Web sémantique, web de données, données liées…

Session Web Sémantique INSIDE : MNHN / CNRS / BRGM

2021-01-22

F. Michel

Université Côte d'Azur, CNRS, Inria, I3S, France

Data Interoperability Challenges

Structural heterogeneity

⇨Uniform representation format

Semantic heterogeneity

⇨Controlled vocabularies, thesaurus, ontologies…

Common way to query the data

The Semantic Web

Publish

Link

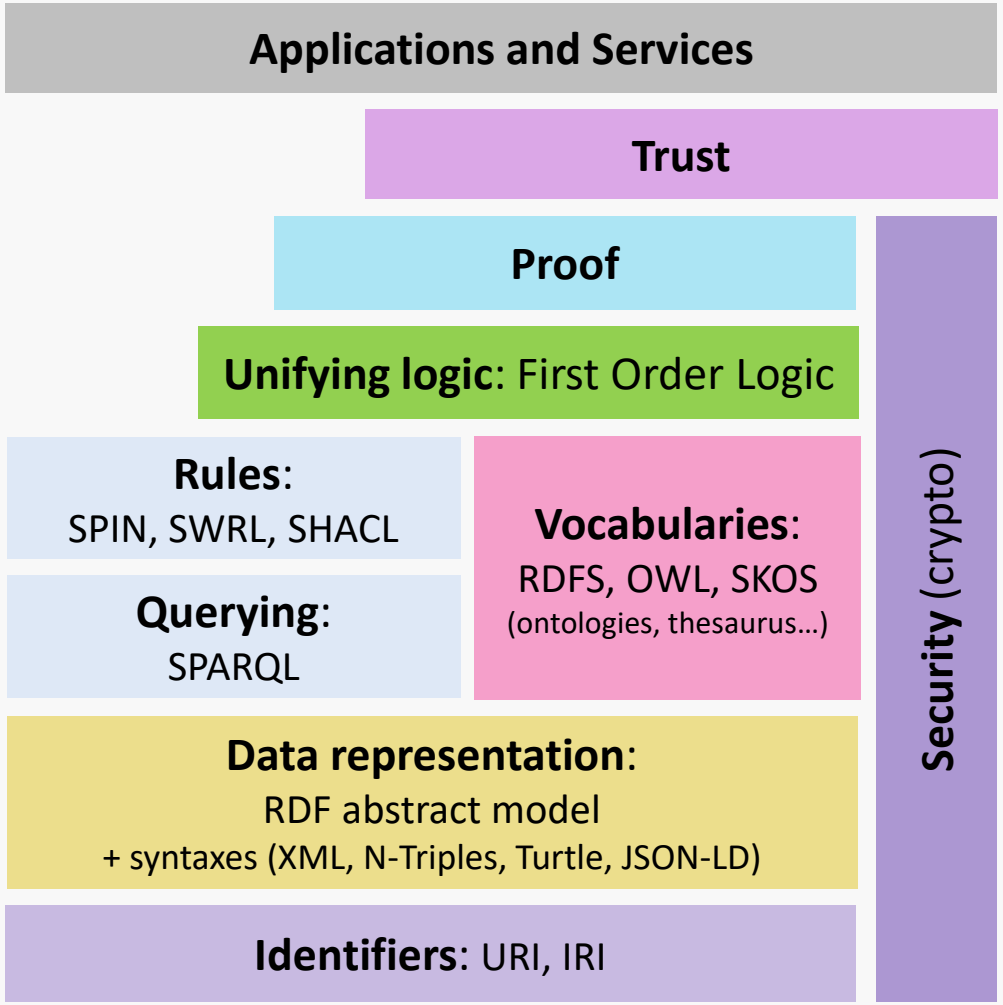
Vocabularies

*"The **Semantic Web** provides an environment where applications can **publish** and **link** data, define **vocabularies**, **query** data at web scale, and draw **inferences**."* *(adapted from W3C website)*

