Ontologies in the Semantic Web

An Introduction to Knowledge Modeling & Sharing

University of Oldenburg, Germany
Wolfram Wingerath
Dec. 07, 2022

Overview

1

Semantic Web

What is the Semantic Web?

2

Ontologies

What is (an) ontology and how does it relate to the Semantic Web?

3

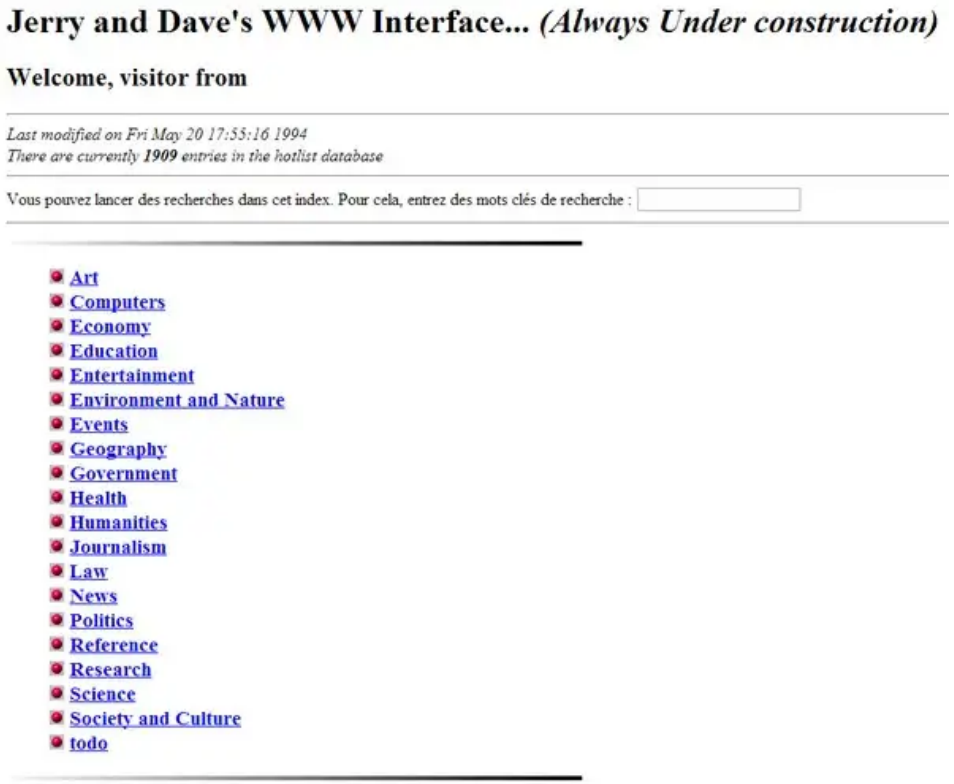
Use Cases & Applications

Where are ontologies used in practice?

The Early Internet and the Web 1.0

- Information access in the **early internet was complicated**:
  - terminal-based (download file ⇨ open file)
  - Required expert knowledge

- In 1990, the **Web 1.0 introduced easy exploration**:
  - Graphical user interface via browsers
  - Linked documents
  - High flexibility (no predefined structure)
  - Keyword search via **search engines**



The Web 2.0: The Social Web

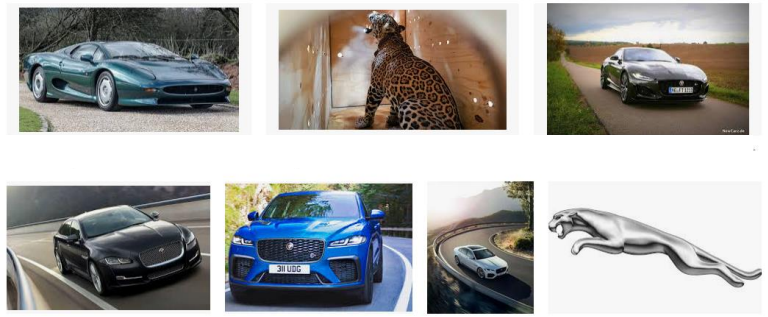
- In the early 2000s, users became **content producers:**
  - Blogs & microblogging
  - Social networks
  - Collaborative projects
  - Dynamic content

Original by Markus Angermeier Vectorised and linked version by Luca Cremonini – Based on web20map.png Vectorised and linked version from Web_2.0_Map.svg

