ISWC 2017 Tutorial:
Semantic Data Management in Practice

Part 1: Introduction

Olaf Hartig
Linköping University
olaf.hartig@liu.se
@olafhartig

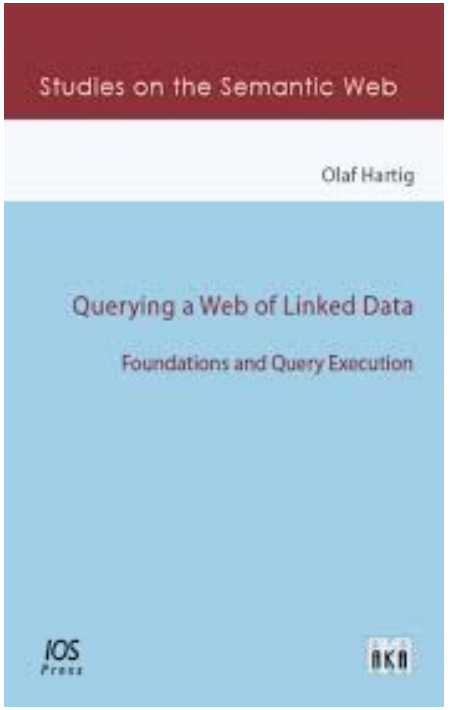
Olivier Curé
University of Paris-Est Marne la Vallée
olivier.cure@u-pem.fr
@oliviercure

LINKÖPING UNIVERSITY

UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

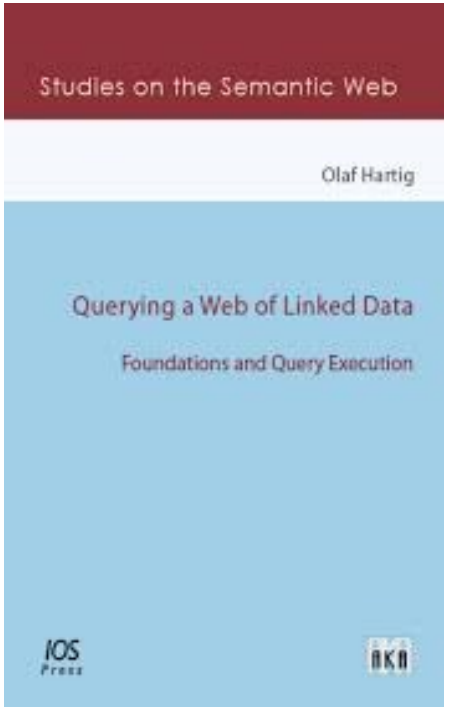
Who Are We?

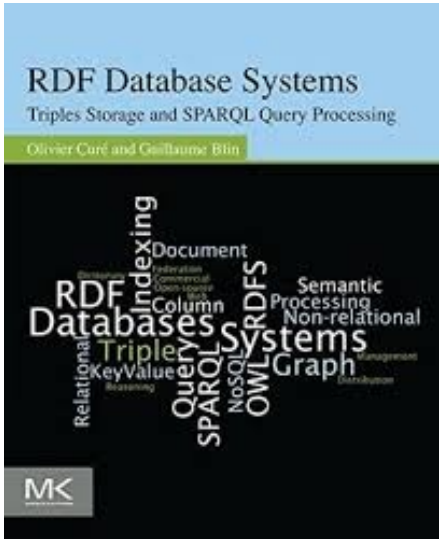
- Olaf Hartig
  - Assistant professor in CS at Linköping University (Sweden)
  - Research on Web data, graph data and semantic data management



Who Are We?

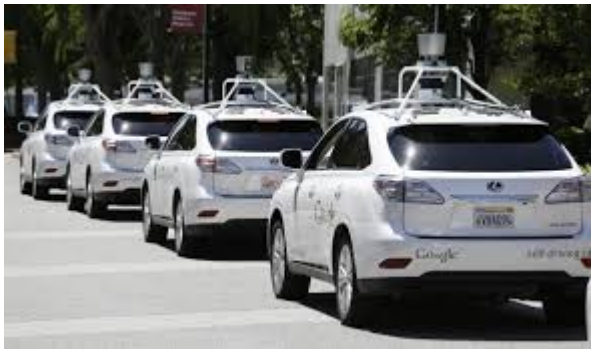
- Olaf Hartig
  - Assistant professor in CS at Linköping University (Sweden)
  - Research on Web data, graph data and semantic data management
- Olivier Curé
  - Associate professor in CS at Université Paris-Est (France)
  - Research on Data and Knowledge management, Reasoning and Big data
  - Creator of a self-medication app with over 6 million clients in France





