Shaping Knowledge Graphs
ISWC'24 Tutorial

Jose Emilio Labra Gayo
WESO Research group
University of Oviedo, Spain

About me…

Main researcher at WESO (Web Semantics Oviedo)

Some books:

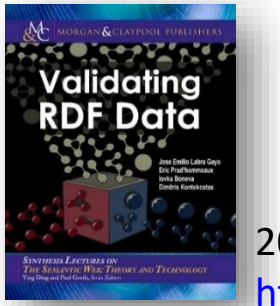
- "*Web semántica*" (in Spanish), 2012
- "*Validating RDF data*", 2017
- "Knowledge Graphs", 2021

…and software:

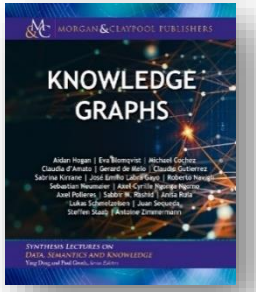
- SHaclEX (Scala library, implements ShEx & SHACL)
- RDFShape (RDF playground)
- rudof (RDF & Shapes library in Rust)

http://labra.weso.es

2012



2017 HTML version:
http://book.validatingrdf.com



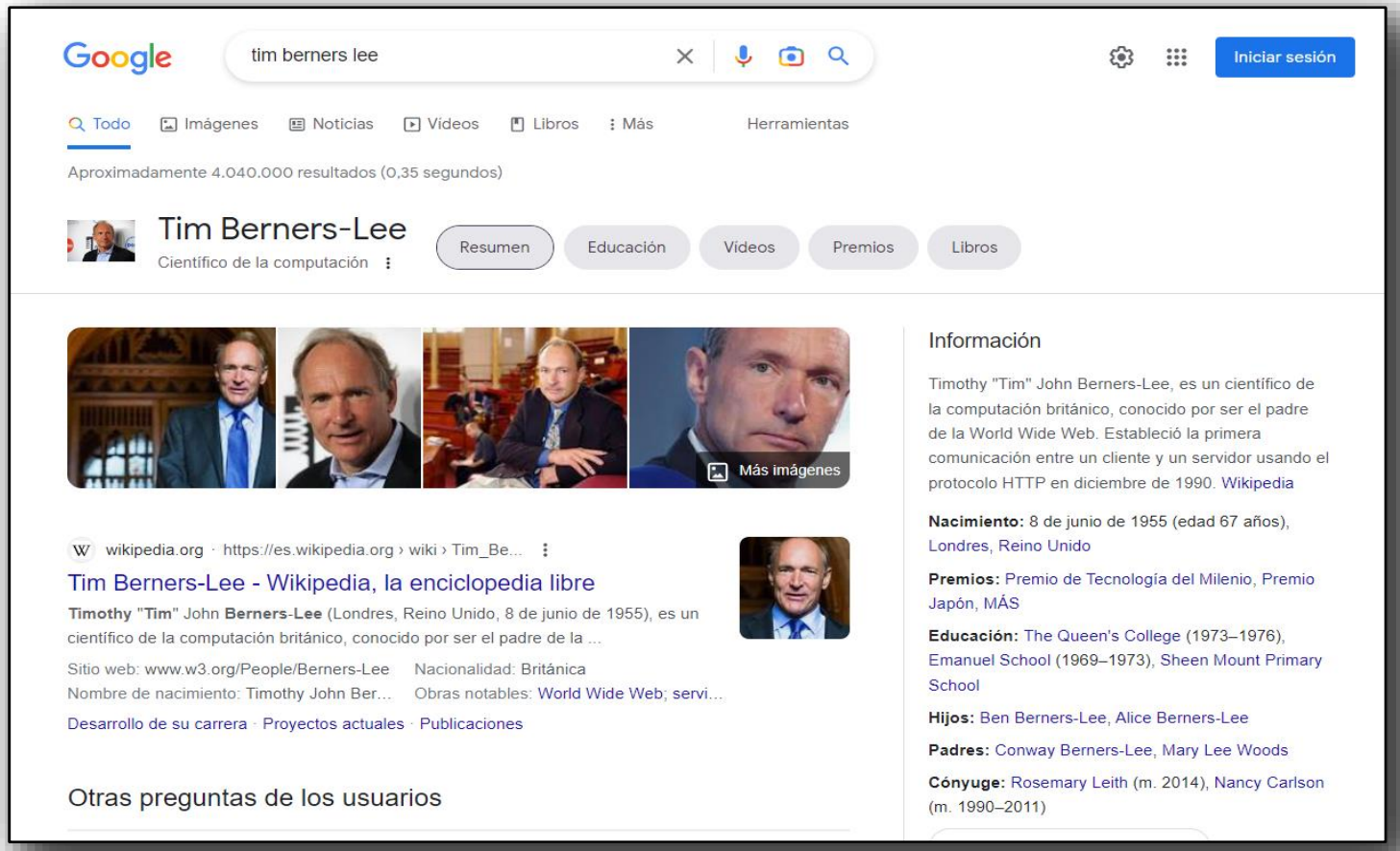
2021, HTML version
https://kgbook.org/

Contents

- Introduction to Knowledge graphs
- Types of Knowledge Graphs:
  - RDF, Property graphs, Wikibase, RDF-Star
- Shaping RDF: ShEx & SHACL
- Shaping Wikibase and Wikidata graphs
- Shaping Property Graphs
- Shaping RDF-Star
- Applications:
  - Inferring shapes from data, Knowledge Graphs Subsets, etc.

