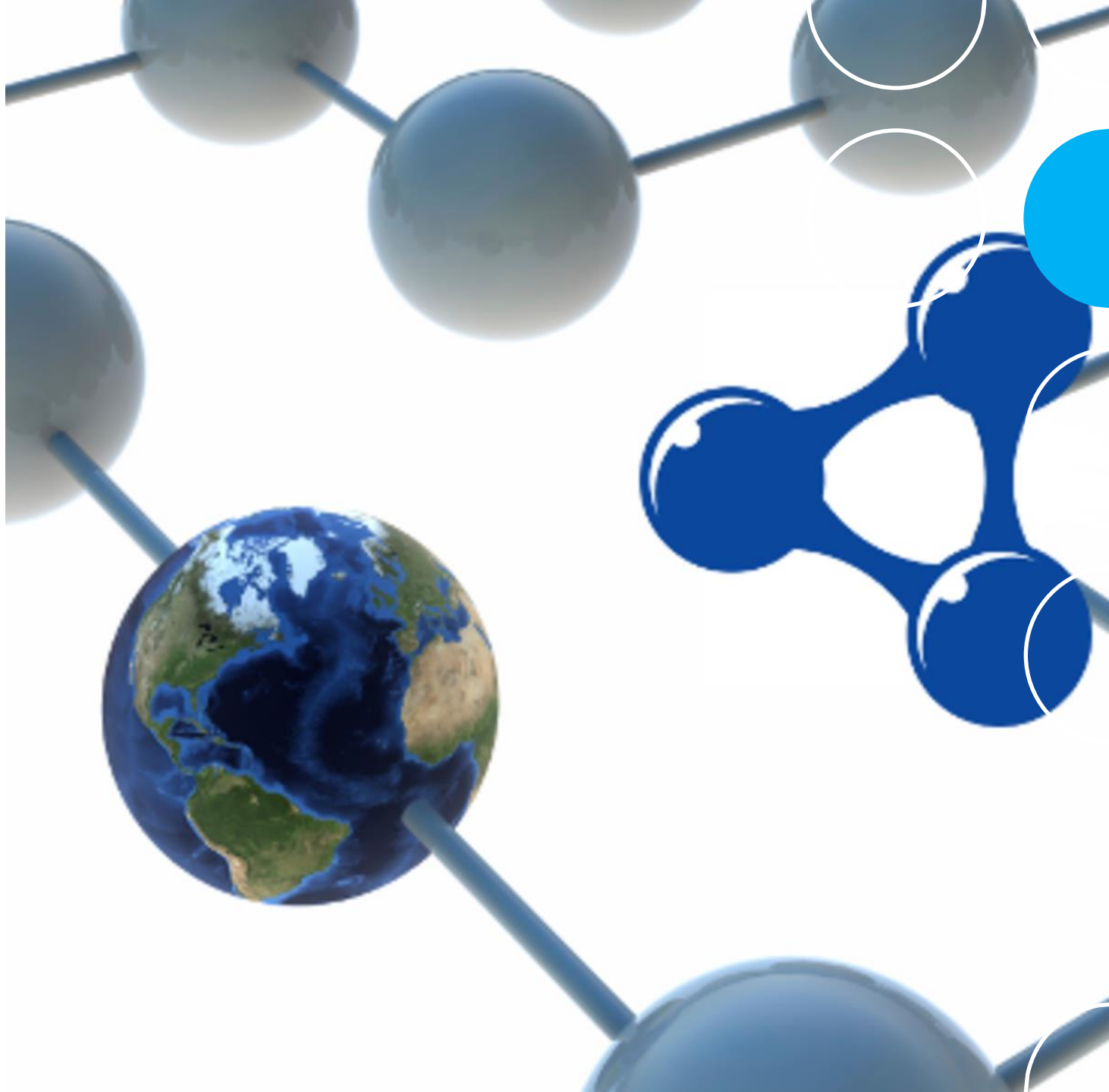# Semantic Web and Linked Data

Liliana Ferreira

2022/23

Class 2: Learning Objectives

- Answer to the question "When is the Semantic Web actually going to happen?"