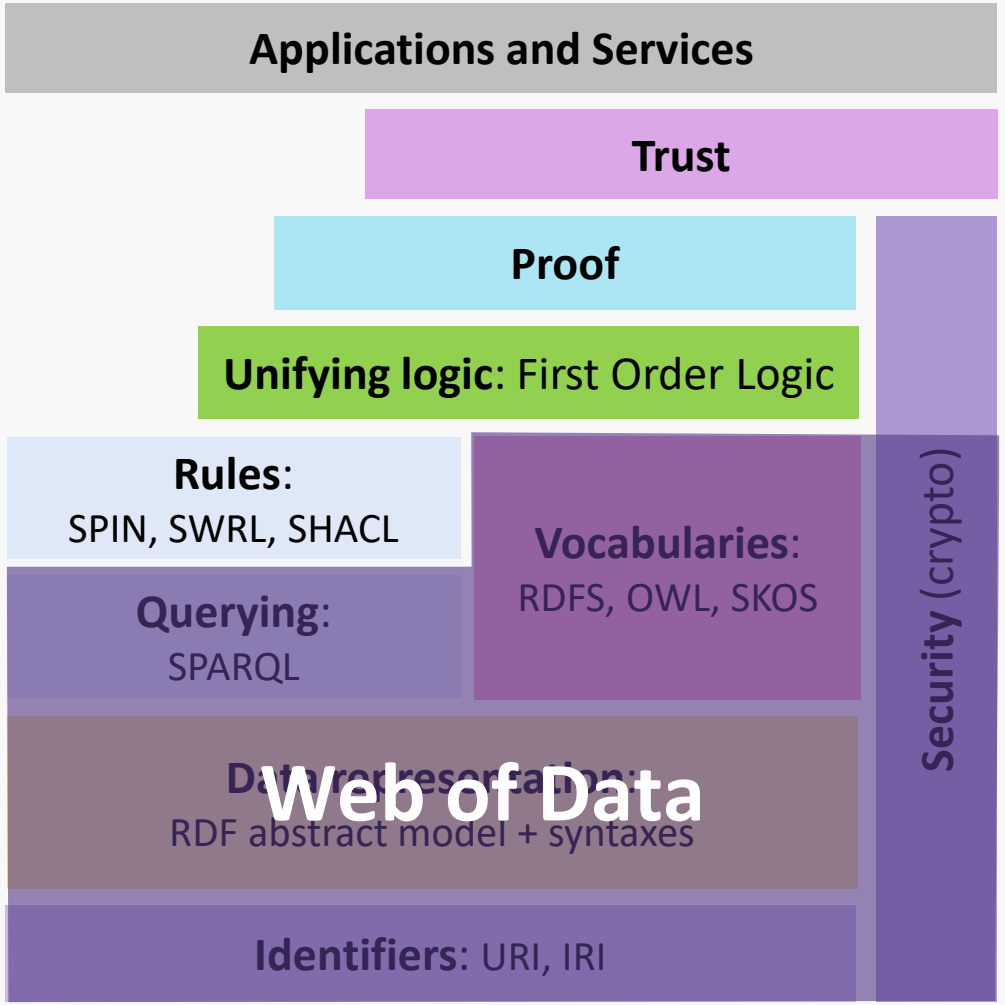
Querying

Inference

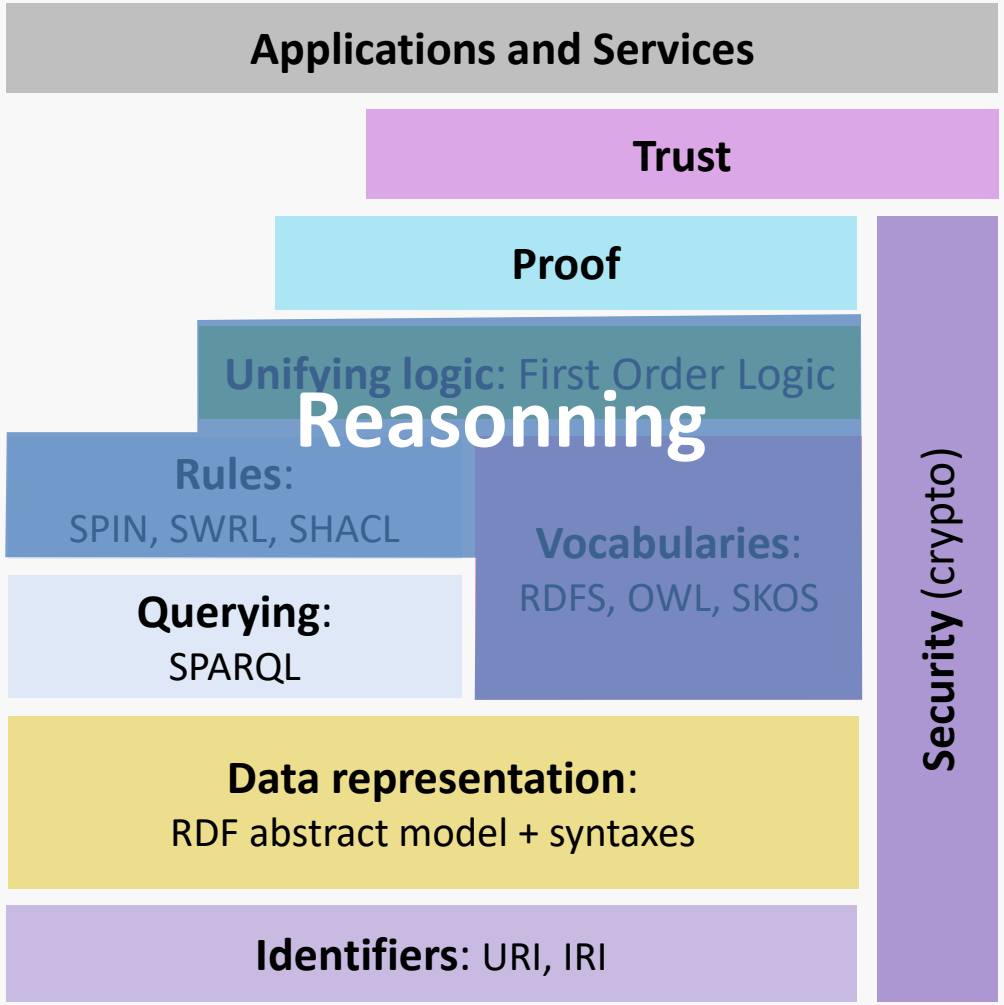
Standards of the Semantic Web



Standards of the Semantic Web



Standards of the Semantic Web



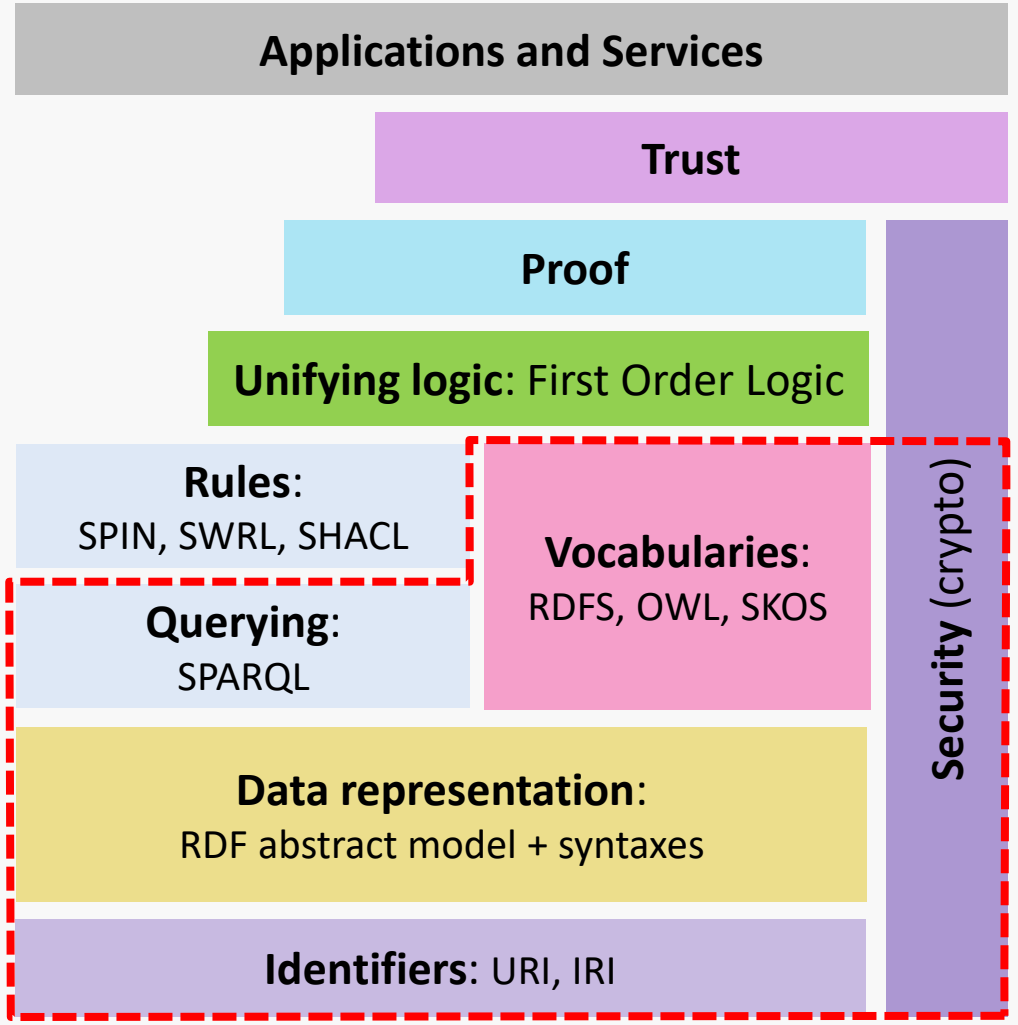
The Web of Data

aka. Data Web, Web 3.0,
Distributed Knowledge Graph…

The Web of Data

First step in the deployment of the **Semantic Web**

Detractors would say *"the part of the Semantic Web that actually works"*…



Publish

Link

Vocabularies

Web of Data

*"The ~~Semantic Web~~ provides an environment where applications can **publish** and **link** data, define **vocabularies**, **query** data at web scale~~, and draw inferences~~."*

Querying

Inference

Vocabularies to name things

Folksonomy

- Collaborative/social tagging, social classification…
- Tag category schemes
- No (not necessarily) hierarchical categorization

Taxonomy

- Practice and science of classification
- Hierarchical categorization of controlled classes/terms
- Nested classes under broader categories

Thesaurus

- Semantic network of controlled vocabulary terms/concepts
- Semantic relationships, e.g. synonyms, antonyms, part of…