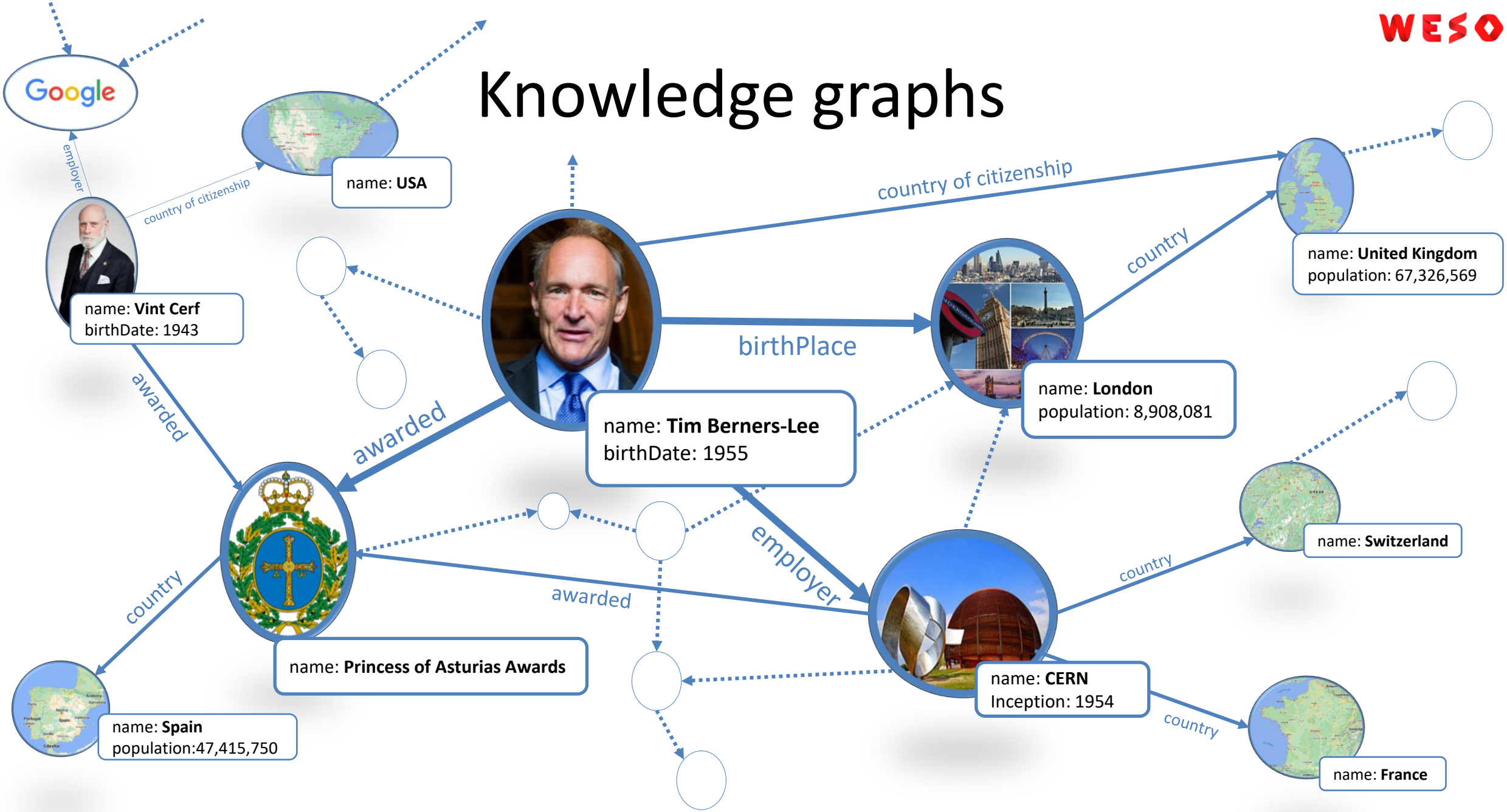
Knowledge Graphs

Current notion of Knowledge Graphs, popular after Google, 2012*



Link: https://www.blog.google/products/search/introducing-knowledge-graph-things-not/