Semantic Data is in the Air

- Artificial Intelligence strikes back



- Big trend is Machine learning but all the discovered models will be handled by Semantic data management tools
- Many initiatives to represent semantic data on the Web, e.g., Schema.org
- A large catalog of linked open data already exists
  - Over 40 billion triples in the Linked Data cloud

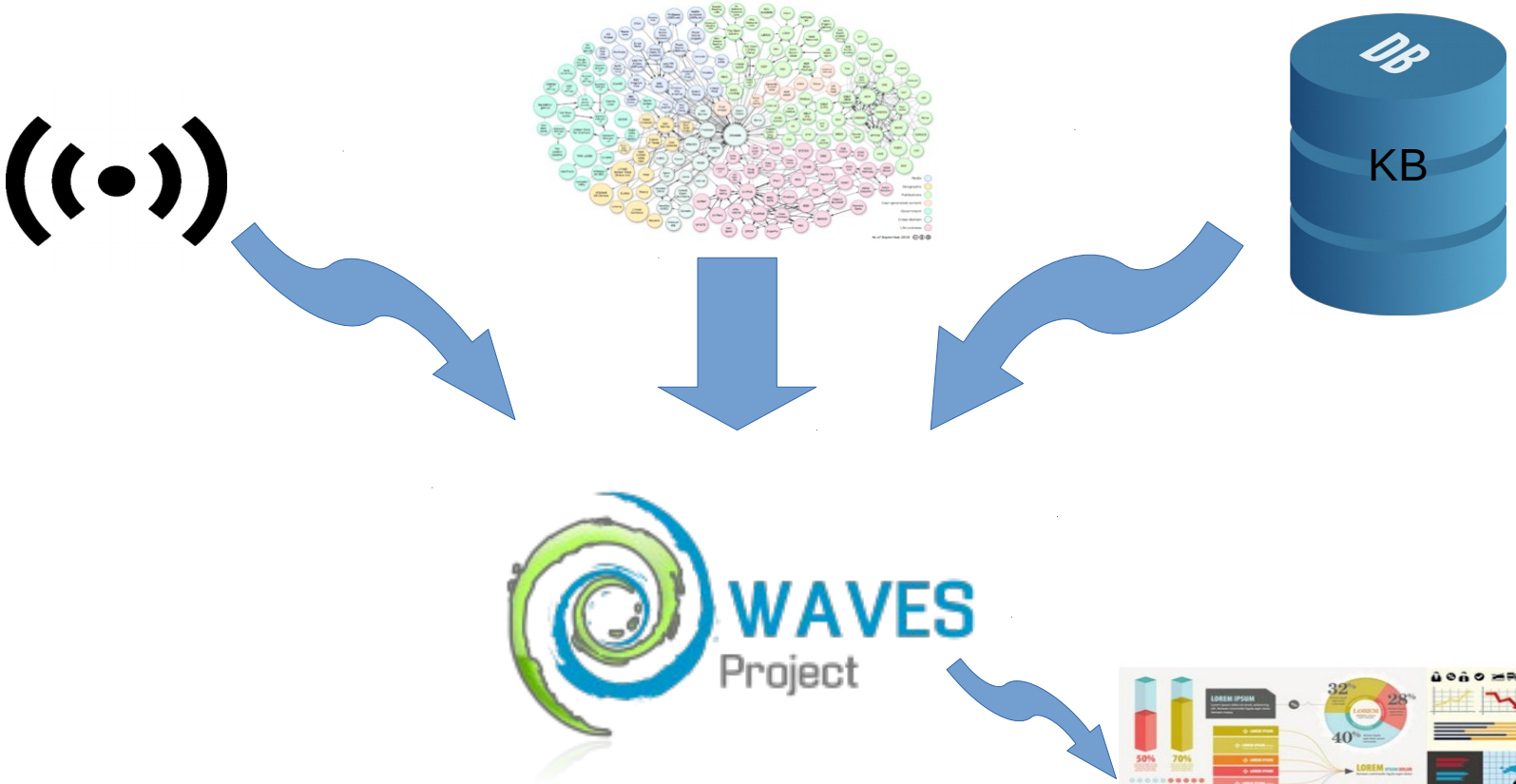
Tutorial’s Goals

- Provide a practitioner's guide to semantic data management

- We consider 7 inescapable aspects:
  - Storing and querying
  - Understanding the data
  - Searching the data
  - Visualizing
  - Automated reasoning
  - Cleaning
  - Integrating

The Waves Running Example

- Semantic data management of potable water networks



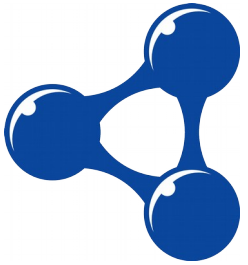
W3C's Semantic Web technologies in a nutshell

