Introduction to Linked Data and SPARQL

Part 1 of the Linked Data Specialization

Lívia Ruback liviaruback@gmail.com

About

Lívia Ruback

P.h.D in Informatics - Databases
PUC-Rio (Pontifical Catholic University of Rio de Janeiro)

liviaruback@gmail.com

The Linked Data Specialization

Part 1 - Introduction to Linked Data and SPARQL

- Linked Data
- Open Data
- SPARQL

Part 2 - Graph Databases and visualiza

- Graph concepts
- Graph storing and visualization with Neo4J

Part 3 - Inferences and Ontologies

- OWL - Web Ontology Language
- Semantics of OWL

Part 4 - Federated queries in the Linked Data

- Federated search
- Federated SPARQL queries

The Linked Data Specialization

Part 1 - Introduction to Linked Data and SPARQL

- Linked Data
- Open Data
- SPARQL

Part 2 - Graph Databases and visualization

- Graph concepts
- Graph storing and visualization with Neo4J

Part 3 - Inferences and Ontologies

- OWL - Web Ontology Language
- Semantics of OWL

Part 4 - Federated queries in the Linked Data

- Federated search
- Federated SPARQL queries

Summary

- Introduction
- Concepts
  - Linked Data
  - Open Data
  - RDF
  - SPARQL query language
- Practice: SPARQL examples
- References

Summary

- Introduction
- Concepts
  - Linked Data
    - Open Data
    - RDF
    - SPARQL query language
- Practice: SPARQL examples
- References

Introduction

HTML - HyperText Markup Language

meaning and structure of web content

HTTP - Hypertext Tranfer Protocol

request-response protocol

URLs - Uniform Resource Locator

location of a web resource

Tim Berners-Lee

"The original idea of the web was that it should be a ***collaborative space*** *where you can communicate through* ***sharing information****"*

https://portalmarketing.digital/Tim-Berners-Lee

http://linkeddata.org/
https://www.w3.org/DesignIssues/LinkedData.html

Introduction

Web of documents

connects documents

Linked Data

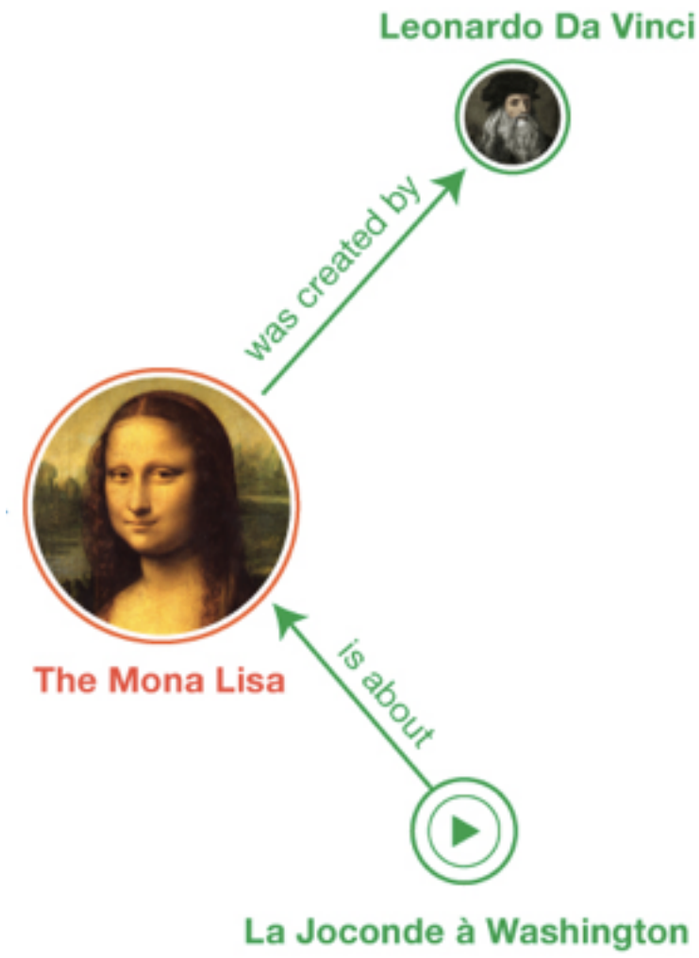
Web of documents

connects documents

Publication of isolated datasets as **interlinked**, **reusable** data graphs

Web of data

connects data!



person

place

event

art pieces

Linked Data rules

Rule 1 Use URIs as names for things

en.wikipedia.org/wiki/Stanford_University

Rule 2 Use HTTP URIs

https://en.wikipedia.org/wiki/Stanford_University

Rule 3 Provide useful information using RDF and SPARQL

Rule 4 Include links to other URIs

T.B-L TED talk: https://www.youtube.com/watch?v=OM6XIICm_qo

Summary

- Introduction
- Concepts
  - Linked Data
  - **Open Data**
  - RDF
  - SPARQL query language
- Practice: SPARQL examples
- References

