Information Visualization

Introduction to RDFLib

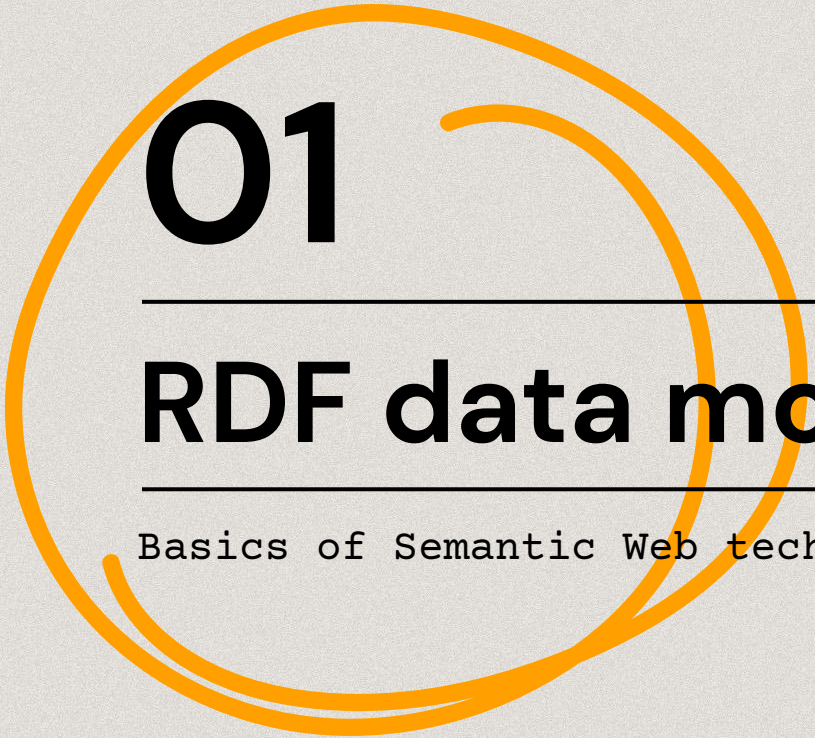
Lesson 3

Marilena Daquino
Assistant Professor

Department of Classical Philology and Italian Studies

marilena.daquino2@unibo.it

Table of contents



01

RDF data model

Basics of Semantic Web technologies

The web we know

01

RDF

URI

UNIFORM RESOURCE IDENTIFIER

Identify the location of documents on the web

HTTP

HYPERTEXT TRANSFER PROTOCOL

It’s the protocol for exchanging data and documents on the web

HTML

HYPERTEXT MARKUP LANGUAGE

It’s the markup language for documents returned via HTTP

The web of data

01

RDF

URI

UNIFORM RESOURCE IDENTIFIER

HTTP

HYPERTEXT TRANSFER PROTOCOL

HTML

HYPERTEXT MARKUP LANGUAGE

In the Web of Data a URI is a ***persistent* conceptual mapping** to a real entity (e.g. a person), not to a HTML page.

- If the **location** of a HTML file describing the entity changes, the URI does not change.
- Many HTML pages can describe the **same entity** identified by the same URI.
- A HTML page can include information about **many entities**, hence being linked to many URIs.

RDF

URI

subject

URI

predicate

URI

object

RESOURCE DESCRIPTION FRAMEWORK

In the WoD a URI identifies both **real entities** and the **relations** (links) between them.

Every piece of information is represented as a **triplet** of URIs, identifying respectively a subject, a predicate, and an object.

RDF example

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

Robert Capa

Has spouse

Gerda Taro

RDF example

URI

http://dbpedia.org/resource/Robert_Capa

URI

http://example.org/hasSpouse

URI

http://dbpedia.org/resource/Gerda_Taro

RDF example

01

RDF

Data

```
<http://dbpedia.org/resource/Robert_Capa>
<http://example.org/hasSpouse>
<http://dbpedia.org/resource/Gerda_Taro> .
```

Save as

```
data.rdf
```

02

Knowledge management

Basics of Knowledge organisation and representation

Classify entities and properties

Person

Entity of the **class** Person

property

A property relating **objects**, e.g. people

Person

Entity of the **class** Person

Vocabulary terms

Person	property	Person
http://dbpedia.org/ontology/Person	http://example.org/ontology/hasSpouse	http://dbpedia.org/ontology/Person

Namespaces

02

Ontologies

Person

http://dbpedia.org/ontology/Person

property

http://example.org/ontology/hasSpouse

Person

http://dbpedia.org/ontology/Person

Ontology example

Data

```
<http://dbpedia.org/ontology/Person>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.w3.org/2002/07/owl#Class> .

<http://dbpedia.org/ontology/Person>
<http://www.w3.org/2000/01/rdf-schema#label> “Person” .

<http://dbpedia.org/ontology/Person>
<http://www.w3.org/2000/01/rdf-schema#comment> “A human being” .

<http://example.org/ontology/hasSpouse>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.w3.org/2002/07/owl#ObjectProperty> .

<http://example.org/ontology/hasSpouse>
<http://www.w3.org/2000/01/rdf-schema#label> “has spouse”.
```

Save as

```
ontology.rdf or
ontology.owl
```

RDF is not a format

02

Serialisations

It’s a framework

Conceptual model to formalise the **logical** structure of data in Graphs (VS tables, hierarchies)

Has n serialisations

While the triplet pattern is always respected, there exist several **syntaxes** to serialise RDF.

Has n formats

A **parser** can read and interpret the same information even if served according to different syntaxes and stored in different file **formats**.