RDF Resource Description Framework

- The most prevalent data approach to represent semantic data and meta-data

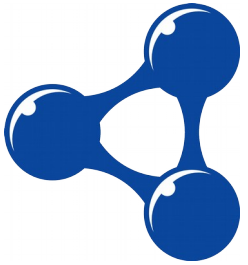
- A key component of W3C's Semantic Web stack together with
  - SPARQL
    - A query language for RDF
  - RDFS
    - A light ontology language
  - OWL
    - A family of expressive ontology languages

RDF in a Nutshell (2)



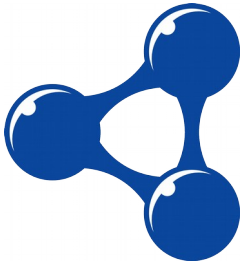
- W3C recommendations in 1999 and 2014 (RDF 1.1)
- A data model for the Web of data based on the notion of triples : (subject, predicate, object)
  - Omnipresence of URIs
  - Several syntaxes available: RDF/XML, JSON-LD, RDFa and Turtle family
- Supports the definition of directed labeled graphs

RDF in a Nutshell (3)

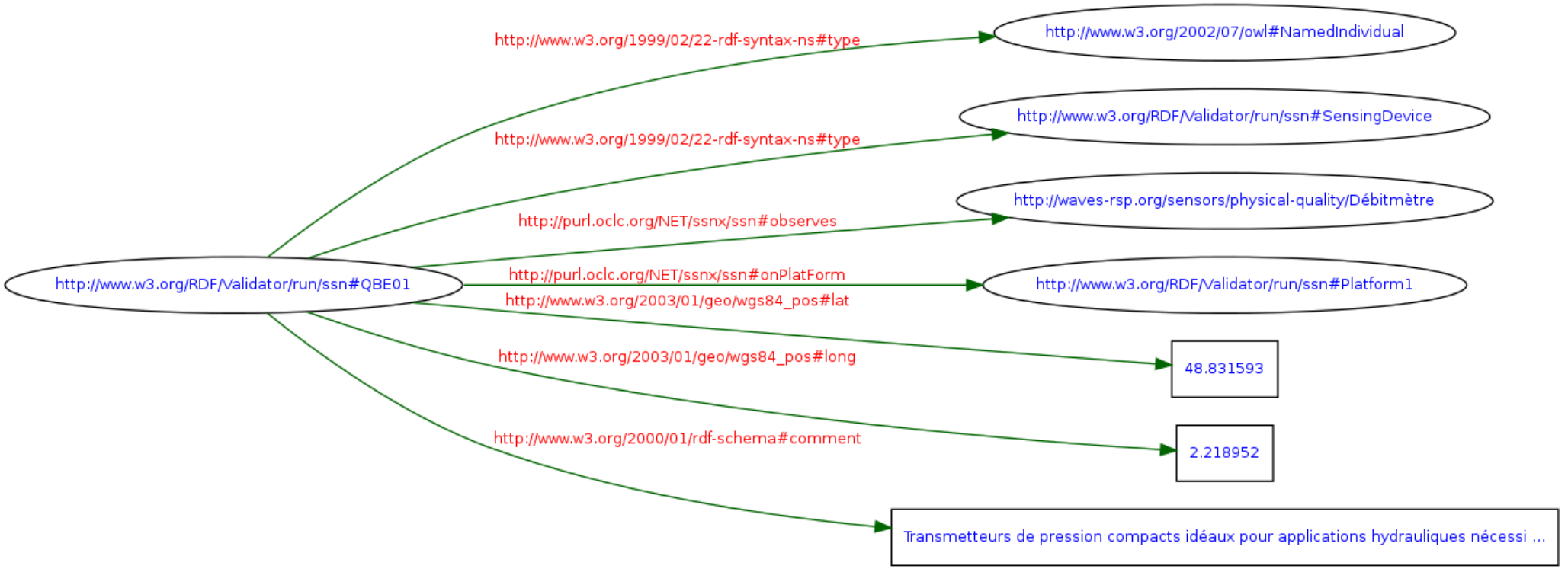


```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix ns0: <http://purl.oclc.org/NET/ssnx/ssn#> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<http://njh.me/ssn#QBE01>
  a owl:NamedIndividual, <http://njh.me/ssn#SensingDevice> ;
  ns0:observes <http://waves-rsp.org/sensors/physical-quality/Débitmètre> ;
  ns0:onPlatForm <http://njh.me/ssn#Platform1> ;
  geo:lat "48.831593" ;
  geo:long "2.218952" ;
  rdfs:comment """Transmetteurs de pression compacts idéaux pour applications hydrauliques
nécessitant une mesure de pression intégrée.
 Les plages jusqu\\'à 10 bar utilisent des éléments de capteurs piézorésistifs. Pour les
plages de 25 → 600 bar, utilisation de capteurs à jauge d\\'extension à film mince.""" .
```