Open Data x Linked Open Data

Data freely available to everyone use without restrictions

- open-source software
- open government
- open access
- open science

when linked to other open data

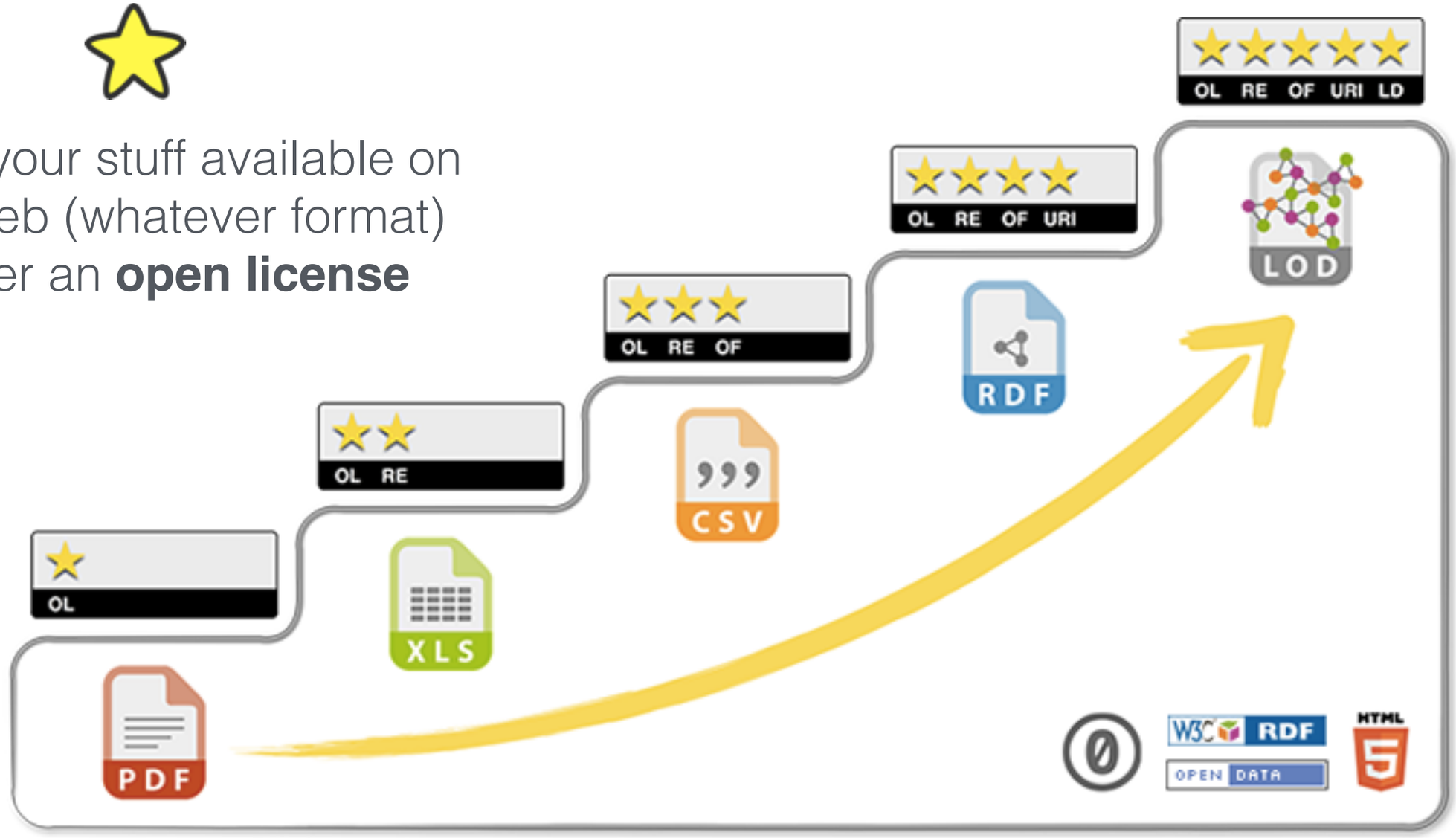
https://www.ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/

https://www.ontotext.com/linked-open-data-cultural-heritage/

Open data

The 5 stars open data

make your stuff available on the Web (whatever format) under an **open license**



https://5stardata.info/en/

Open data

The 5 stars open data

make it available as **structured data** (e.g., Excel instead of image scan of a table)

https://5stardata.info/en/

Open data

The 5 stars open data

make it available in a **non-proprietary open format** (e.g., CSV instead of Excel)

