Semantic Web Ontologies

Karl Hammar

2019-02-01

WHAT'S THIS THEN?

- The Semantic Web Vision
- Linked Data and RDF
  - RDF vs Property Graphs
- OWL Ontologies
- (possibly, if there's time: OWL reasoning and its challenges)

- Quid pro quo: survey to be filled out (and ideally shared) after the workshop
  - 15 minutes at most
  - Chance to win cinema tickets

WHO'S THIS THEN?

- Lecturer and until recently program manager at Jönköping University (JU).
- BSc (2008) / MSc (2009) at JU, PhD (2017) at Linköping University.
- Consultant in software development and IT since 1998
  - Primary focuses/skill sets: web and application development in high-level languages for customers in the public policy sector
- Hacker (not very good one?) turned academic.
- Co-founder Jönköping AI Lab
- Research interests:
  - Ontology engineering methods and tools
  - Architectures for integrating AI/ML techniques and knowledge graphs
- Fully funded 2019-2020 to explore the above interests in a Smart Buildings / Smart Cities context w/ REC project.

ERRATA

- IRI = URI = URL (more or less)

THE SEMANTIC WEB VISION

- Berners-Lee, Tim, James Hendler, and Ora Lassila. "The semantic web." *Scientific American* 284.5 (2001): 28-37.
- *”The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation”*
  – Tim Berners-Lee.

… BUT ISN'T THIS OBVIOUS?

- It may be obvious NOW, but it has taken some time:
- The Semantic Web article was published in 2001, ten years prior to Siri being introduced in iOS 5.
- Guha, Ramanathan, Rob McCool, and Eric Miller. "Semantic search." *Proceedings of the 12th international conference on World Wide Web*. ACM, 2003.
- Google introduced Knowledge Graph hits in search results in 2012, based on the above article, 9 years later.

