Unraveling the Semantic Web

Pedro Reis

https://github.com/pmreis/semweb

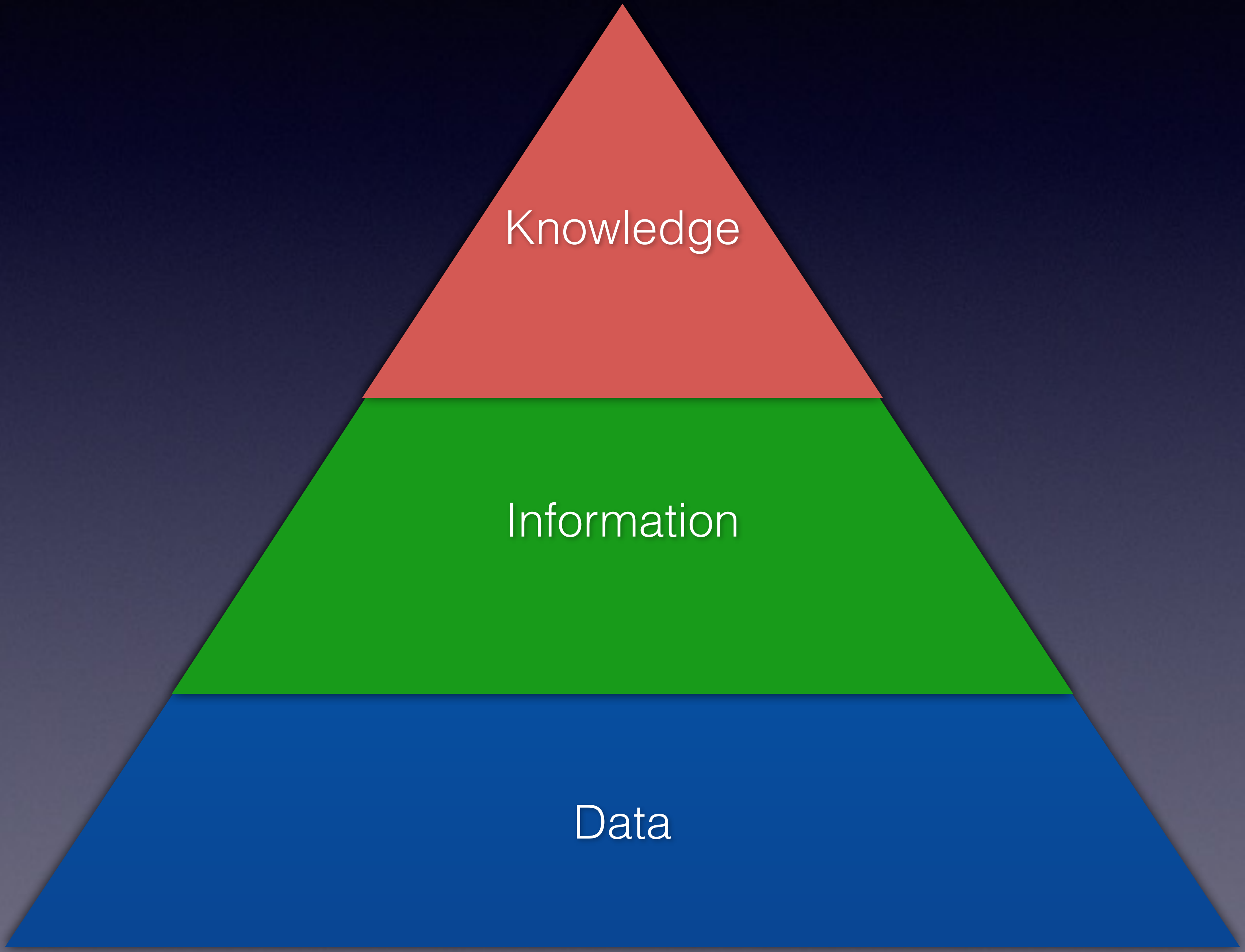
The Web

- The Web
  - It is not the Internet
  - Invented by Tim Berners-Lee @ CERN
  - First site created in 1991
  - Accessing documents on any machine connected to the internet (HTTP), while using a common representation format (HTML)

The Web 2.0

- Dynamic documents (contents / data)
- Data produced by users
- First used in 1999
- JavaScript; AJAX; XML; JSON…
- Rich Internet Applications; Web Oriented Architecture; Social Web…

Semantic Web! Why?