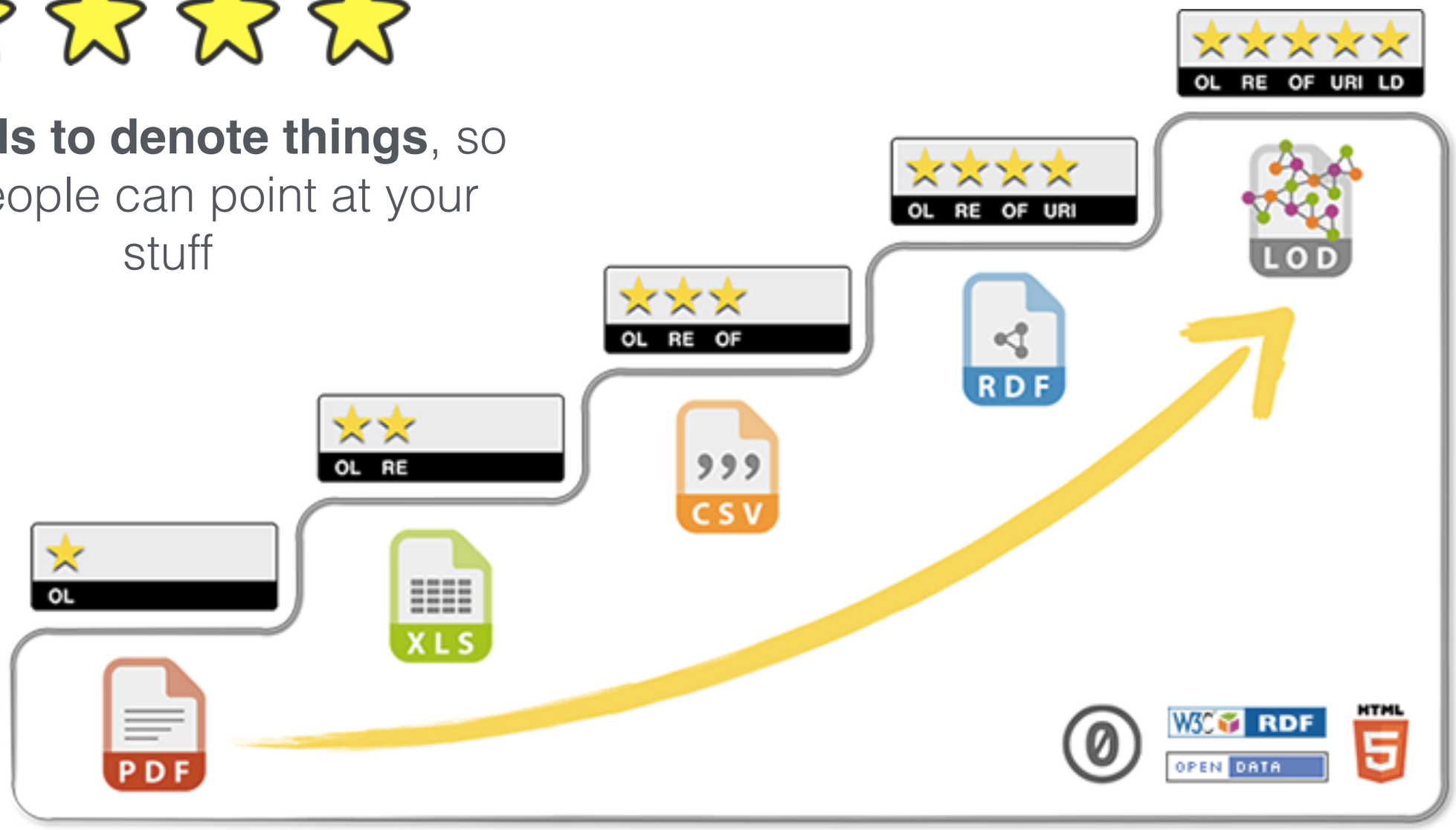
https://5stardata.info/en/

Open data

The 5 stars open data

★ ★ ★ ★

use URIs to denote things, so that people can point at your stuff



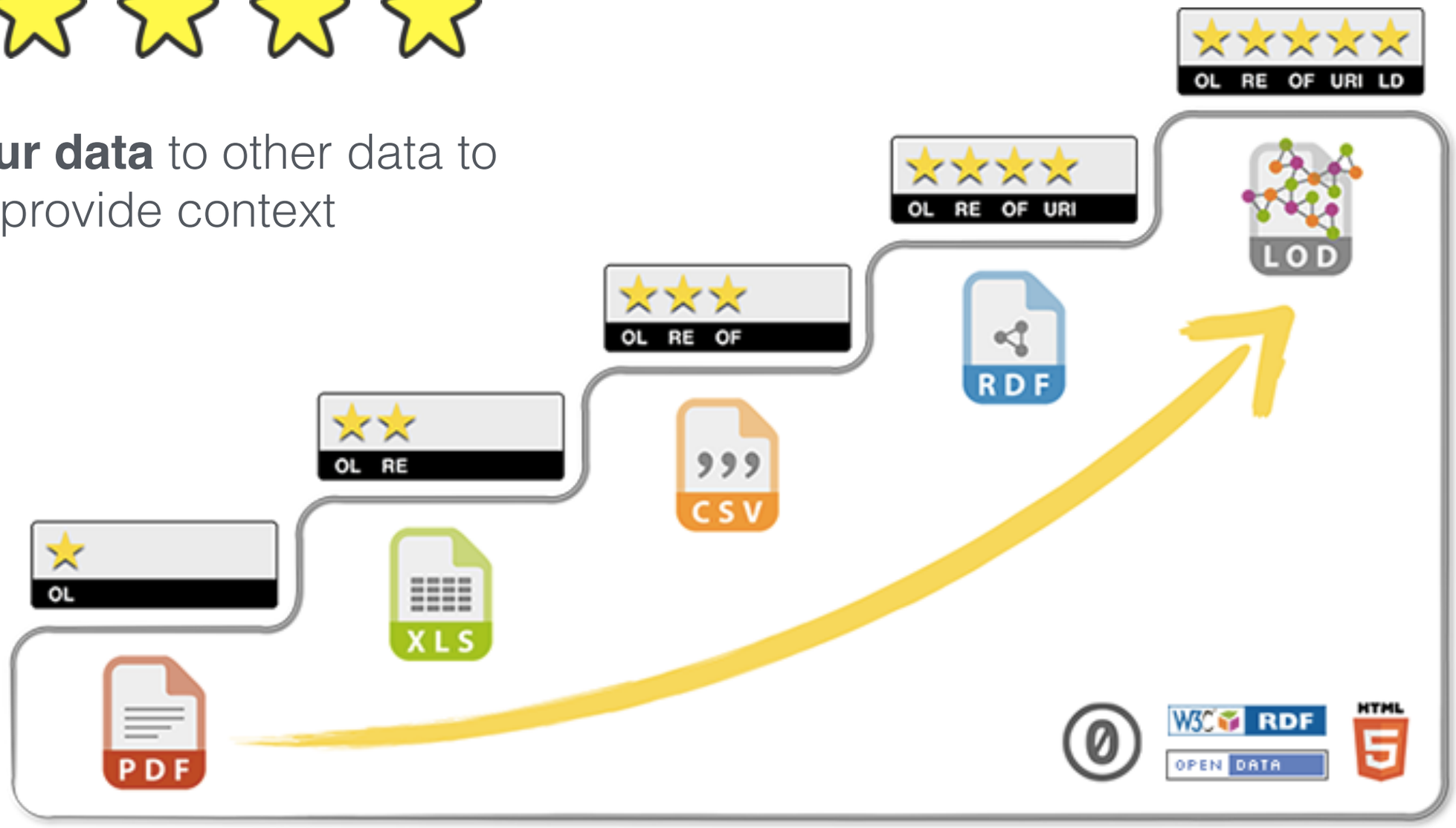
https://5stardata.info/en/

Open data

The 5 stars open data