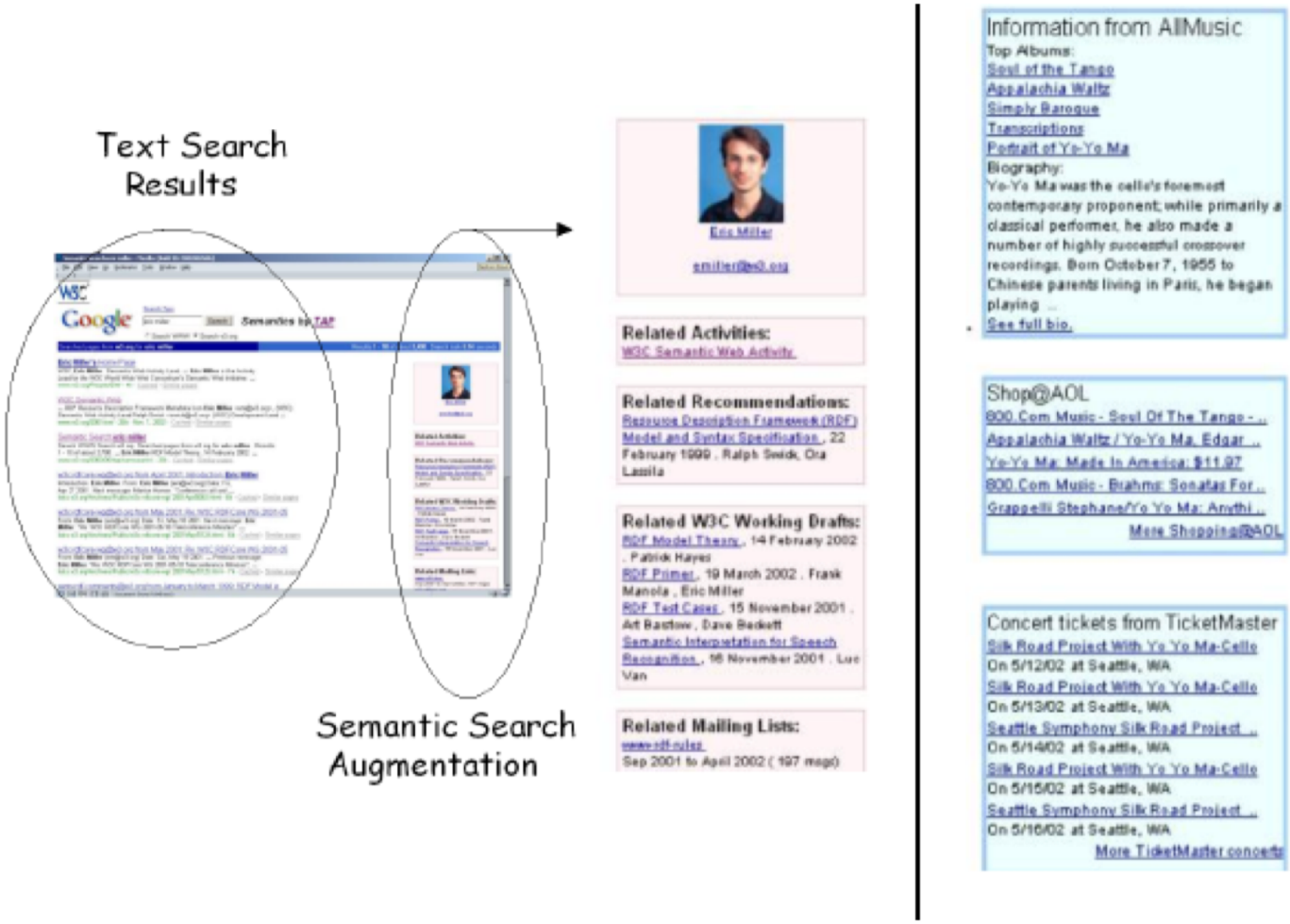


Figure 3: Text Search Results Augmentated with Data from the DataWeb. Left : Search for 'Eric Miller' on the W3C Web site showing overall page and data augmentation. Right: Data augmentation alone for search on 'Yo-Yo Ma'

SEMANTIC WEB REQUIREMENTS

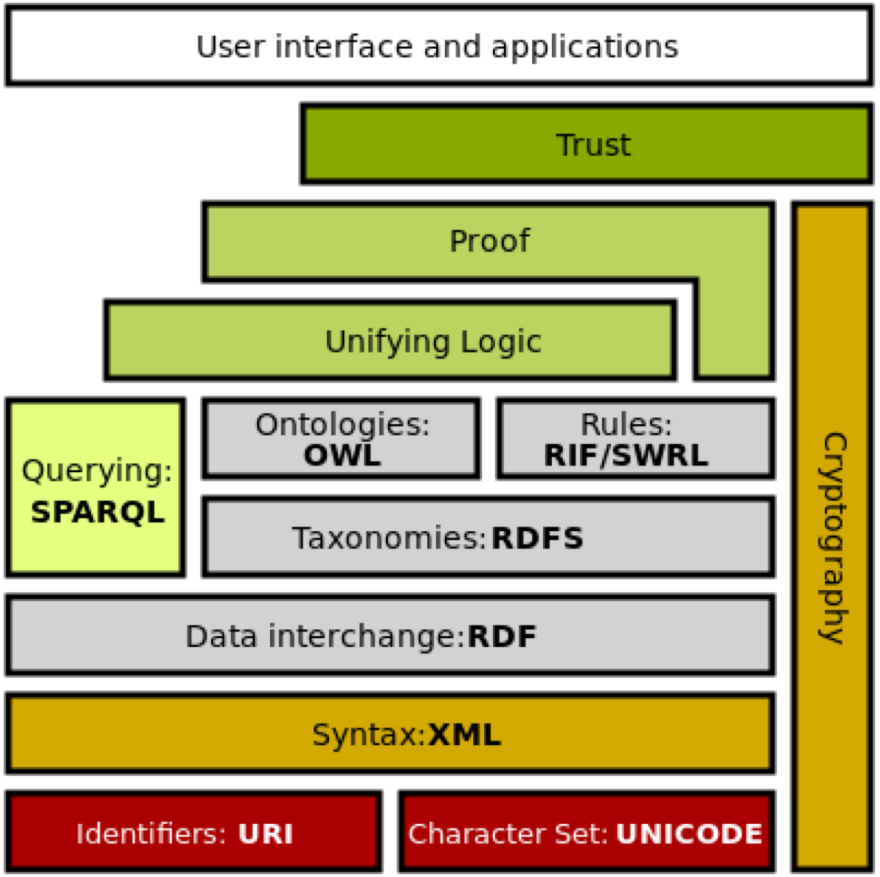
- Implementing this vision requires (at least) two things:
  1. Standard formats for open exchange of data between systems on the net (XML, **RDF**, JSON, …?)
  2. Standard formats for expressing semantics of that data (**RDFS**, Dublin Core, SKOS, **OWL**)
- But also:
  - Knowledge/data to be made available (business cases)
  - Agent software that supports use of the above in a user-friendly manner