Knowledge

Information

Data

Data

- What is this?
    - 38 44 58.5

Data

- What is this?
  - 38 44 58.5

OK! It is a number!

But it represents what exactly?

Data

- What about now?
  - 38°44'58.5"N

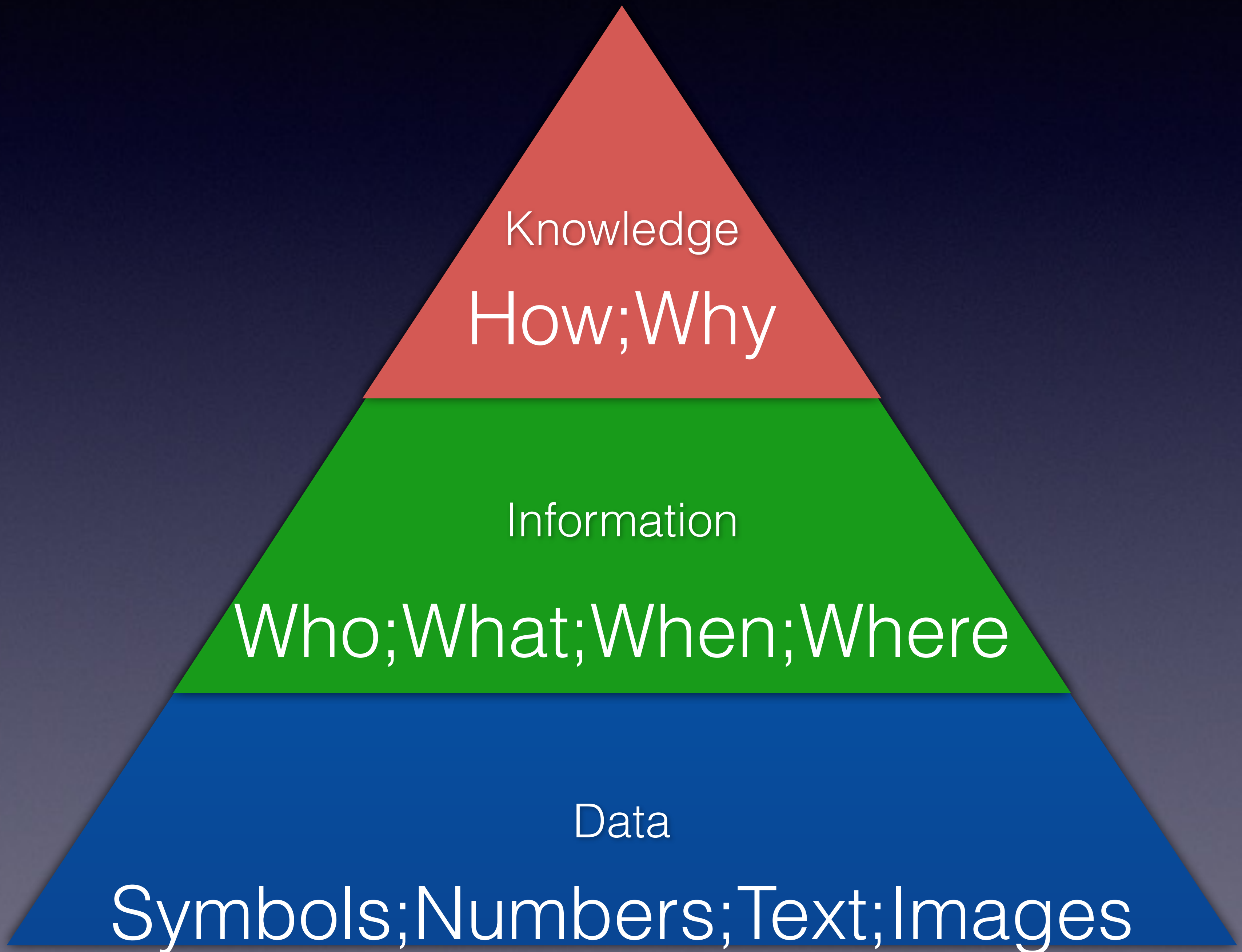
Data

- Completely useless without context
- Has different types (heterogeneous)
    - Text
    - Number
    - Image
    - Sound

Information

- Allows us to make decisions
- Provides context to data, making it useful
    - Describes the data

Semantic Web! Why?