Knowledge graphs



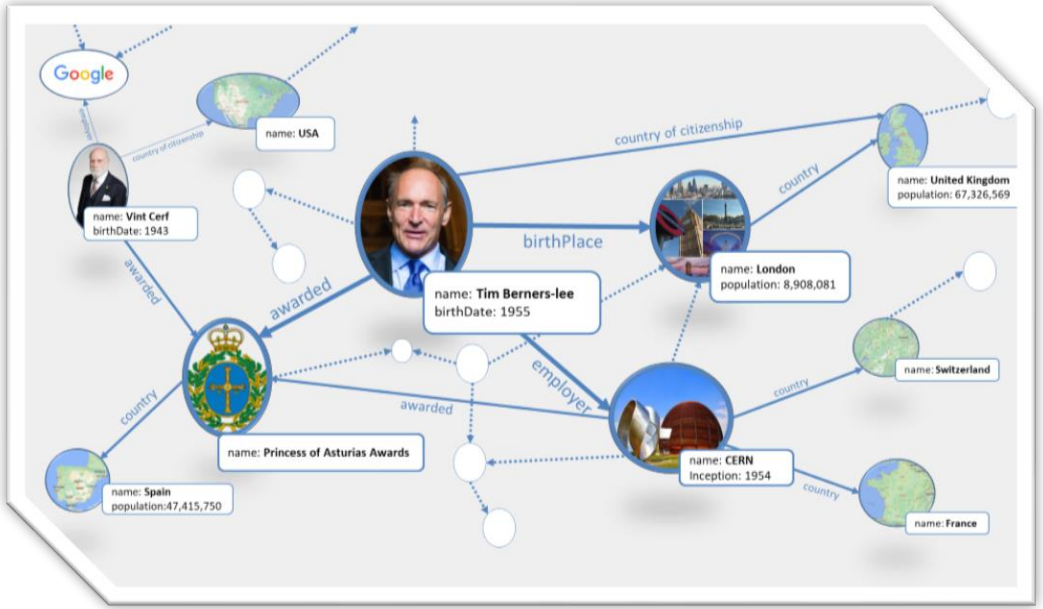
Knowledge graphs

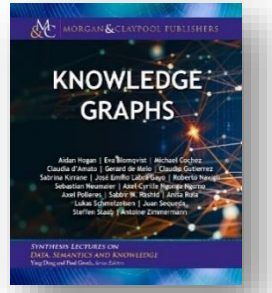
Knowledge graph = *a **graph of data***

*intended to accumulate and convey **knowledge** of the real world*

*whose nodes represent **entities** of interest and*

*whose edges represent **relations** between these entities.*





https://kgbook.org/

Applications of Knowledge graphs

Improve search results

Question answering

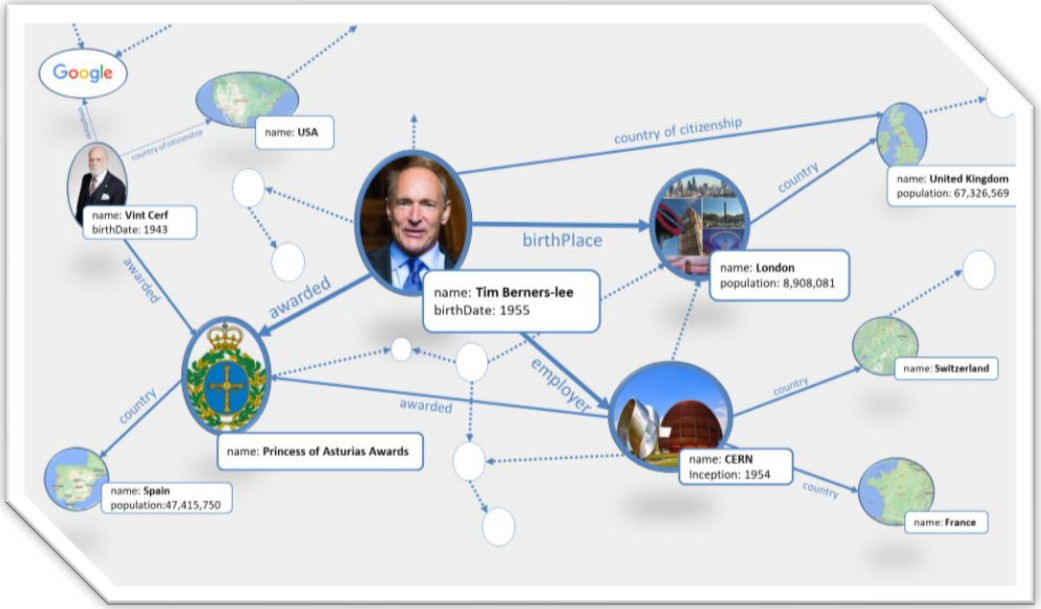
Data governance

Handling heterogenous data

Recommender systems

Chatbots and NLP

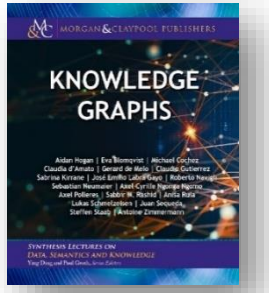
. . .



Contents

Types of Knowledge graphs

- Open Knowledge graphs
  - Cross-domain: Wikidata, Dbpedia, Freebase, YAGO, ...
  - Domain specific
    - Academic: Open citations, SciGraph, Microsoft Academic Knowledge Graph, ...
    - Life sciences: UniProt, PubChem, PDB, ...
    - Government: EU Knowledge graph, ...
    - ...
- Enterprise Knowledge graphs
  - Web search: Google, Bing...
  - Commerce: AirBnb, Amazon, eBay, Uber,...
  - Social networks: Linkedin, Facebook,...
  - Finance: Banca d'Italia, Bloomberg, Wells Fargo, Capital One,...
  - ...