- Know and explore available demos of websites using semantic web and link data approaches;

- Understand RDF principles and how to provide useful RDF information;

Group exercise

- 4 groups; 1 topic per group

- 20 mins to read and discuss the topic:

  - **Individually:** read the provided links and search for aditional information on the topic; try to give preliminar answers to the questions bellow (10 mins);

  - **In group:** discuss with each other the individual and aditional information found. Answer and prepare together the following questions (10mins):

    - Context and motivation;

    - Semantic Web Technologies used;

    - Key Benefits of Using Semantic Web Technology;

    - If possible, a short demo.

- 10 mins to prepare a demo and presentation;

- 7 minutes to present your topic/demo to the class

The BBC Website

- www.bbc.com
- https://www.w3.org/2001/sw/sweo/public/UseCases/BBC/
- https://www.youtube.com/watch?v=gFbp7Mpzmzs

DBpedia

http://dbpedia.org

http://dbpedia.org/resource/Paris

Open Calais: From natural language to linked data

- Intelligent Tagging

- https://www.refinitiv.com/en/products/intelligent-tagging-text-analytics

- https://developers.refinitiv.com/en/api-catalog/open-perm-id/intelligent-tagging-restful-api

European Data Portal

- https://www.fokus.fraunhofer.de/en/dps/projects/opendataportal

- https://data.europa.eu/en

Do you SEARCH or do you FIND?

- We need to help machines understanding the web so that machines can help us understanding things:

  - They can learn what we are interested in;

  - They can help us better find what we want;

- How can we do that?

  - Besides publishing documents on the web – which computers can't understand easily – let's publish something that computers can understand:

    RAW DATA!

Current Data on the Web

- We do publish raw data already:

    - Relational databases; XML; CSV; APIs; …

- But in all different formats and data models!

- The data in different data sources aren't linked:

    How do we know that the Liliana Ferreira in LinkedIn is the same as the Liliana Ferreira in Twitter?

Current Data on the Web

- Wouldn't it be great if we had a standard way of publishing data on the Web?

- *YES – there is one!*

Semantic Web

The Semantic Web

"The **Semantic Web** is an extension of the current web in which information is given well-defined **meaning,** better enabling computers and people to **work in co-operation**."

[Berners-Lee *et al*, 2001]

The Web today is still...

Currently most of the Web content is suitable for human use.

Typical uses of the Web today are information seeking, publishing, and using, searching for people and products, shopping, reviewing catalogues, etc.

Dynamic pages generated based on information from databases but without original information structure found in databases.

Limitations of the Web Search today

The Web search results are high recall, low precision.

Results are highly sensitive to vocabulary.

Results are single Web pages.

Most of the publishing contents are not structured to allow logical reasoning and query answering.

Resource Description Framework (RDF)

- RDF is a language for the representation of resources:

    - A resource can be anything;

- A standard of W3C;

- RDF is a data model;

- One of main applications: data integration.

    - Relationships *between* documents;