LINKED DATA

- Berners-Lee, Tim. ”Linked Data.”, W3C, https://www.w3.org/DesignIssues/LinkedData.html:
  1. Use URIs as names for 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 that they can discover more things.

RDF - RESOURCE DESCRIPTION FRAMEWORK

- Cyganiak, Richard, David Wood, and Markus Lanthaler, ” RDF 1.1 Concepts and Abstract Syntax”, W3C, https://www.w3.org/TR/rdf11-concepts/



RDF DATA MODEL

- RDF describes *data graphs*
- Three (but actually two) datatypes:
  - IRI reference (anything that is a resource/entity)
  - Blank nodes (same as IRIs, but without identity)
  - Literal (anything else: strings, integers, floats, etc)
- Basic building block: the RDF *statement* or *triple*:
  - Structure: <subject> <predicate> <object>
  - States that some relationship, indicated by the predicate, holds between the resources denoted by the subject and object.

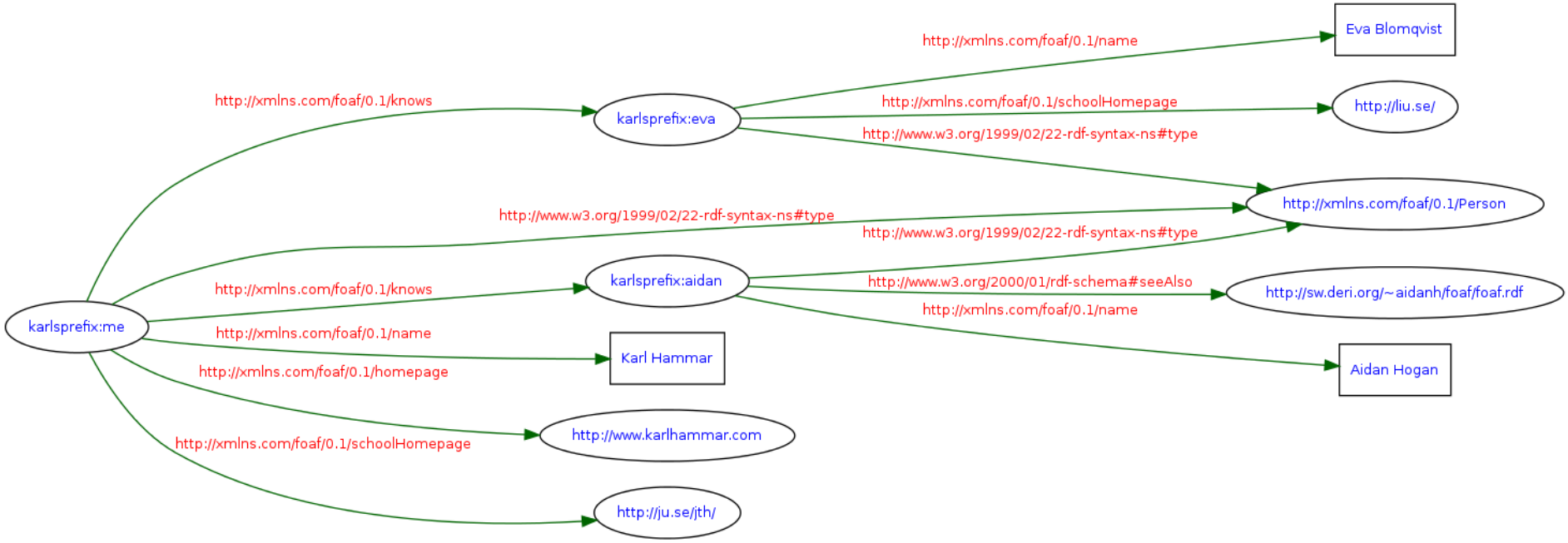
RDF EXAMPLE

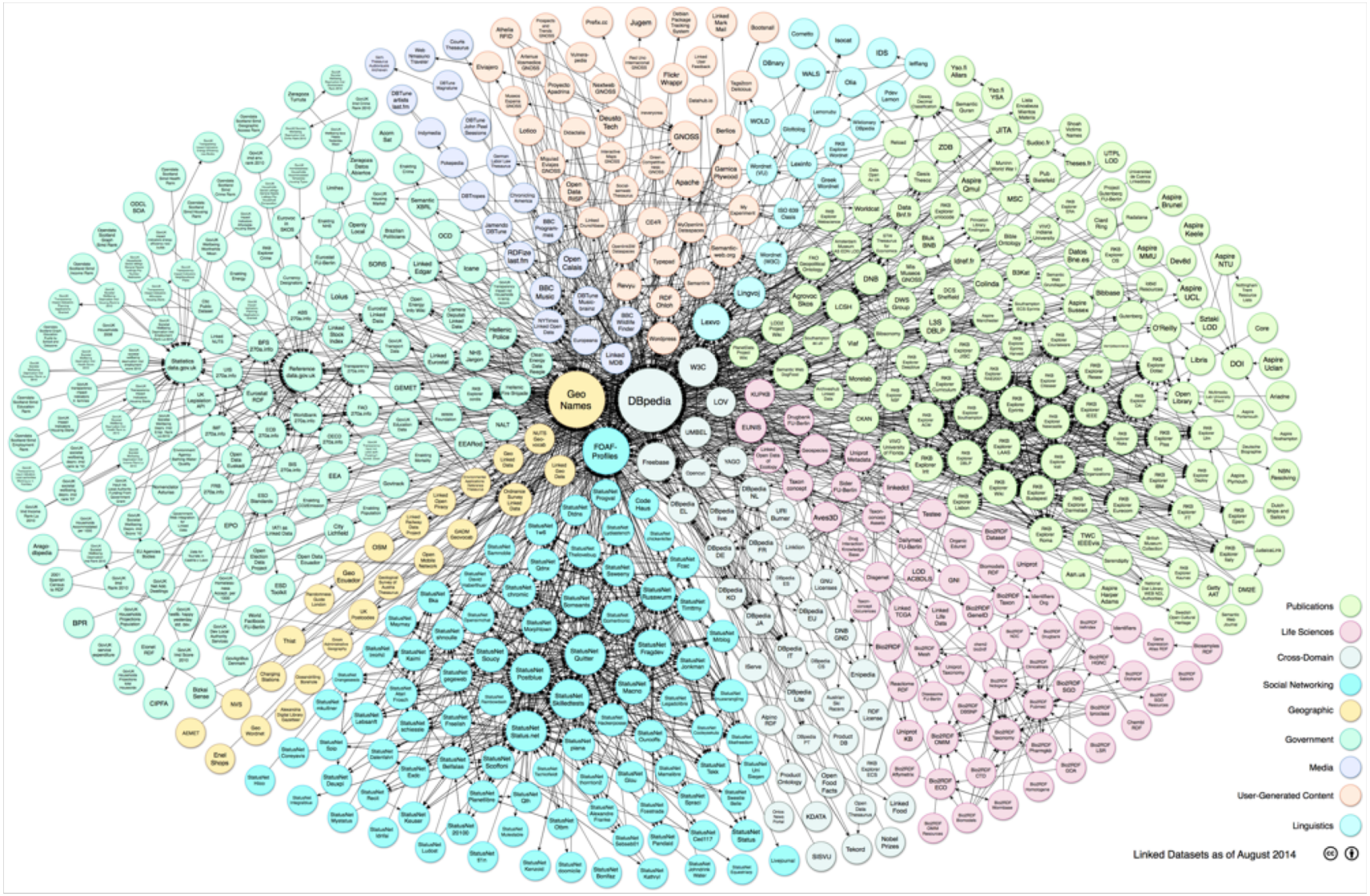
```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix karlsprefix: <http://purl.karlhammar.com/MyFoafFile.rdf#> .

<karlsprefix:me> <foaf:name> "Karl Hammar" .
<karlsprefix:me> <rdf:type> <foaf:Person> .
<karlsprefix:me> <foaf:homepage> <http://www.karlhammar.com> .
<karlsprefix:me> <foaf:schoolHomepage> <http://ju.se/jth/> .
<karlsprefix:me> <foaf:knows> <karlsprefix:eva> .
<karlsprefix:me> <foaf:knows> <karlsprefix:aidan> .

<karlsprefix:eva> <foaf:name> "Eva Blomqvist" .
<karlsprefix:eva> <foaf:schoolHomepage> <http://liu.se/> .
<karlsprefix:eva> <rdf:type> <foaf:Person> .

<karlsprefix:aidan> <foaf:name> "Aidan Hogan" .
<karlsprefix:aidan> <rdf:type> <foaf:Person> .
<karlsprefix:aidan> <rdfs:seeAlso> <http://sw.deri.org/~aidanh/foaf/foaf.rdf> .
```