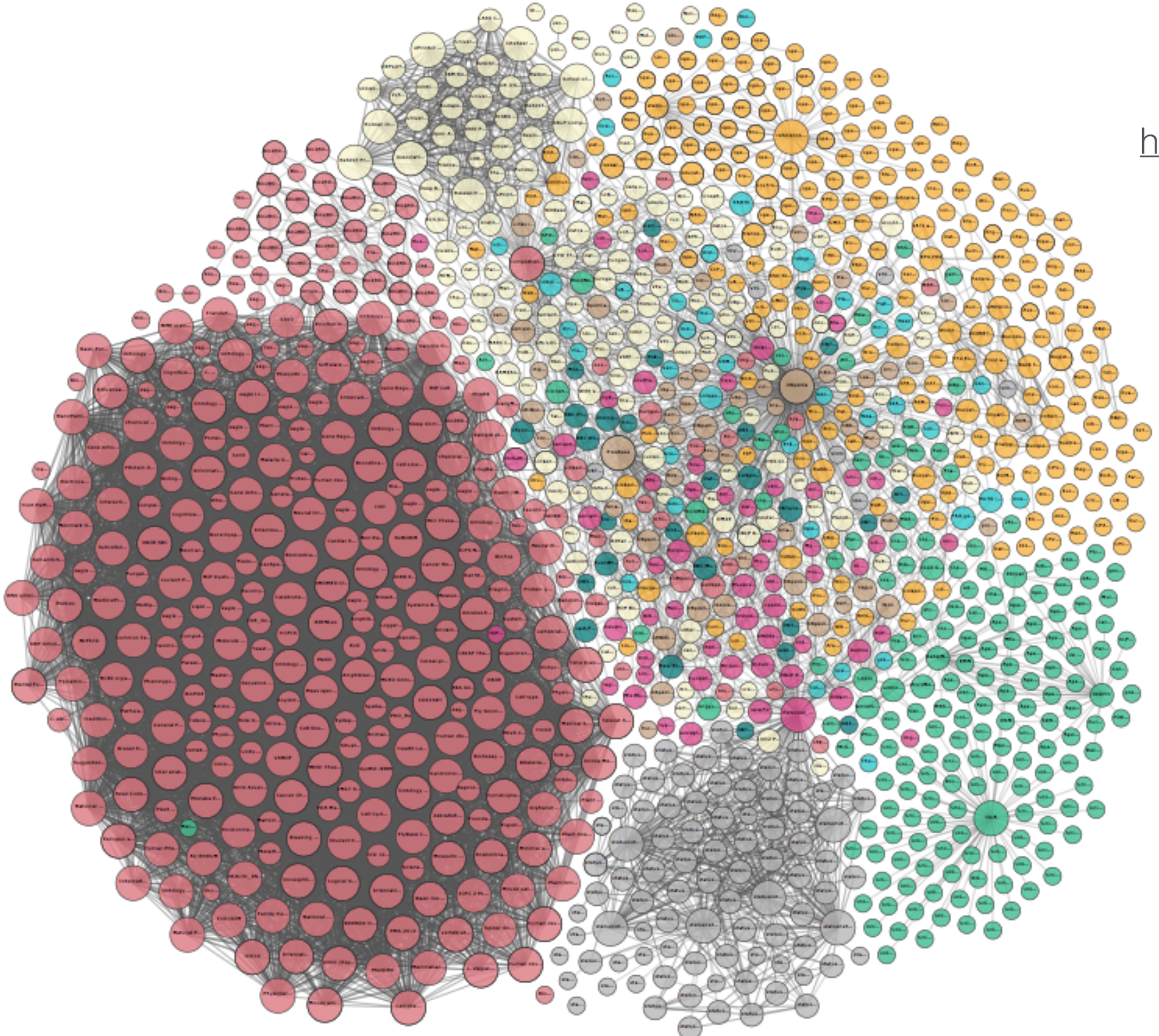
★ ★ ★ ★ ★

link your data to other data to provide context



https://5stardata.info/en/

The Linked Open Data cloud



https://lod-cloud.net/

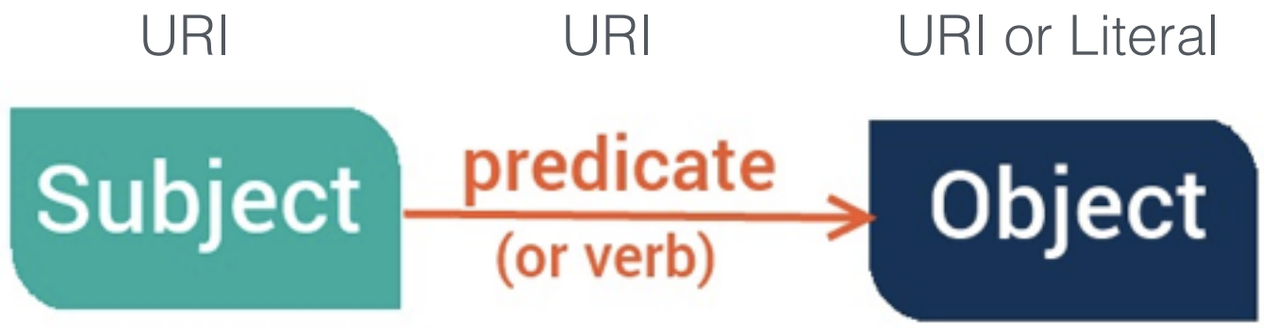
Legend
- Cross Domain
- Geography
- Government
- Life Sciences
- Linguistics
- Media
- Publications
- Social Networking
- User Generated
- Incoming Links
- Outgoing Links

Summary

- Introduction
- Concepts
  - Linked Data
  - Open Data
  - **RDF**
  - SPARQL query language
- Practice: SPARQL examples
- References

