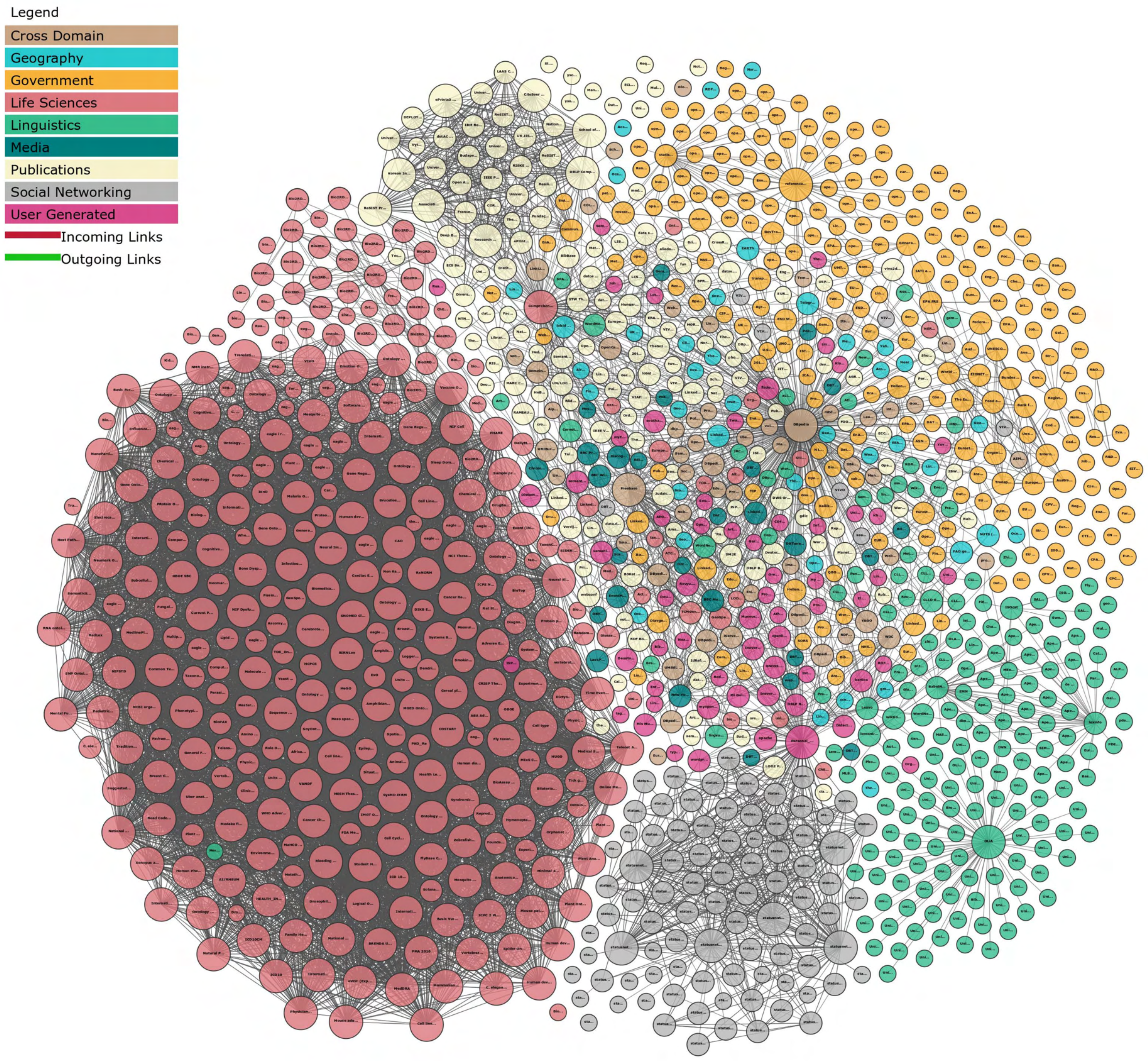
People also search for

Despicable Me 2010
Monsters University 2013
The Lone Ranger 2013
Man of Steel 2013
The Smurfs 2 2013



“Knowledge Graph” became a popular term for Semantic Web data.

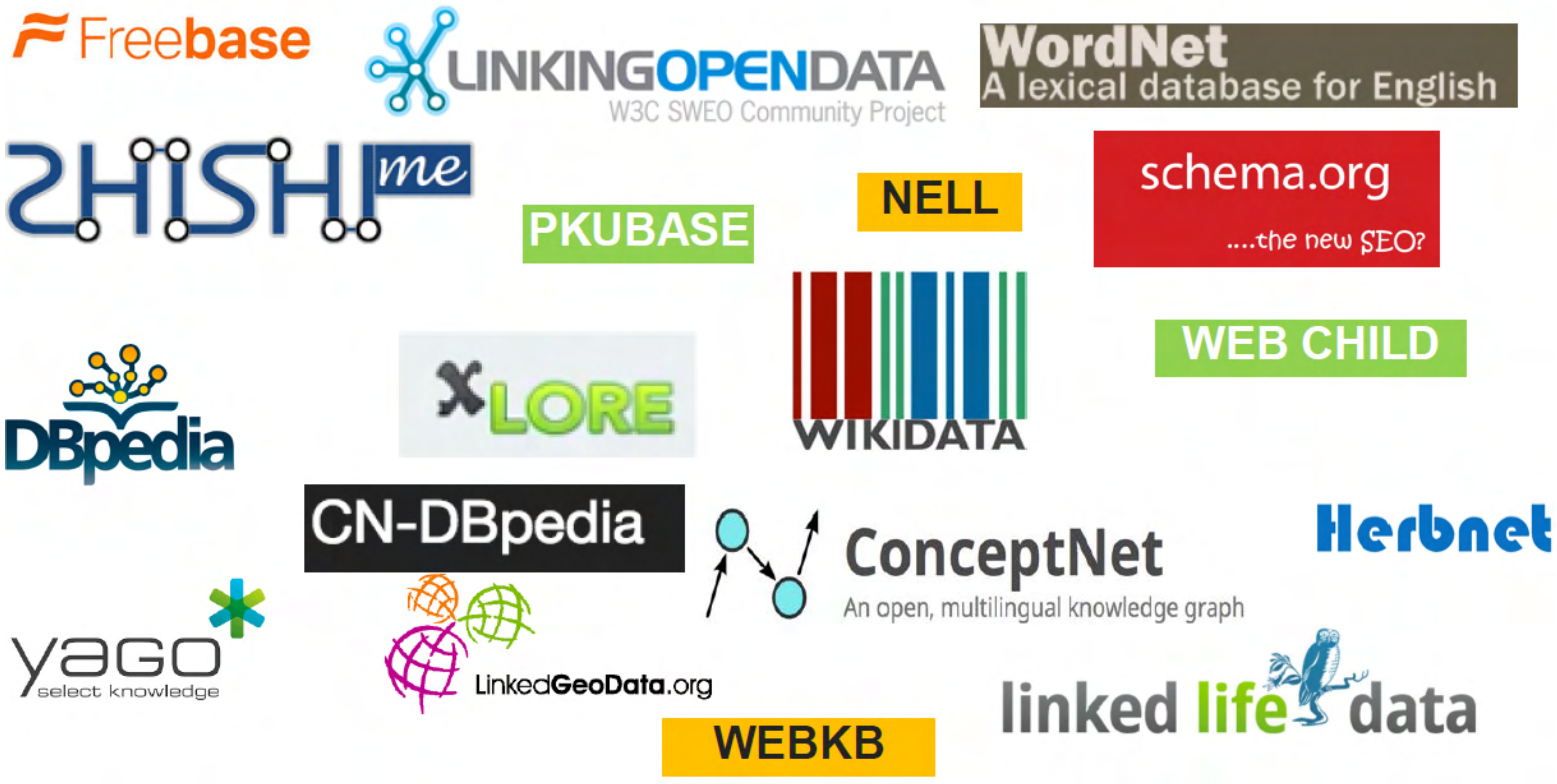
Linked Data (Open Linked KGs) in 2017



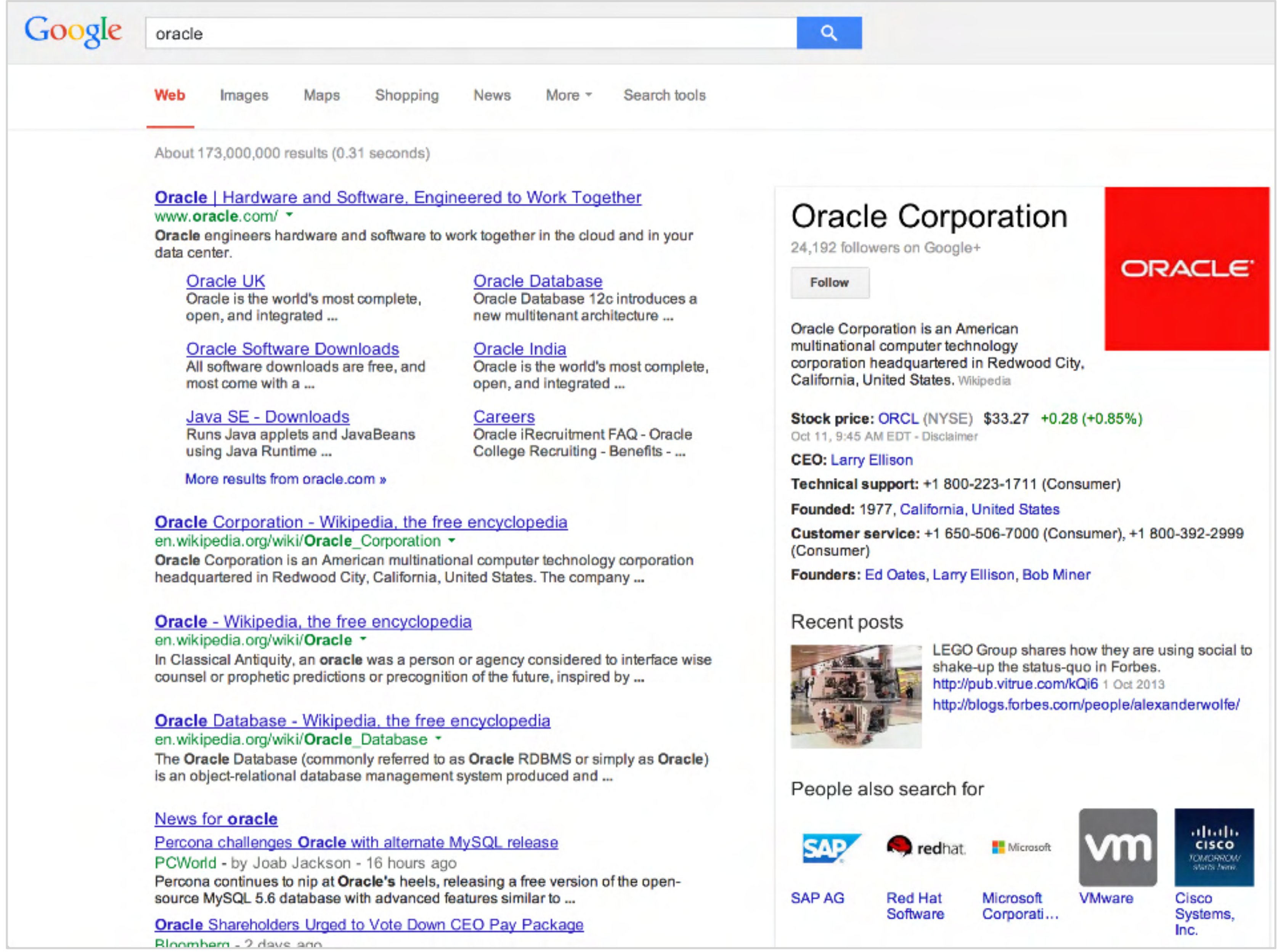
1163 datasets,
40 billion RDF triples

"Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/"

Typical Knowledge Graph Projects



General Applications: Semantic Search



About 173,000,000 results (0.31 seconds)

Oracle | Hardware and Software, Engineered to Work Together
www.**oracle**.com/
Oracle engineers hardware and software to work together in the cloud and in your data center.

Oracle UK
Oracle is the world's most complete, open, and integrated ...

Oracle Database
Oracle Database 12c introduces a new multitenant architecture ...

Oracle Software Downloads
All software downloads are free, and most come with a ...

Oracle India
Oracle is the world's most complete, open, and integrated ...

Java SE - Downloads
Runs Java applets and JavaBeans using Java Runtime ...

Careers
Oracle iRecruitment FAQ - Oracle College Recruiting - Benefits - ...

More results from oracle.com »

Oracle Corporation - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Oracle**_Corporation
Oracle Corporation is an American multinational computer technology corporation headquartered in Redwood City, California, United States. The company ...

Oracle - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Oracle**
In Classical Antiquity, an **oracle** was a person or agency considered to interface wise counsel or prophetic predictions or precognition of the future, inspired by ...

Oracle Database - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Oracle**_Database
The **Oracle** Database (commonly referred to as **Oracle** RDBMS or simply as **Oracle**) is an object-relational database management system produced and ...

News for **oracle**
Percona challenges **Oracle** with alternate MySQL release
PCWorld - by Joab Jackson - 16 hours ago
Percona continues to nip at **Oracle's** heels, releasing a free version of the open-source MySQL 5.6 database with advanced features similar to ...
Oracle Shareholders Urged to Vote Down CEO Pay Package
Bloomberg - 2 days ago

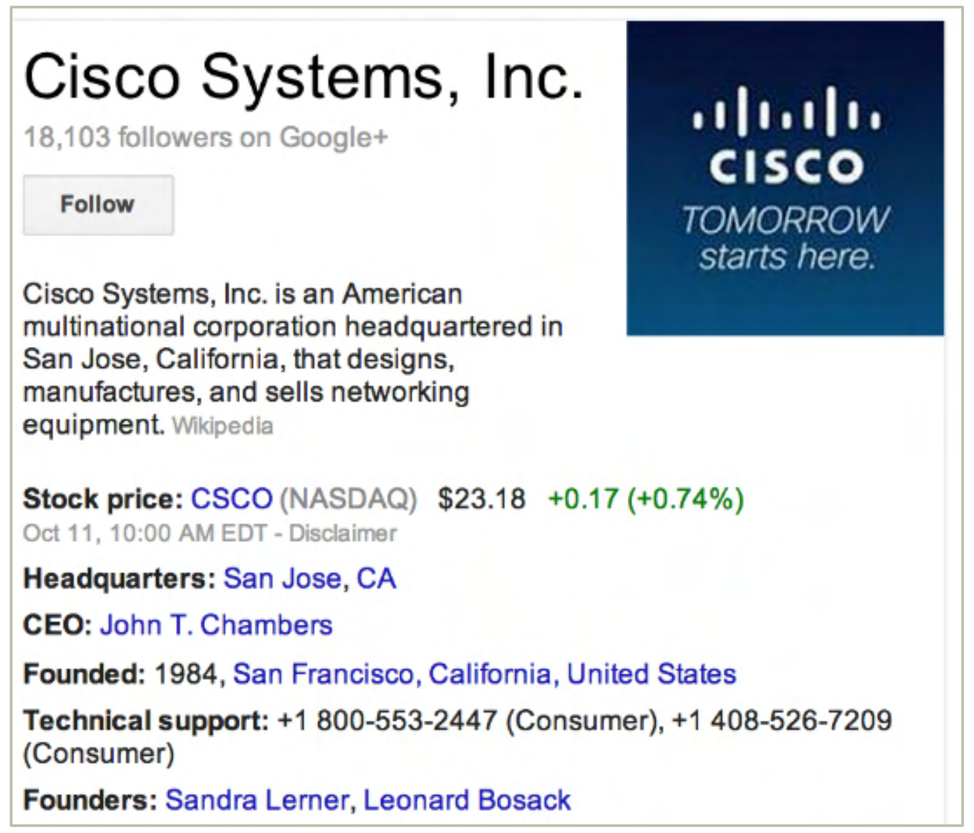
Oracle Corporation
24,192 followers on Google+
Follow

Oracle Corporation is an American multinational computer technology corporation headquartered in Redwood City, California, United States. Wikipedia

Stock price: ORCL (NYSE) $33.27 +0.28 (+0.85%)
Oct 11, 9:45 AM EDT - Disclaimer
CEO: Larry Ellison
Technical support: +1 800-223-1711 (Consumer)
Founded: 1977, California, United States
Customer service: +1 650-506-7000 (Consumer), +1 800-392-2999 (Consumer)
Founders: Ed Oates, Larry Ellison, Bob Miner

Recent posts
LEGO Group shares how they are using social to shake-up the status-quo in Forbes. http://pub.vitrue.com/kQi6 1 Oct 2013
http://blogs.forbes.com/people/alexanderwolfe/

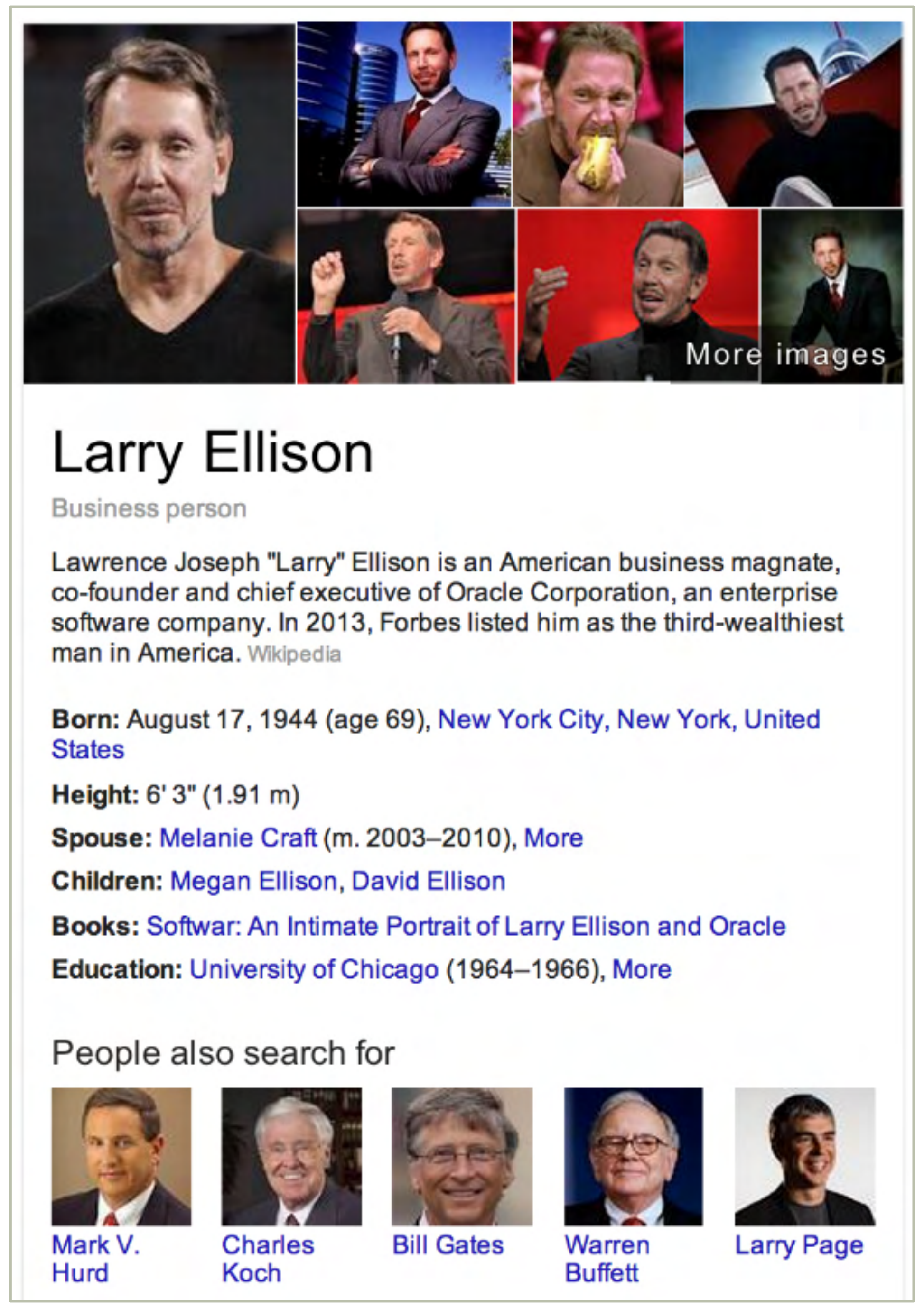
People also search for
SAP AG
Red Hat Software
Microsoft Corporati...
VMware
Cisco Systems, Inc.