Eva Blomqvist
http://xmlns.com/foaf/0.1/name
http://xmlns.com/foaf/0.1/knows
karlsprefix:eva
http://xmlns.com/foaf/0.1/schoolHomepage
http://liu.se/
http://www.w3.org/1999/02/22-rdf-syntax-ns#type
http://xmlns.com/foaf/0.1/Person
http://www.w3.org/1999/02/22-rdf-syntax-ns#type
http://www.w3.org/1999/02/22-rdf-syntax-ns#type
http://xmlns.com/foaf/0.1/knows
karlsprefix:aidan
http://www.w3.org/2000/01/rdf-schema#seeAlso
http://sw.deri.org/~aidanh/foaf/foaf.rdf
http://xmlns.com/foaf/0.1/name
karlsprefix:me
http://xmlns.com/foaf/0.1/name
Karl Hammar
Aidan Hogan
http://xmlns.com/foaf/0.1/homepage
http://www.karlhammar.com
http://xmlns.com/foaf/0.1/schoolHomepage
http://ju.se/jth/


Geo Names
DBpedia
FOAF-Profiles
W3C
Publications
Life Sciences
Cross-Domain
Social Networking
Geographic
Government
Media
User-Generated Content
Linguistics
Linked Datasets as of August 2014

WHAT ABOUT PROPERTY GRAPHS?

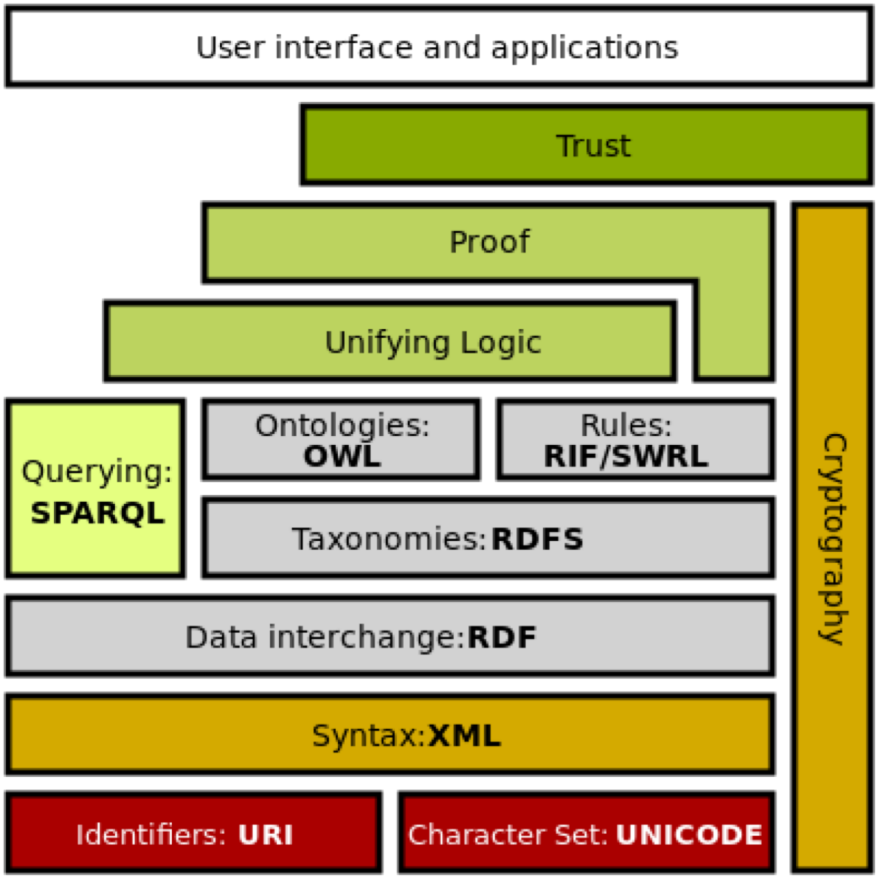
- Alternative method to graph data modelling, employed by Neo4J, Amazon Neptune, Apache TinkerPop, (Cosmos DB?), etc
- Nodes and edges are first-order structures, each of which can be associated with key-value pairs (properties).
- RDF advantages: IRI:s as identifiers, ease of serialization/deserialization, W3C standard, standardized query language (SPARQL), schemas for consistency checking (OWL or SHACL), research community buy-in, built in multilinguality
- PG advantages: richer modelling, simple statement-level metadata, implementations supporting graph-centric use cases (path-finding),