Challenges for Information Management

Internet Scale

automation is required for today‘s amounts of data



Implicit Knowledge

meaning often depends on context and “common sense”

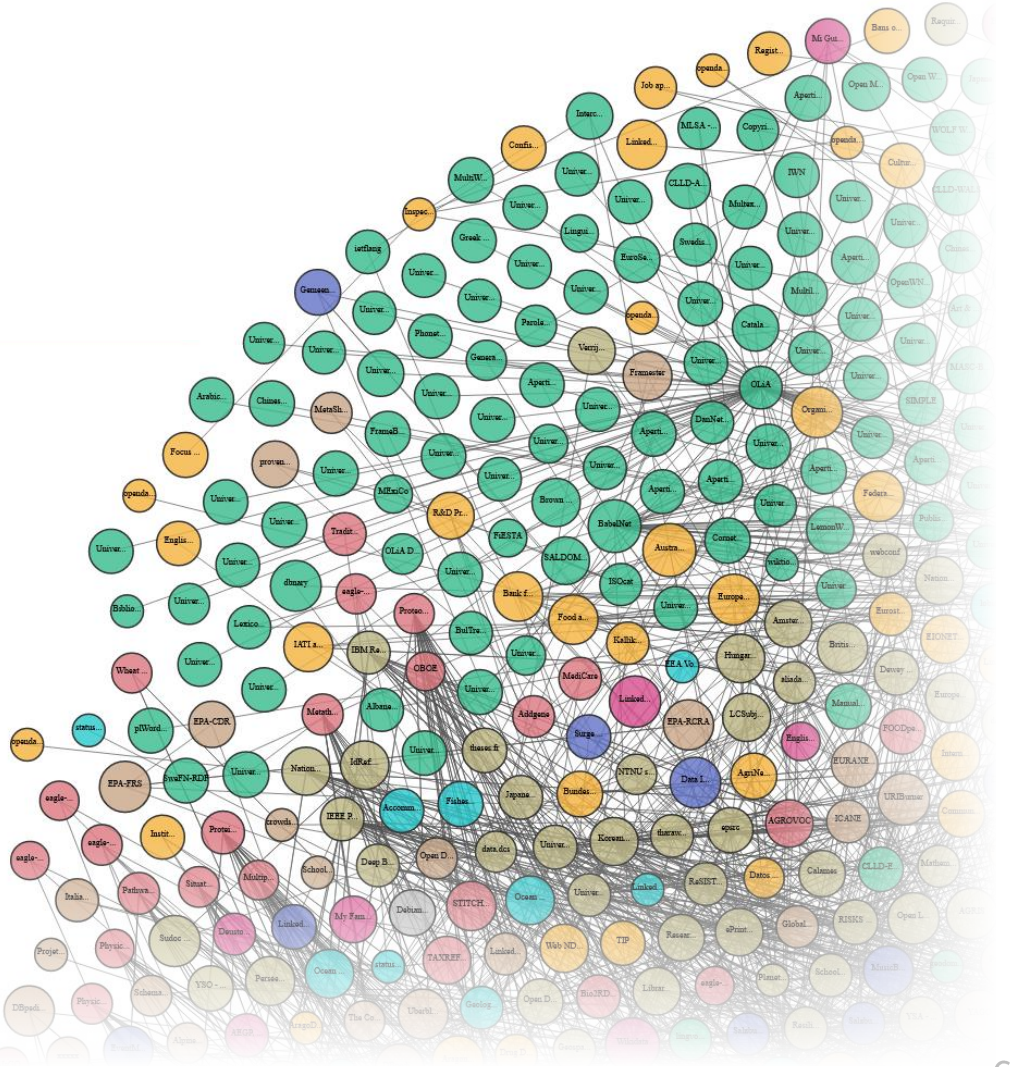
Polysemy

natural language is ambiguous (synonyms, metaphors, etc.)

Harald Sack. Knowledge Engineering With Semantic Web Technologies, Lecture 1: The Limits of the Web, OpenHPI Tutorials (2019)

The Web 3.0: The Semantic Web

- Organized to be **interpretable by machines**
- Semantics can be derived through syntactic rules
  - ⇨ **Meaning** can be computed!
- **World Wide Web ⇨ Giant Global Graph**
  - Standards for knowledge exchange
  - Linked datasets
  - “Smart Data”



An Ontology in Computer Science

We are not talking about philosophy!

consensual knowledge

An abstract model referring to a specific domain

„An ontology is a formal, explicit specification of a shared conceptualization.