Cisco Systems, Inc.
18,103 followers on Google+
Follow

Cisco Systems, Inc. is an American multinational corporation headquartered in San Jose, California, that designs, manufactures, and sells networking equipment. Wikipedia

Stock price: CSCO (NASDAQ) $23.18 +0.17 (+0.74%)
Oct 11, 10:00 AM EDT - Disclaimer
Headquarters: San Jose, CA
CEO: John T. Chambers
Founded: 1984, San Francisco, California, United States
Technical support: +1 800-553-2447 (Consumer), +1 408-526-7209 (Consumer)
Founders: Sandra Lerner, Leonard Bosack

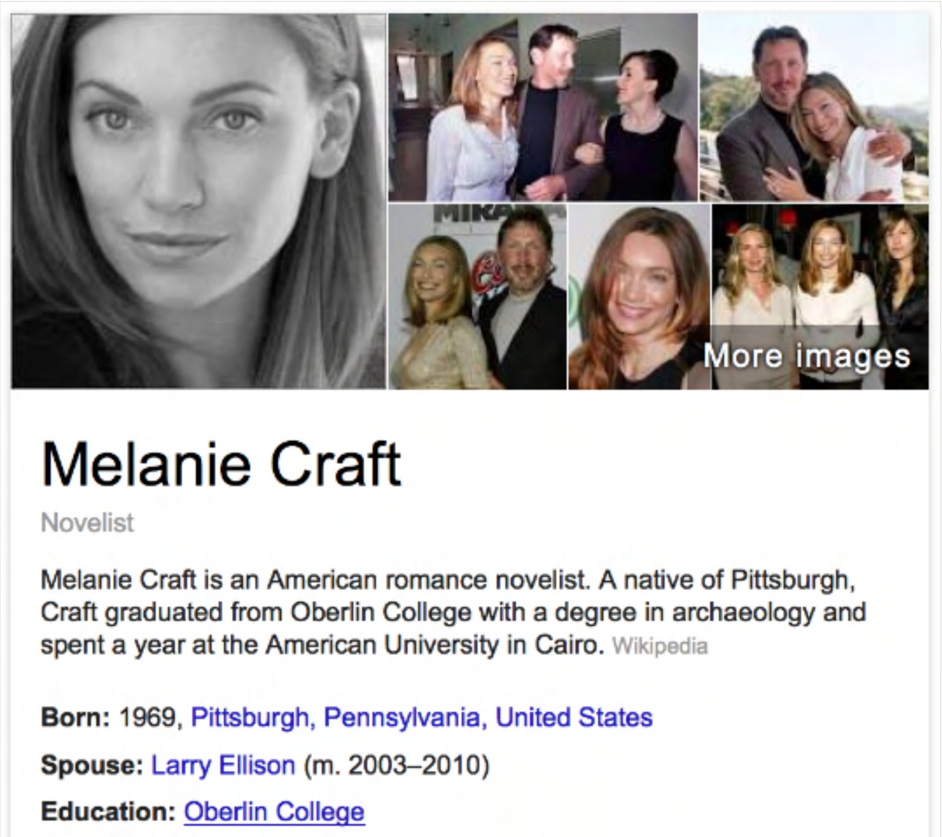


Larry Ellison
Business person

Lawrence Joseph "Larry" Ellison is an American business magnate, co-founder and chief executive of Oracle Corporation, an enterprise software company. In 2013, Forbes listed him as the third-wealthiest man in America. Wikipedia

Born: August 17, 1944 (age 69), New York City, New York, United States
Height: 6' 3" (1.91 m)
Spouse: Melanie Craft (m. 2003–2010), More
Children: Megan Ellison, David Ellison
Books: Softwar: An Intimate Portrait of Larry Ellison and Oracle
Education: University of Chicago (1964–1966), More

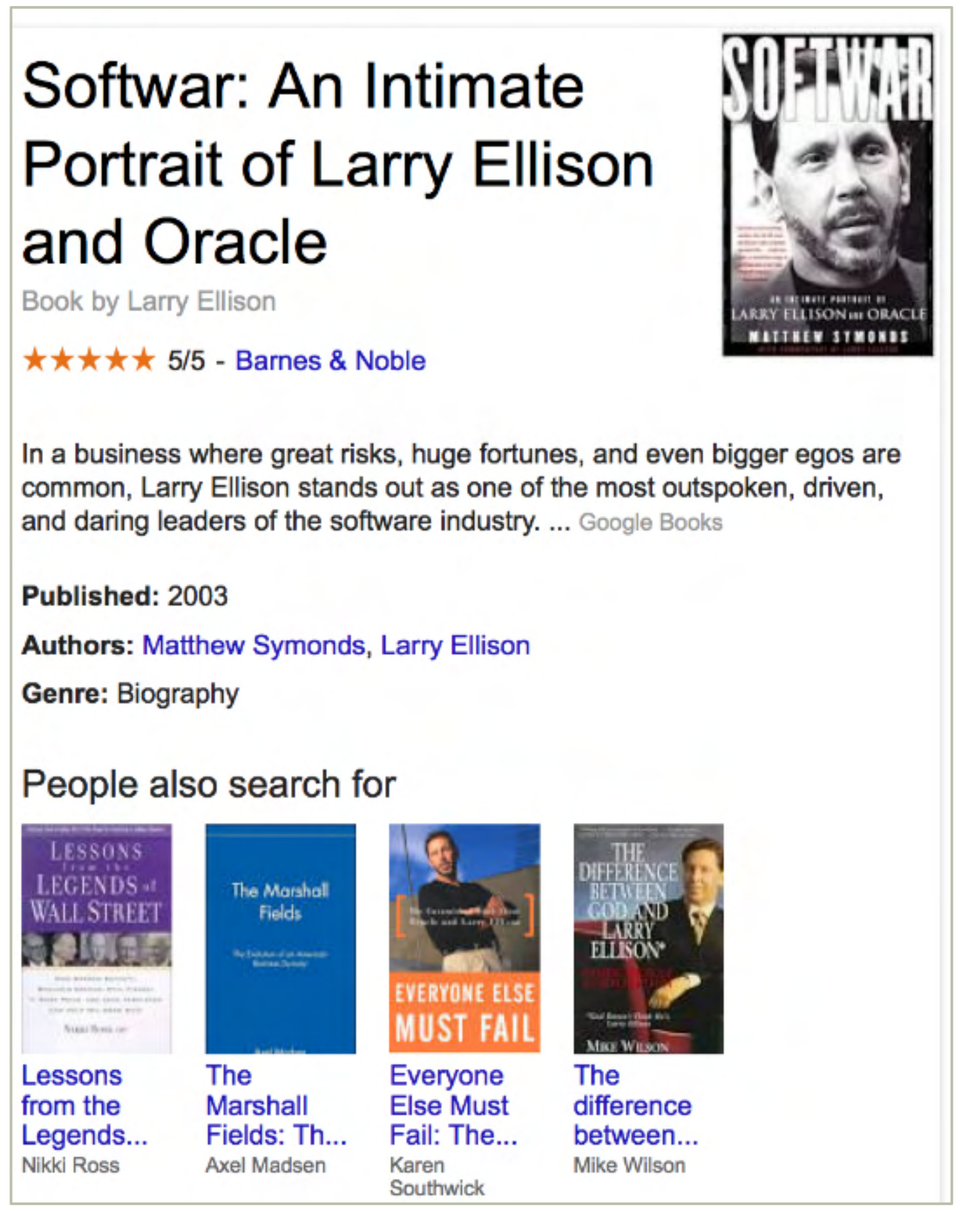
People also search for
Mark V. Hurd
Charles Koch
Bill Gates
Warren Buffett
Larry Page



Melanie Craft
Novelist

Melanie Craft is an American romance novelist. A native of Pittsburgh, Craft graduated from Oberlin College with a degree in archaeology and spent a year at the American University in Cairo. Wikipedia

Born: 1969, Pittsburgh, Pennsylvania, United States
Spouse: Larry Ellison (m. 2003–2010)
Education: Oberlin College



Softwar: An Intimate Portrait of Larry Ellison and Oracle
Book by Larry Ellison

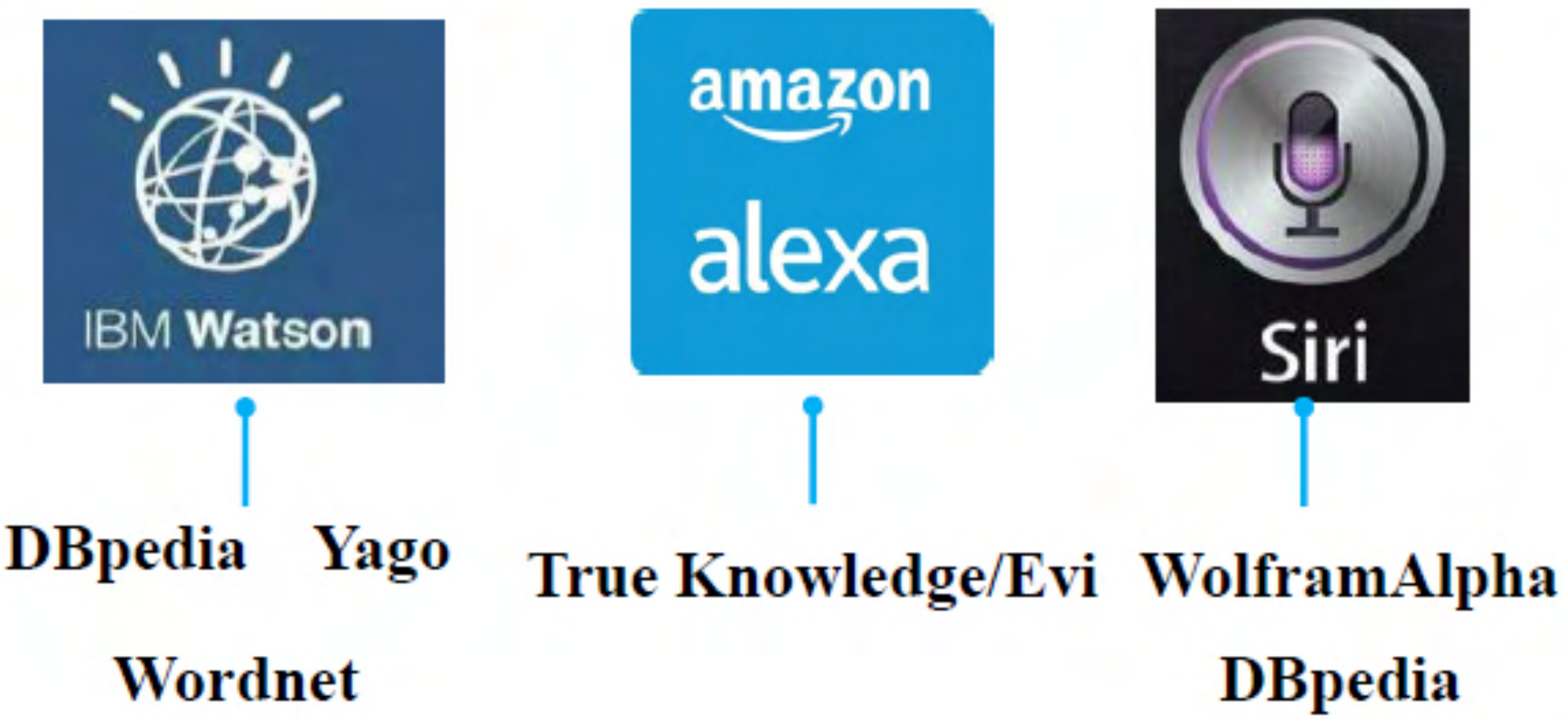
★★★★★ 5/5 - Barnes & Noble

In a business where great risks, huge fortunes, and even bigger egos are common, Larry Ellison stands out as one of the most outspoken, driven, and daring leaders of the software industry. ... Google Books

Published: 2003
Authors: Matthew Symonds, Larry Ellison
Genre: Biography

People also search for
Lessons from the Legends... Nikki Ross
The Marshall Fields: Th... Axel Madsen
Everyone Else Must Fail: The... Karen Southwick
The difference between... Mike Wilson

General Applications: Question Answering



DBpedia Yago

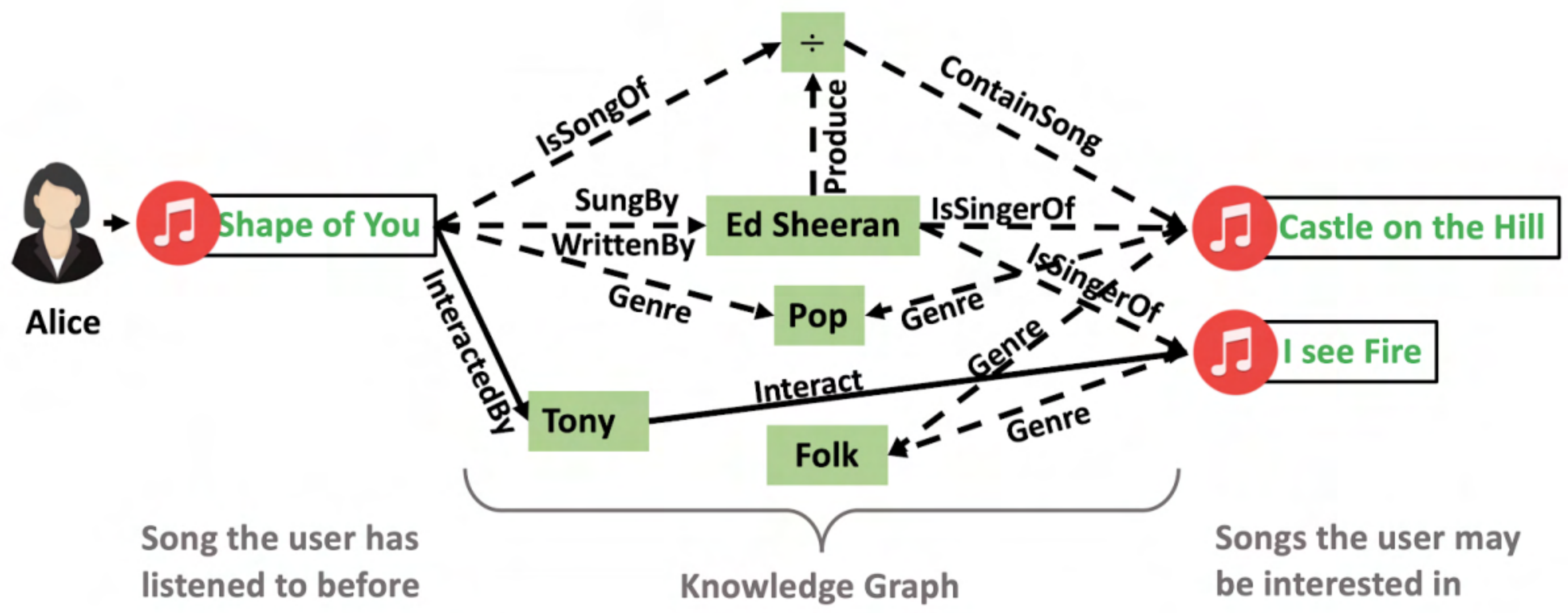
Wordnet

True Knowledge/Evi

WolframAlpha

DBpedia

General Application: Recommender system

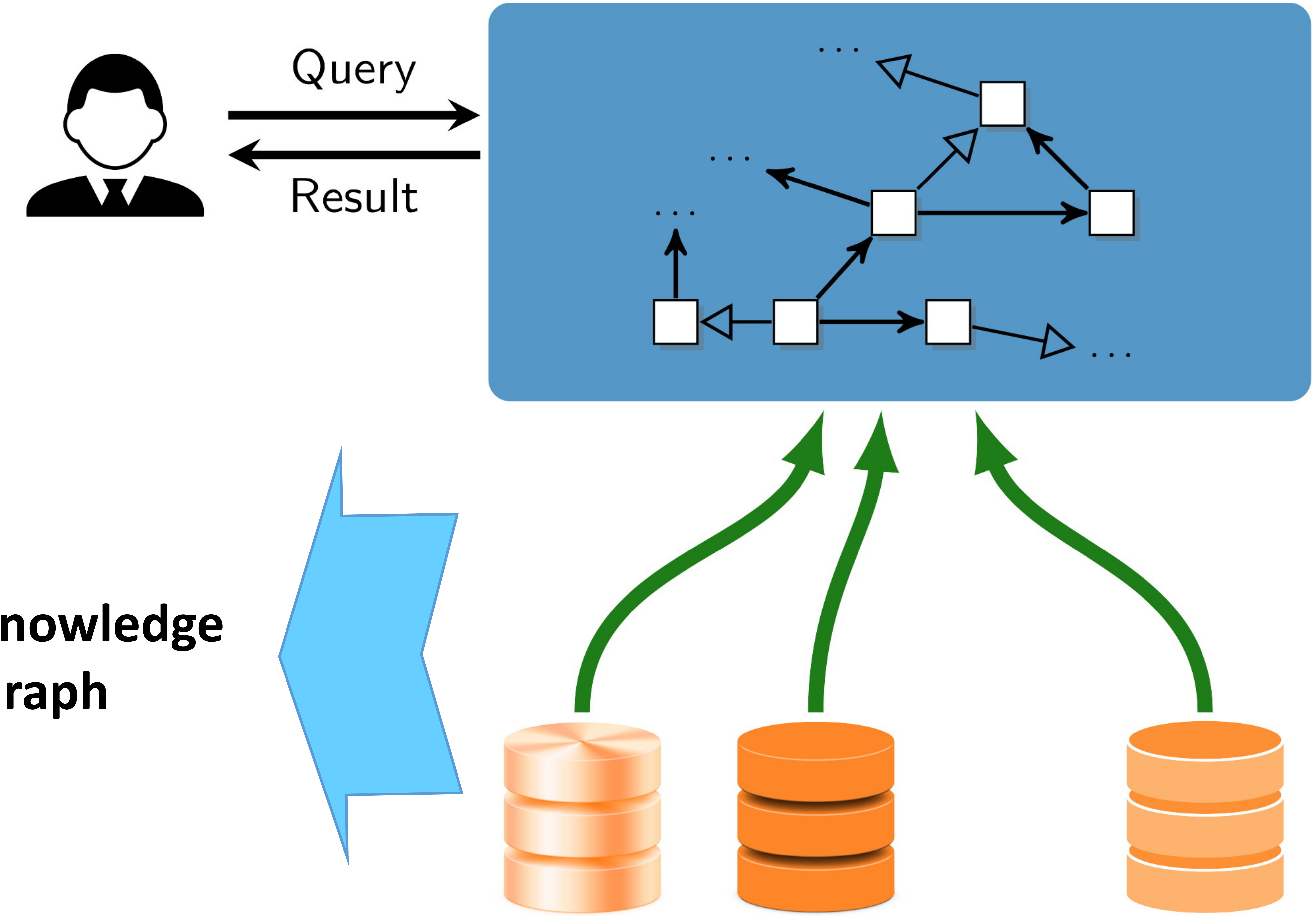


2. The Virtual Knowledge Graph Framework

Reality

- Knowledge Graphs are great, but how to get them?
- Many organization and companies are still relying on large relational databases to manage their internal data.
- It is not realistic to ask them completely migrate to knowledge graphs

Virtual Knowledge Graphs



Ontology
provides
global vocabulary
and
conceptual view

Mappings
how to populate
the ontology
from the data

Data Sources
external and
heterogeneous

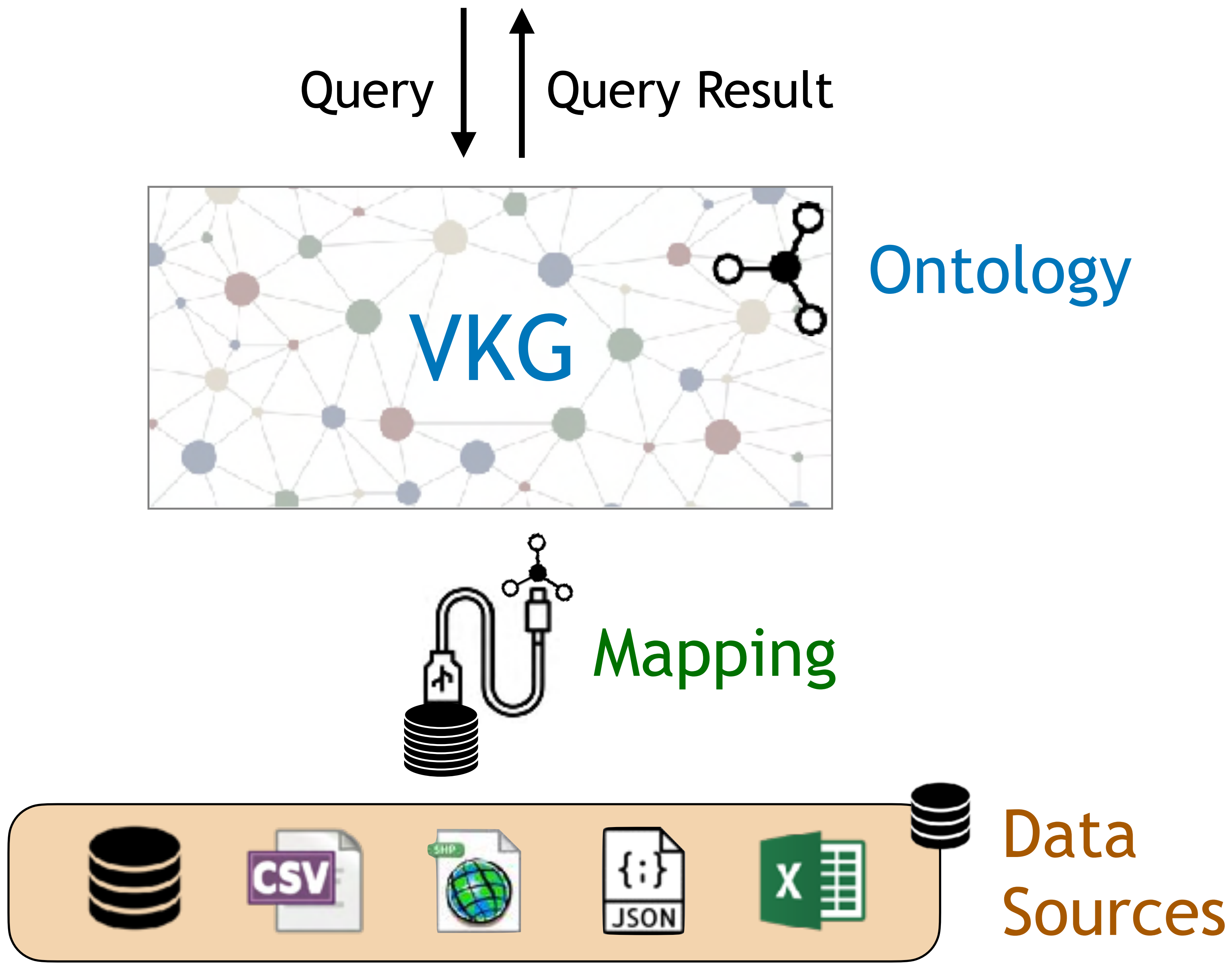
Why virtualization?

- The data stays in the sources and is only accessed at query time
- There is no need to construct a large and potentially costly materialization pipeline
- The data is always fresh wrt the latest updates at the sources
- One can rely on the existing data infrastructure and expertise
- Better supports an incremental approach to integration

Components of the VKG framework

- We consider now the main components that make up the VKG framework, and the languages used to specify them.

- The W3C has standardized languages suitable for VKGs:
- Knowledge graph: expressed in RDF [W3C Rec. 2014] (v1.1)
- Ontology O: expressed in OWL2QL [W3C Rec. 2012]
- MappingM: expressed in R2RML [W3C Rec. 2012]
- Query: expressed in SPARQL [W3C Rec. 2013] (v1.1)



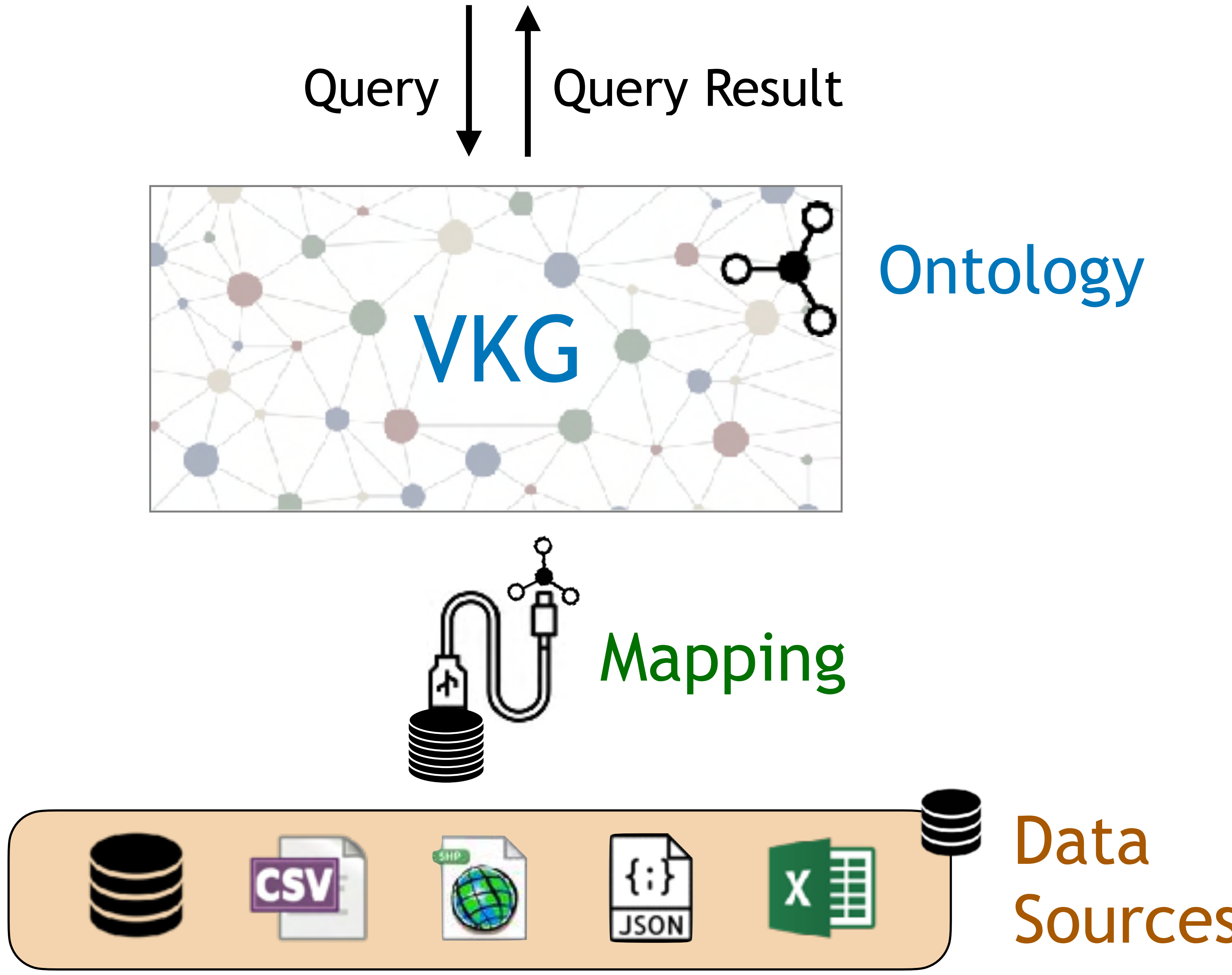
Use of mappings

In the VKG framework, the mapping encodes how the data in the sources should be used to create the Virtual Knowledge Graph, which is formulated in the vocabulary of the ontology.