Knowledge

How;Why

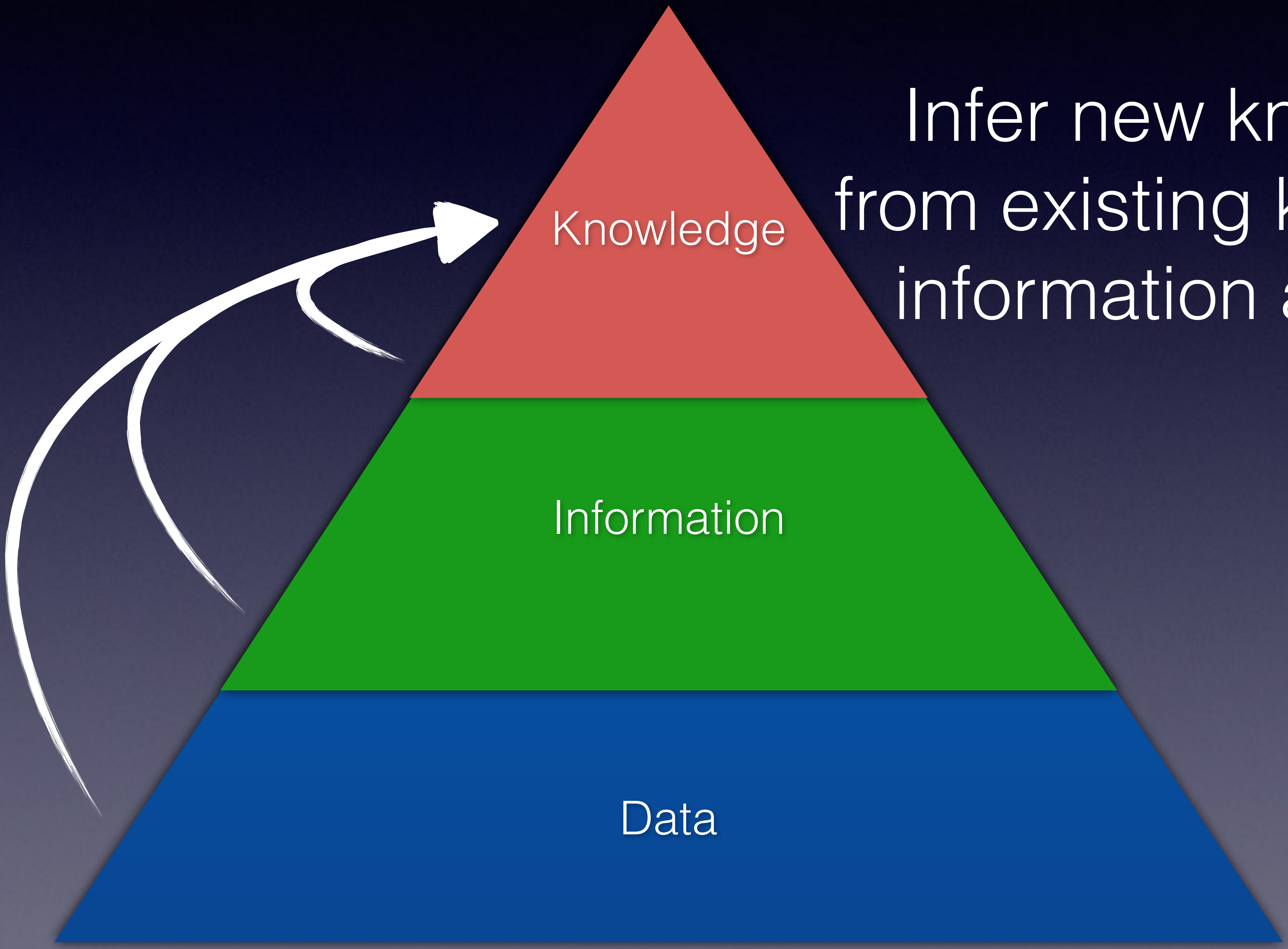
Information

Who;What;When;Where

Data

Symbols;Numbers;Text;Images

Semantic Web! How?



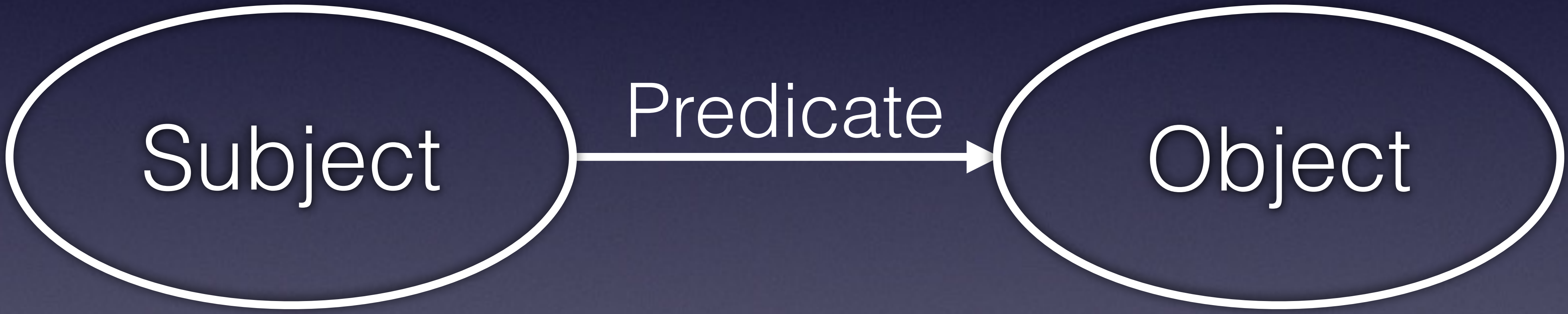
Infer new knowledge from existing knowledge, information and data.

Linked Data

- Linked Data is the basic element
    - Data has built-in context
    - The atomic form is a Statement
    - Not to be confused with Open Data
        - Open Data may be Linked Data