N. Guarino, D. Oberle, S. Staab. What Is an Ontology? Handbook on Ontologies (2009)

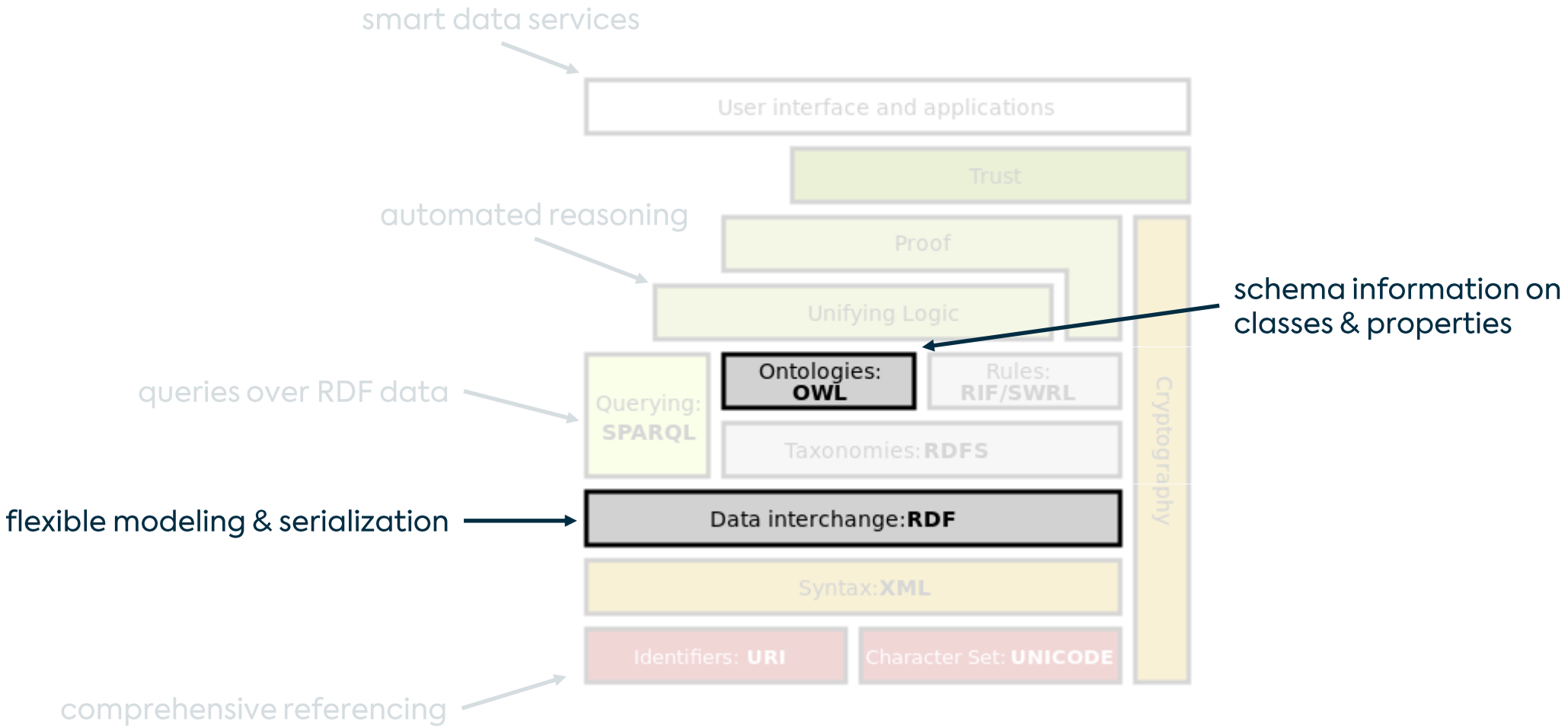
Must be interpretable for machines

All relevant aspects must be defined (concepts, properties, constraints, etc.)