RDF

RDF triple



RDF -> Resource Description Framework

a generic model for triples

Serialization formats: RDF-XML, Turtle, N-Triples

RDF

RDF triple example

subject

predicate

creator

object

http://dbpedia.org/ontology/author

http://dbpedia.org/page/Mona_Lisa

http://dbpedia.org/page/Leonardo_da_Vinci

RDF

the rdf:type predicate

subject

http://dbpedia.org/page/Mona_Lisa

predicate

creator

http://dbpedia.org/ontology/author

object

http://dbpedia.org/page/Leonardo_da_Vinci

predicate

rdf:type

http://dbpedia.org/ontology/Artwork

RDF

the rdf:type predicate

subject

predicate

artist

http://dbpedia.org/ontology/author

object

http://dbpedia.org/page/Mona_Lisa

http://dbpedia.org/page/Leonardo_da_Vinci

predicate

rdf:type

http://dbpedia.org/ontology/Artwork

predicate

rdf:type

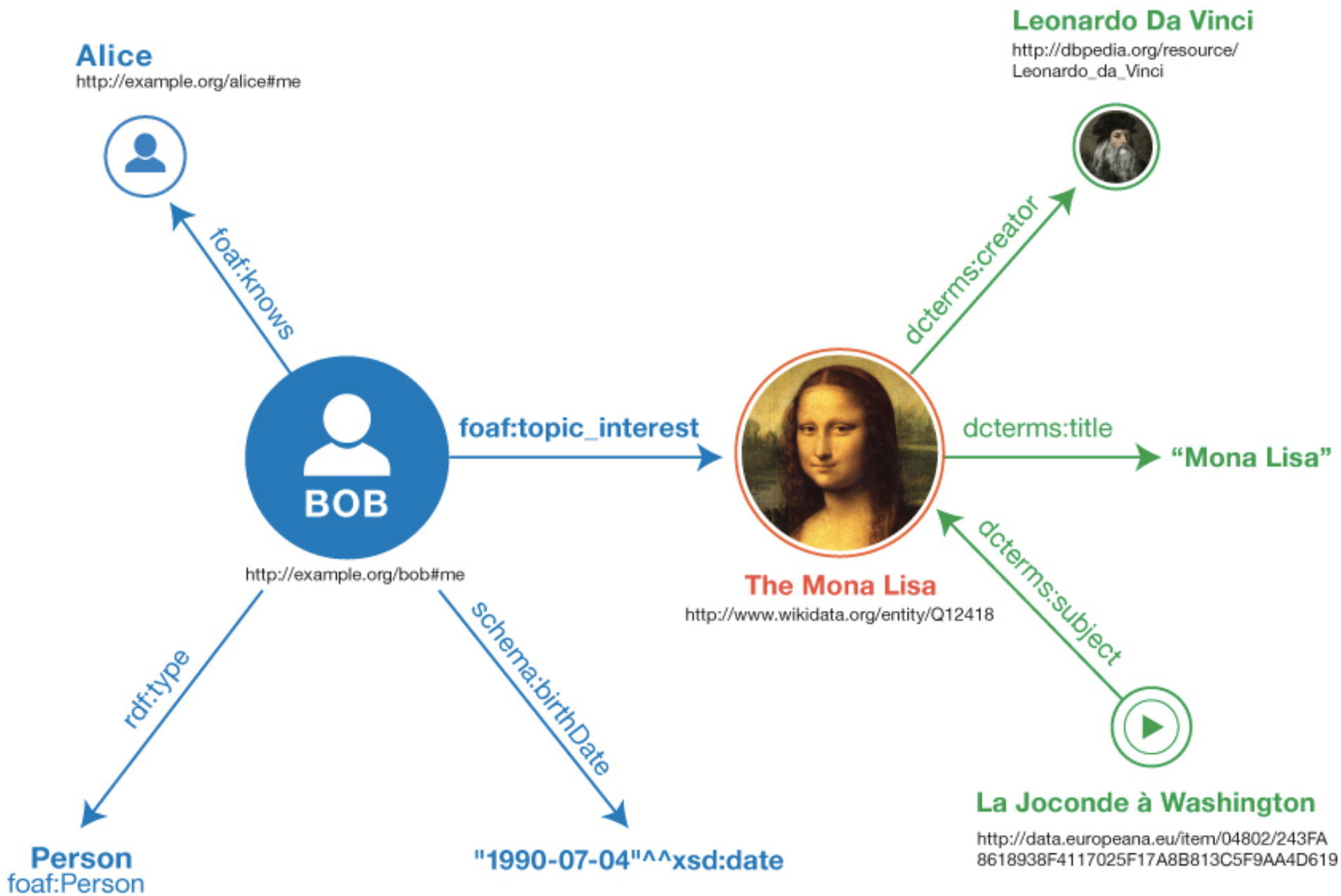
predicate

http://dbpedia.org/ontology/Person

http://dbpedia.org/class/yago/Painter110391653

RDF

RDF graph example