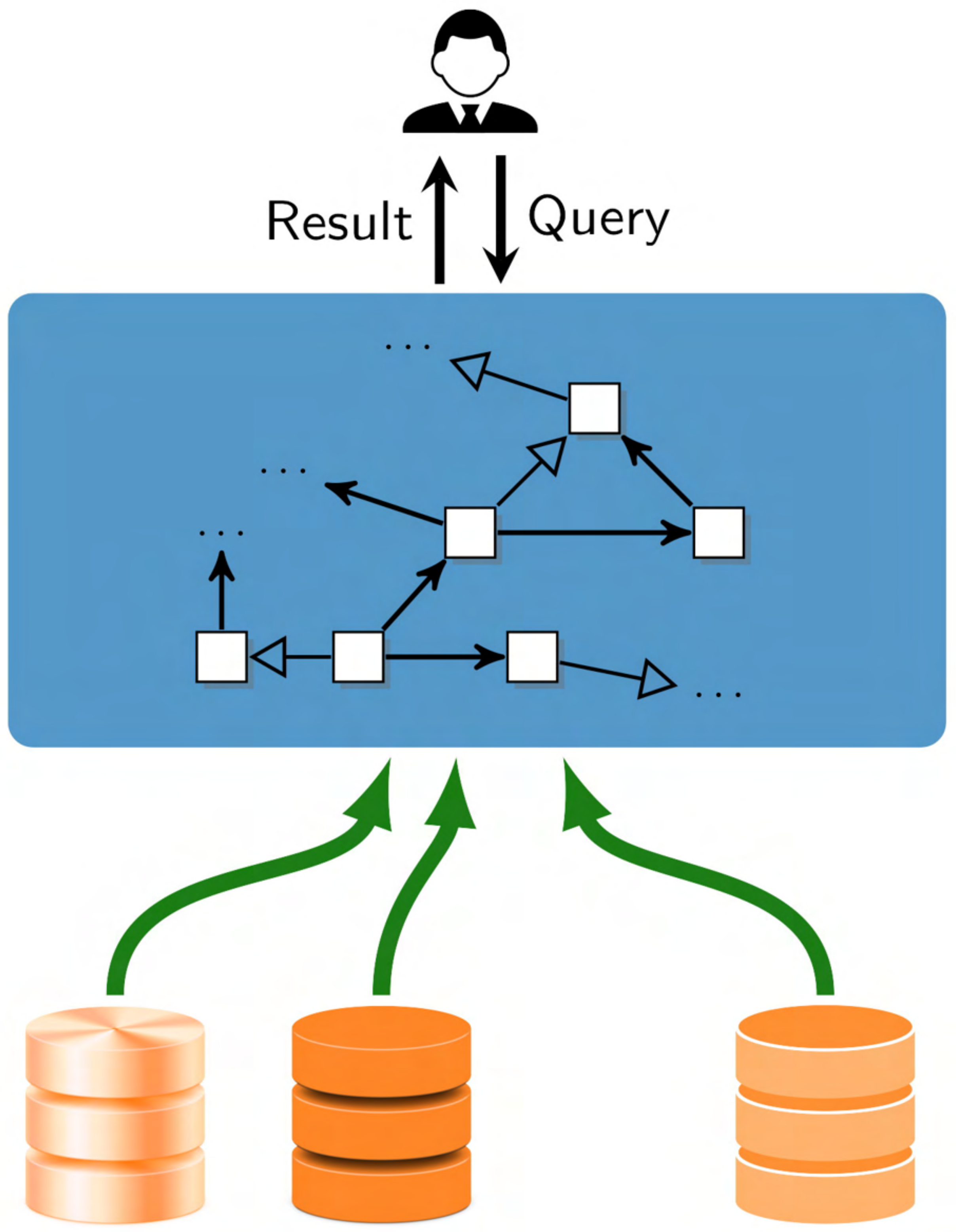
VKG defined from the mapping and the data.

- Queries are answered with respect to the ontology and the data of the VKG.
- The data of the VKG is not materialized (it is virtual!).
- Instead, the information in the ontology and the mapping is used to translate queries over the ontology into queries formulated over the sources.

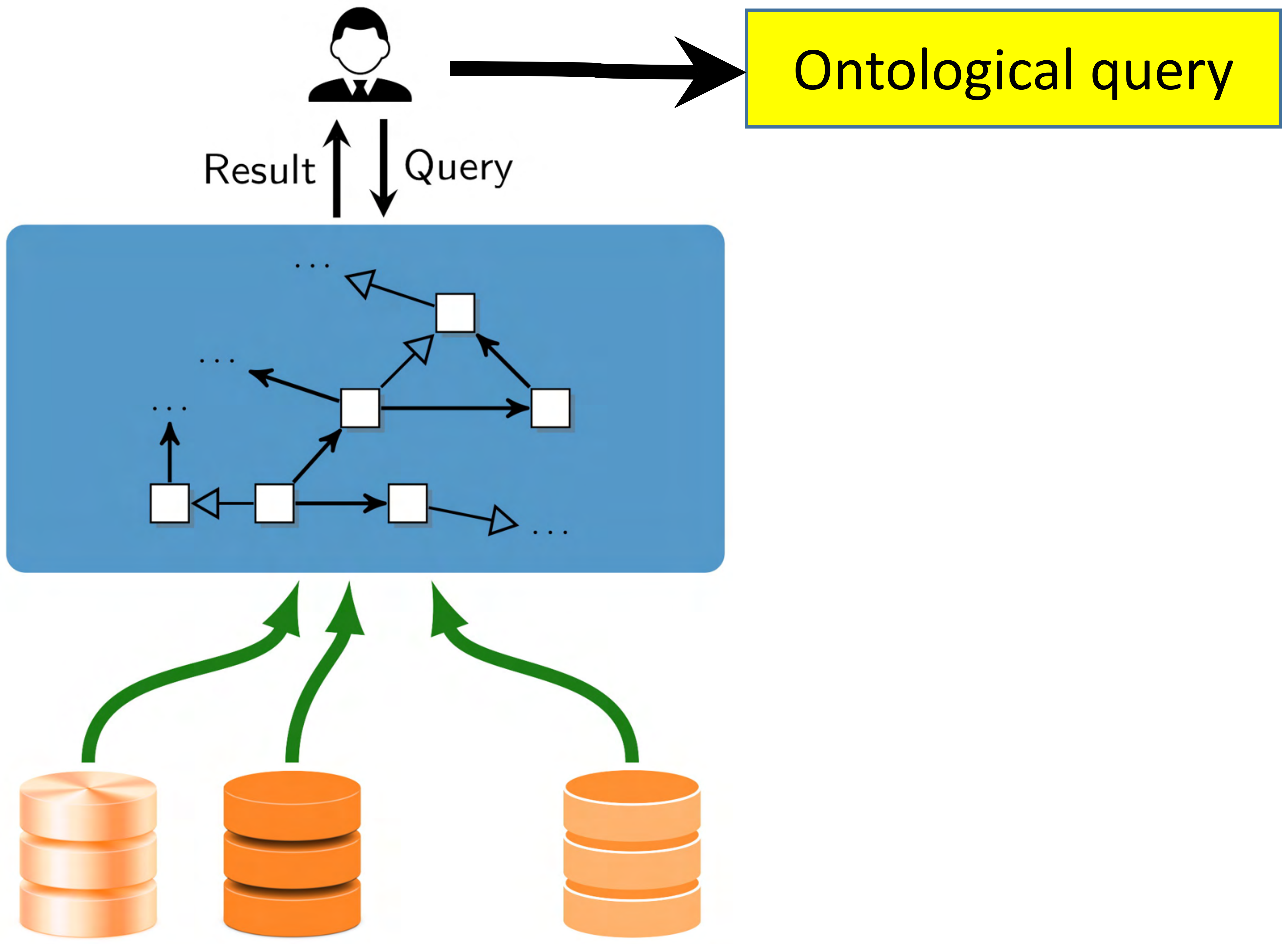
Note: The graph is always up to date wrt the data sources.



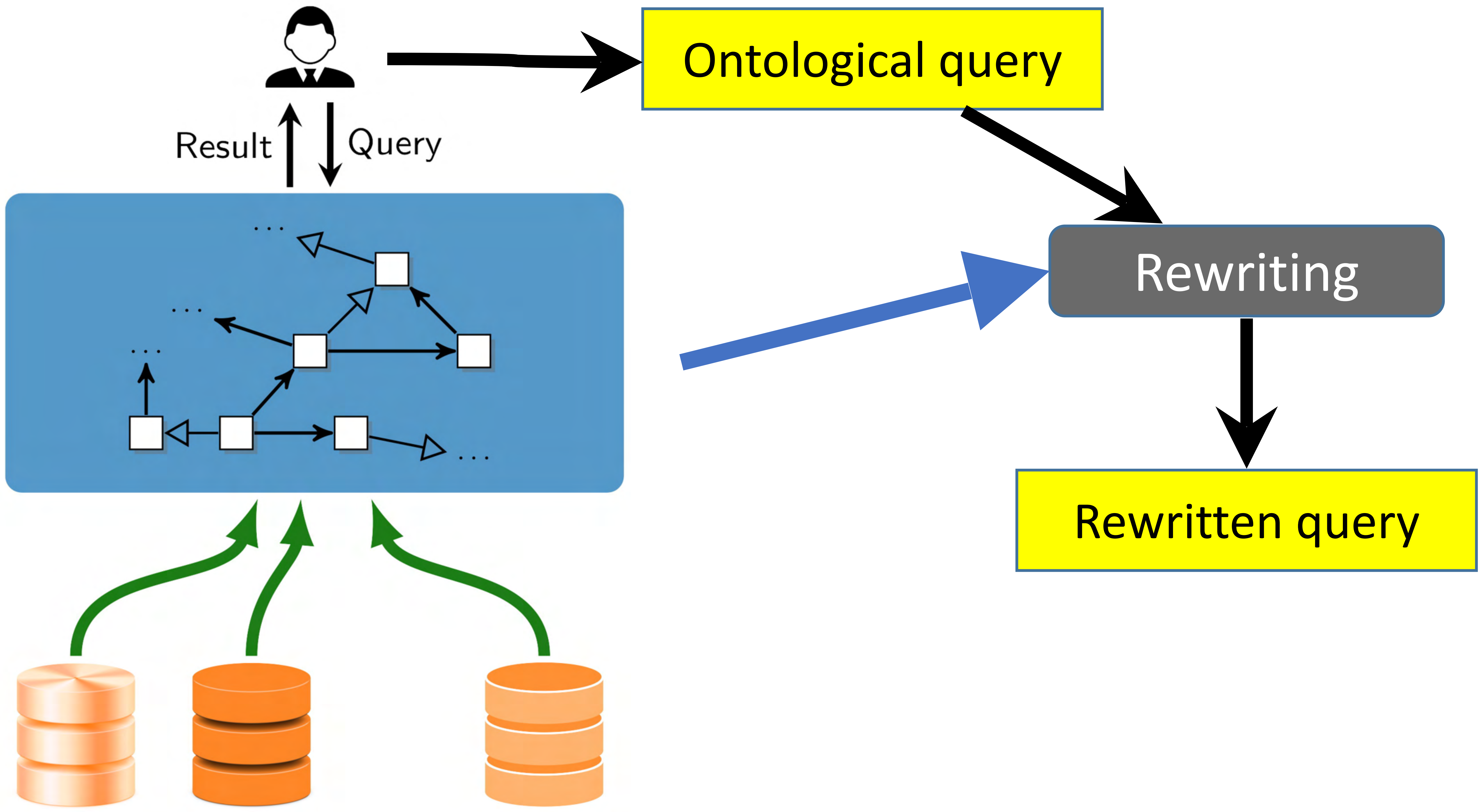
VKG Approach: Query answering by rewriting