https://kgbook.org/

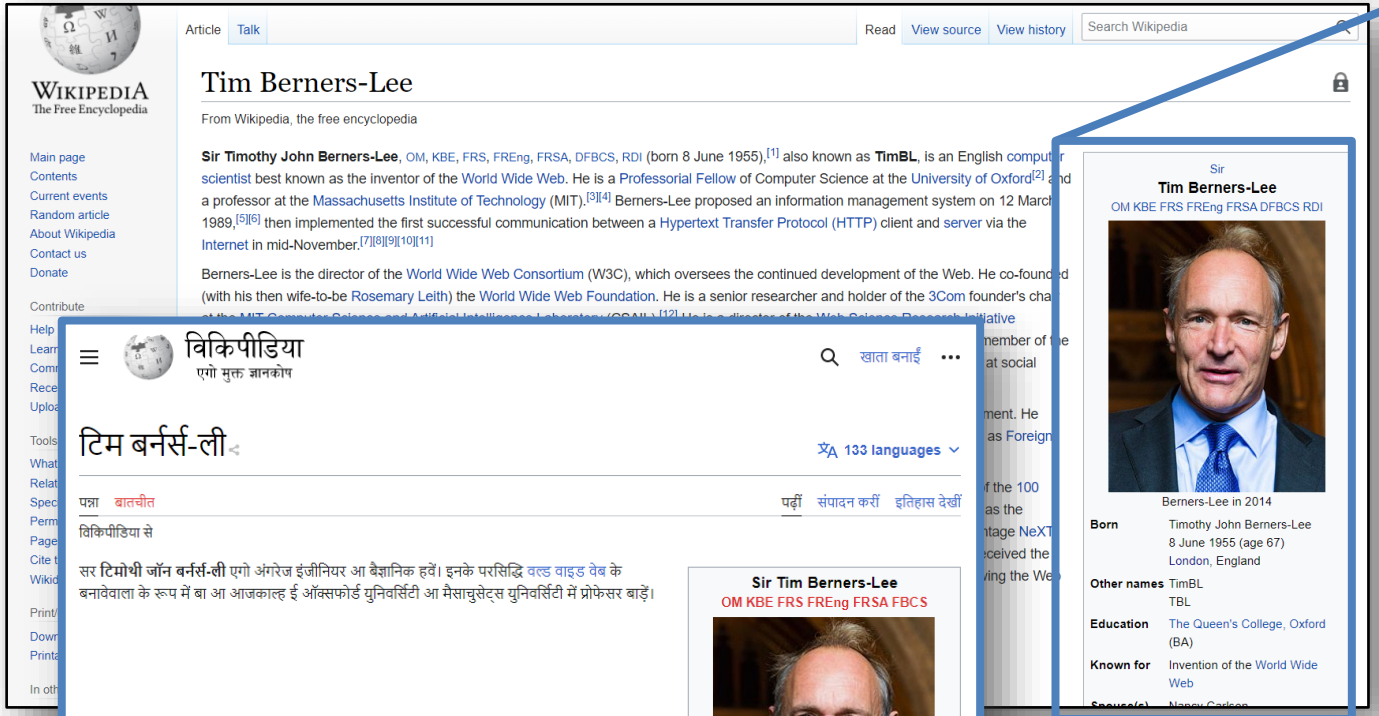
Wikidata as an example

Wikidata created in 2012 as a collaborative knowledge graph

- Initial goal:
  - Support multilingual infoboxes in Wikipedia

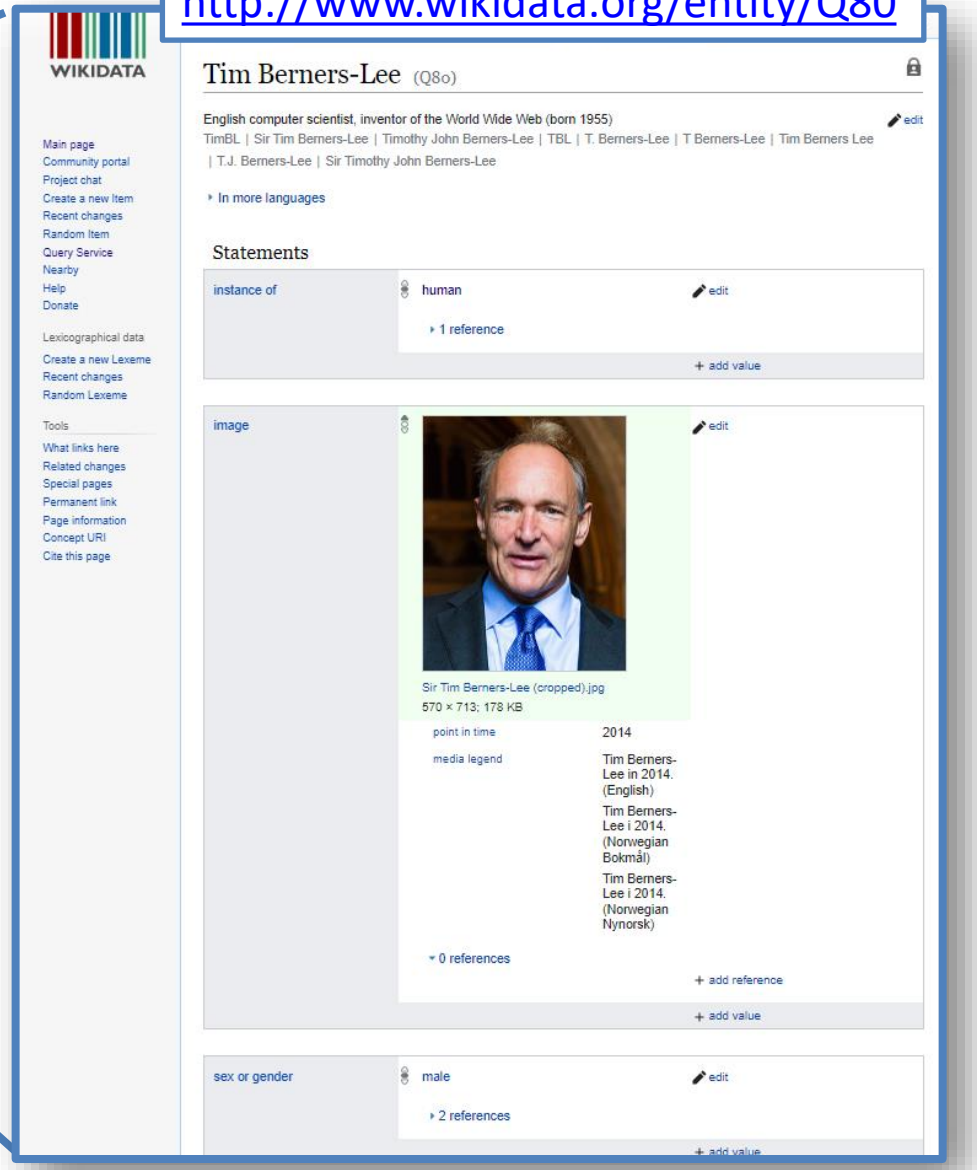
Wikidata as an example

English Wikipedia page of Tim Berners-Lee



Bihari Wikipedia of Tim Berners-Lee

http://www.wikidata.org/entity/Q80



Wikidata: some features

Collaborative: anyone can edit

Free and open license

Currently (01/2023): 101m items, 1,8b edits

Co-edited by humans and bots: 23k active users, 343 bots

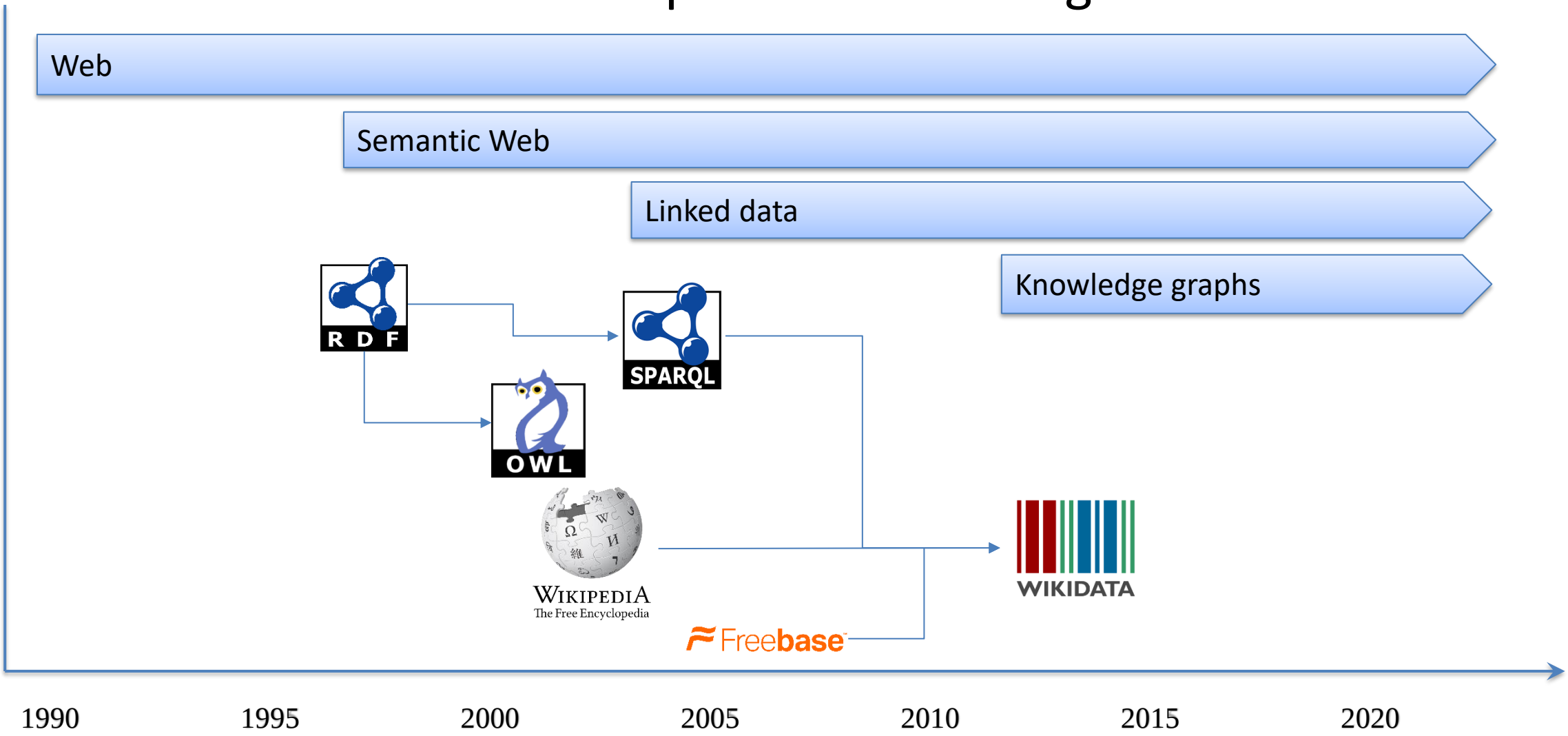
Open Wikidata Query Service: Public SPARQL endpoint

Dumps freely available: 109Gb compressed

Software that supports Wikidata = Wikibase

Evolution

Timeline with some concepts and technologies...