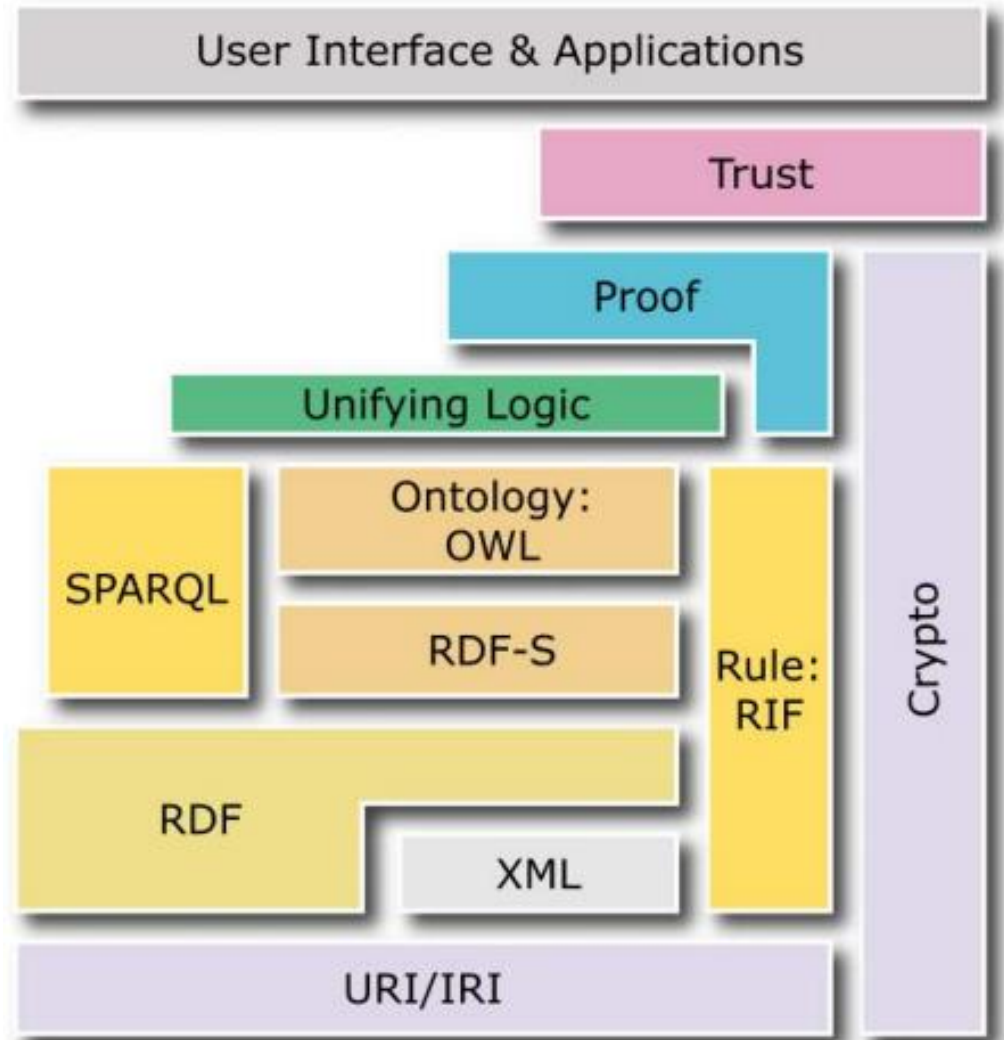- Basic building block: **triples** or statements

```
<subject, property, object>

<"Mozart", composed, "The Magic Flute">
```

RDF means

- Resource:
  - Everything that can have an URI: pages, chairs, persons, pens, ideas
- Description:
  - Attributes, characteristics and relations between resources
- Framework:
 - Model, language and syntaxes for these descriptions

Semantic Web Stack of standards W3C®

Resource Description Framework (RDF)

Resource Description Framework (RDF)

- Universal, machine readable exchange format;

- Data structured in graphs (vertices, edges).