VKG Approach: Query answering by rewriting



VKG Approach: Query answering by rewriting



Result

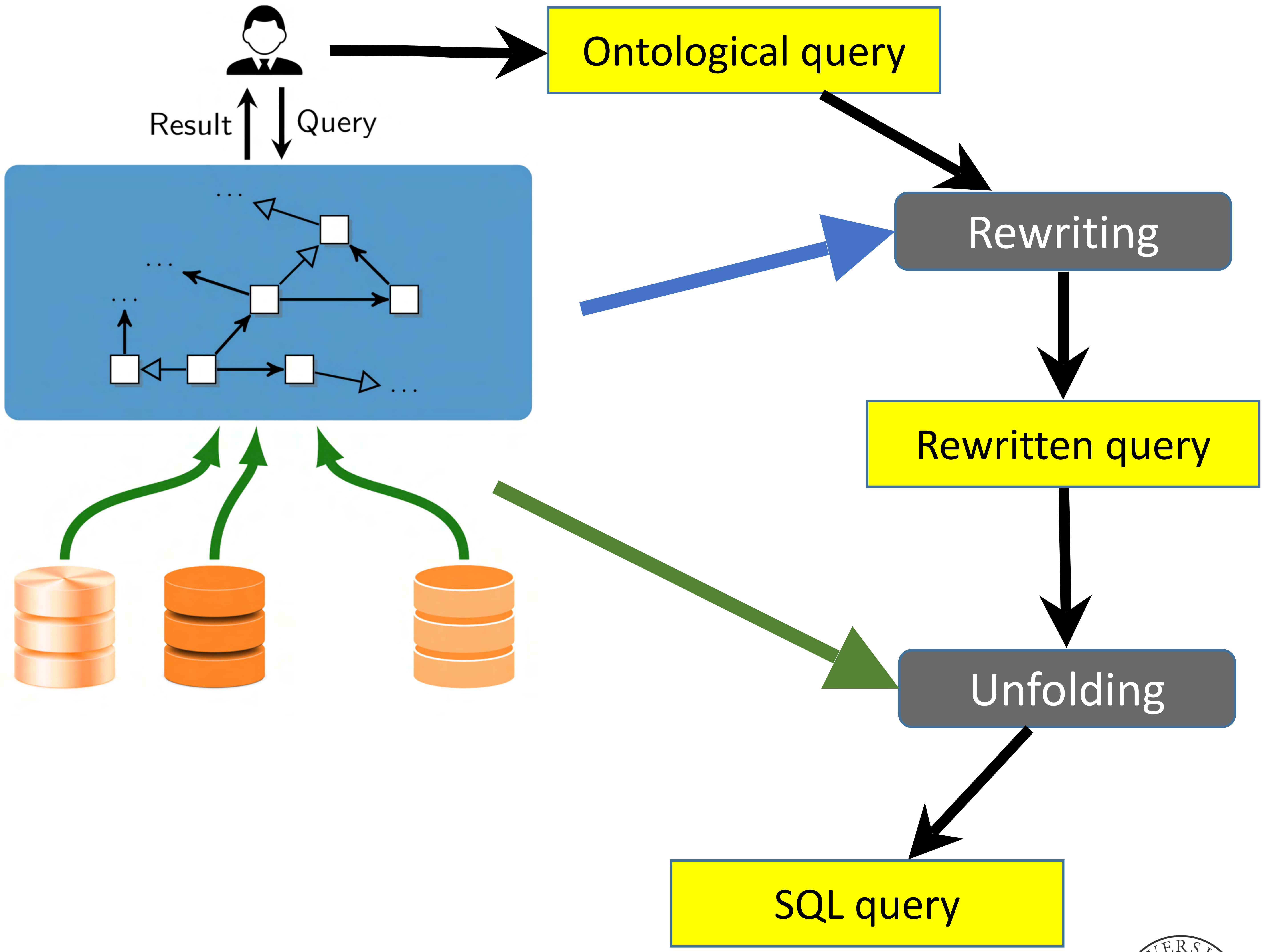
Query

Ontological query

Rewriting

Rewritten query

VKG Approach: Query answering by rewriting



Result

Query

Ontological query

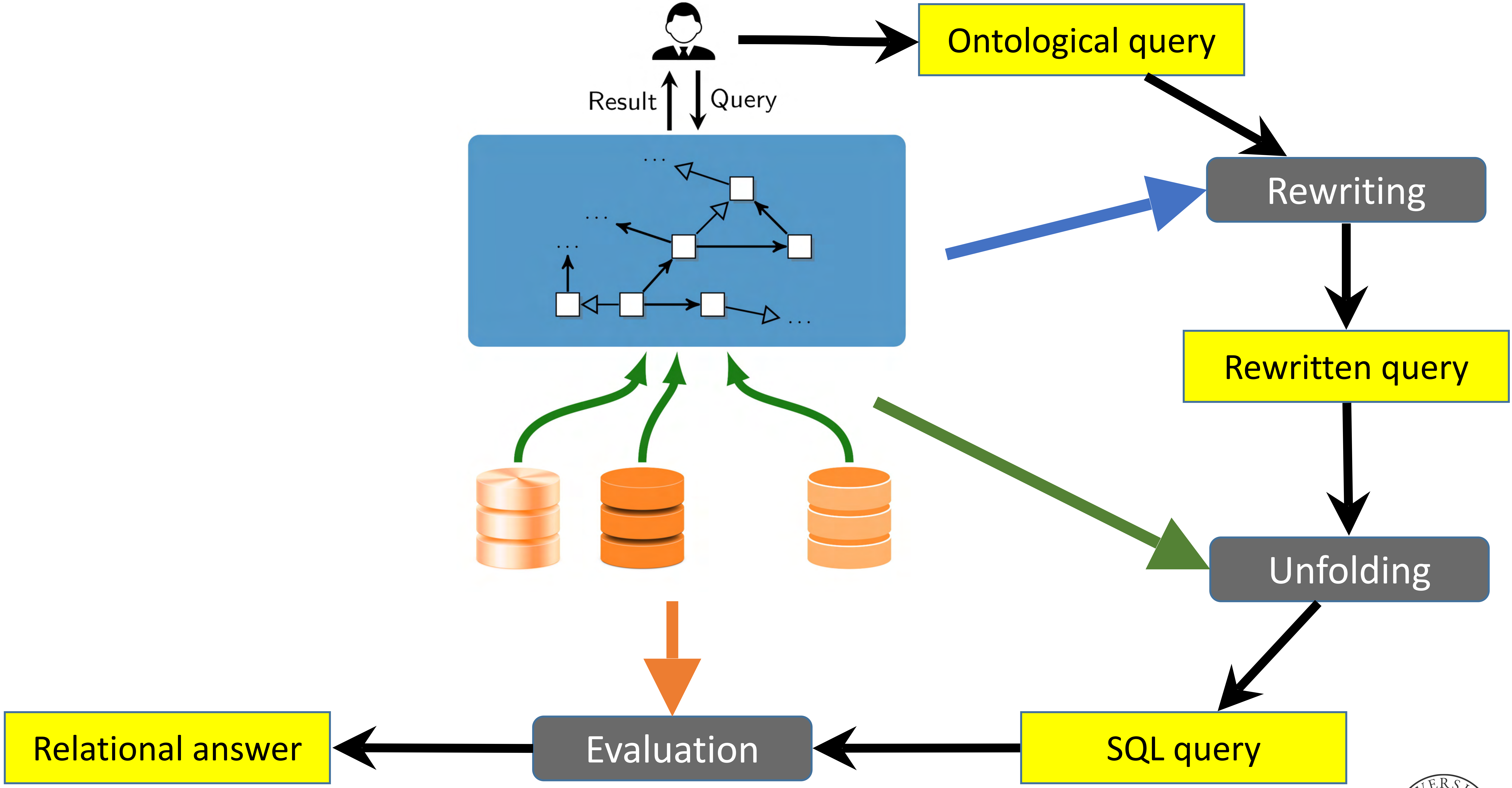
Rewriting

Rewritten query

Unfolding

SQL query

VKG Approach: Query answering by rewriting



Ontological query

Rewriting

Rewritten query

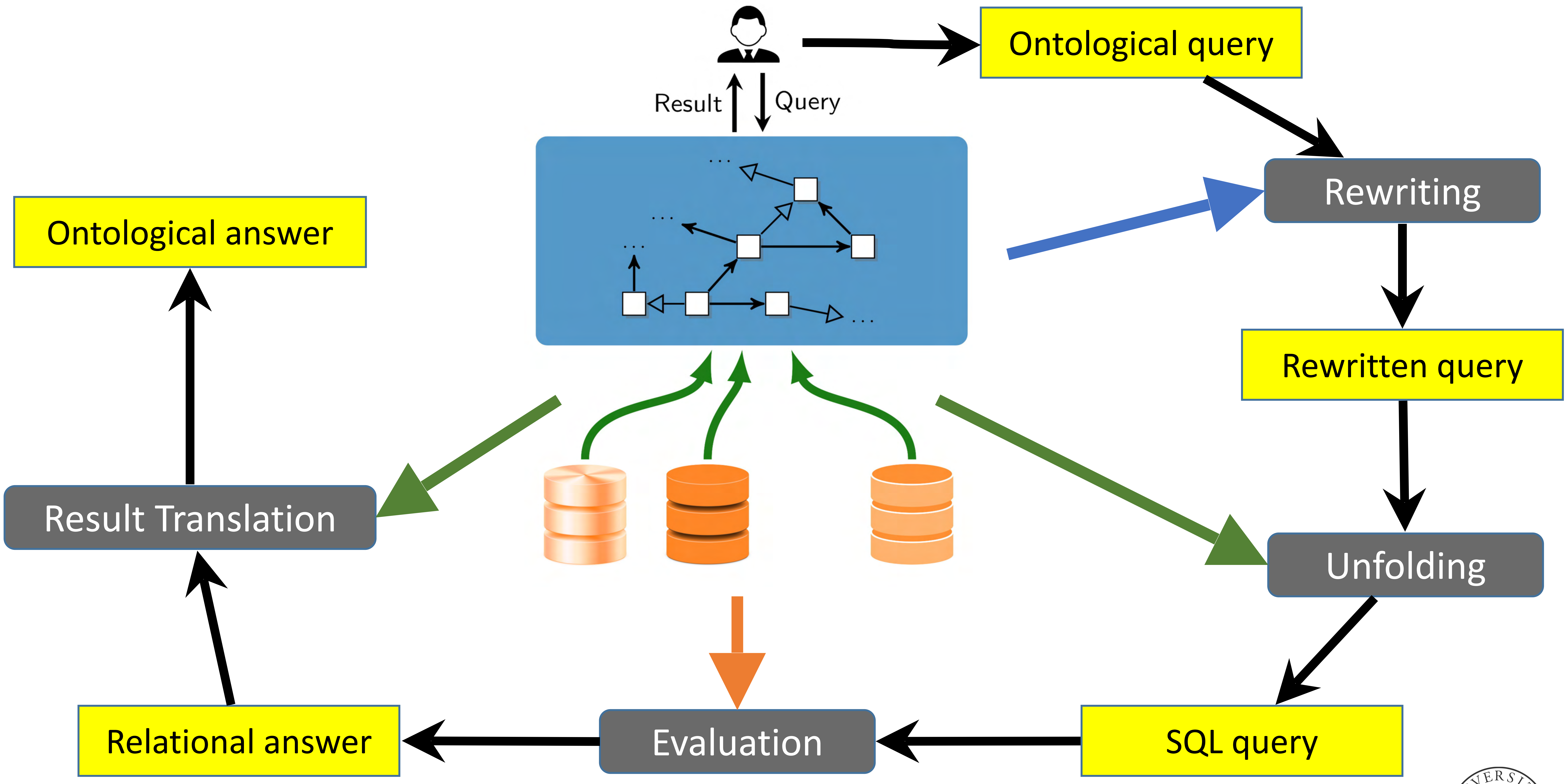
Unfolding

SQL query

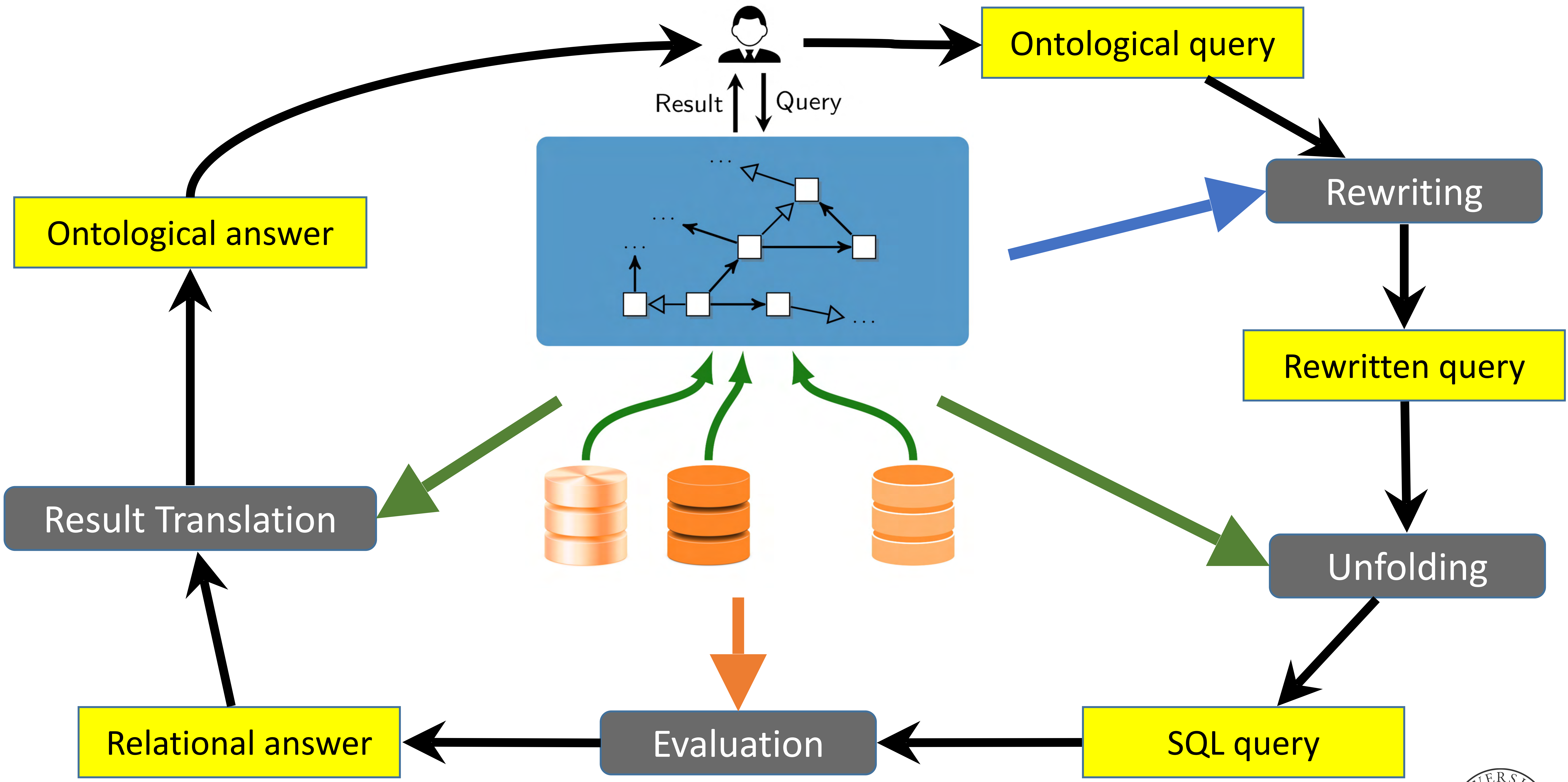
Evaluation

Relational answer

VKG Approach: Query answering by rewriting



VKG Approach: Query answering by rewriting



3. VKG Systems and Applications

VKG systems

- Mastro, Sapienza Università di Roma & OBDA systems SRL, Italy
- Morph, Technical University of Madrid, Spain
- Ontop, Free Uni. of Bozen-Bolzano, Italy & Ontopic SRL, Italy
- Stardog, Stardog Union, US
  - powered by Ontop in the first version, later own implementation
- GraphDB, OntoText, US & EU
  - powered by Ontop
- Allegro Graph, Franz Inc., US
  - powered by Ontop
- Oracle Spatial and Graph, US

The VKG system Ontop

ontop

https://ontop-vkg.org/

- State-of-the-art VKG system developed at the Free University of Bozen-Bolzano.
- Compliant with all relevant Semantic Web standards:

  RDF, RDFS, OWL 2 QL, R2RML, and SPARQL
- Supports all major relational DBs:

  Oracle, DB2, MS SQL Server, Postgres, MySQL, Teiid, Dremio, Denodo, etc.
- Open-source and released under Apache 2 license.
- Development: started in 2009
  - Already well established (e.g., ~5000 downloads / year, +200 members in mailing list)
  - Recently: major release of v4, with several improvements and new features.

Developer Community

unibz

ON TOP IC

Birkbeck UNIVERSITY OF LONDON

UNIVERSITAS BERGENSIS

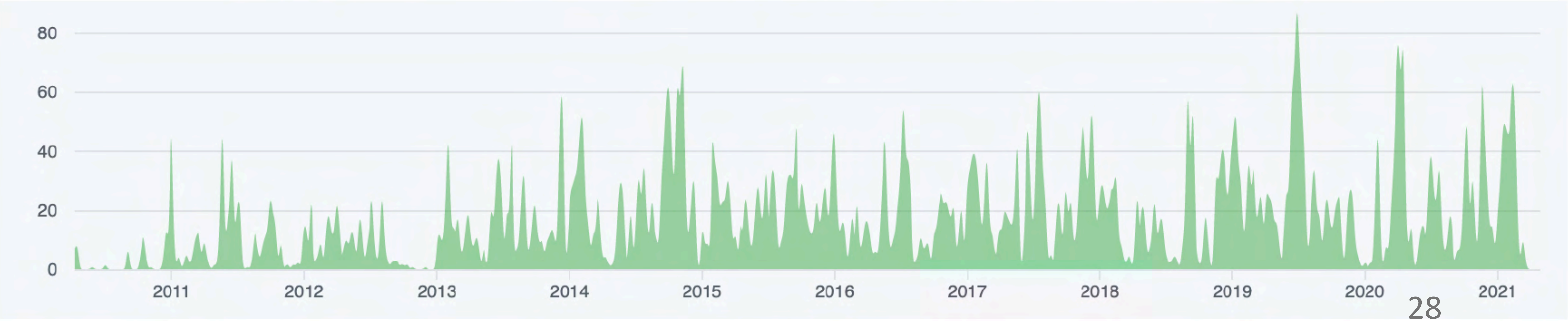
UiO University of Oslo

HELLENIC REPUBLIC
National and Kapodistrian
University of Athens

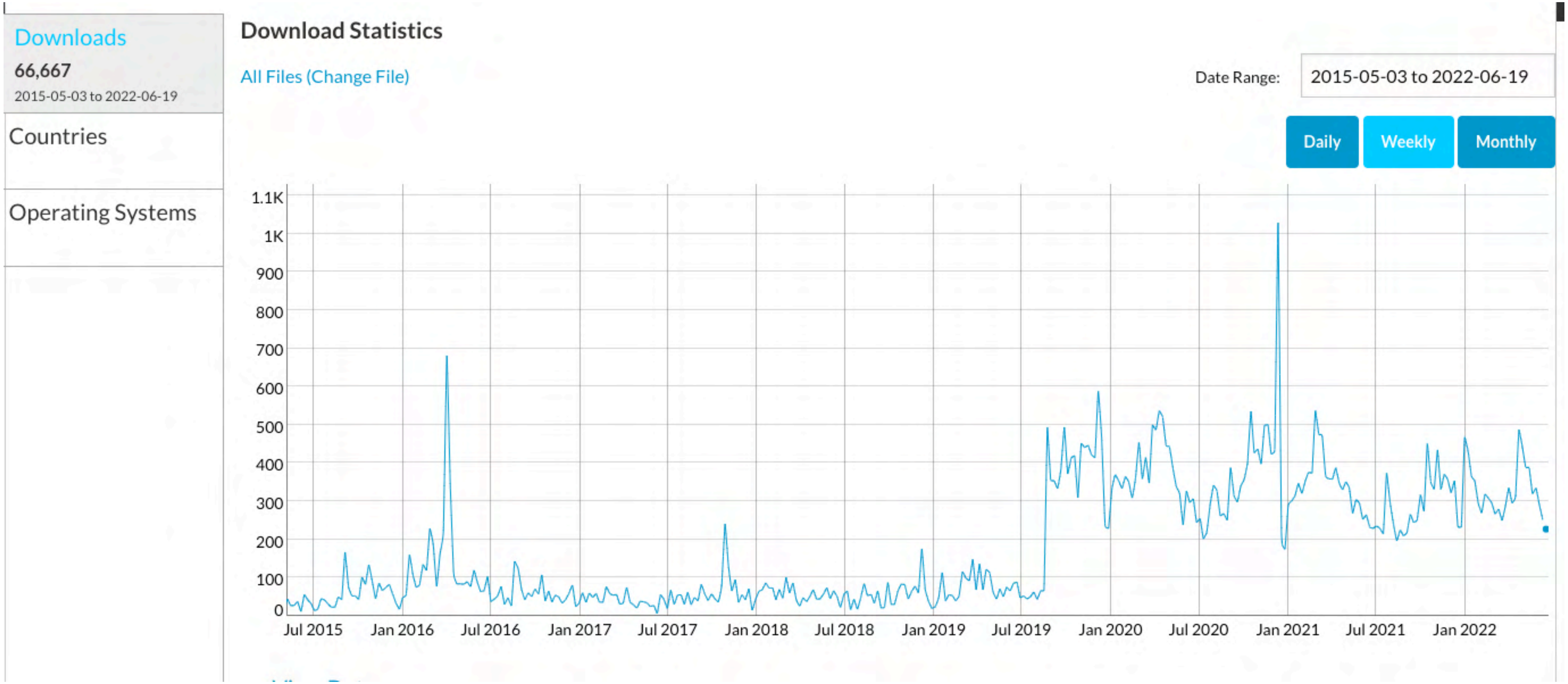
UNIVERSITÄT LEIPZIG

ontotext

POLITECNICO MILANO 1863



Ontop Downloads



Downloads

66,667

2015-05-03 to 2022-06-19

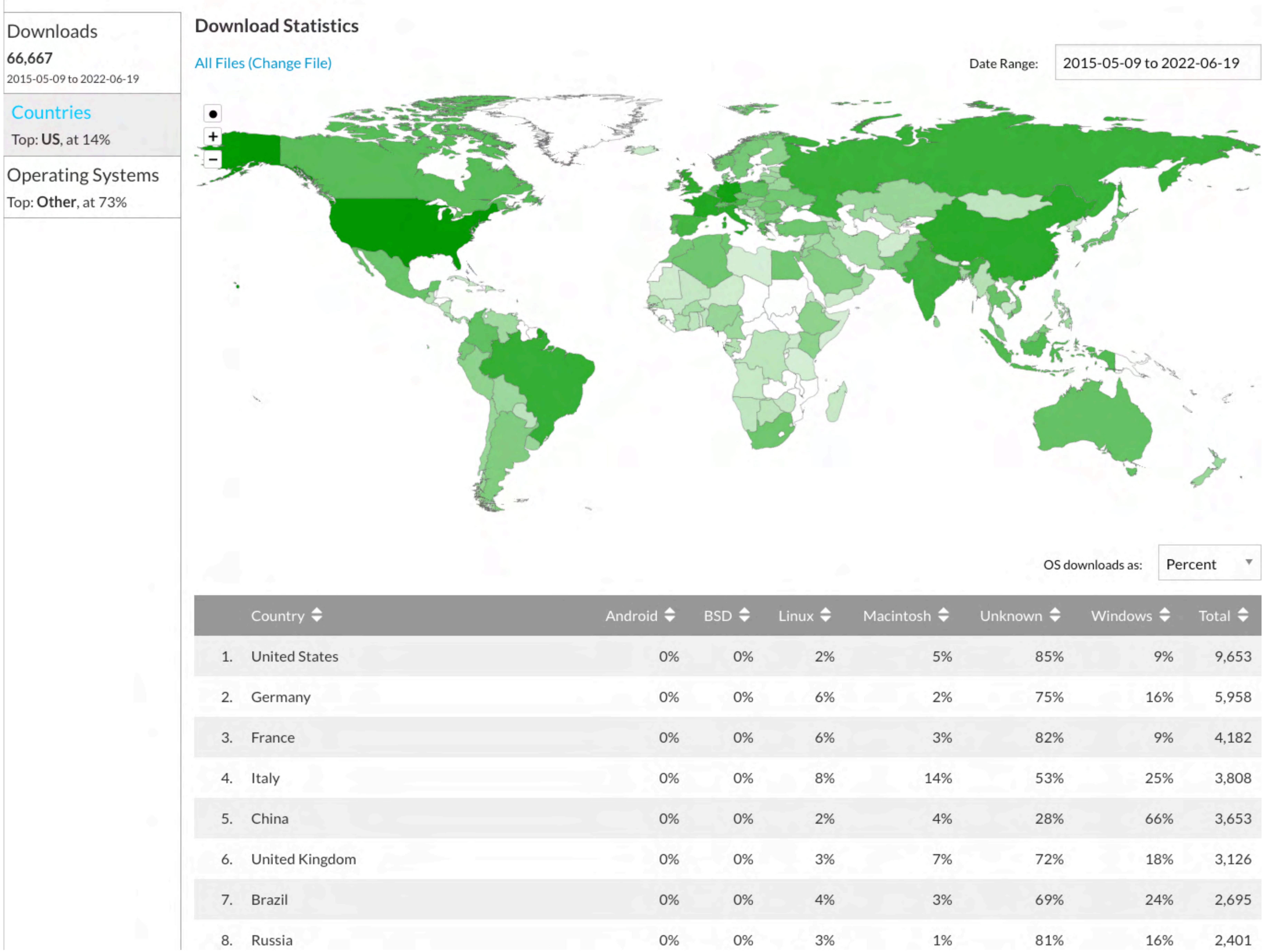
Countries

Operating Systems

Download Statistics

All Files (Change File)

Date Range: 2015-05-03 to 2022-06-19



Downloads
66,667
2015-05-09 to 2022-06-19

Countries
Top: **US**, at 14%

Operating Systems
Top: **Other**, at 73%

Download Statistics

All Files (Change File)

Date Range: 2015-05-09 to 2022-06-19

OS downloads as: Percent ▾

	Country ⇕	Android ⇕	BSD ⇕	Linux ⇕	Macintosh ⇕	Unknown ⇕	Windows ⇕	Total ⇕
1.	United States	0%	0%	2%	5%	85%	9%	9,653
2.	Germany	0%	0%	6%	2%	75%	16%	5,958
3.	France	0%	0%	6%	3%	82%	9%	4,182
4.	Italy	0%	0%	8%	14%	53%	25%	3,808
5.	China	0%	0%	2%	4%	28%	66%	3,653
6.	United Kingdom	0%	0%	3%	7%	72%	18%	3,126
7.	Brazil	0%	0%	4%	3%	69%	24%	2,695
8.	Russia	0%	0%	3%	1%	81%	16%	2,401

ONTOPIC

- https://ontopic.ai
- First spin-off of the Free University of Bozen-Bolzano.
- Incorporated in April 2019.
- Product: *Ontopic Studio* based on the Ontop engine.
- Delivers services around Ontop and the Ontopic Suite.
- Consultancy for VKG-based data integration projects.

Selected Use Cases

- Oil & Gas: Statoil [Kharlamov et al. 2017a]
- Turbine Diagnoses: Siemens [Kharlamov et al. 2017b]
- Log Extraction in Process Mining [Calvanese, Kalayci, et al. 2017]
- Cultural heritage [Calvanese, Liuzzo, et al. 2016]
- Maritime security [Brüggemann et al. 2016]
- Manufacturing [Petersen et al. 2017]
- Health care: electronic health records [Rahimi et al. 2014]
- Public debt: the Italian Ministry of Economy and Finance [Antonioli et al. 2014]
- Smart cities: IBM Ireland [Lopez et al. 2015]
- Knowledge resource management: Huawei & Ontopic
- and many more … (see our survey [Xiao et al . 2019])

VKG over LinStar in Huawei

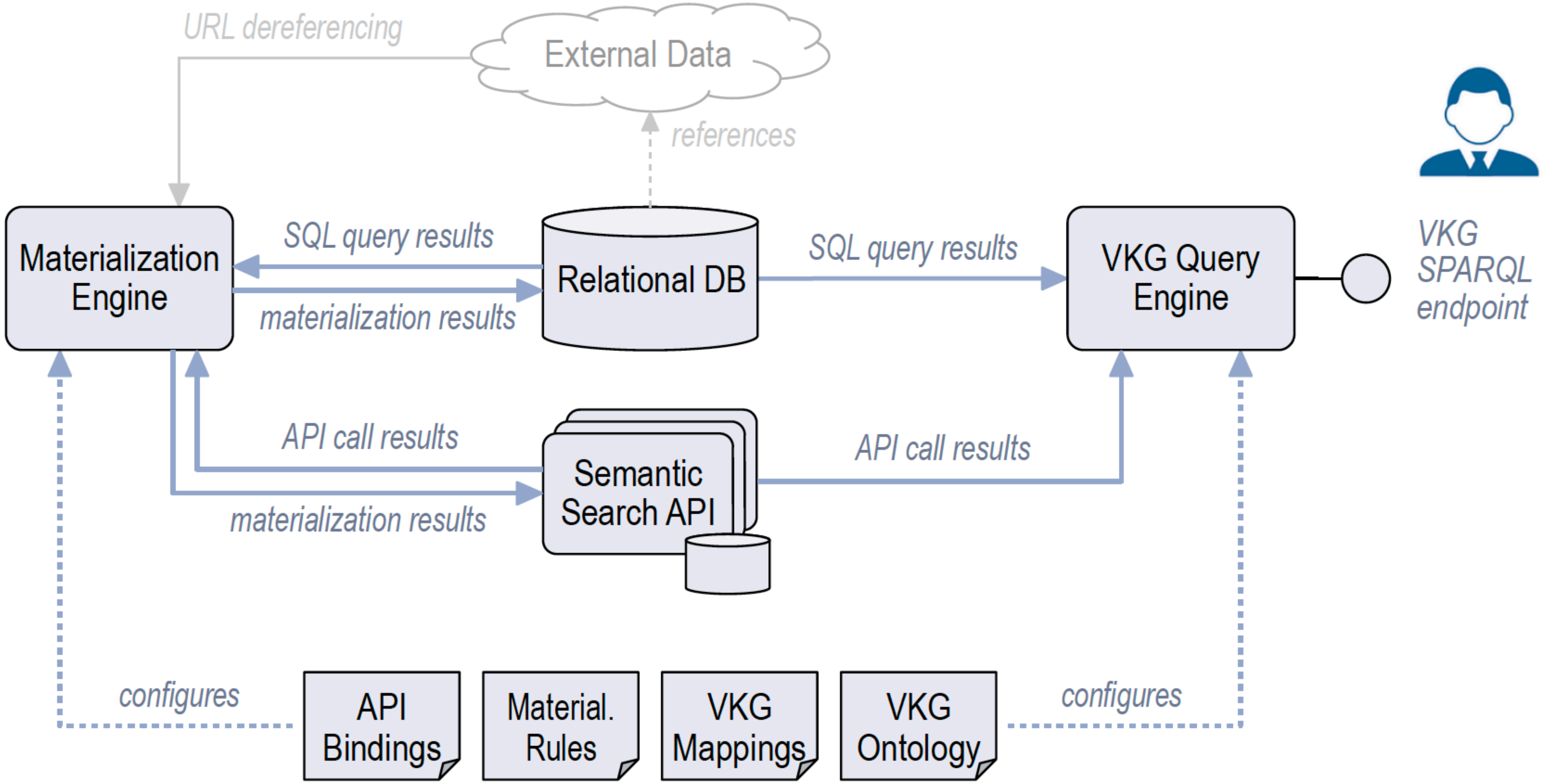
Goal: build a virtual knowledge graph integrating structured data in a proprietary platform (LinStar) and the results of information extraction from related semi-structured data.

- Structured data is provided by a relational database.
- Semi-structured data consisting of text with little (if any) structure or markup, such as natural language text, HTML documents, PDF files.
- Information extraction (IE) aims at extracting structured information from semi-structured data, possibly leveraging natural language processing (NLP) techniques.

Motivations:

- Provide an unambiguous formalization of the knowledge in the platform, to ease exploitation.
- Provide an integrated, queryable, up-to-date view over all available information.
- Enable more advanced services, such as intelligent search and intelligent recommendation.