- **Knowledge model**: defines a set of concepts and the relationship between them
- Purpose: capture human knowledge and **common sense**

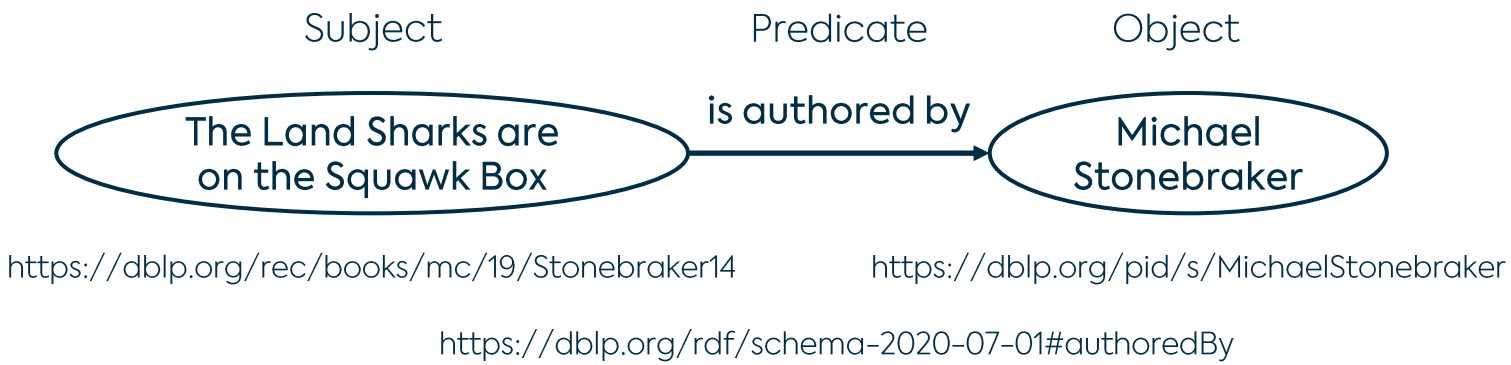
⇨ enable **automated reasoning** and logical inference

The Semantic Web Stack (Simplified)



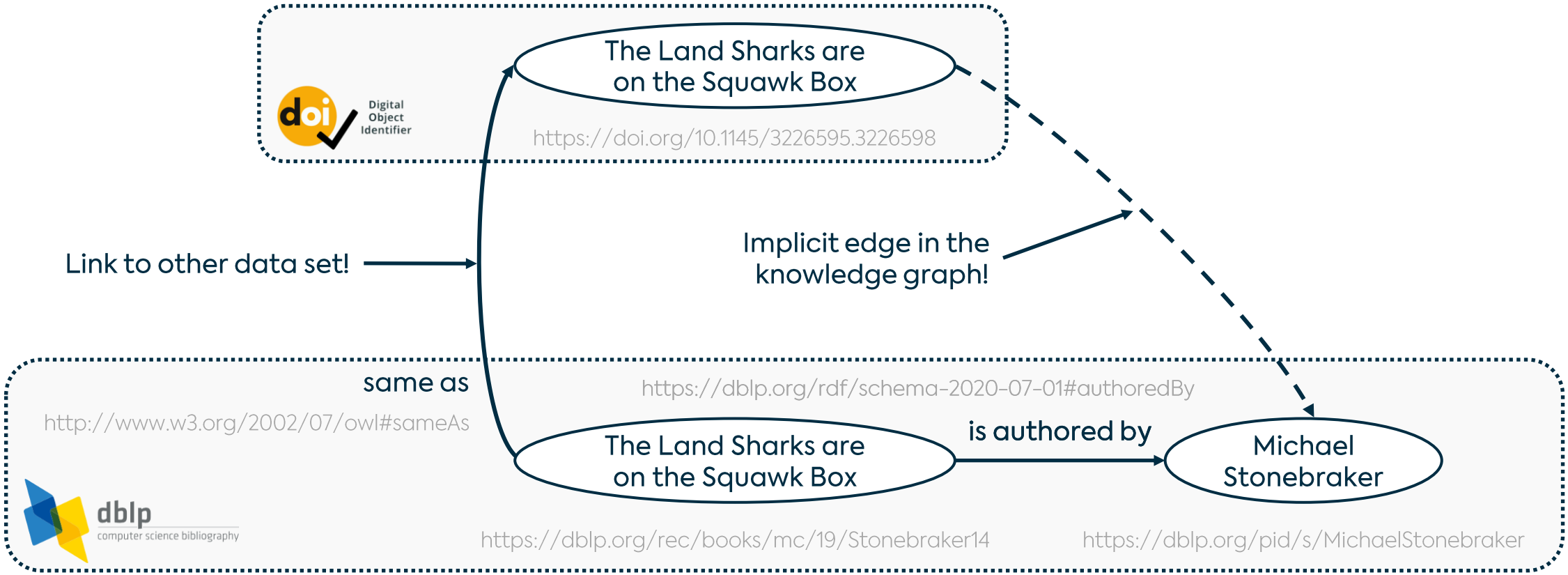
RDF: Resource Description Framework

- The standard method for describing information in a **knowledge graph**
- **Triple-based** data model
- **No schema**
  (everything is a triple)



- Mostly everything has a **unique identifier** to allow reusing data
- Various serialization formats for **easy exchange** (XML, JSON, Turtle, ...)