Linked Data

Structure of a Statement, aka Triple



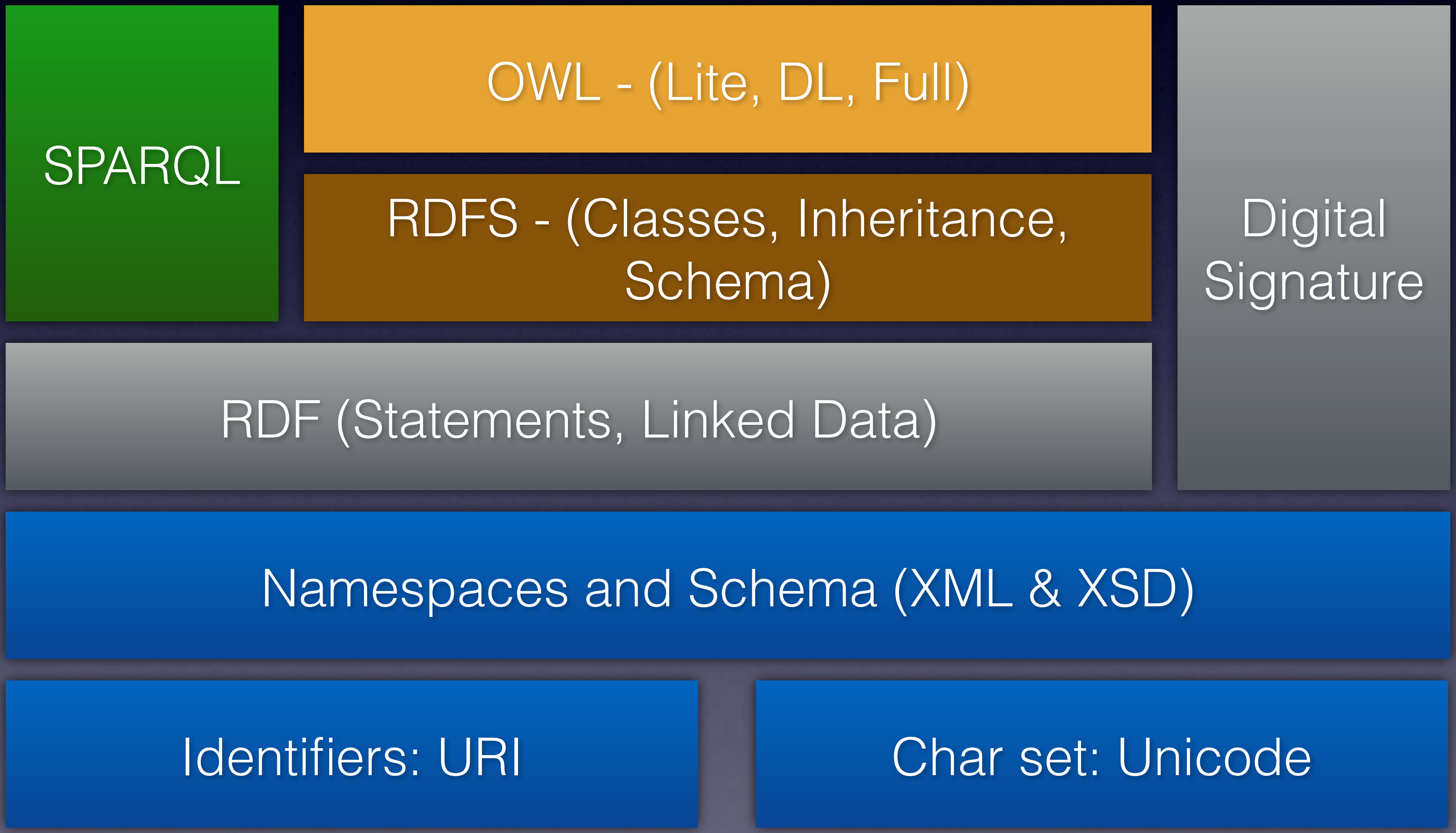
5 Stars rating (1-2)

- 1 star
    - Make data available on the web
    - Any format
    - Open licence
- 2 stars
    - Format must be machine-readable

5 Stars rating (3-5)

- 3 stars
    - Publish it in a non-proprietary format
- 4 stars
    - Use W3C standards
- 5 stars
    - Link the published data with existing data

The Semantic Web Stack



SPARQL

OWL - (Lite, DL, Full)

RDFS - (Classes, Inheritance, Schema)

Digital Signature

RDF (Statements, Linked Data)

Namespaces and Schema (XML & XSD)

Identifiers: URI

Char set: Unicode

The Semantic Web Stack

- Other layers are still in research / development
    - Logic and Proof
    - Reasoners allow inference, but inferences should be validated against trusted sources
- Digital Signature layer
    - Enables trust. Filters only trustable sources of data / ontologies