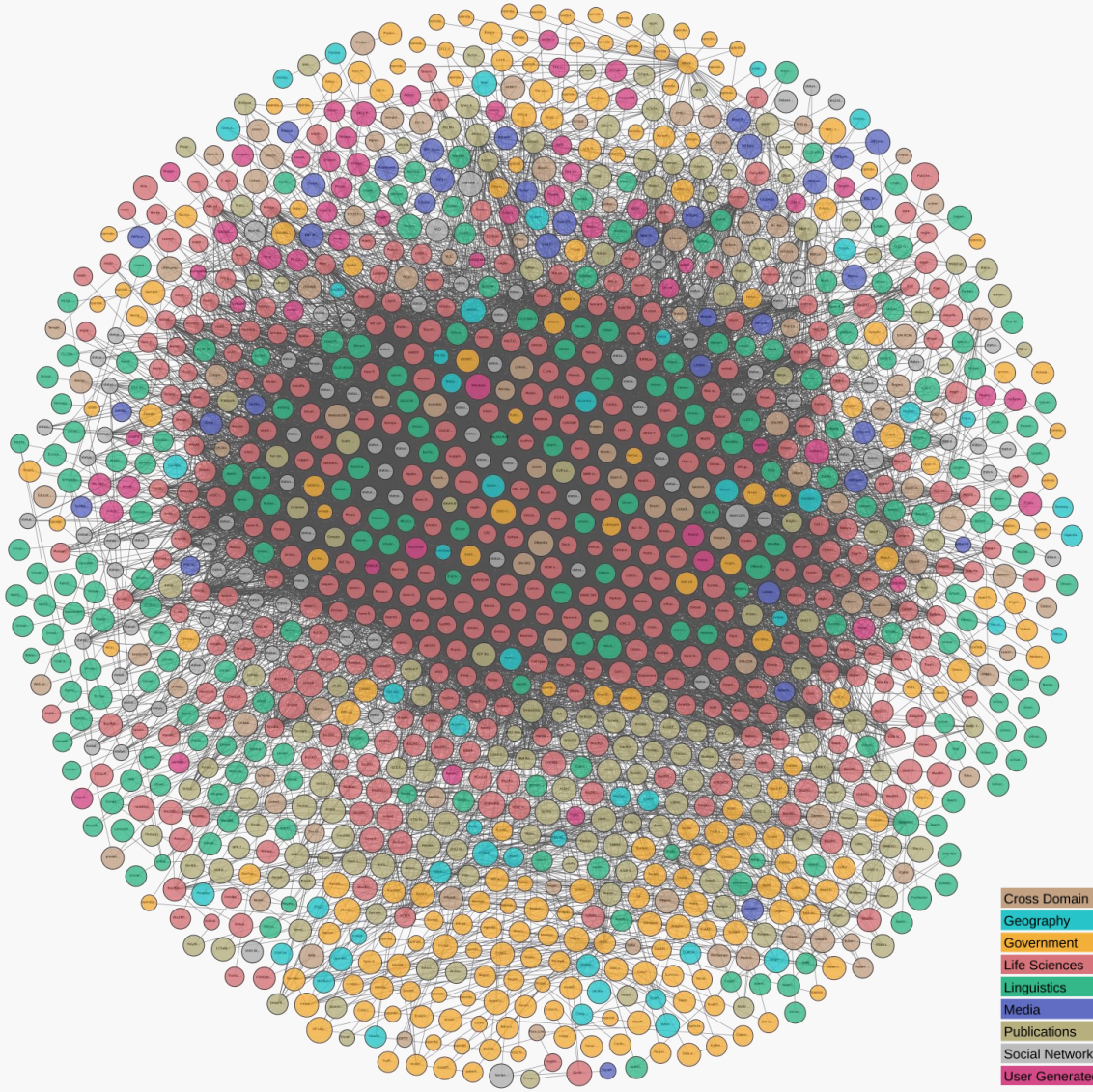
Ontology

- Formal semantic description for the classes, properties and their relationships in a domain of discourse, to facilitate conceptual search and reasoning.

Linked Data principles (Tim Berners Lee)

1. Use URIs to name things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information using the standards (RDF, SPARQL)
4. Include links to other URIs, so they can discover more things

Linked Open Data Cloud: 1200+ linked datasets



Linking Open Data cloud diagram, 2020. J.P. McCrae, A. Abele, P. Buitelaar, R. Cyganiak, A. Jentzsch, V. Andryushechkin, J. Debattista, J. Nasir. http://lod-cloud.net/

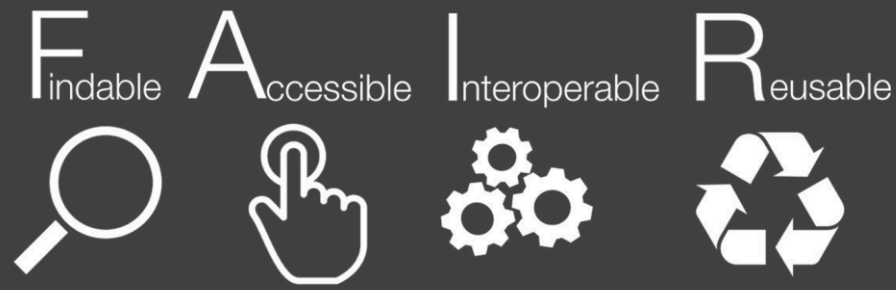
- ✓ On the web, under open licenses
- ✓ Machine-readable (RDF)
- ✓ URIs to name things
- ✓ Use common vocabularies
- ✓ Linked with each other
- ✓ Queryable

Iconic but partial view of the Web of Data
LOD Atlas: 25,000+ datasets

Take-aways

- The SW provides explicit, machine-processable data semantics
- Multiples technologies to address new data integration scenarios
- SW as a toolbox: no need to apply all at once
- Pay-as-you-go approach: use only what you need, now.
  - Assign HTTP URIs to reference vocabularies to need to be shared
  - Make URIs dereferenceable (to web page, RDF… whatever!),
  - Reuse other's well-adopted vocabularies when relevant
  - Assign HTTP URIs to anything!
  - Links, links, links…

- These technics are a way (best ?) to achieve Open Data, Open Science, FAIRness



Thank you