Semantic Data Integration With RDF



OWL: Web Ontology Language

- **De facto standard** language for expressing ontologies
- **Schema** layer for (RDF) knowledge graphs for specifying
  - (Hierarchies of) classes and properties
  - Equivalence
  - Inverse relations
  - Value ranges
  - …

OWL extends RDF

```
<owl:Class rdf:ID="Class">
  <rdfs:subClassOf rdf:resource="OtherClass" />
  ...
</owl:Class>
```

- Includes **entailment rules** to enable inferring new triples from existing ones
- Different versions (current: 2) and sublanguages (OWL Lite < OWL DL < OWL Full)
- Various serialization formats (XML, JSON, Turtle, …)

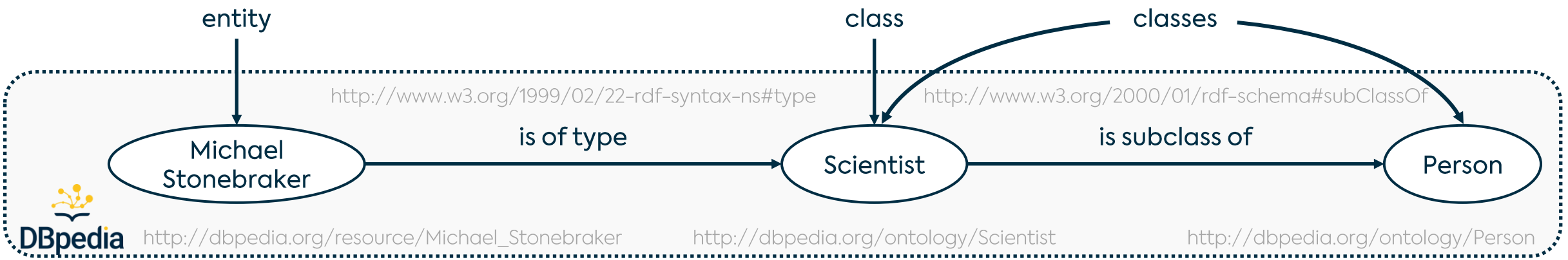
Applying Domain Knowledge to Data

- **Consistency** checks: Are their inconsistencies?
- **Satisfiability** checks: Can all classes have instances?
- **Relationships** between entities and classes (same/different, intersection, disjunctness, …)

Schema info on `Person`:

- Can have a `Mother`
- Disjoint with `Activity`
- …

⇨ `Michael` is an instance of `Scientist` (and by proxy an instance of `Person`), so the same implicitly applies to him!