Knowledge Graphs models

3 popular knowledge graphs models

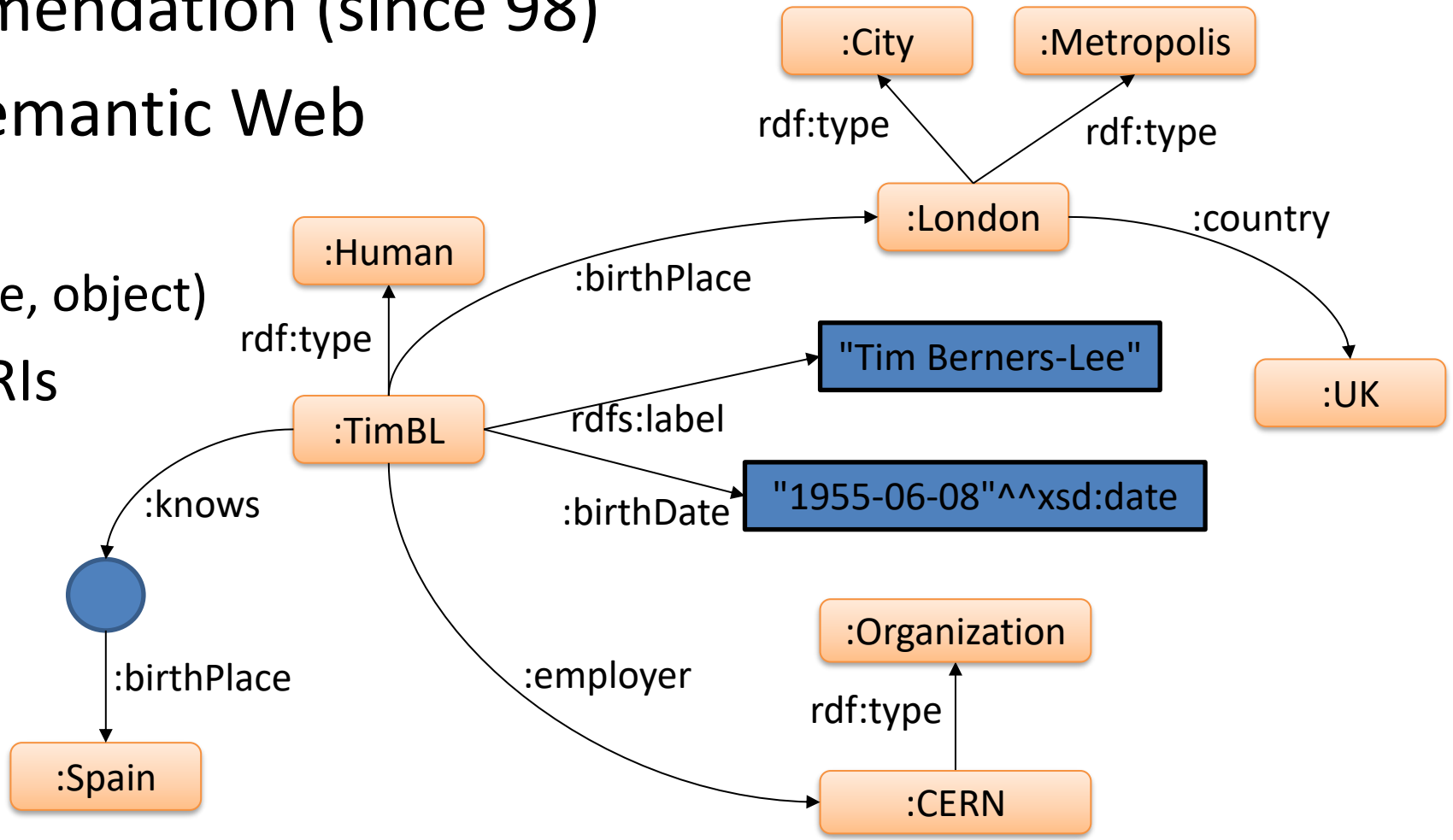
- RDF based
- Property graphs
- Wikibase graphs

RDF graphs

RDF = W3C recommendation (since 98)

Lingua franca of Semantic Web

- Based on triples
  - (subject, predicaje, object)
- Most nodes are URIs
  - Interoperability