ONTOLOGIES

- *“explicit specification of a shared conceptualization that holds in a particular context”*
  (several authors)
- In layman’s terms: the schemas for Linked Data:
  - Concepts
  - Relations
  - Restrictions
  - Induction
- In layman’s terms: a set of reserved IRIs that carry particular meaning when used in RDF graphs

RDFS AND OWL

- RDFS = RDF Schema
  - Adds a little bit of structure to RDF: typing of nodes, subclasses, domain and range of properties
- OWL = Web Ontology Language
  - Equivalent classes and properties
  - Equivalent and disjoint entities
  - Inverse, transitive, functional, etc. properties
  - Value and cardinality constraints



EXAMPLES

- The property *”has father”* is an object property that has as domain *”Person”* and as range *”Person”*. It is also a functional property.
- The property *”has uncle”* is equivalent to a property chain *”has father -> has brother”*.
- The property *”has ancestor”* is transitive.
- The class *”Car”* is a subclass of the cardinality restriction *”has wheels minimum 3”*
- The class ”Car” is a subclass of the value restriction *”has owner Person OR Company”*

VALUE OF OWL REASONING

- Classification of entities by necessary and sufficient conditions.
- Keeping logic pertaining to the knowledge declaratively with the knowledge itself, rather than in executable program code somewhere else.
- Detection of inconsistency in data or in ontology model.
  - Example:”Car hasOwner only Person” + ”HAX999 isA Car” + ”HAX999 hasOwner JönköpingUniversity” => Either JU isA Person OR there is an inconsistency in the dataset.

THE OPEN WORLD ASSUMPTION

- Traditional databases operate under closed world assumption, i.e., that in our database we have all the knowledge concerning the domain that is of any importance.
  - Therefore, the absence of fact implies the negation of fact: ”There is no car with the license plate HAX-999”
- Semantic Web ontology languages operate under assumption that data is fluid and may or may not be available at any given time:
  - ”We do not know that there exists a car with the license plate HAX-999”

OPEN WORLD ASSUMPTION EFFECTS

- No reasoning by default – i.e., no way of expressing default state
- Unexpected classification can occur:
  - Example from last slide: JU isA Person?
- Computability characteristics unfavourable (though workarounds exist)
- Other options: use techniques for RDF schemas (e.g., SHACL) that are not open world, possibly implementing reasoning as needed for specific problems or use cases.

AAA SLOGAN

- Anyone can say Anything about Any topic.
- I.e., do not assume that only controller of subject namespace can express facts about subject.
- Makes data integration trickier

JÖNKÖPING UNIVERSITY

School of Engineering