RDF in a Nutshell (4)



```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix ns0: <http://purl.oclc.org/NET/ssnx/ssn#> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<http://njh.me/ssn#QBE01>
  a owl:NamedIndividual, <http://njh.me/ssn#SensingDevice> ;
  ns0:observes <http://waves-rsp.org/sensors/physical-quality/Débitmètre> ;
```



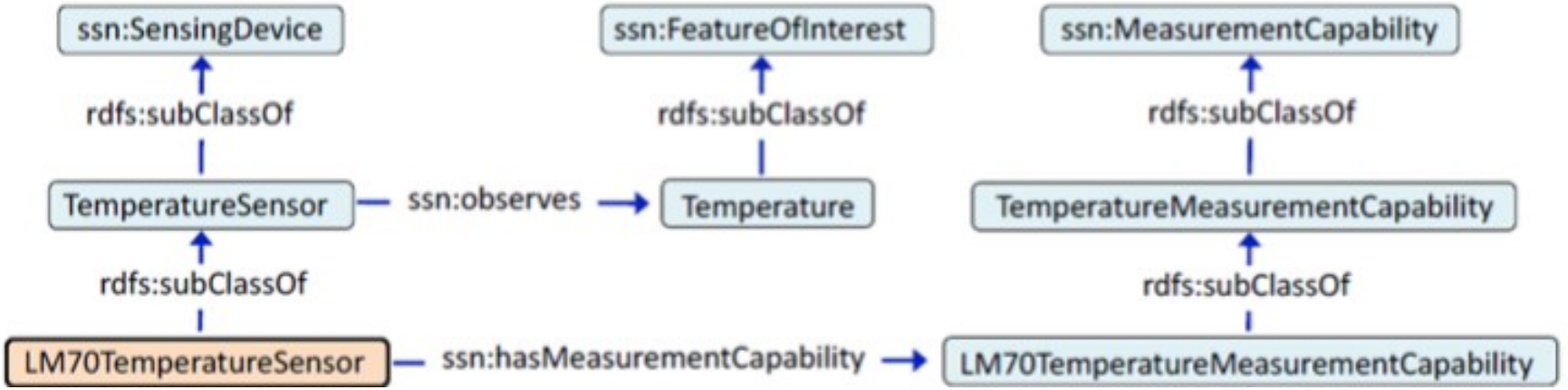
SPARQL

- A query language for RDF data
- Conjunctive queries based on **graph pattern matching**
- Triples that possibly contain variables (prefixed with ? Symbol)
  - Example: retrieve the platform of sensing devices observing flow measurement:

```
PREFIX rdf : <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ns0 : <http://purl.oclc.org/NET/ssnx/ssn#>
SELECT ?s ?p
WHERE {
    ?s rdf:type <http://njh.me/ssn#SensingDevice>.
    ?s ns0:observes <http://waves-rsp.org/sensors/physical-quality/Débitmètre>
    ?s ns0:onPlatForm ?p. }
```

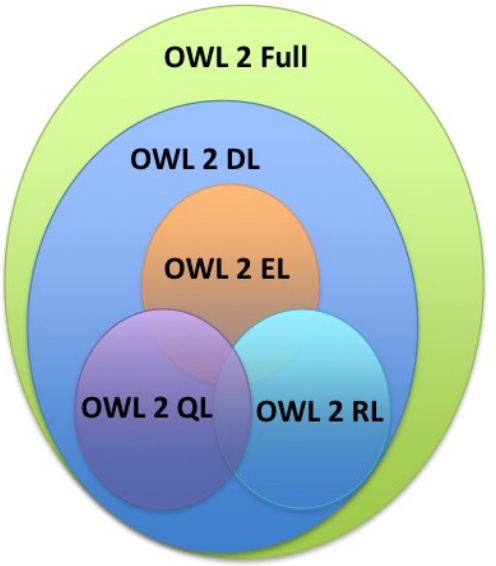
Ontology Languages of the W3C

- For a given domain, an ontology describes a set of concepts and their relationships
- RDF Schema (RDFS) is the language with the lowest expressiveness : sub-classes, sub-properties, domain and range of properties, instance typing



Ontology Languages of the W3C (2)

- OWL (Web Ontology Language) is a family of increasing expressive ontology languages
- Expressiveness comes at a computational cost justifying a trade-off
- OWL2 QL, RL and EL are tractable
- OWL2 DL is not tractable
- OWL2 Full is not decidable



Tutorial's goals

- We consider 7 inescapable aspects:
  - Storing and querying
  - Understanding the data
  - Searching the data
  - Visualizing
  - Automated reasoning
  - Cleaning
  - Integrating

- Tutorial web site:

  http://www.ida.liu.se/research/semanticweb/events/TutorialAtISWC2017.shtml