- Any relational data can be represented as triples:
  - Triples are statements about things (resources), using URIs or literal values
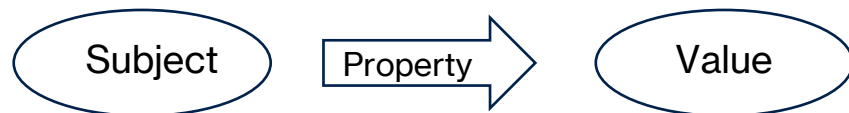
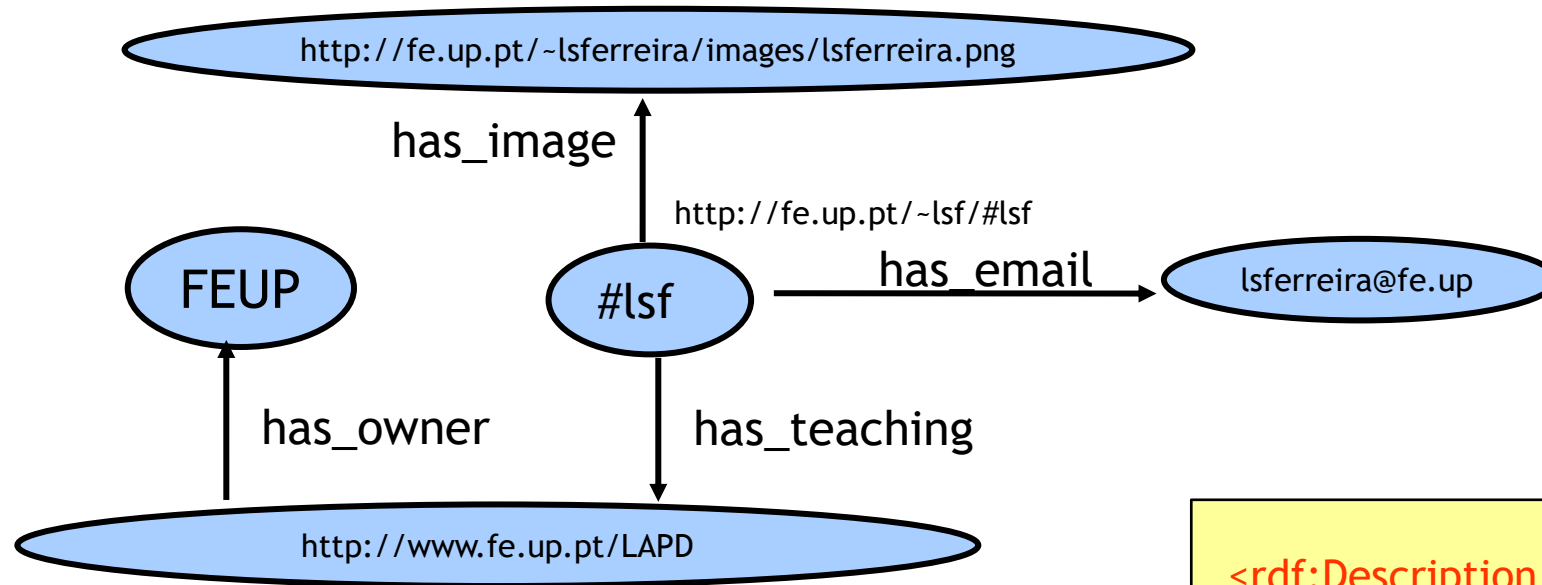# RDF decomposes descriptions into triples

(subject, predicate, object)

E.g.: "SWDL" has as students João, Sara and Miguel and as topic the Semantic Web.

Composition Rules for RDF Triples

1. The subject is always a resource (and not a literal)

2. The type of the binary property is identified by a URI

3. The value is a resource or a literal

RDF triples form graphs


```
<rdf:Description rdf:about="#lsf">
    <has_email>lsferreira@fe.up</has_email>
</rdf:Description>
```

RDF: triples form graph edges

(subject, predicate, object)

->

(node, edge, node)

RDF is an oriented labeled multigraph model

- RDF is an oriented labeled multigraph model
    1. Several edges can connect the same two nodes;
    2. Edges are oriented: the head is the object, the tail is the subject;
    3. Edges and nodes are labeled.

Resource Description Framework (RDF)

Resources

- A *resource* can be anything describable using RDF.

- Every resource has a URI (Universal Resource Identifier);
- A URI can be a URL (a web address) or some other kind of identifier;
- An identifier does not necessarily enable access to a resources;
- We can think of a resources as an **object** that we want to describe. For example:
  - Books
  - Person
  - Places, etc.

Resource Description Framework (RDF)

Properties

- A ***property*** is a specific aspect of a resource.

- It can be a characteristic that belongs to a resource, or a relationship that links one resource with another.

- Properties describe relations between resources;

  - For example: "written by", "composed by", "title", "topic", etc;

- Properties in RDF are also identified by URIs. This provides a global, unique naming scheme.

Resource Description Framework (RDF)

Statements

- A ***statement*** is a piece of description about a particular resource in the RDF format: an object-attribute-value triple;

- It consists of a resources, a property, and a value.

http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=10140

Resource Description Framework (RDF)

- A statement about a resource instance has:
  - the resource's identifier
  - one of the resource's property (defined in an RDF schema)
  - the value for that property (can be either a literal, or a resource)

```
<rdf:RDF
 xmlns:wc="http://www.lapd.fe.up.pt/~exRDF/wc/schema">
 <rdf:Description about="http://www.cnn.com/2000/HEALTH/cancer/12/06/
             colon.cancer.ap/index.html">
 <wc:Title>Cigarette smoking linked to colorectal cancer </wc:Title>
 </rdf:Description>
</rdf:RDF>
```

Further reading

- Semantic Web Stack
- RDF 1.1 Primer
- RDF 1.1 Concepts and Abstract