RDF ecosystem

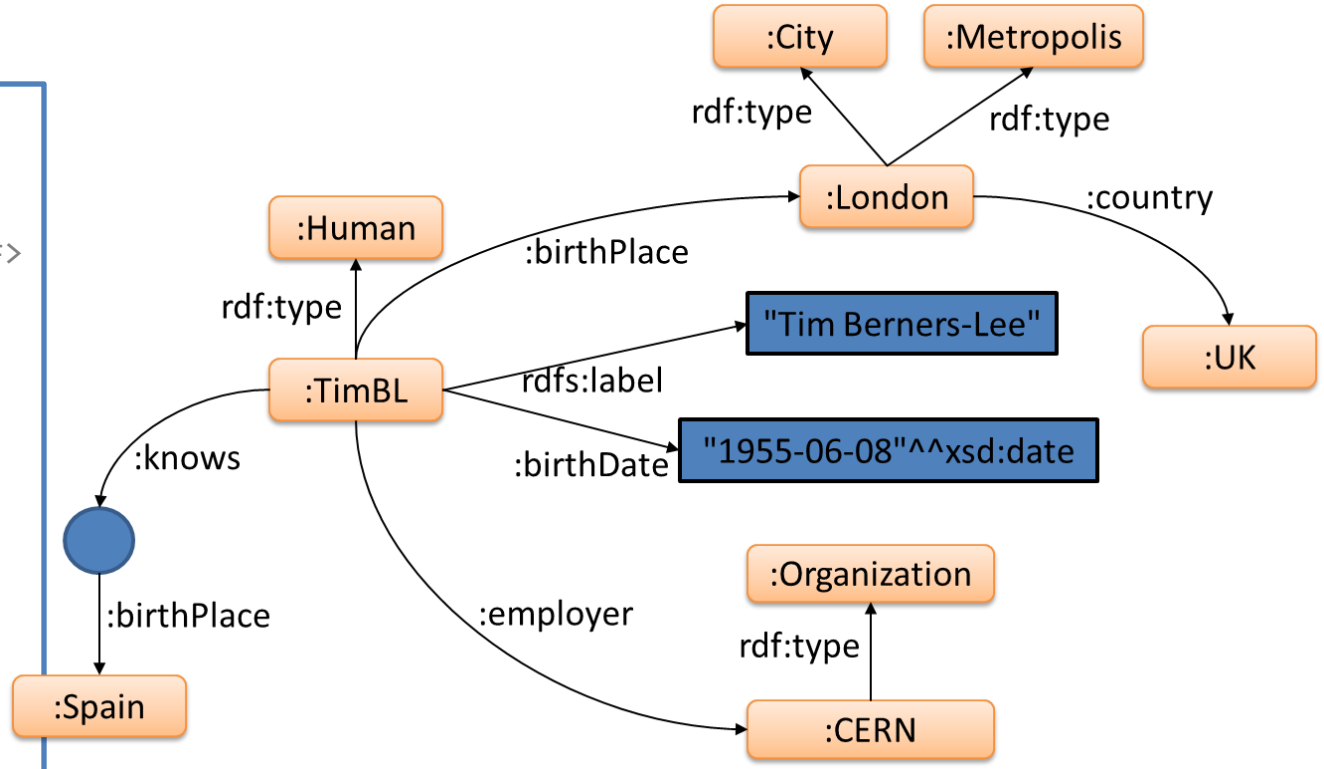
One data model, several syntaxes: Turtle, N-Triples, JSON-LD

Vocabularies: RDF Schema, OWL, SKOS, etc.

Turtle

```
prefix :       <http://example.org/>
prefix rdfs:   <http://www.w3.org/2000/01/rdf-schema#>
prefix xsd:    <http://www.w3.org/2001/XMLSchema#>
prefix rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

:timbl  rdf:type     :Human ;
        :birthPlace  :london ;
        rdfs:label   "Tim Berners-Lee" ;
        :birthDate   "1955-06-08"^^xsd:date ;
        :employer    :CERN ;
        :knows       _:1    .
:london rdf:type     :City, :Metropolis ;
        :country     :UK .
:CERN   rdf:type     :Organization .
_:1     :birthPlace  :Spain .
```



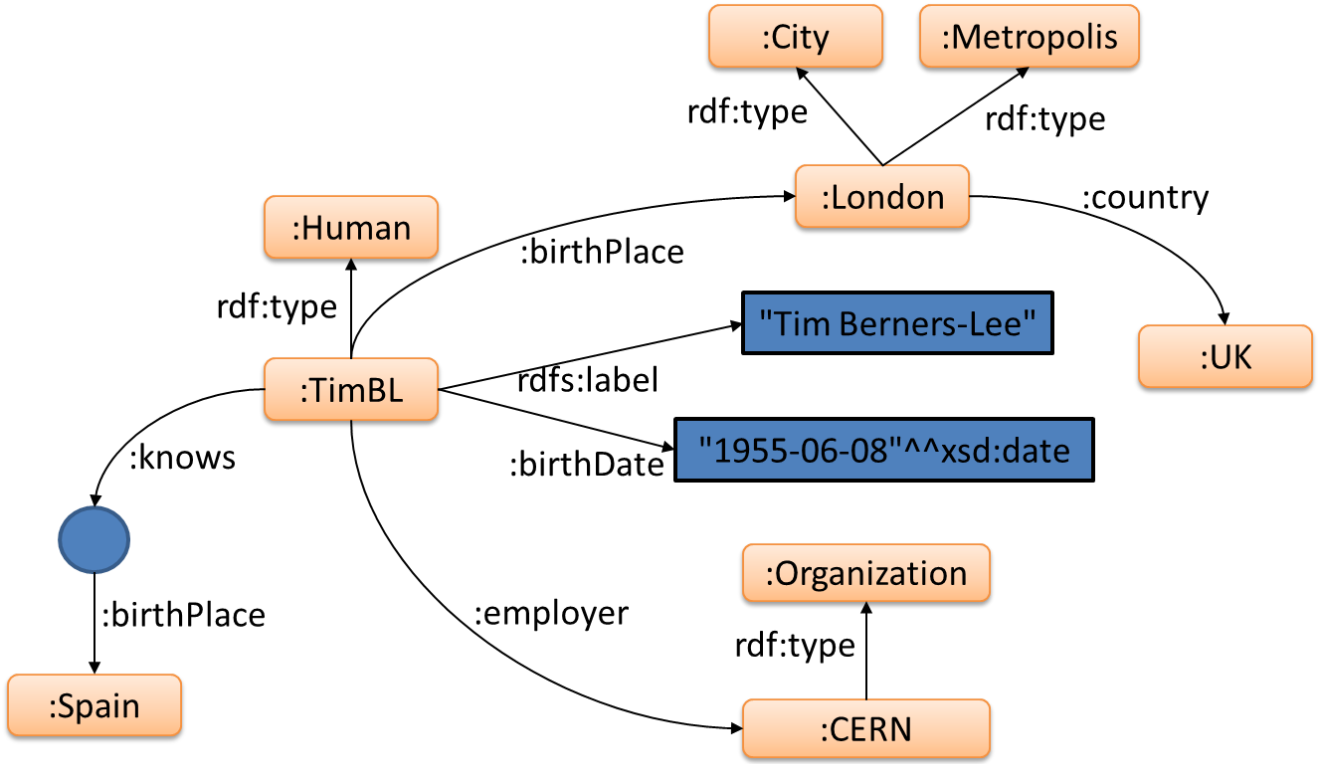
RDF ecosystem: SPARQL

SPARQL is an RDF query language and protocol

It enables the creation of SPARQL endpoints

```
select ?person ?date ?country where {
    ?person :birthDate  ?date .
    ?person :birthPlace ?p .
    ?p      :country    ?country
}
```

?person	?date	?country
:timbl	1955-06-08	:UK



RDF1.2 (RDF-Star)