RDF serializations

02

Serialisations

data.nt – ntriples

```
<http://dbpedia.org/resource/Robert_Capa>
<http://example.org/hasSpouse>
<http://dbpedia.org/resource/Gerda_Taro> .
```

data.ttl – turtle

```
@prefix db: <http://dbpedia.org/resource/> .
@prefix ex: <http://example.org/> .
db:Robert_Capa ex:hasSpouse db:Gerda_Taro .
```

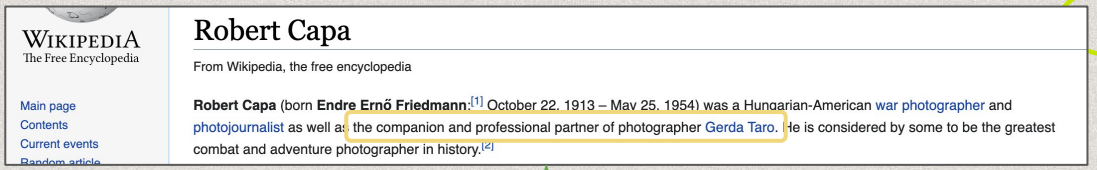
data.xml – XML

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF xmlns:db="http://dbpedia.org/resource/" xmlns:ex="http://example.org/">
<rdf:Description rdf:about="http://dbpedia.org/resource/Robert_Capa">
          <ex:hasSpouse rdf:resource="http://dbpedia.org/resource/Gerda_Taro" />
     </rdf:Description>
</rdf:RDF>
```

What if triples are not enough

Graphs

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



Has source

Robert Capa

Gerda Taro

Has spouse

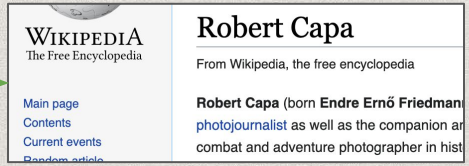
Named graphs

Graphs

We wrap triples in a graph, a container of triples. We can assign a URI to the **graph** and use the latter as subject of further triples.

Robert Capa has spouse Gerda Taro

Has source



http://example.org/resource/Robert_Capa_hasSpouse_Gerda_Taro

http://example.org/hasSource

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

Named graphs serializations

data.nq – nQuads

```
<http://dbpedia.org/resource/Robert_Capa>
<http://example.org/hasSpouse>
<http://dbpedia.org/resource/Gerda_Taro>
<http://example.org/resource/Robert_Capa_hasSpouse_Gerda_Taro> .

<http://example.org/resource/Robert_Capa_hasSpouse_Gerda_Taro>
<http://example.org/hasSource>
<https://en.wikipedia.org/wiki/Robert_Capa>
<http://example.org/resource/Robert_Capa_hasSpouse_Gerda_Taro> .
```

Named graphs serializations

data.trig – Trig

```
@prefix db: <http://dbpedia.org/resource/> .
@prefix ex: <http://example.org/> .
@prefix wiki: <https://en.wikipedia.org/wiki/> .

ex:Robert_Capa_hasSpouse_Gerda_Taro {
      db:Robert_Capa ex:hasSpouse db:Gerda_Taro .
      ex:Robert_Capa_hasSpouse_Gerda_Taro ex:hasSource wiki:Robert_Capa .
}
```

03

Case study

ARTchives

Scope and content

Art historians

The creators of archival collections.

Focus on historians of **Italian Modern Art** (15-16th centuries).

Collections

Materials produced by art historians (**letters, photos**, etc.)

<30 collections.

Keepers

Cultural institutions (archives) **preserving** collections.

6 institutes promoting the project.

Types of entities

Actors

Artists and artistic movements studied by the art historian.

Other art historians.
Universities and research centres.

Contents

Types of materials.
Dates.

Contents: Artists and movements.
Correspondants.
Places. **Bibliography**.

Descriptions (biographies, scope and content)

Other

Cities and contact addresses.

Vocabularies and data

03

ARTchives

Wikidata vocab.

Whenever applicable classes and properties are taken from Wikidata.

Wikidata entities

Likewise, entities reuse Wikidata URIs when existing, otherwise new URIs are minted.

Named graphs

All information about a collection, its creator, and the keeper, are stored in a named graph.

Aby Warburg's collection	`<https://w3id.org/artchives/collectionfondo-aby-warburg>`
Main subject	`<http://www.wikidata.org/prop/direct/P921>`
Federico Zeri	`<http://www.wikidata.org/entity/Q1089074>`
The graph of A.W.'s collection	`<https://w3id.org/artchives/1598630286-3009102/> .`

Some problems

User data

Duplication of data (e.g. **labels**).

Wrong reconciliation.

Misspelling.

Design choices

Data about historians and keepers appear only in the **graph** of the first created collection

(if there are two collections for the same historian, only one includes data about the historian).

Mistakes

URIs of Wikidata properties use the wrong **namespace** in the online version.

The data dump you will use, has the correct namespaces.

04

Hands-on

Jupyter notebook, RDFLib

Hands-on

04

Get all the materials

Download data

Download the data (resources/artchives.nq) in a folder for the exercise

Tutorial

Open the tutorial:
in GitHub, Colab
or Jupyter (download)

Install packages

In the terminal/shell (if IDE or Jupyter)

```
pip install rdflib
pip install pprint
```

Practice

Choose your environment:
IDE: create a .py file
Jup: create .ipynb file
Colab: new notebook

Exercise

04

Assignment

Review

Review the tutorial

Exercise

Solve the problems
(time to code!)

Answer

Fill in the form
with your answers

Thanks!

Do you have any questions?

marilena.daquino2@unibo.it

https://github.com/marilenadaquino/information_visualization