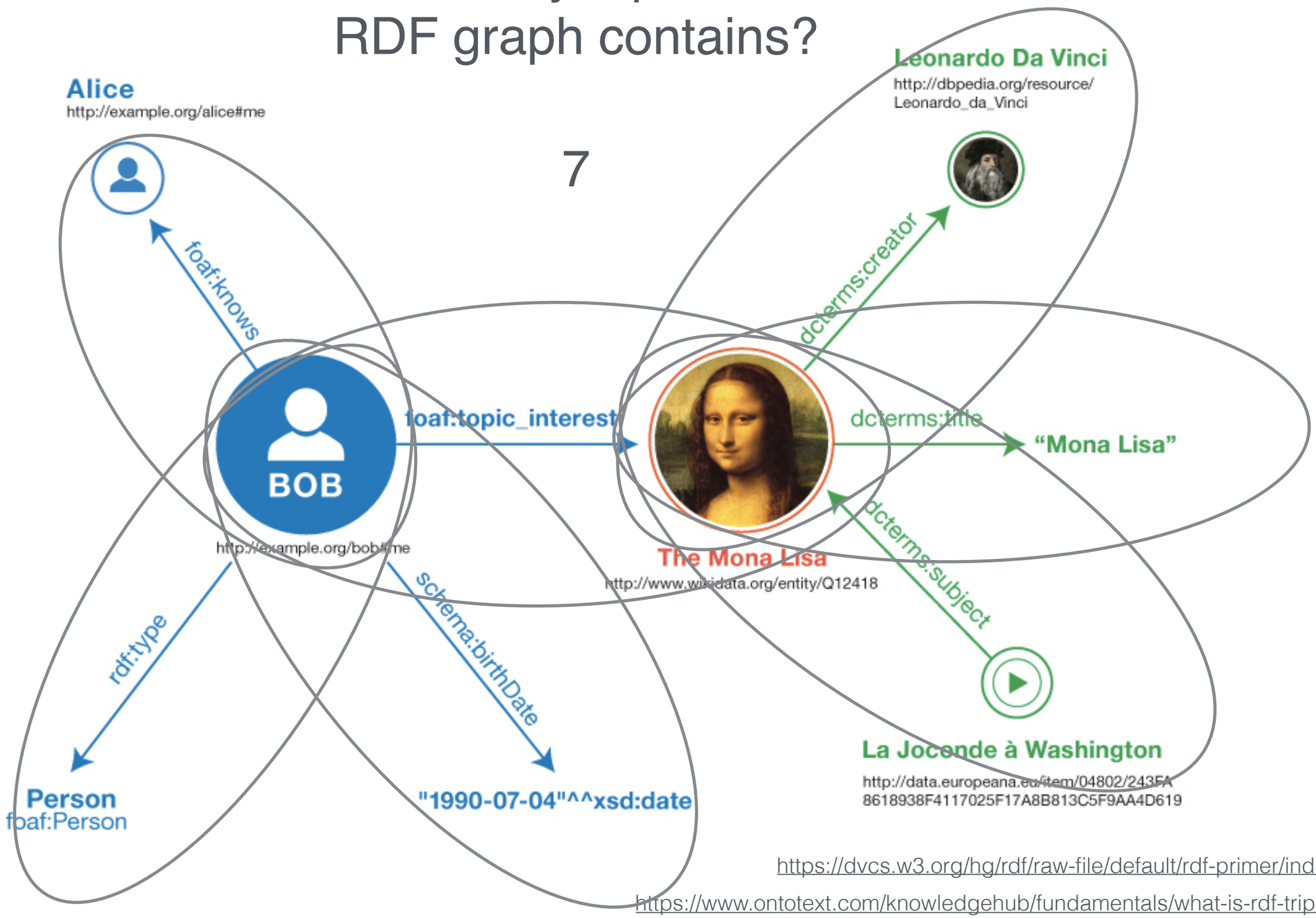
https://dvcs.w3.org/hg/rdf/raw-file/default/rdf-primer/index.html

https://www.ontotext.com/knowledgehub/fundamentals/what-is-rdf-triplestore/

RDF

How many triples this RDF graph contains?

7



https://dvcs.w3.org/hg/rdf/raw-file/default/rdf-primer/index.html

https://www.ontotext.com/knowledgehub/fundamentals/what-is-rdf-triplestore/

Summary

- Introduction
- Concepts
  - Linked Data
  - Open Data
  - RDF
  - **SPARQL query language**
- Practice: SPARQL examples
- References

SPARQL query language

recursive acronym for SPARQL Protocol and RDF Query Language

```
SELECT ...
FROM ...
WHERE{
  ...
  FILTER ()
}
ORDER BY ...
GROUP BY ...
LIMIT ...
```

type of

dataset

graph patterns - filter is optional

query modifiers

https://www.w3.org/TR/rdf-sparql-query/

Summary

- Introduction

- Concepts
  - Linked Data
  - Open Data
  - RDF
  - SPARQL query language

- **Practice: SPARQL examples**

- Conclusions
- References

Let’s practice?