Triple terms

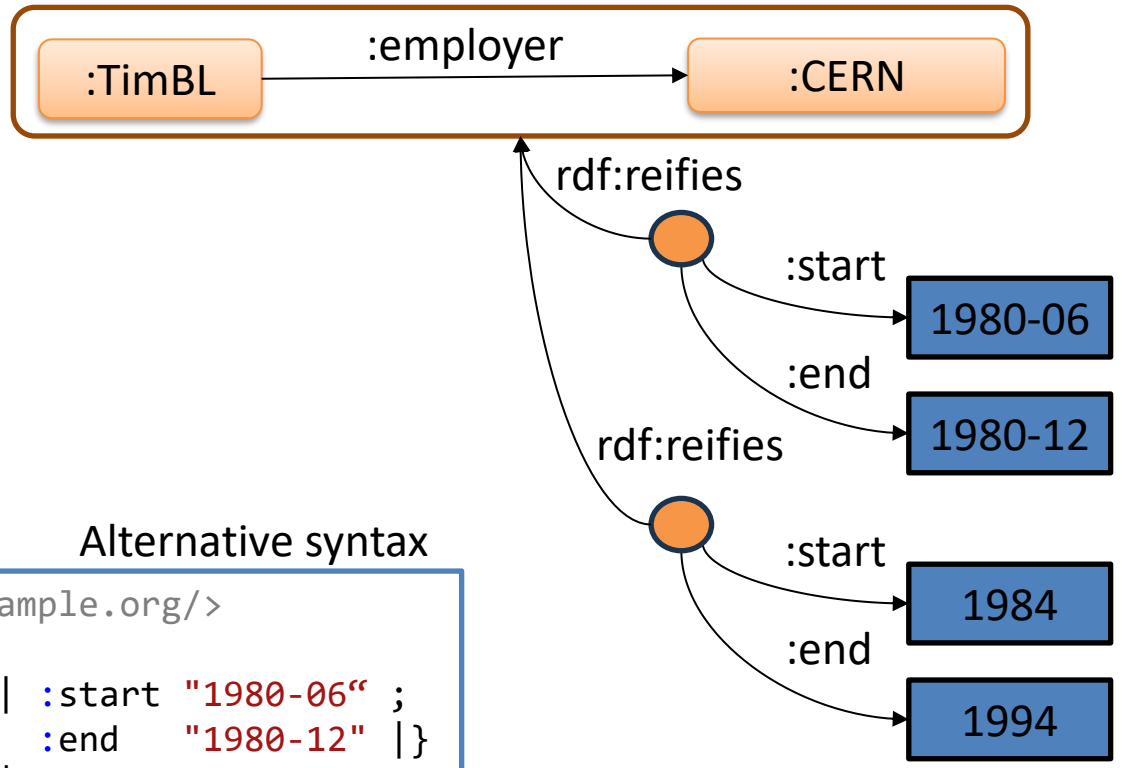
Add statements about triples

Reifiers

```
prefix :      <http://example.org/>

_:r1 rdf:reifies << :timbl :employer :CERN >> .
_:r1 :start "1980-06";
     :end   "1980-12" .
```

```
_:r1 rdf:reifies << :timbl :employer :CERN >> .
_:r2 :start "1984;
     :end   "1994" .
```



Alternative syntax

```
prefix :      <http://example.org/>

:timbl :employer :CERN {| :start "1980-06“ ;
                          :end   "1980-12" |}
                       {| :start "1984“ ;
                          :end   "1994" |} .
```