RDF - Analogy & Example

- RDF or Linked Data
  - Similar to a table on a RDBMS
  - Table = Subject; Field = Predicate; Value = Object
  - Serialization formats: Turtle; N-Triples; JSON-LD; RDF/XML

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:mus="http://ontomusic#">

<rdf:Description rdf:about=“http://albums/thebluenotebooks”>
  <mus:title>The Blue Notebooks</mus:title>
  <mus:artist>Max Richter</mus:artist>
  <mus:year>2004</mus:year>
</rdf:Description>
```

Namespace

Subject

Predicate = Year ; Object = 2004

RDFS - Analogy

- RDFS
  - Similar to a database model (schema) on a RDBMS
  - Fields are know as Properties
  - There are no:
    - Explicit triggers
    - Constraints on data values

RDFS - Features

- Enables creation of Classes
- Enables creation of Taxonomies
  - Classes may have subClasses
  - Properties are declared as standalone elements
  - Properties may have subProperties
  - Properties have Domains (classes)
  - Properties have Ranges (data type)

RDFS - Example (N3)

```
@prefix :     <http://www.example.org/sample.rdfs#> .
@prefix rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

    :Animal   rdf:type        :rdfs:Class.
    :Dog      rdfs:subClassOf :Animal.
    :Person   rdfs:subClassOf :Animal.

    :hasChild rdfs:range :Animal;
              rdfs:domain :Animal.
    :hasSon   rdfs:subPropertyOf :hasChild.

    :Max   a :Dog.
    :Abel  a :Person.
    :Adam  a :Person;
     :hasSon :Abel.
```

Simple ontology definition.

Essentially it is a **taxonomy**, because it only defines an **hierarchy**.

Instances of Classes. Same as data.

A semicolon denotes a new statement with the same subject of the previous statement.

Web Ontology Language (OWL)

- Extends the RDFS vocabulary
  - Functional property
  - Transitive property
  - Symmetric property
  - Equivalent class / property
- Ontologies may be linked just like data

OWL

- Leading modelling tool is Protégé.
- Ontology design requires expertise
- Inconsistencies may result in absurd inferences
- Ontology as data provides fast insight into new knowledge domains.

Practical Example

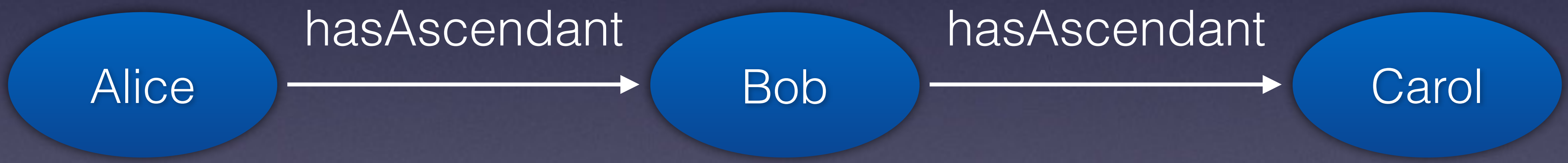
- Typical examples are
  - Governmental data
  - Research data
- Lets try something different…

Similar alternatives

- Before 2012 searching on Google for **boot repair** would only return results related with boot sectors! Not boots (footwear)
  - Google Knowledge Graph
    - Collects data from Wikipedia and Freebase
  - Google Knowledge Vault
    - Uses machine learning to rank untrusted sources of data
- Wikidata

Not all roses

- Ontology design is crucial to inference
  - A property both transitive and functional will result in absurdity



  - Infers that Bob and Carol are the same person

Not all roses

- Classes are assumed to overlap
- Open World Assumption
  - Assumes not all knowledge is available
  - Some questions result in unknown as answer
- Federated queries using multiple sources are difficult to construct

References

- http://www.w3.org/standards/semanticweb/
- http://protege.stanford.edu/
- http://virtuoso.openlinksw.com/
- http://rdf4j.org/
- http://www.bbc.co.uk/ontologies

Questions