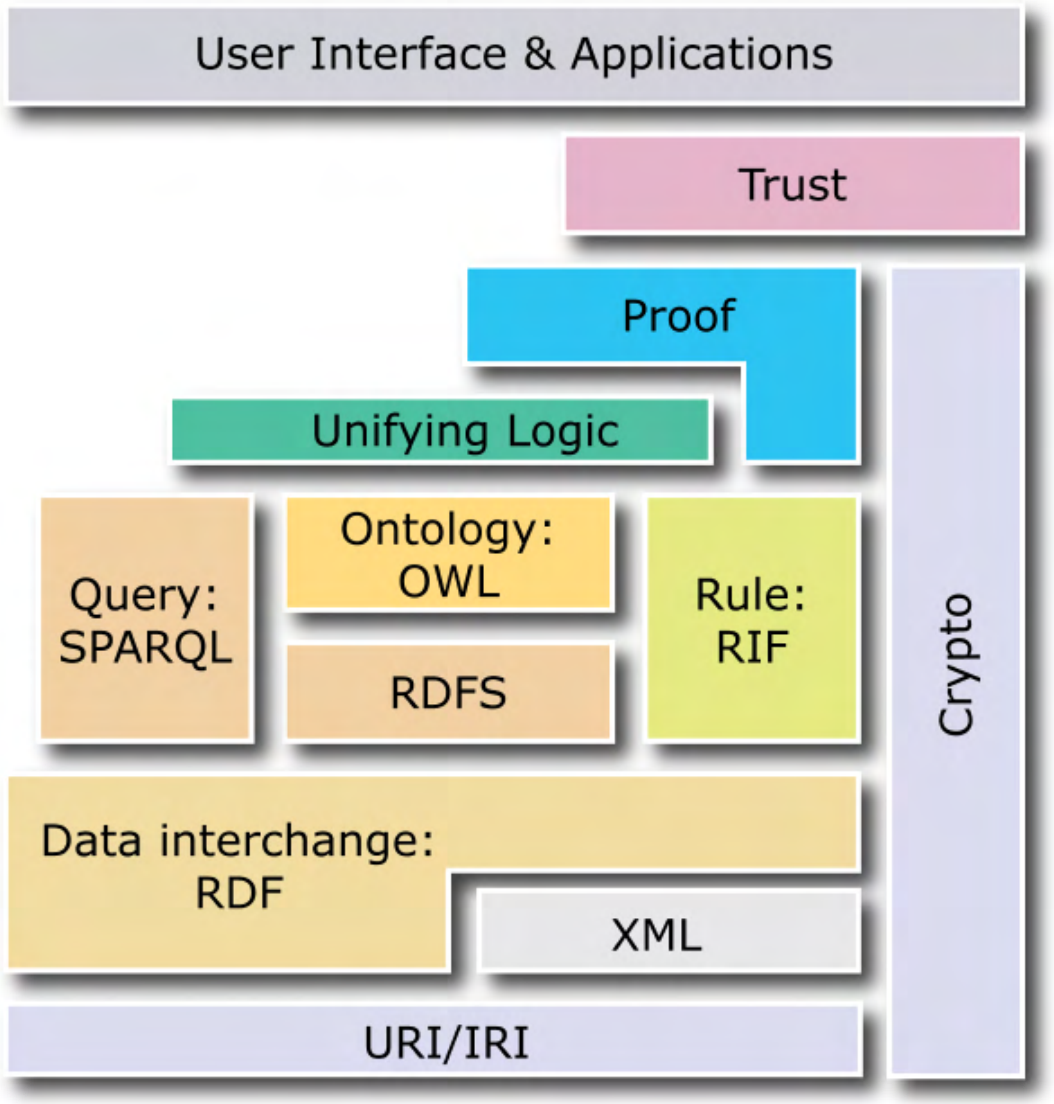
High-Level architecture of VKG over semi-structured data



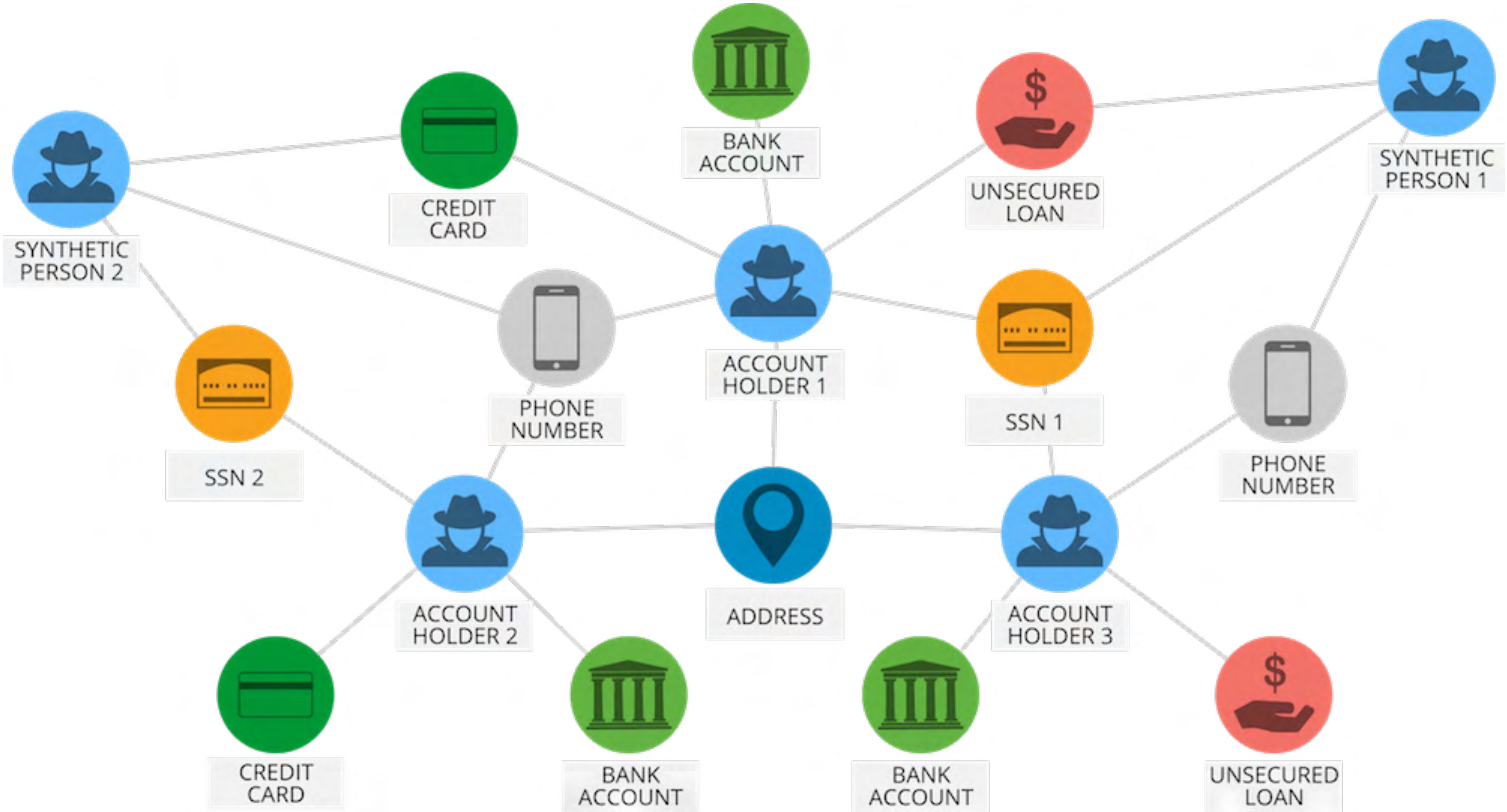
3. Key Technologies in Virtual Knowledge Graphs

Some Knowledge Graph technologies

- RDF Graph Language
- SPARQL Query Language
- OWL Ontology Language
- R2RML Mapping Language



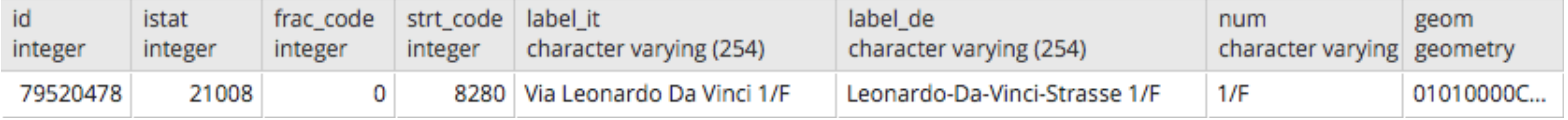
RDF — Data is represented as a graph



The graph consists of a set of subject-predicate-object triples.

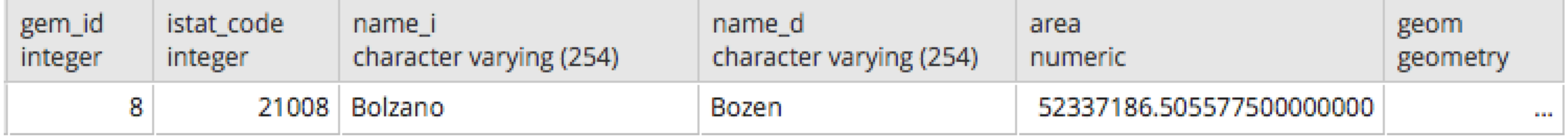
Relational DB vs. graph data — Example

Addresses



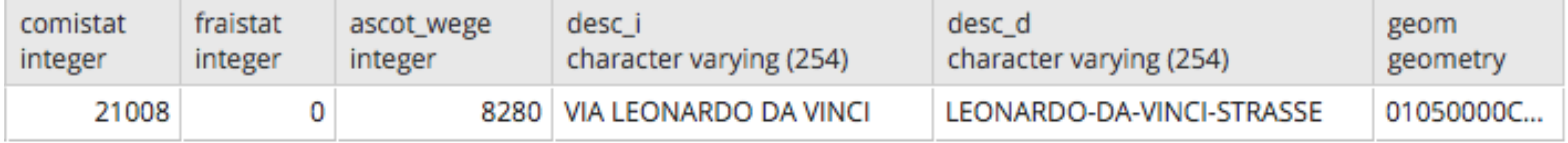
id	istat	frac_code	strt_code	label_it	label_de	num	geom
integer	integer	integer	integer	character varying (254)	character varying (254)	character varying	geometry
79520478	21008	0	8280	Via Leonardo Da Vinci 1/F	Leonardo-Da-Vinci-Strasse 1/F	1/F	01010000C...

Municipalities



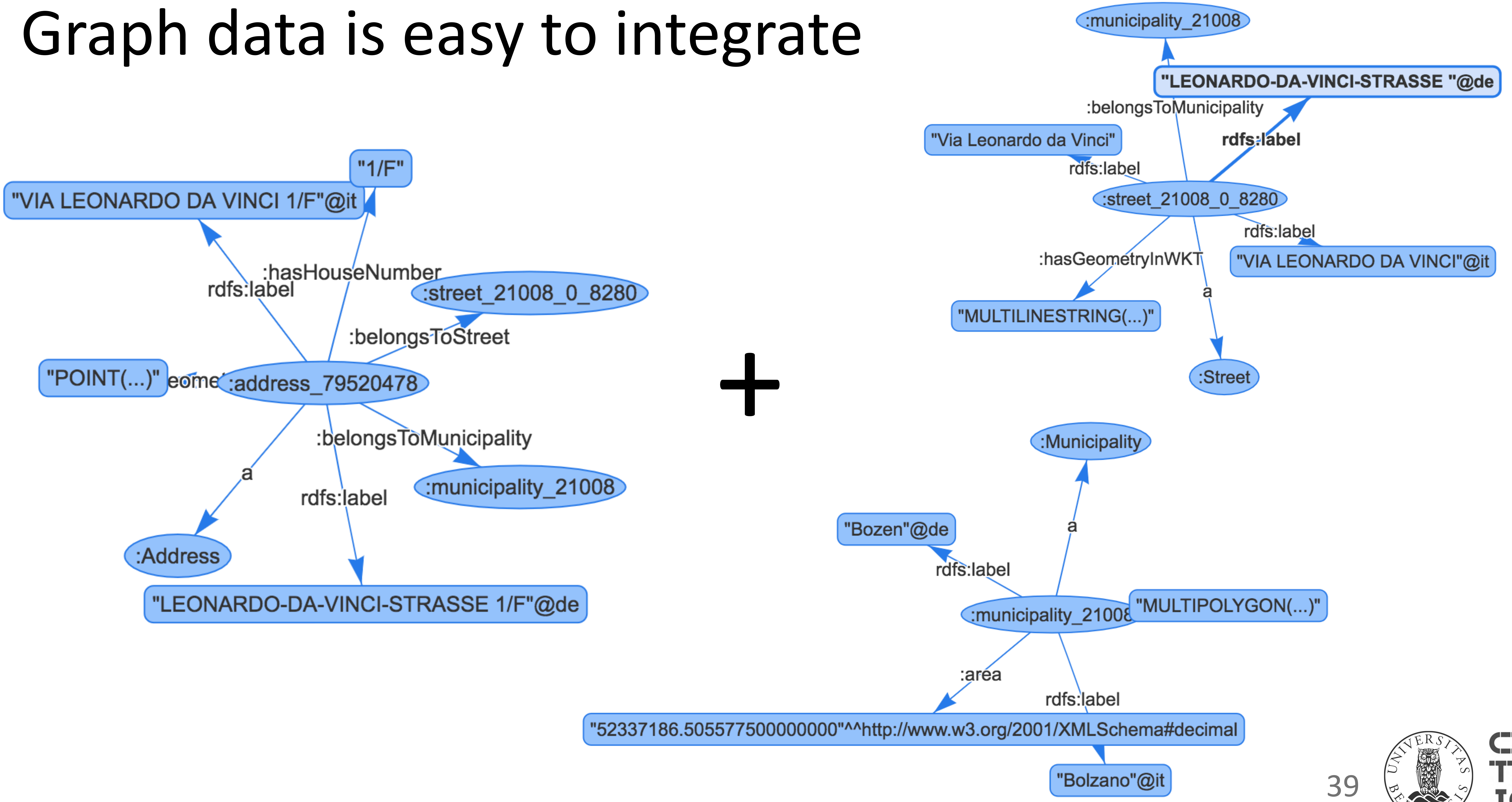
gem_id	istat_code	name_i	name_d	area	geom
integer	integer	character varying (254)	character varying (254)	numeric	geometry
8	21008	Bolzano	Bozen	52337186.505577500000000	...

Streets

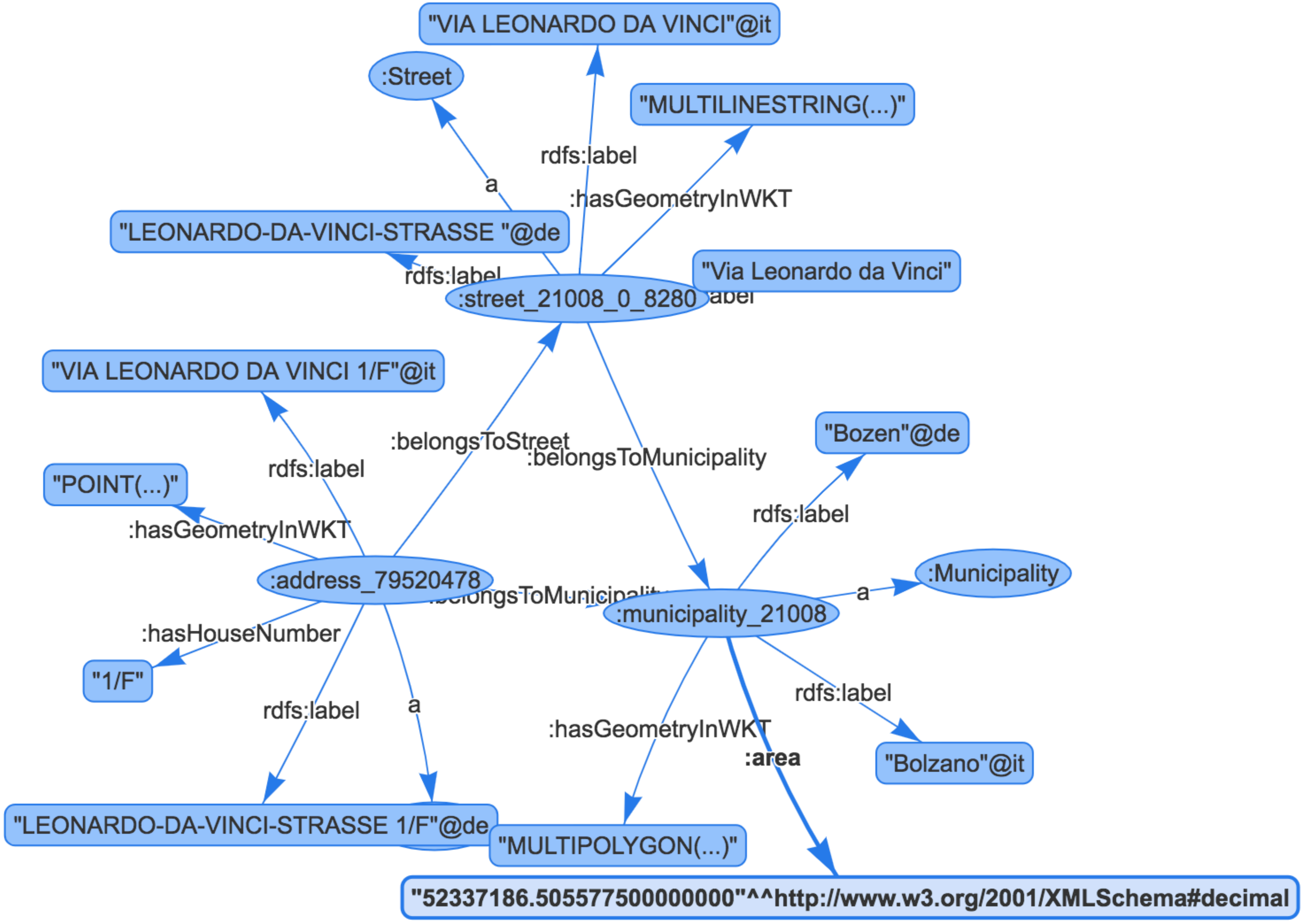


comistat	fraistat	ascot_wege	desc_i	desc_d	geom
integer	integer	integer	character varying (254)	character varying (254)	geometry
21008	0	8280	VIA LEONARDO DA VINCI	LEONARDO-DA-VINCI-STRASSE	01050000C...

Graph data is easy to integrate



=

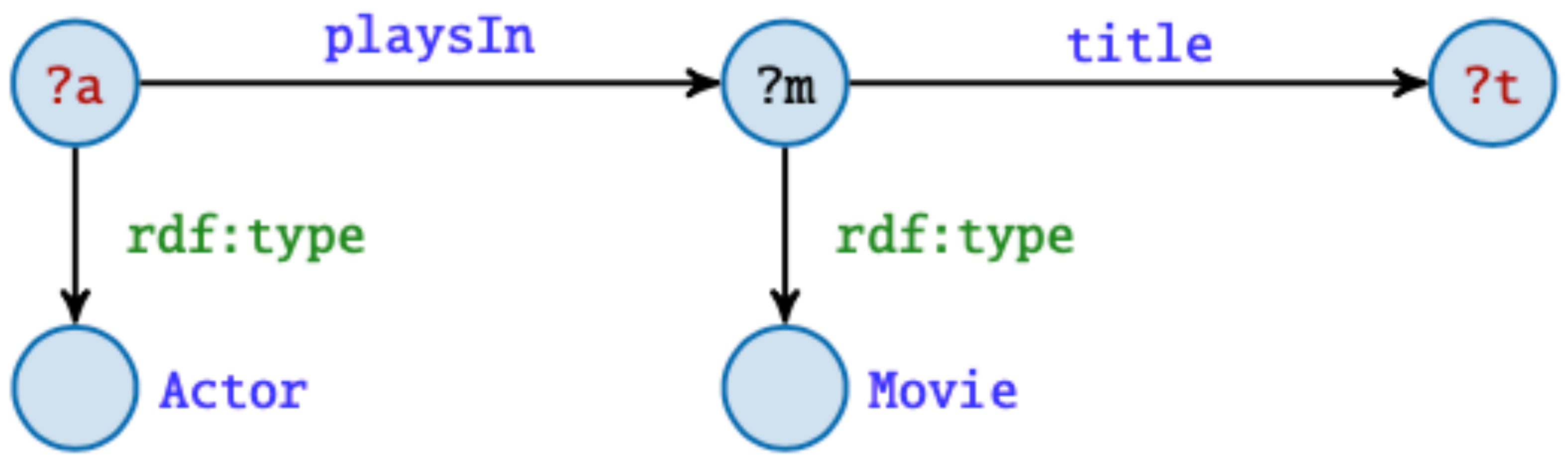


"VIA LEONARDO DA VINCI"@it
:Street
"MULTILINESTRING(...)"
rdfs:label
a
:hasGeometryInWKT
"LEONARDO-DA-VINCI-STRASSE "@de
rdfs:label
"Via Leonardo da Vinci"
:street_21008_0_8280
label
"VIA LEONARDO DA VINCI 1/F"@it
:belongsToStreet
:belongsToMunicipality
"Bozen"@de
rdfs:label
"POINT(...)"
rdfs:label
:hasGeometryInWKT
:Municipality
:address_79520478
:belongsToMunicipality
a
:municipality_21008
:hasHouseNumber
"1/F"
rdfs:label
rdfs:label
a
:hasGeometryInWKT
:area
"Bolzano"@it
"LEONARDO-DA-VINCI-STRASSE 1/F"@de
"MULTIPOLYGON(...)"
"52337186.505577500000000"^^http://www.w3.org/2001/XMLSchema#decimal

The SPARQL query language

- Is the standard query language for RDF data. [W3C Rec. 2008, 2013]
- Core query mechanism is based on graph matching:

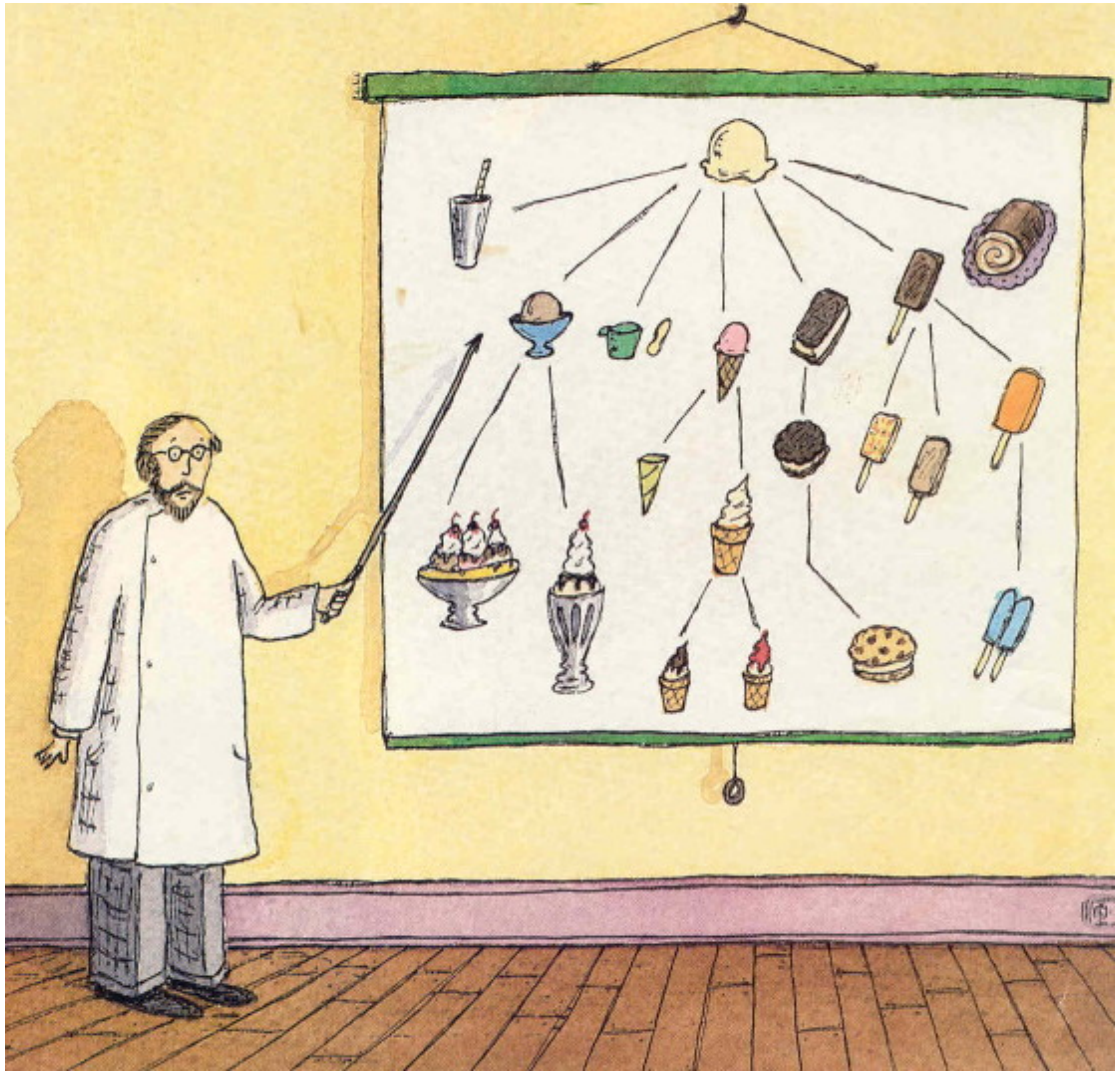
```
SELECT ?a ?t
WHERE { ?a rdf:type Actor .
        ?a playsIn ?m .
        ?m rdf:type Movie .
        ?m title ?t .
      }
```



- Has many additional features:
  - **UNION**: matches one of alternative graph patterns
  - **OPTIONAL**: produces a match even when part of the pattern is missing
  - complex **FILTER** conditions
  - **GROUP BY**: to express aggregations
  - **MINUS**: to remove possible solutions
  - property paths (regular expressions)

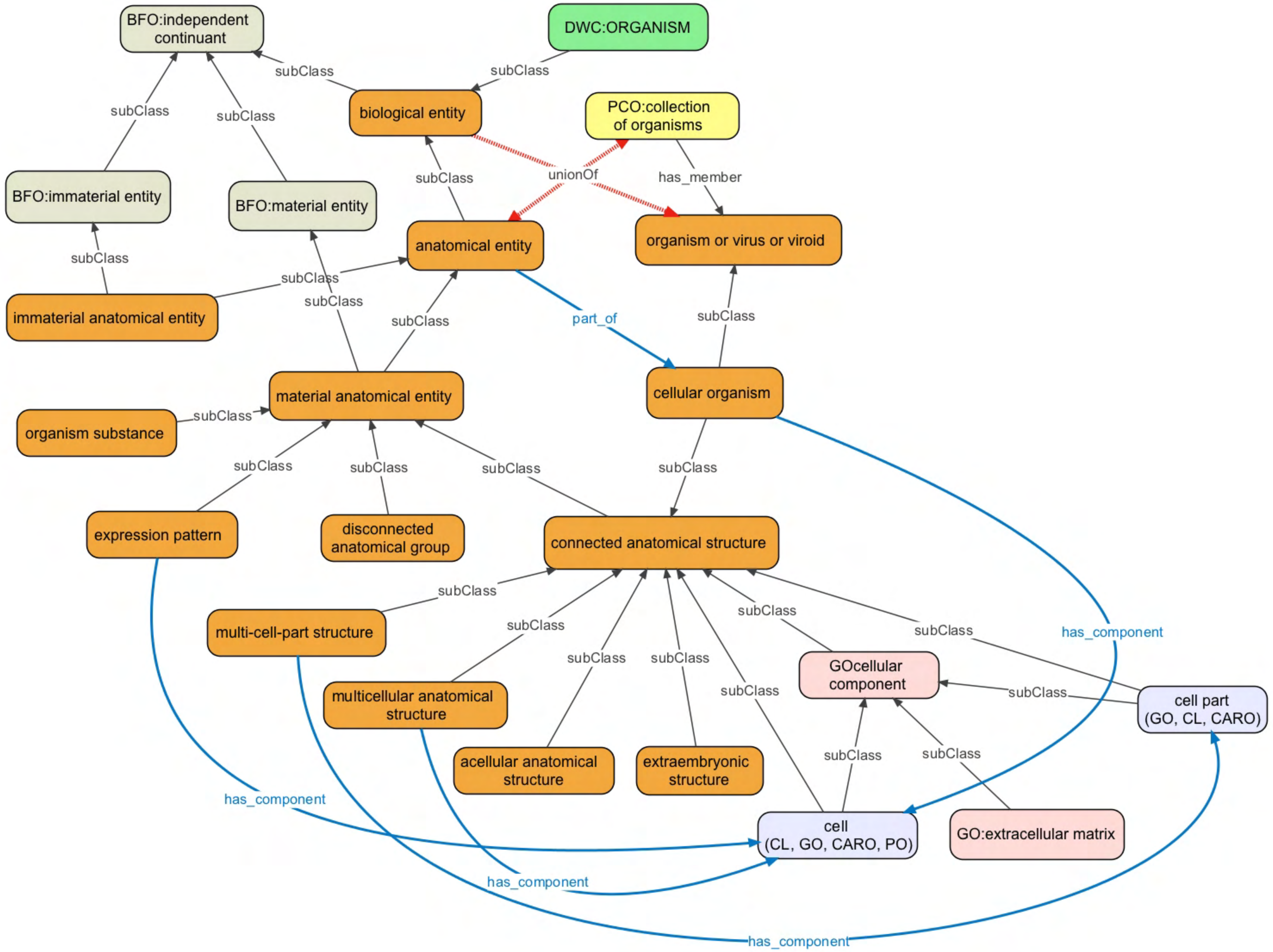
What is an ontology?

- An ontology conceptualizes a domain of interest in terms of concepts/classes, (binary) relations, and their properties.
- It typically organizes the concepts in a hierarchical structure.



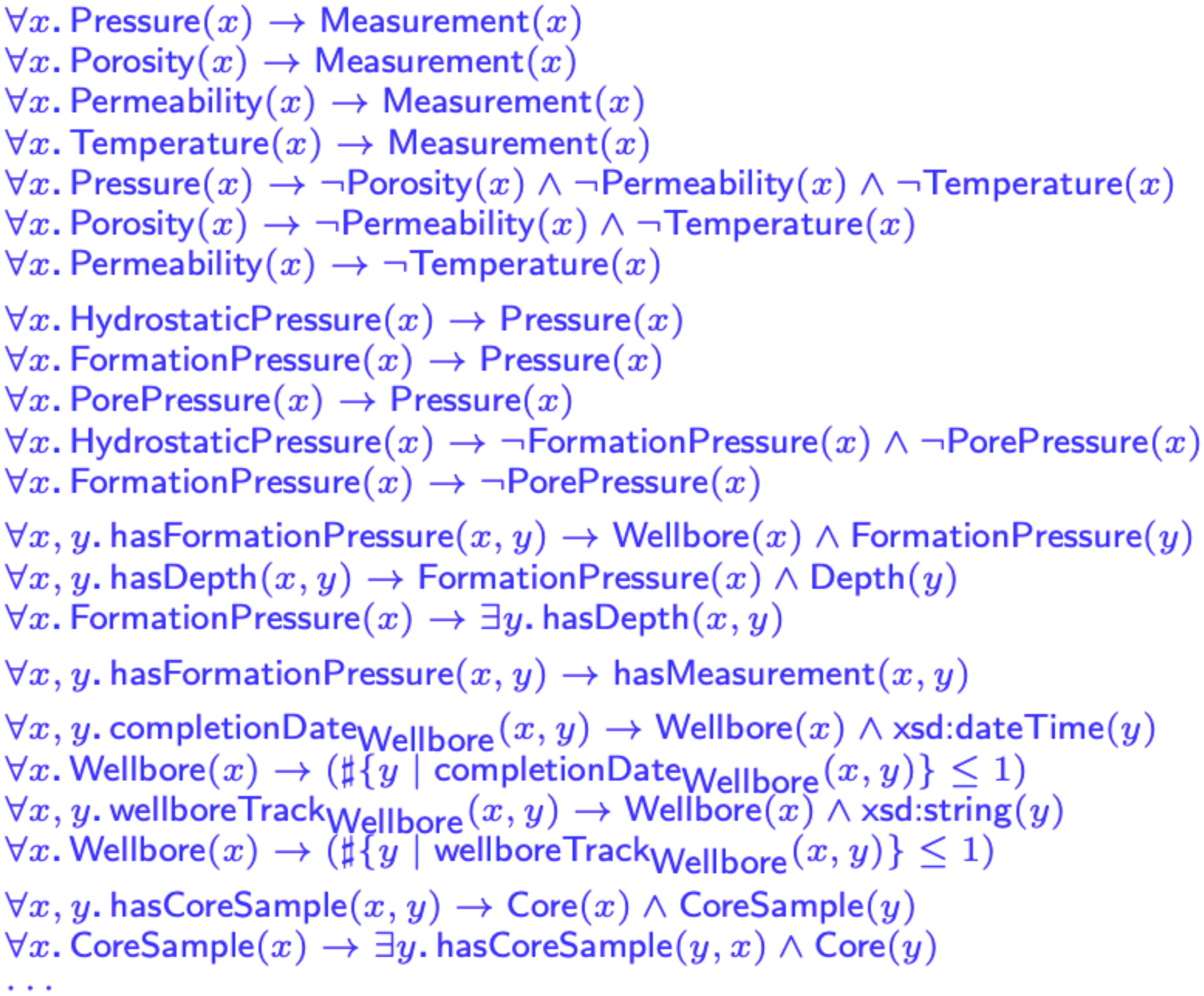
What is an ontology?

- An ontology conceptualizes a domain of interest in terms of concepts/classes, (binary) relations, and their properties.
- It typically organizes the concepts in a hierarchical structure.
- Ontologies are often represented as graphs. Sometimes they are also considered as Knowledge Graphs.



What is an ontology?

- An ontology conceptualizes a domain of interest in terms of concepts/classes, (binary) relations, and their properties.
- It typically organizes the concepts in a hierarchical structure.
- Ontologies are often represented as graphs. Sometimes they are also considered as Knowledge Graphs.
- However, an ontology is actually a logical theory, expressed in a suitable fragment of first-order logic.



$$
\begin{array}{l}
\forall x.\, \mathsf{Pressure}(x) \to \mathsf{Measurement}(x)\\
\forall x.\, \mathsf{Porosity}(x) \to \mathsf{Measurement}(x)\\
\forall x.\, \mathsf{Permeability}(x) \to \mathsf{Measurement}(x)\\
\forall x.\, \mathsf{Temperature}(x) \to \mathsf{Measurement}(x)\\
\forall x.\, \mathsf{Pressure}(x) \to \neg\mathsf{Porosity}(x) \land \neg\mathsf{Permeability}(x) \land \neg\mathsf{Temperature}(x)\\
\forall x.\, \mathsf{Porosity}(x) \to \neg\mathsf{Permeability}(x) \land \neg\mathsf{Temperature}(x)\\
\forall x.\, \mathsf{Permeability}(x) \to \neg\mathsf{Temperature}(x)\\[6pt]
\forall x.\, \mathsf{HydrostaticPressure}(x) \to \mathsf{Pressure}(x)\\
\forall x.\, \mathsf{FormationPressure}(x) \to \mathsf{Pressure}(x)\\
\forall x.\, \mathsf{PorePressure}(x) \to \mathsf{Pressure}(x)\\
\forall x.\, \mathsf{HydrostaticPressure}(x) \to \neg\mathsf{FormationPressure}(x) \land \neg\mathsf{PorePressure}(x)\\
\forall x.\, \mathsf{FormationPressure}(x) \to \neg\mathsf{PorePressure}(x)\\[6pt]
\forall x, y.\, \mathsf{hasFormationPressure}(x, y) \to \mathsf{Wellbore}(x) \land \mathsf{FormationPressure}(y)\\
\forall x, y.\, \mathsf{hasDepth}(x, y) \to \mathsf{FormationPressure}(x) \land \mathsf{Depth}(y)\\
\forall x.\, \mathsf{FormationPressure}(x) \to \exists y.\, \mathsf{hasDepth}(x, y)\\[6pt]
\forall x, y.\, \mathsf{hasFormationPressure}(x, y) \to \mathsf{hasMeasurement}(x, y)\\[6pt]
\forall x, y.\, \mathsf{completionDate}_{\mathsf{Wellbore}}(x, y) \to \mathsf{Wellbore}(x) \land \mathsf{xsd{:}dateTime}(y)\\
\forall x.\, \mathsf{Wellbore}(x) \to (\sharp\{y \mid \mathsf{completionDate}_{\mathsf{Wellbore}}(x, y)\} \leq 1)\\
\forall x, y.\, \mathsf{wellboreTrack}_{\mathsf{Wellbore}}(x, y) \to \mathsf{Wellbore}(x) \land \mathsf{xsd{:}string}(y)\\
\forall x.\, \mathsf{Wellbore}(x) \to (\sharp\{y \mid \mathsf{wellboreTrack}_{\mathsf{Wellbore}}(x, y)\} \leq 1)\\[6pt]
\forall x, y.\, \mathsf{hasCoreSample}(x, y) \to \mathsf{Core}(x) \land \mathsf{CoreSample}(y)\\
\forall x.\, \mathsf{CoreSample}(x) \to \exists y.\, \mathsf{hasCoreSample}(y, x) \land \mathsf{Core}(y)\\
\ldots
\end{array}
$$

What is an ontology?

- An ontology conceptualizes a domain of interest in terms of concepts/classes, (binary) relations, and their properties.
- It typically organizes the concepts in a hierarchical structure.
- Ontologies are often represented as graphs.
- However, an ontology is actually a logical theory, expressed in a suitable fragment of first-order logic, or better, in description logics, the theoretical foundation of the OWL standard.

$$
\begin{array}{rcl}
\mathsf{Pressure} & \sqsubseteq & \mathsf{Measurement} \\
\mathsf{Porosity} & \sqsubseteq & \mathsf{Measurement} \\
\mathsf{Permeability} & \sqsubseteq & \mathsf{Measurement} \\
\mathsf{Temperature} & \sqsubseteq & \mathsf{Measurement} \\
\mathsf{Pressure} & \sqsubseteq & \neg\mathsf{Porosity} \sqcap \neg\mathsf{Permeability} \sqcap \neg\mathsf{Temperature} \\
\mathsf{Porosity} & \sqsubseteq & \neg\mathsf{Permeability} \sqcap \neg\mathsf{Temperature} \\
\mathsf{Permeability} & \sqsubseteq & \neg\mathsf{Temperature} \\[6pt]
\mathsf{HydrostaticPressure} & \sqsubseteq & \mathsf{Pressure} \\
\mathsf{FormationPressure} & \sqsubseteq & \mathsf{Pressure} \\
\mathsf{PorePressure} & \sqsubseteq & \mathsf{Pressure} \\
\mathsf{HydrostaticPressure} & \sqsubseteq & \neg\mathsf{FormationPressure} \sqcap \neg\mathsf{PorePressure} \\
\mathsf{FormationPressure} & \sqsubseteq & \neg\mathsf{PorePressure} \\[6pt]
\exists\mathsf{hasFormationPressure} & \sqsubseteq & \mathsf{Wellbore} \\
\exists\mathsf{hasFormationPressure}^{-} & \sqsubseteq & \mathsf{FormationPressure} \\
\exists\mathsf{hasDepth} & \sqsubseteq & \mathsf{FormationPressure} \\
\exists\mathsf{hasDepth}^{-} & \sqsubseteq & \mathsf{Depth} \\
\mathsf{FormationPressure} & \sqsubseteq & \exists\mathsf{hasDepth} \\[6pt]
\mathsf{hasFormationPressure} & \sqsubseteq & \mathsf{hasMeasurement} \\[6pt]
\exists\mathsf{completionDate} & \sqsubseteq & \mathsf{Wellbore} \\
\exists\mathsf{completionDate}^{-} & \sqsubseteq & \mathsf{xsd{:}dateTime} \\
\mathsf{Wellbore} & \sqsubseteq & (\leq 1\ \mathsf{completionDate}) \\
\exists\mathsf{wellboreTrack} & \sqsubseteq & \mathsf{Wellbore} \\
 & \cdots &
\end{array}
$$

Reasoning in Ontology

Class hierarchy: rdfs:subClassOf
Example: :MovieActor rdfs:subClassOf :Actor .
Inference: <person/2> rdf:type :MovieActor . => <person/2> rdf:type :Actor .

Domain of properties: rdfs:domain
Example: :playsIn rdfs:domain :MovieActor .
Inference: <person/2> :playsIn <movie/3> . => <person/2> rdf:type :MovieActor .

Range of properties: rdfs:range
Example: :playsIn rdfs:range :Movie .
Inference: <person/2> :playsIn <movie/3> . => <movie/3> rdf:type :Movie .

Reasoning in Ontology

Class disjointness: owl:disjointWith
Example: :Actor owl:disjointWith :Movie .
Inference: <person/2> rdf:type :Actor . <person/2> rdf:type :Movie .
=> RDF graph inconsistent with the ontology

Inverse properties: owl:inverseOf
Example: :actsIn owl:inverseOf :hasActor .
Inference: <person/2> :actsIn <movie/3> . => <movie/3> :hasActor <person/2> .

Property hierarchy
Property disjointness
Mandatory participation
…

Mapping Language

The mapping consists of a set of assertions of the form

SQL Query ⇝ Class membership assertion

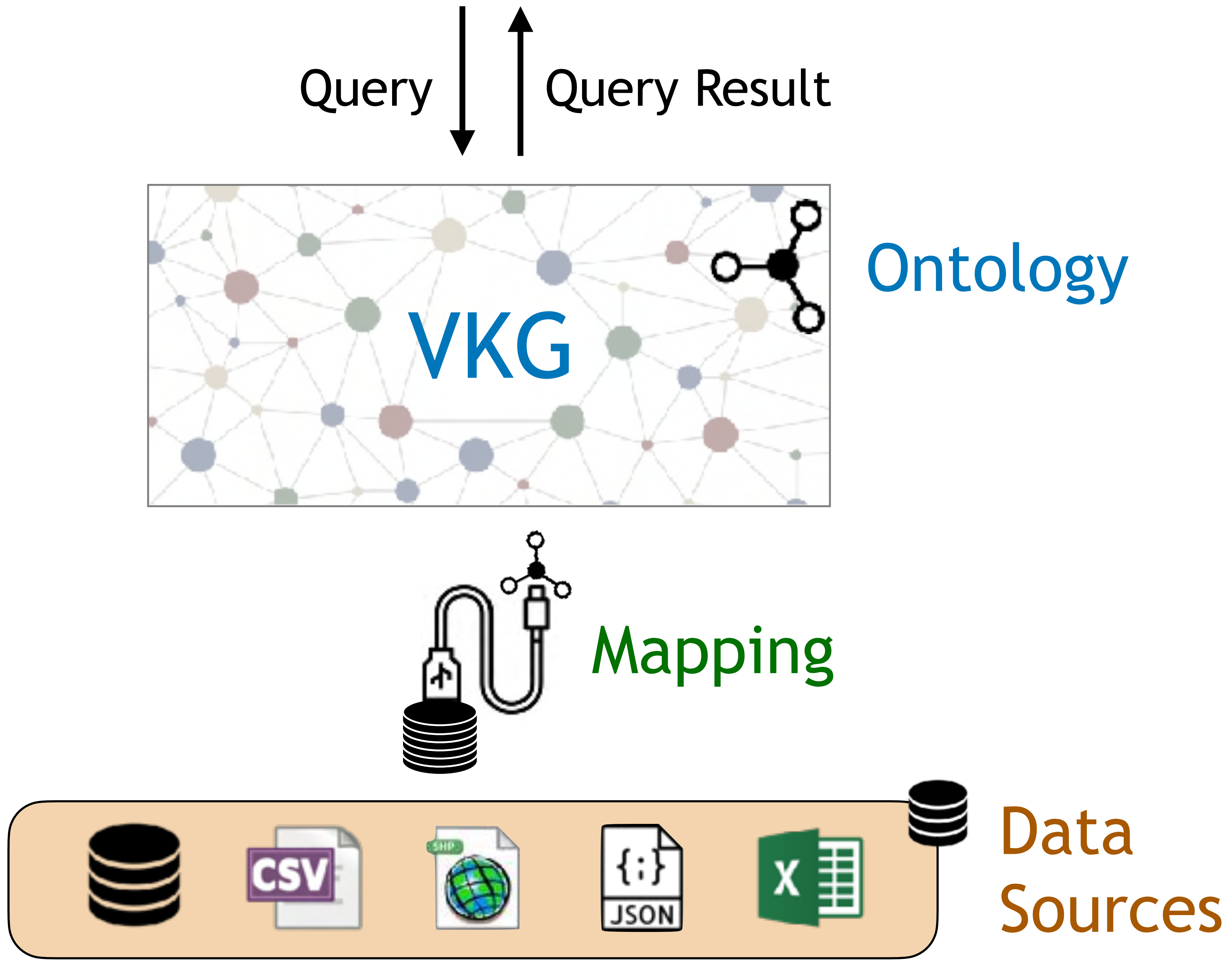
SQL Query ⇝ Property membership assertion

Intuition behind the mapping

The answers returned by the SQL Query in the left-hand side are used to create the objects (and values) that populate the Class / Property in the right-hand side.

Note: The mapping contains also a mechanism to transform values retrieved from the database into objects of the VKG (thus solving the so-called impedance mismatch).

Why mappings?



The traditional approach to data integration relies on mediators, which are specified through complex code.

Mappings, instead:

- Provide a declarative specification, and not code.
- Are easier to understand, and hence to design and to maintain.
- Support an incremental approach to integration.
- Are machine processable, hence can be used for query optimization.