Find all material, references, links and exercices:

https://github.com/liviaruback/intro_linkeddata

Let’s practice?

1 *Take a look in the Stanford University entry in DBPedia (RDF version of wikipedia)*

http://dbpedia.org/resource/Stanford_University

Let's practice?

2 *Give me all the properties about Stanford University*

Go to the DBPedia SPARQL endpoind:

https://dbpedia.org/sparql

variables

```
SELECT ?p ?o
WHERE{
  <http://dbpedia.org/resource/Stanford_University> ?p ?o
}
```

triple pattern

Let's practice?

3 *How many universities are in the world?*

Go to the DBPedia SPARQL endpoind:
https://dbpedia.org/sparql

aggregation

```
SELECT (COUNT(?univ) AS ?count)
WHERE {
   ?univ rdf:type <http://dbpedia.org/ontology/University>.
}
```

triple pattern

Let's practice?

4 *Give me the universities with more than 20k postgraduate students*

```
SELECT
   ?univ ?numPgs
WHERE {
   ?univ rdf:type <http://dbpedia.org/ontology/University>.
   ?univ <http://dbpedia.org/ontology/numberOfPostgraduateStudents> ?numPgs.
   FILTER (?numPgs > 20000)
 }
ORDER BY
   desc(?numPgs)
```

variables

filter condition

triple pattern

Let’s practice?

What are the top 5 universities with more postgraduate students are in the world?

In which cities that universities are located in?

What is the population of the cities with the higher number of postgraduate students?

References

Linked Data references:

- linkeddata.org/
- https://www.w3.org/DesignIssues/LinkedData.html

Tim Berners-Lee TED talk

The 5 star open data

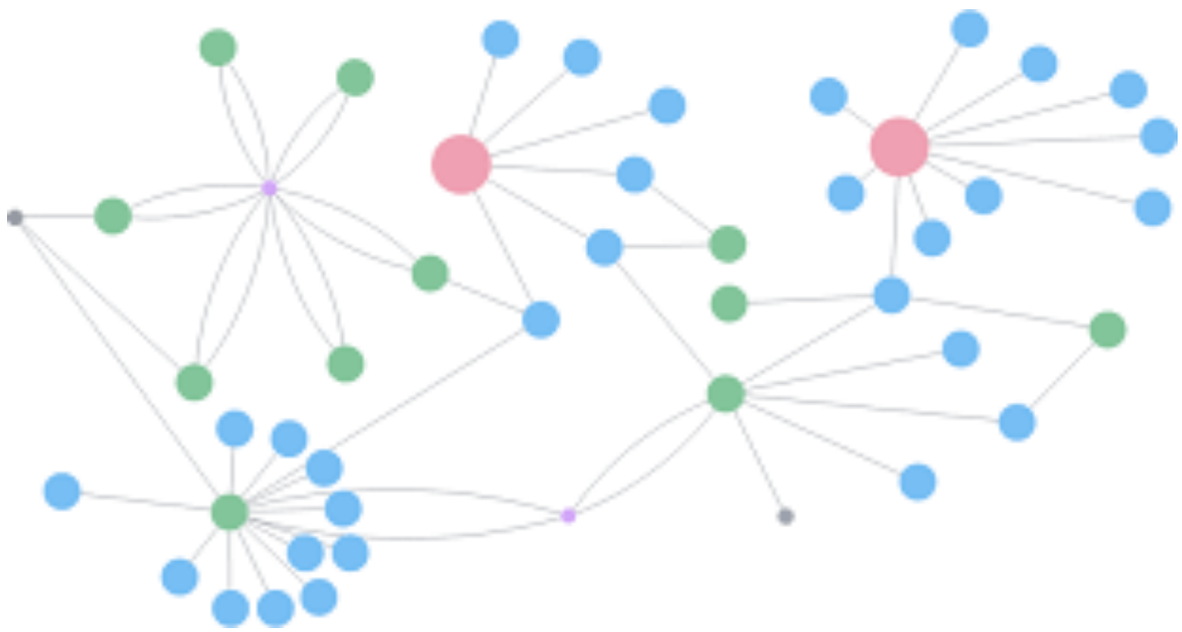
The Linked Open Data cloud

SPARQL query language specification

Coming next..

Part 2 - Graph Databases and visualization

- Graph concepts
- Graph storing and visualization with Neo4J



Questions?

liviaruback@gmail.com