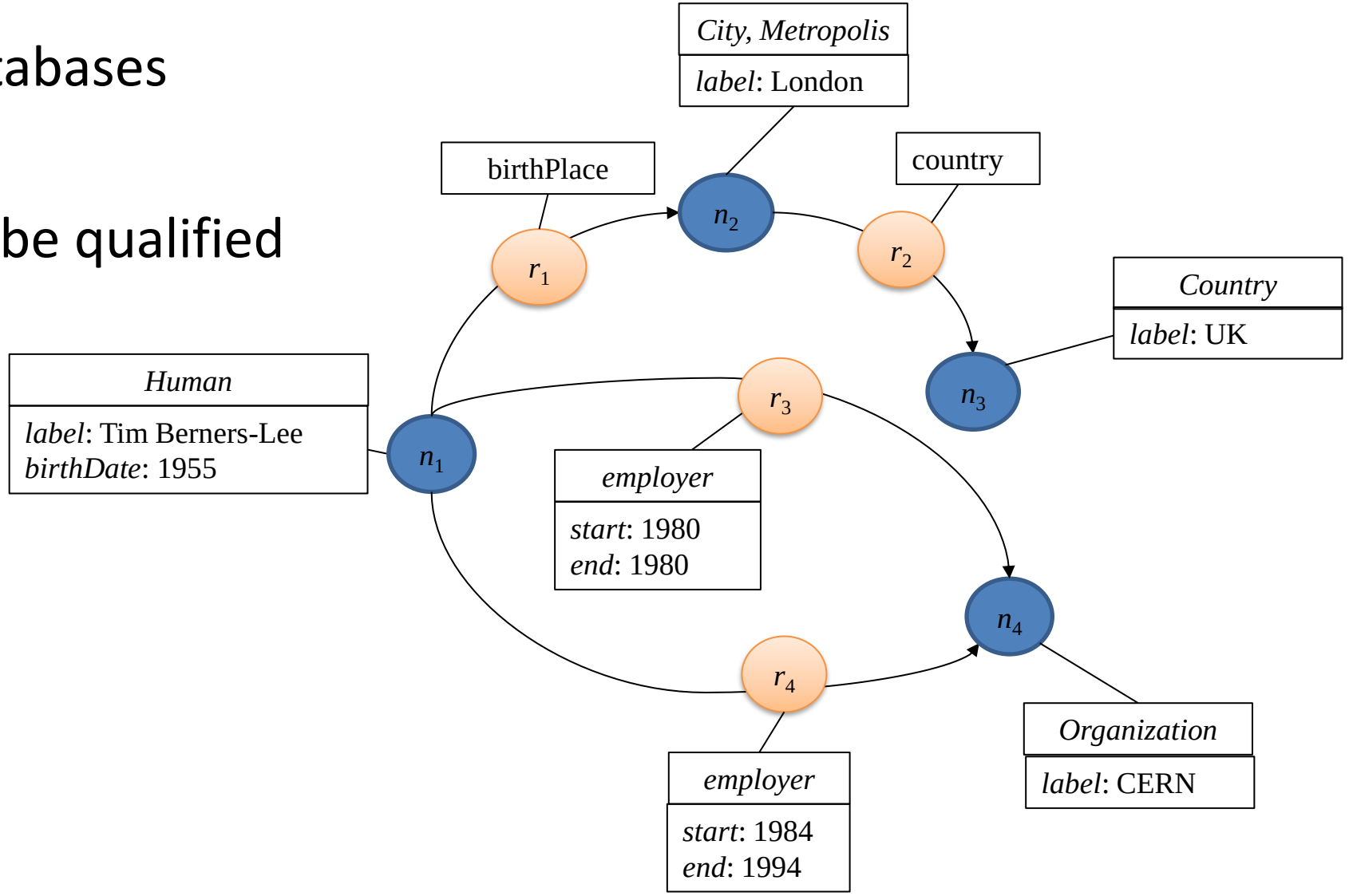
Property graphs

Popularized by graph databases

Example: Neo4j

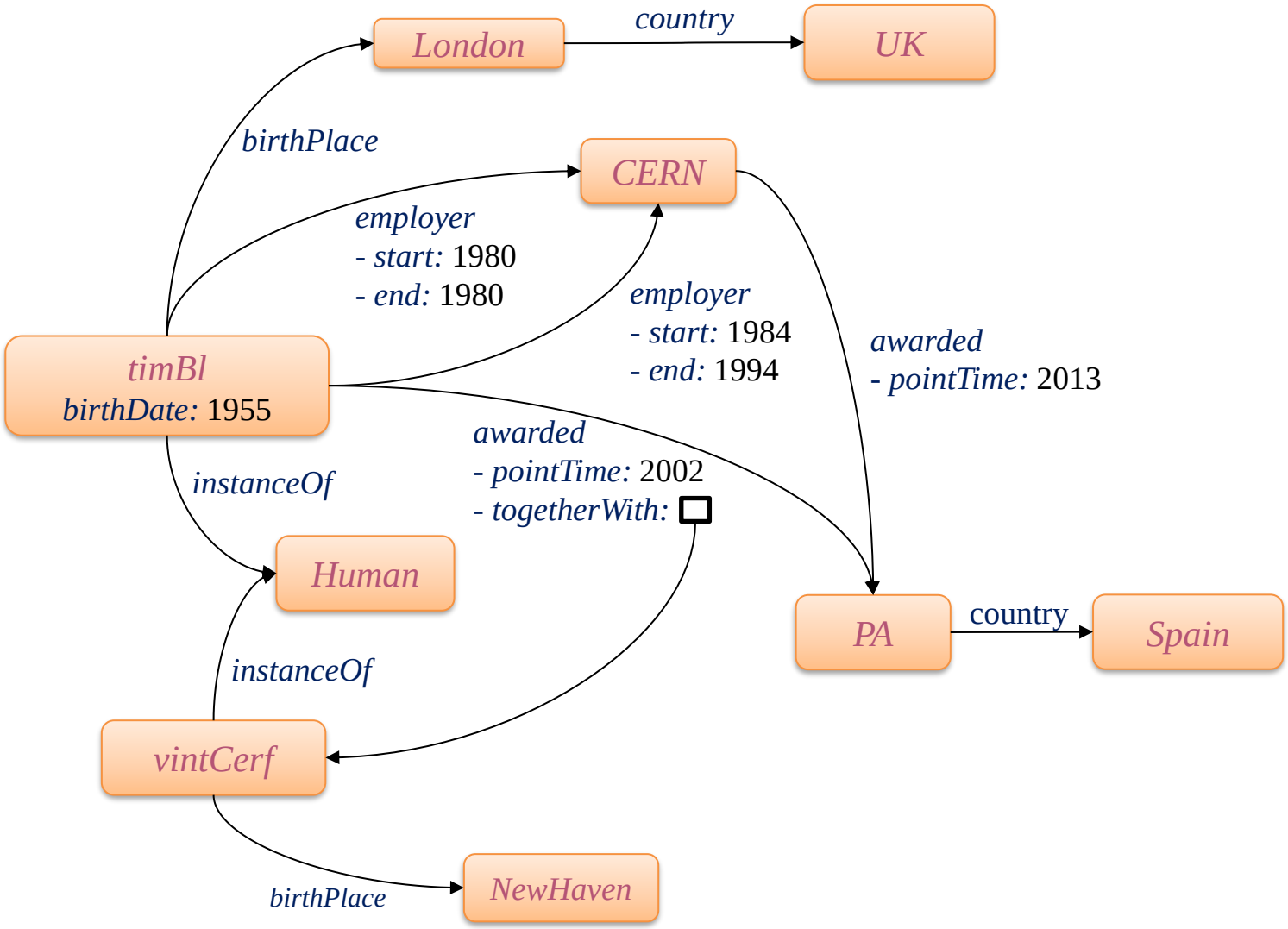
Nodes and relations can be qualified

Nodes can have types



Wikibase graphs

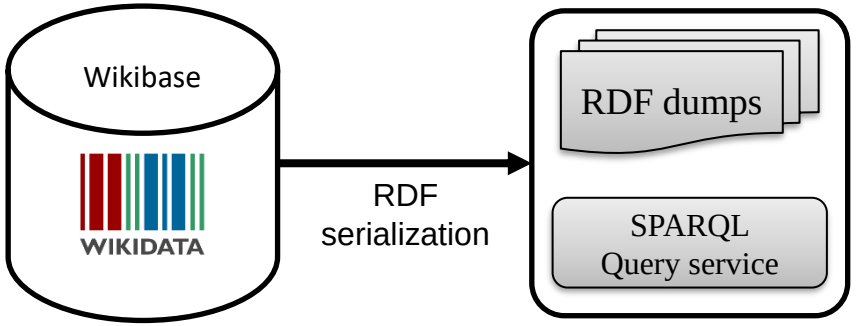
Popularized by Wikidata
Wikibase = software supporting Wikidata
The values can be nodes in the graph
Example:
Tim Berners Lee
http://www.wikidata.org/entity/Q80



Wikibase graphs and SPARQL

Wikibase graphs generate RDF serializations for each item

SPARQL endpoint and Query service available




```
select ?name ?date?country where {
 wd:Q80 wdt:P1559 ?name .
 wd:Q80 wdt:P569  ?date .
 wd:Q80 wdt:P19   ?place .
 ?place wdt:P17   ?country
}
```

?name	?date	?country
Tim Berners-lee	1955-06-08	:UK

Try it: https://w.wiki/5yGu

Contents

- Introduction to Knowledge graphs
- Types of Knowledge Graphs:
  - RDF, Property graphs, Wikibase, RDF-Star
- Shaping RDF: ShEx & SHACL
- Shaping other types of Knowledge graphs:
  - Wikibase and Wikidata graphs
  - Property Graphs
  - RDF-Star
- Applications:
  - Inferring shapes from data, Knowledge Graphs Subsets, etc.