5. A Demo for Data Integration using VKGs

Demo: Using Ontop for data integration

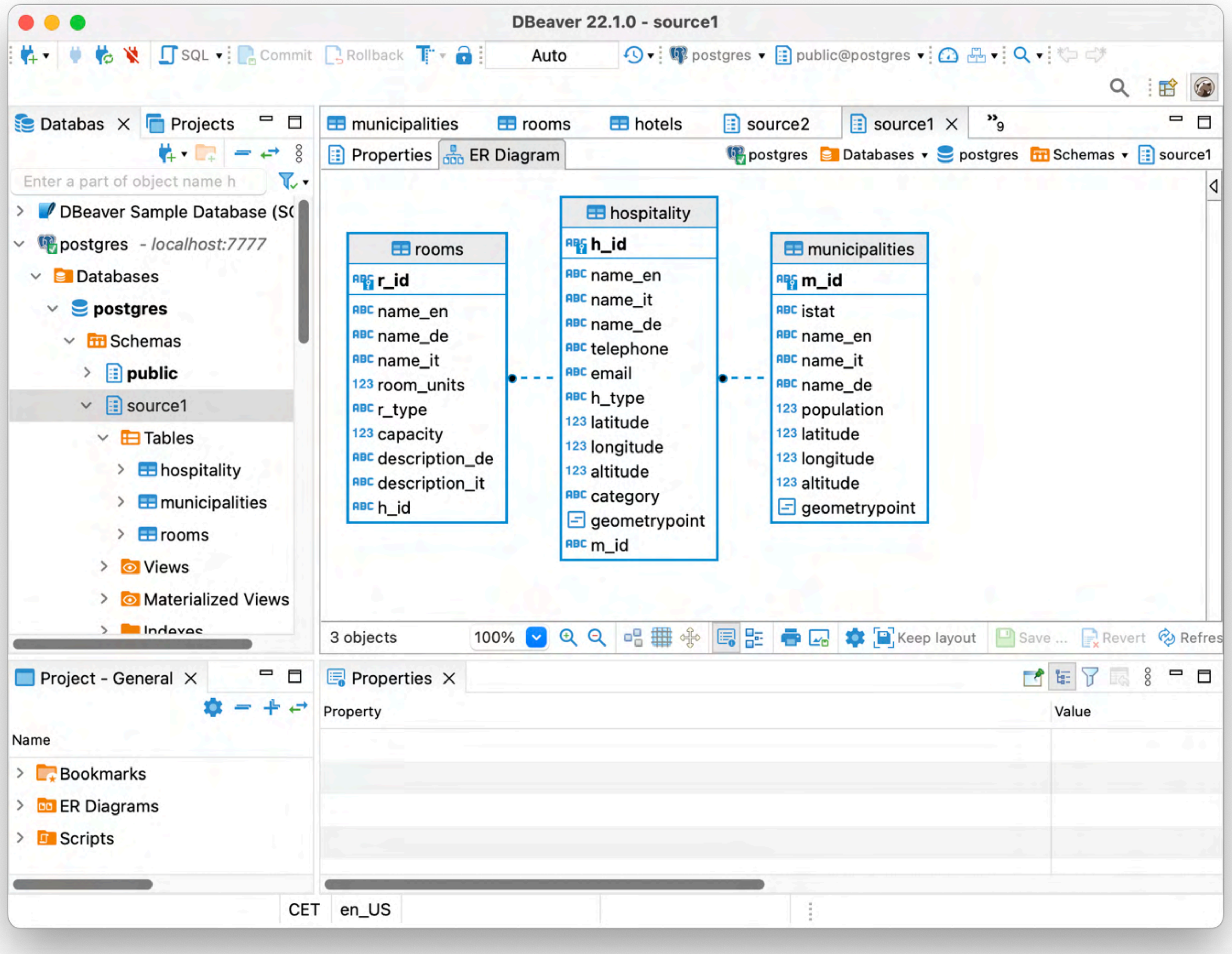
```
$ git clone \
https://github.com/ontopic-vkg/destination-tutorial \
--config core.autocrlf=input # important for Windows
$ cd destination-tutorial
$ docker-compose -f docker-compose.solution.yml up
```

1. Check the database in DBeaver
2. Open vkg/dest-solution.ttl in Protégé
3. Open http://localhost:8080 in the browser

Demo Database

Three data sources:

- Sources 1 & 2: hotels and rooms in different areas
- Source 3: weather information



DBeaver 22.1.0 - source1

- DBeaver Sample Database (SQ
- postgres - localhost:7777
  - Databases
    - postgres
      - Schemas
        - public
        - source1
          - Tables
            - hospitality
            - municipalities
            - rooms
          - Views
          - Materialized Views

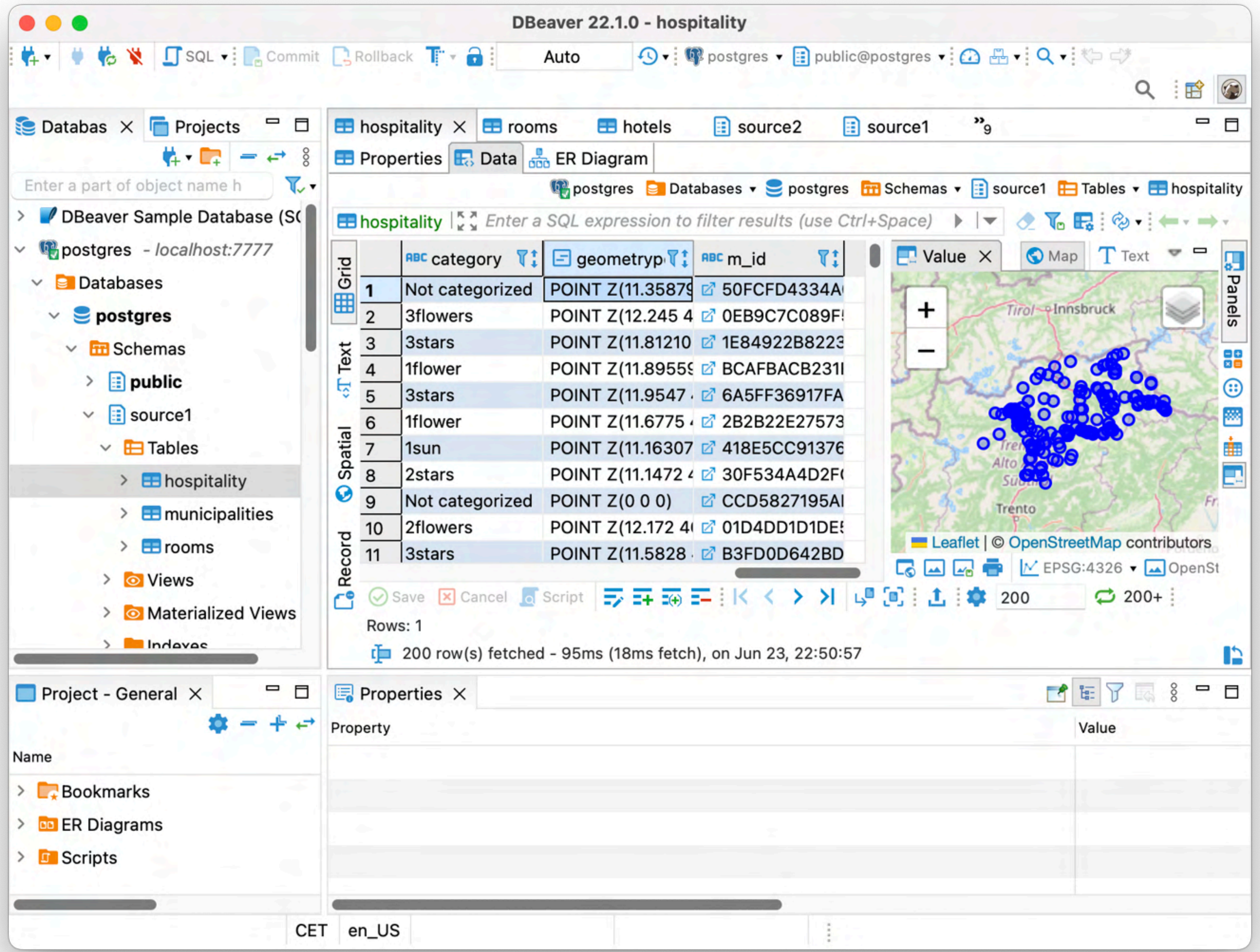
ER Diagram:

rooms: r_id, name_en, name_de, name_it, room_units, r_type, capacity, description_de, description_it, h_id

hospitality: h_id, name_en, name_it, name_de, telephone, email, h_type, latitude, longitude, altitude, category, geometrypoint, m_id

municipalities: m_id, istat, name_en, name_it, name_de, population, latitude, longitude, altitude, geometrypoint

3 objects



DBeaver 22.1.0 - hospitality

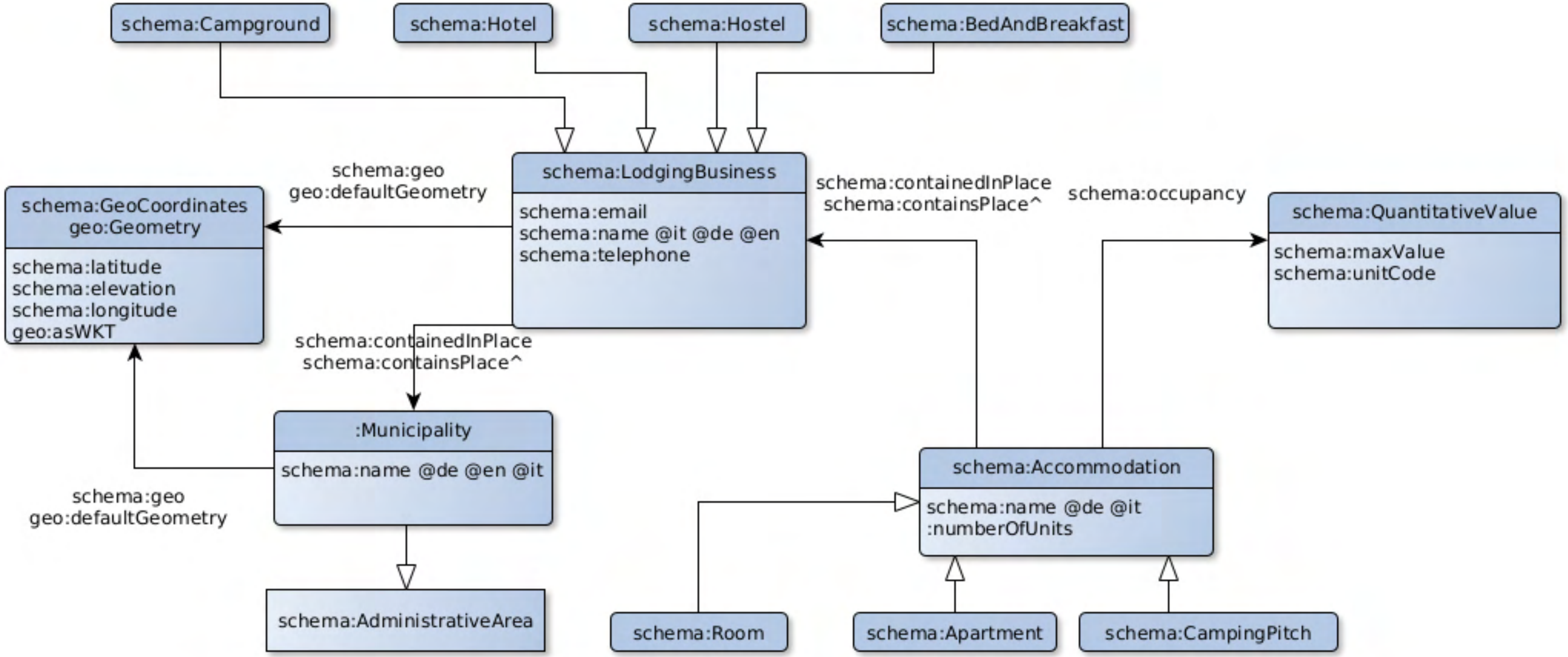
hospitality | Enter a SQL expression to filter results (use Ctrl+Space)

	category	geometryp	m_id
1	Not categorized	POINT Z(11.35879	50FCFD4334A(
2	3flowers	POINT Z(12.245 4	0EB9C7C089F!
3	3stars	POINT Z(11.81210	1E84922B8223
4	1flower	POINT Z(11.89559	BCAFBACB231I
5	3stars	POINT Z(11.9547	6A5FF36917FA
6	1flower	POINT Z(11.6775	2B2B22E27573
7	1sun	POINT Z(11.16307	418E5CC91376
8	2stars	POINT Z(11.1472	30F534A4D2F(
9	Not categorized	POINT Z(0 0 0)	CCD5827195AI
10	2flowers	POINT Z(12.172 4(	01D4DD1D1DE!
11	3stars	POINT Z(11.5828	B3FD0D642BD

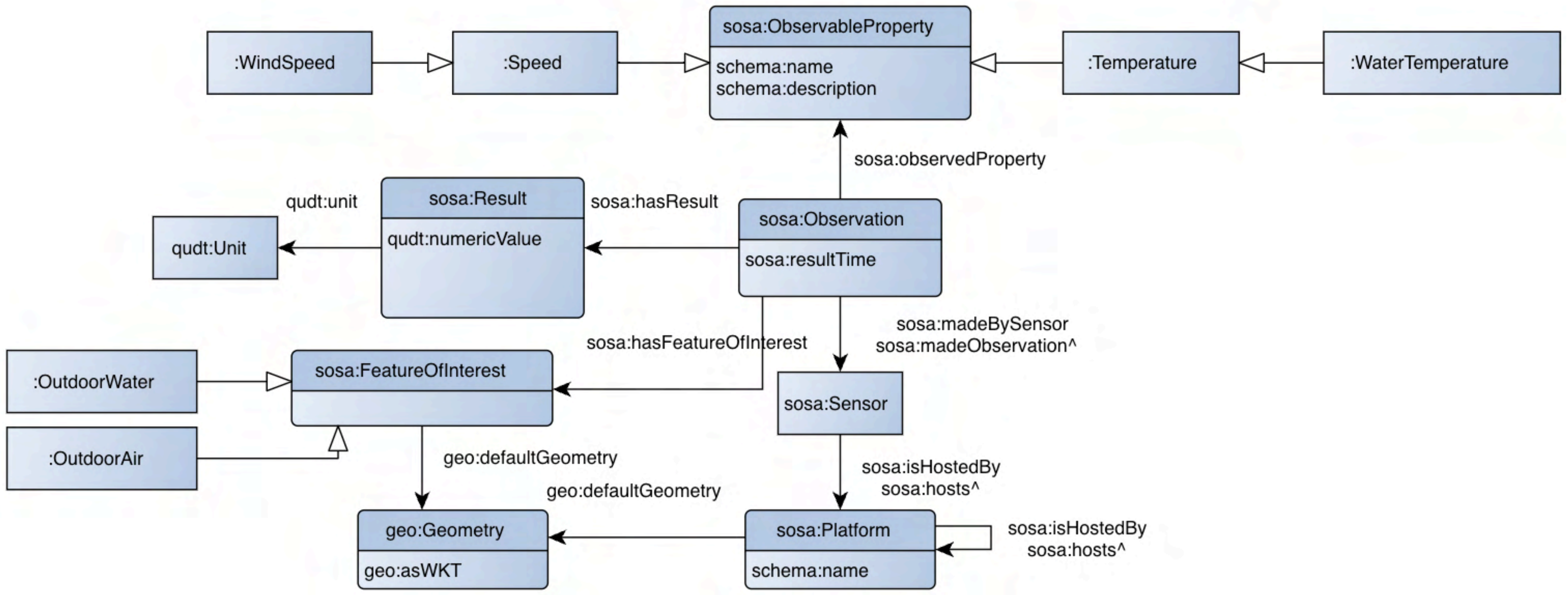
Rows: 1

200 row(s) fetched - 95ms (18ms fetch), on Jun 23, 22:50:57

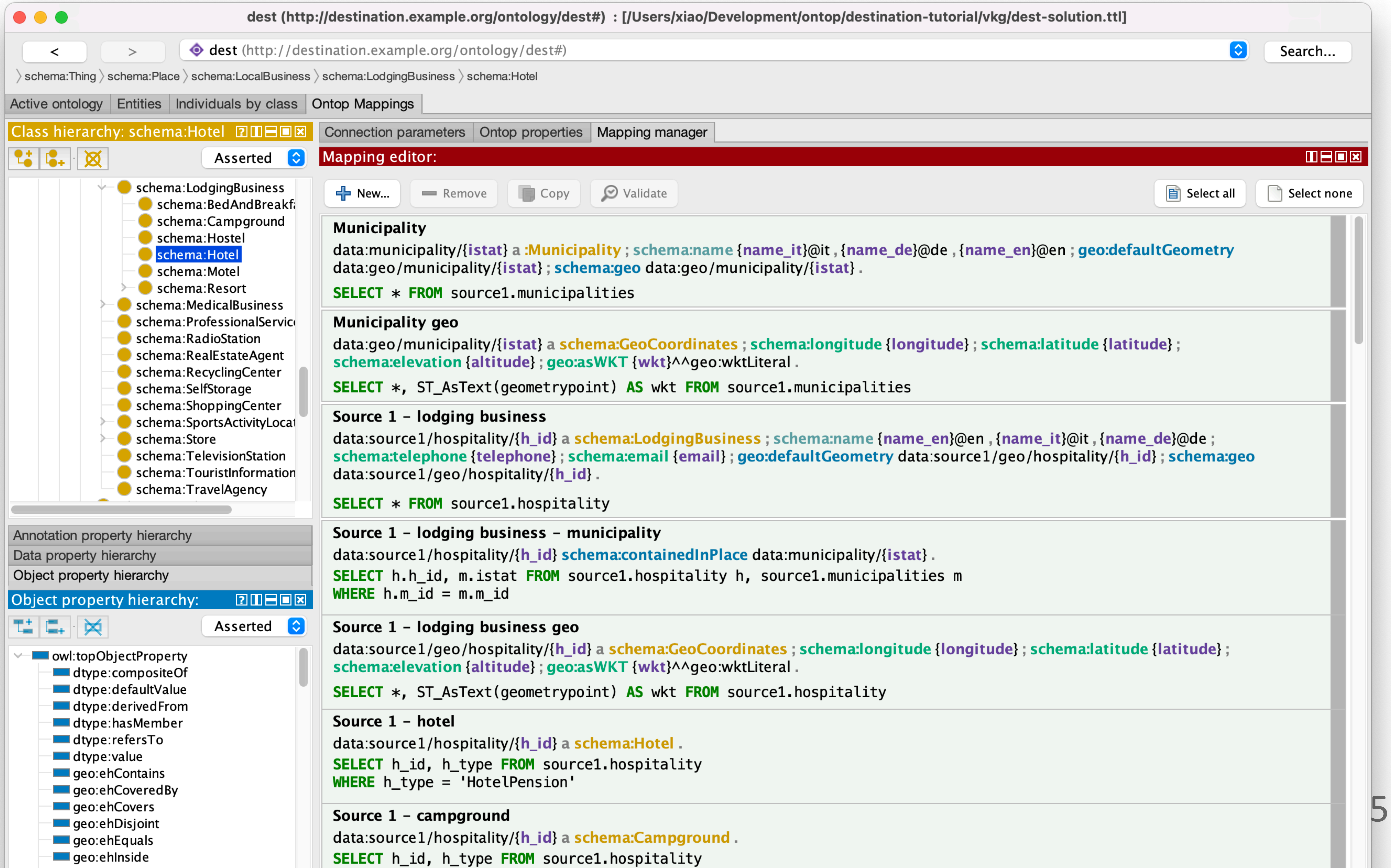
Ontology for tables in the schemas source1 and source2



Ontology for tables in the schema source3



Protege Editor with Ontop Plugin



Municipality

data:municipality/{istat} a :Municipality ; schema:name {name_it}@it , {name_de}@de , {name_en}@en ; geo:defaultGeometry data:geo/municipality/{istat} ; schema:geo data:geo/municipality/{istat} .

```
SELECT * FROM source1.municipalities
```

Municipality geo

data:geo/municipality/{istat} a schema:GeoCoordinates ; schema:longitude {longitude} ; schema:latitude {latitude} ; schema:elevation {altitude} ; geo:asWKT {wkt}^^geo:wktLiteral .

```
SELECT *, ST_AsText(geometrypoint) AS wkt FROM source1.municipalities
```

Source 1 – lodging business

data:source1/hospitality/{h_id} a schema:LodgingBusiness ; schema:name {name_en}@en , {name_it}@it , {name_de}@de ; schema:telephone {telephone} ; schema:email {email} ; geo:defaultGeometry data:source1/geo/hospitality/{h_id} ; schema:geo data:source1/geo/hospitality/{h_id} .

```
SELECT * FROM source1.hospitality
```

Source 1 – lodging business – municipality

data:source1/hospitality/{h_id} schema:containedInPlace data:municipality/{istat} .

```
SELECT h.h_id, m.istat FROM source1.hospitality h, source1.municipalities m
WHERE h.m_id = m.m_id
```

Source 1 – lodging business geo

data:source1/geo/hospitality/{h_id} a schema:GeoCoordinates ; schema:longitude {longitude} ; schema:latitude {latitude} ; schema:elevation {altitude} ; geo:asWKT {wkt}^^geo:wktLiteral .

```
SELECT *, ST_AsText(geometrypoint) AS wkt FROM source1.hospitality
```

Source 1 – hotel

data:source1/hospitality/{h_id} a schema:Hotel .

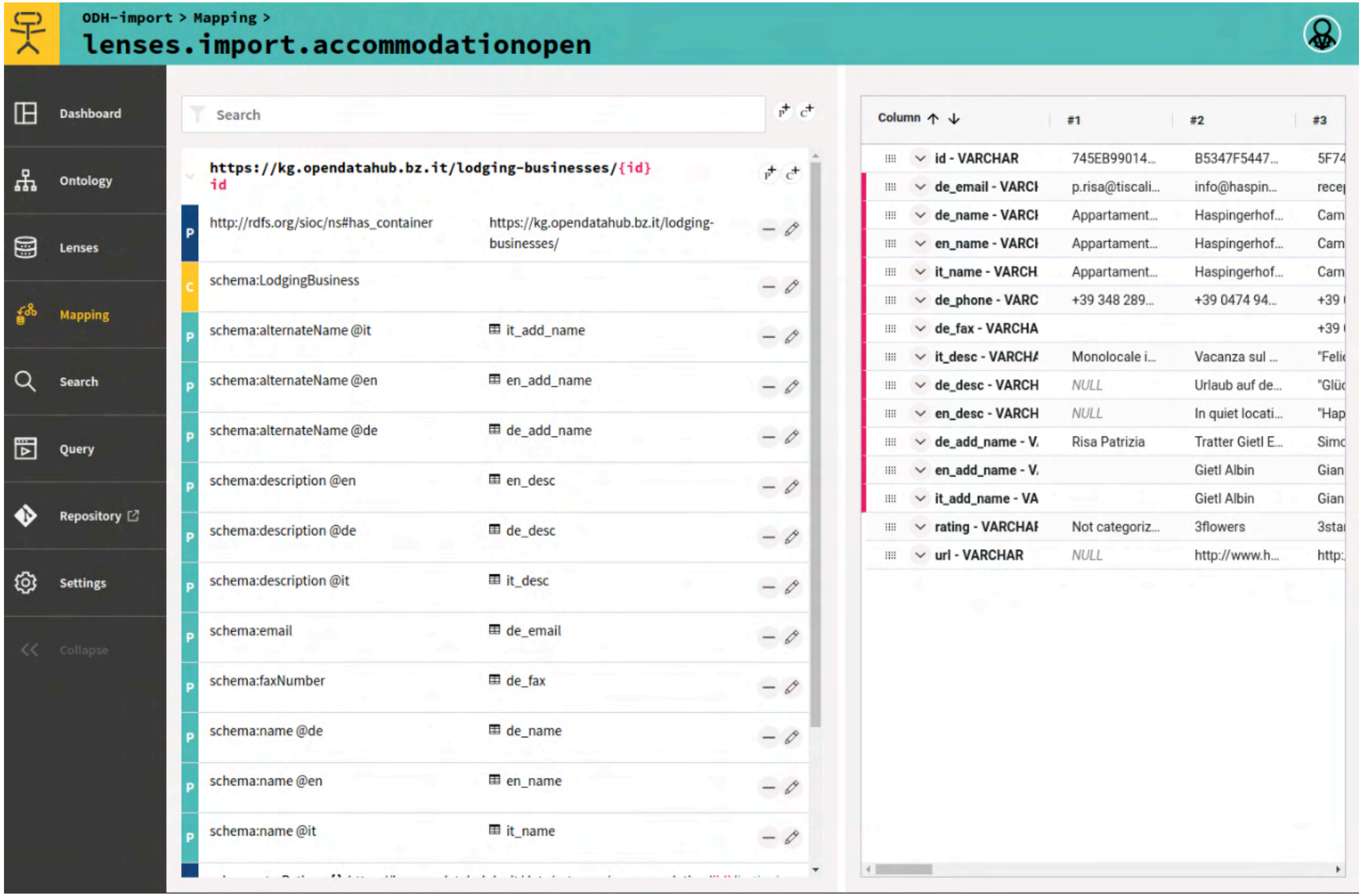
```
SELECT h_id, h_type FROM source1.hospitality
WHERE h_type = 'HotelPension'
```

Source 1 – campground

data:source1/hospitality/{h_id} a schema:Campground .

```
SELECT h_id, h_type FROM source1.hospitality
```

Ontopic Studio



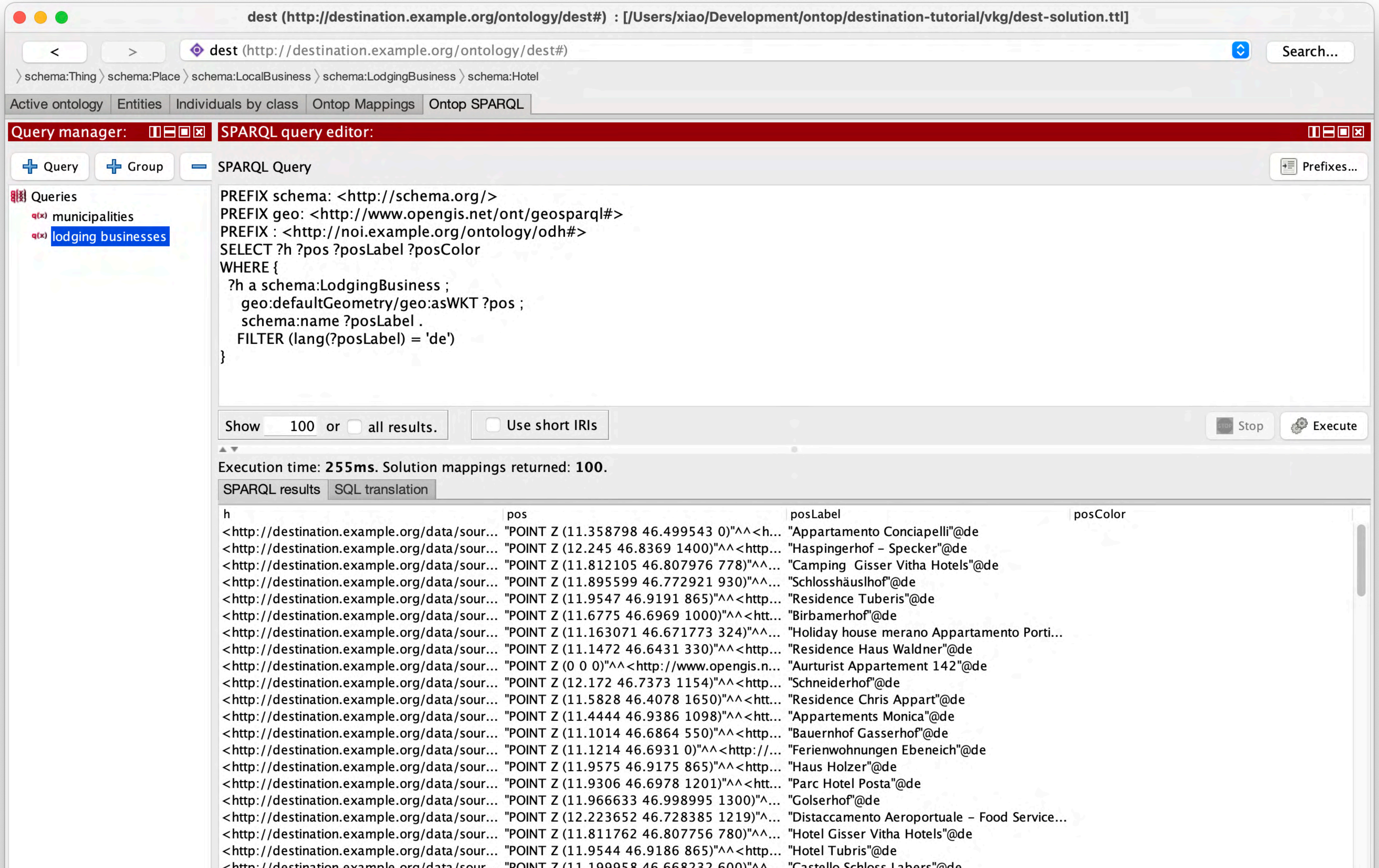
lenses.import.accommodationopen

https://kg.opendatahub.bz.it/lodging-businesses/{id}
id

	Property	Value
P	http://rdfs.org/sioc/ns#has_container	https://kg.opendatahub.bz.it/lodging-businesses/
C	schema:LodgingBusiness	
P	schema:alternateName @it	it_add_name
P	schema:alternateName @en	en_add_name
P	schema:alternateName @de	de_add_name
P	schema:description @en	en_desc
P	schema:description @de	de_desc
P	schema:description @it	it_desc
P	schema:email	de_email
P	schema:faxNumber	de_fax
P	schema:name @de	de_name
P	schema:name @en	en_name
P	schema:name @it	it_name

Column ↑ ↓	#1	#2	#3
id - VARCHAR	745EB99014...	B5347F5447...	5F74
de_email - VARCI	p.risa@tiscali...	info@haspin...	rece
de_name - VARCI	Appartament...	Haspingerhof...	Cam
en_name - VARCI	Appartament...	Haspingerhof...	Cam
it_name - VARCH	Appartament...	Haspingerhof...	Cam
de_phone - VARC	+39 348 289...	+39 0474 94...	+39
de_fax - VARCHA			+39
it_desc - VARCHA	Monolocale i...	Vacanza sul ...	"Feli
de_desc - VARCH	NULL	Urlaub auf de...	"Glü
en_desc - VARCH	NULL	In quiet locati...	"Hap
de_add_name - V.	Risa Patrizia	Tratter Gietl E...	Simo
en_add_name - V.		Gietl Albin	Gian
it_add_name - VA		Gietl Albin	Gian
rating - VARCHAI	Not categoriz...	3flowers	3sta
url - VARCHAR	NULL	http://www.h...	http:

SPARQL Query Answering



dest (http://destination.example.org/ontology/dest#) : [/Users/xiao/Development/ontop/destination-tutorial/vkg/dest-solution.ttl]

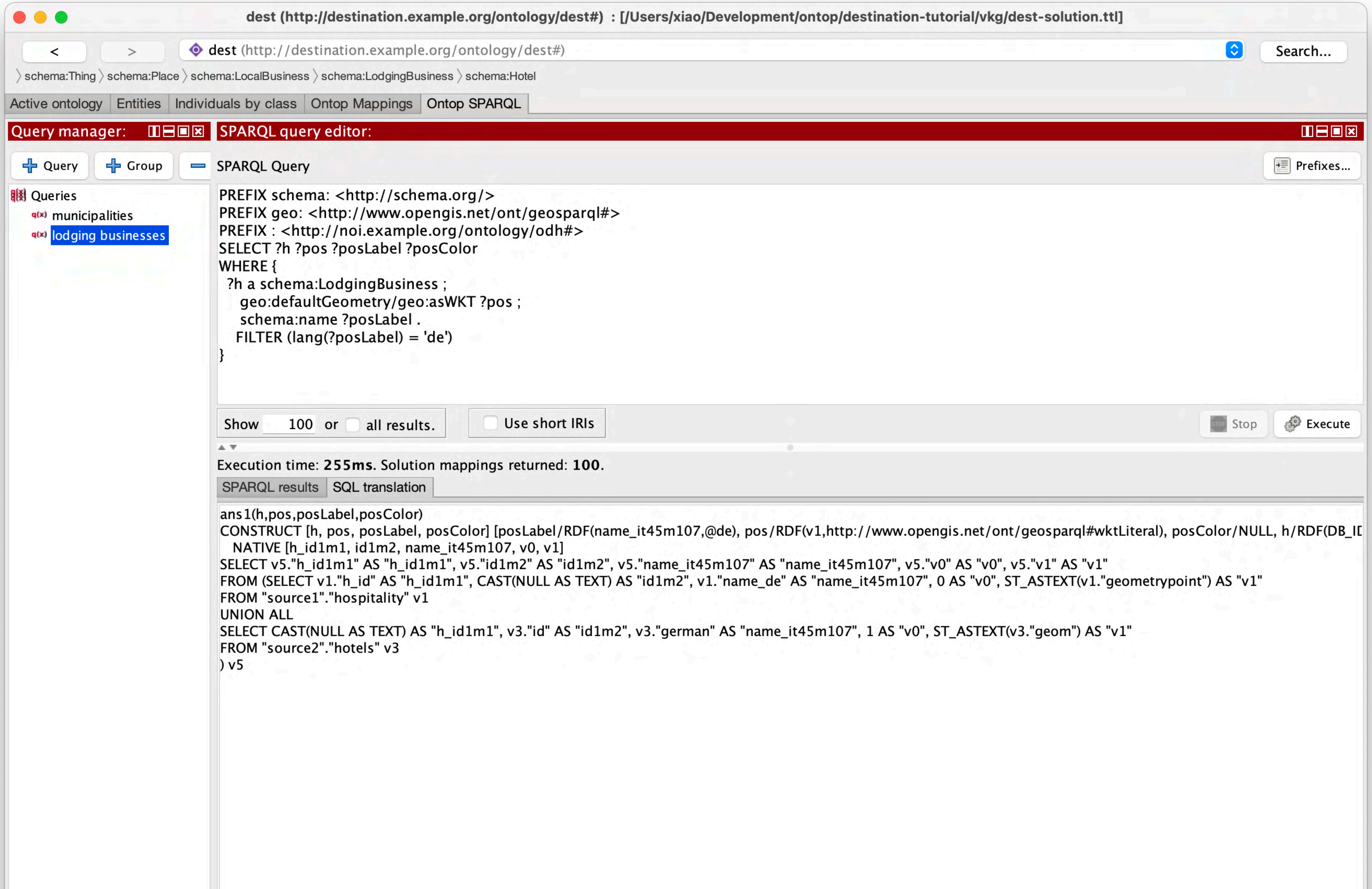
```
PREFIX schema: <http://schema.org/>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX : <http://noi.example.org/ontology/odh#>
SELECT ?h ?pos ?posLabel ?posColor
WHERE {
 ?h a schema:LodgingBusiness ;
    geo:defaultGeometry/geo:asWKT ?pos ;
    schema:name ?posLabel .
   FILTER (lang(?posLabel) = 'de')
}
```

Show 100 or □ all results. □ Use short IRIs

Execution time: **255ms**. Solution mappings returned: **100**.

h	pos	posLabel	posColor
<http://destination.example.org/data/sour...	"POINT Z (11.358798 46.499543 0)"^^<h...	"Appartamento Conciapelli"@de	
<http://destination.example.org/data/sour...	"POINT Z (12.245 46.8369 1400)"^^<http...	"Haspingerhof – Specker"@de	
<http://destination.example.org/data/sour...	"POINT Z (11.812105 46.807976 778)"^^...	"Camping Gisser Vitha Hotels"@de	
<http://destination.example.org/data/sour...	"POINT Z (11.895599 46.772921 930)"^^...	"Schlosshäuslhof"@de	
<http://destination.example.org/data/sour...	"POINT Z (11.9547 46.9191 865)"^^<http...	"Residence Tuberis"@de	
<http://destination.example.org/data/sour...	"POINT Z (11.6775 46.6969 1000)"^^<htt...	"Birbamerhof"@de	
<http://destination.example.org/data/sour...	"POINT Z (11.163071 46.671773 324)"^^...	"Holiday house merano Appartamento Porti...	
<http://destination.example.org/data/sour...	"POINT Z (11.1472 46.6431 330)"^^<http...	"Residence Haus Waldner"@de	
<http://destination.example.org/data/sour...	"POINT Z (0 0 0)"^^<http://www.opengis.n...	"Aurturist Appartement 142"@de	
<http://destination.example.org/data/sour...	"POINT Z (12.172 46.7373 1154)"^^<http...	"Schneiderhof"@de	
<http://destination.example.org/data/sour...	"POINT Z (11.5828 46.4078 1650)"^^<htt...	"Residence Chris Appart"@de	
<http://destination.example.org/data/sour...	"POINT Z (11.4444 46.9386 1098)"^^<htt...	"Appartements Monica"@de	
<http://destination.example.org/data/sour...	"POINT Z (11.1014 46.6864 550)"^^<http...	"Bauernhof Gasserhof"@de	
<http://destination.example.org/data/sour...	"POINT Z (11.1214 46.6931 0)"^^<http://...	"Ferienwohnungen Ebeneich"@de	
<http://destination.example.org/data/sour...	"POINT Z (11.9575 46.9175 865)"^^<http...	"Haus Holzer"@de	
<http://destination.example.org/data/sour...	"POINT Z (11.9306 46.6978 1201)"^^<htt...	"Parc Hotel Posta"@de	
<http://destination.example.org/data/sour...	"POINT Z (11.966633 46.998995 1300)"^...	"Golserhof"@de	
<http://destination.example.org/data/sour...	"POINT Z (12.223652 46.728385 1219)"^...	"Distaccamento Aeroportuale – Food Service...	
<http://destination.example.org/data/sour...	"POINT Z (11.811762 46.807756 780)"^^...	"Hotel Gisser Vitha Hotels"@de	
<http://destination.example.org/data/sour...	"POINT Z (11.9544 46.9186 865)"^^<http...	"Hotel Tubris"@de	
<http://destination.example.org/data/sour...	"POINT Z (11.199958 46.668232 600)"^^...	"Castello Schloss Labers"@de	

SPARQL Query Answering

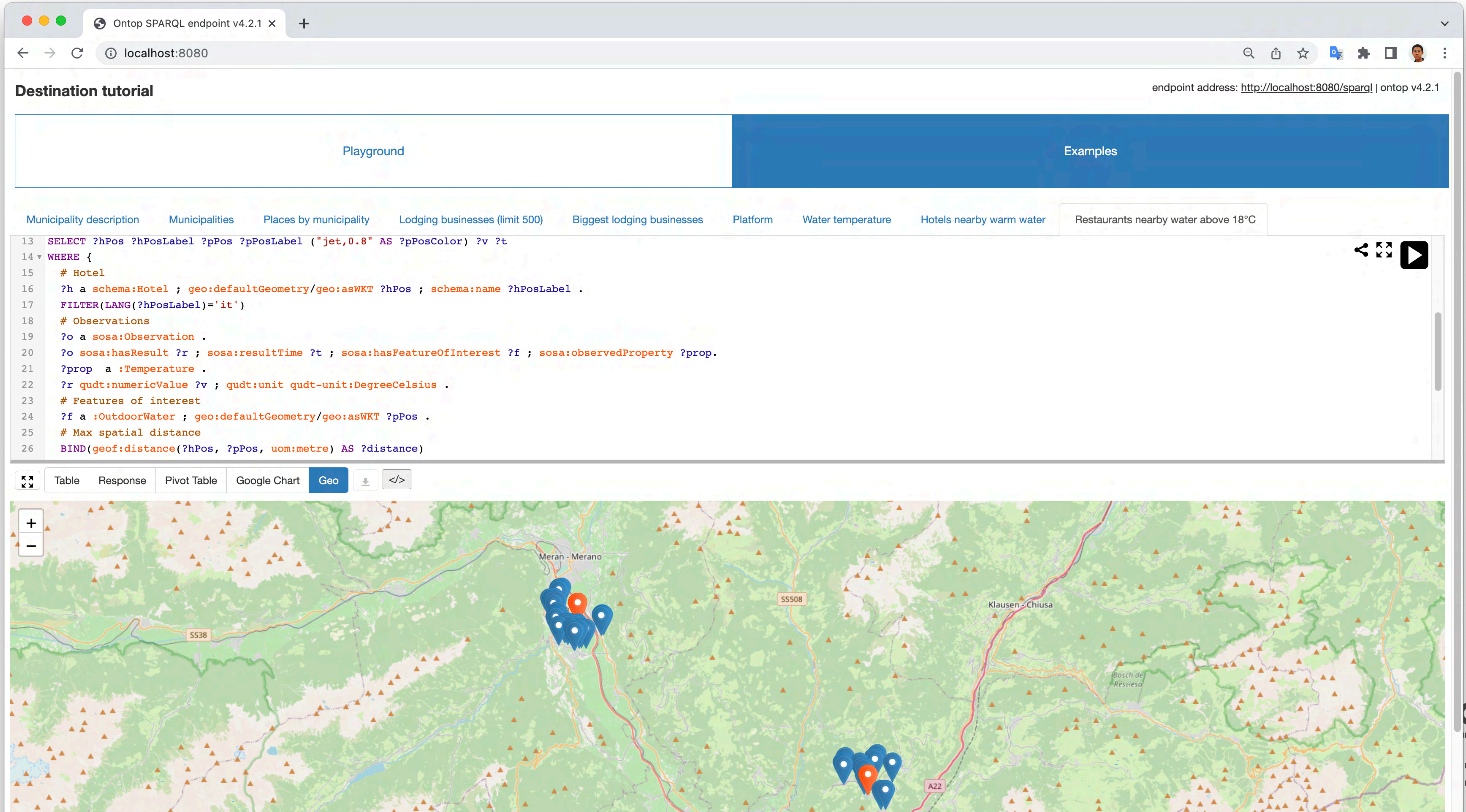

dest (http://destination.example.org/ontology/dest#) : [/Users/xiao/Development/ontop/destination-tutorial/vkg/dest-solution.ttl]
<
>
dest (http://destination.example.org/ontology/dest#)
Search...
schema:Thing 〉 schema:Place 〉 schema:LocalBusiness 〉 schema:LodgingBusiness 〉 schema:Hotel
Active ontology
Entities
Individuals by class
Ontop Mappings
Ontop SPARQL
Query manager:
SPARQL query editor:
Query
Group
SPARQL Query
Prefixes...
Queries
municipalities
lodging businesses
Show 100 or □ all results.
□ Use short IRIs
Stop
Execute
SPARQL results
SQL translation

```
PREFIX schema: <http://schema.org/>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX : <http://noi.example.org/ontology/odh#>
SELECT ?h ?pos ?posLabel ?posColor
WHERE {
 ?h a schema:LodgingBusiness ;
    geo:defaultGeometry/geo:asWKT ?pos ;
    schema:name ?posLabel .
   FILTER (lang(?posLabel) = 'de')
}
```

Execution time: **255ms**. Solution mappings returned: **100**.

```
ans1(h,pos,posLabel,posColor)
CONSTRUCT [h, pos, posLabel, posColor] [posLabel/RDF(name_it45m107,@de), pos/RDF(v1,http://www.opengis.net/ont/geosparql#wktLiteral), posColor/NULL, h/RDF(DB_ID
   NATIVE [h_id1m1, id1m2, name_it45m107, v0, v1]
SELECT v5."h_id1m1" AS "h_id1m1", v5."id1m2" AS "id1m2", v5."name_it45m107" AS "name_it45m107", v5."v0" AS "v0", v5."v1" AS "v1"
FROM (SELECT v1."h_id" AS "h_id1m1", CAST(NULL AS TEXT) AS "id1m2", v1."name_de" AS "name_it45m107", 0 AS "v0", ST_ASTEXT(v1."geometrypoint") AS "v1"
FROM "source1"."hospitality" v1
UNION ALL
SELECT CAST(NULL AS TEXT) AS "h_id1m1", v3."id" AS "id1m2", v3."german" AS "name_it45m107", 1 AS "v0", ST_ASTEXT(v3."geom") AS "v1"
FROM "source2"."hotels" v3
) v5
```

SPARQL Endpoint



Ontop SPARQL endpoint v4.2.1
localhost:8080
Destination tutorial
endpoint address: http://localhost:8080/sparql | ontop v4.2.1
Playground
Examples
Municipality description
Municipalities
Places by municipality
Lodging businesses (limit 500)
Biggest lodging businesses
Platform
Water temperature
Hotels nearby warm water
Restaurants nearby water above 18°C

```
SELECT ?hPos ?hPosLabel ?pPos ?pPosLabel ("jet,0.8" AS ?pPosColor) ?v ?t
WHERE {
  # Hotel
  ?h a schema:Hotel ; geo:defaultGeometry/geo:asWKT ?hPos ; schema:name ?hPosLabel .
  FILTER(LANG(?hPosLabel)='it')
  # Observations
  ?o a sosa:Observation .
  ?o sosa:hasResult ?r ; sosa:resultTime ?t ; sosa:hasFeatureOfInterest ?f ; sosa:observedProperty ?prop.
  ?prop  a :Temperature .
  ?r qudt:numericValue ?v ; qudt:unit qudt-unit:DegreeCelsius .
  # Features of interest
  ?f a :OutdoorWater ; geo:defaultGeometry/geo:asWKT ?pPos .
  # Max spatial distance
  BIND(geof:distance(?hPos, ?pPos, uom:metre) AS ?distance)
```

Take-home messages

- KG is a very simple but powerful data structure
- KGs give semantics to the data: things, not strings
- KGs have been applied in many applications
- VKGs are by now a mature technology to address the challenges related to data access and integration.
- It has been well-investigated and applied in many different scenarios mostly for the case of relational data sources.

Further readings

1. Tim Berners-Lee. **Information management: A proposal**. Technical report, 1989.
2. Tim Berners-Lee, James Hendler, and Ora Lassila. **The semantic web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities**. Scientific American, 2001.
3. A. Hogan, et al. **Knowledge graphs**. ACM Computer Survey, 54(4):71:1–71:37, 2021.
4. G. Xiao, et al. **Ontology-based data access: A survey**. IJCAI, 2018.
5. G. Xiao, Linfang Ding, Benjamin Cogrel, and Diego Calvanese. **Virtual knowledge graphs: An overview of systems and use cases**. Data Intelligence, 1:201–223, 2019.

Thank you for your attention

Questions?

Guohui Xiao – guohui.xiao@ontopic.ai