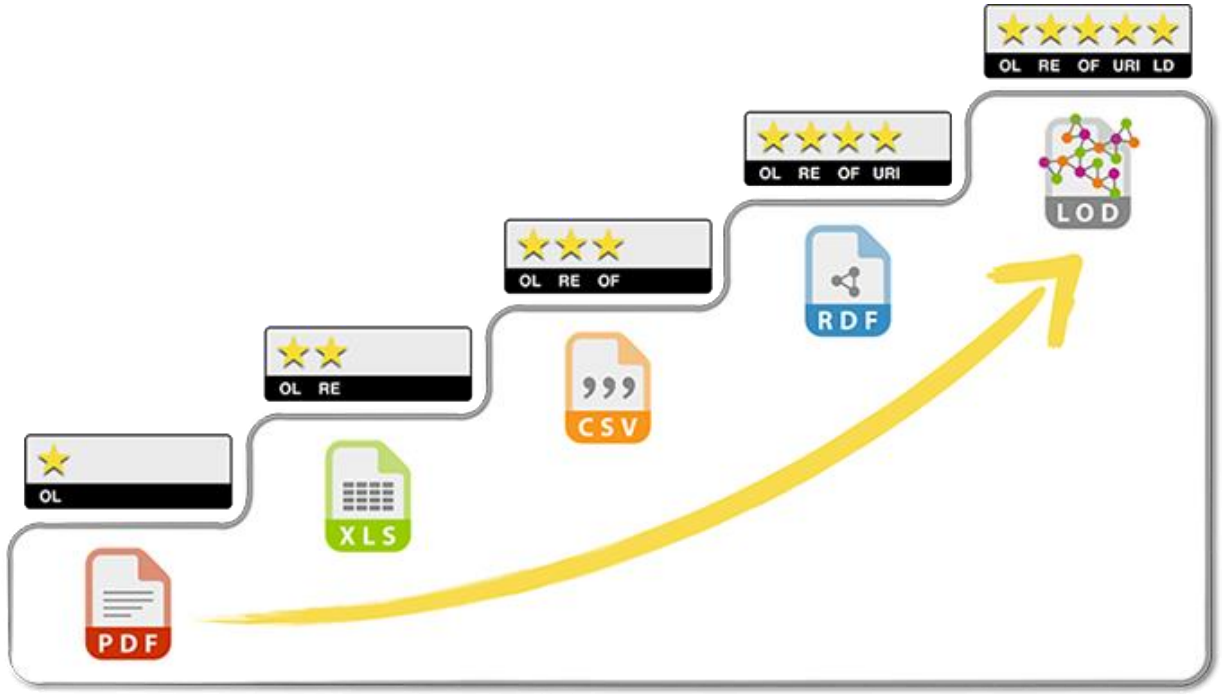


Linked (Open) Data

- **Open Data** is available for free and may be used by anyone for any purpose

- **Linked Data** follows three principles:
  1. Use HTTP URIs for referencing entities and concepts
  2. Use standards (e.g. RDF format & SPARQL query API)
  3. Provide links to other URIs

- **5-Star Grading Scheme**:
  1. Open data
  2. Available as machine-readable structure data
  3. In a non-proprietary format
  4. Follows W3C standards (RDF, URIs)
  5. Links to other LOD sources



Linked Open Data (LOD)

Tim Berners-Lee. Linked Data (Design Issues), W3C (2019)

Linked Open Data Cloud

Challenge: How to avoid **terminology bloat**?

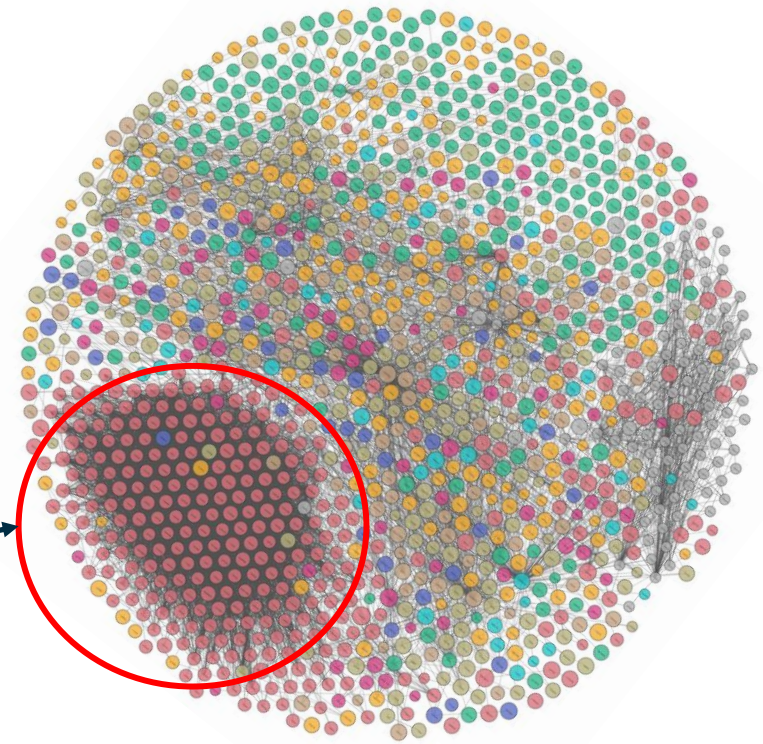
⇨ **Upper Ontologies** to provide shared vocabulary

Notable datasets:

- **DBpedia**: crawled Wikipedia content
- **Wikidata**: metadata to amend Wikipedia
- **BFO**: upper ontology for life sciences

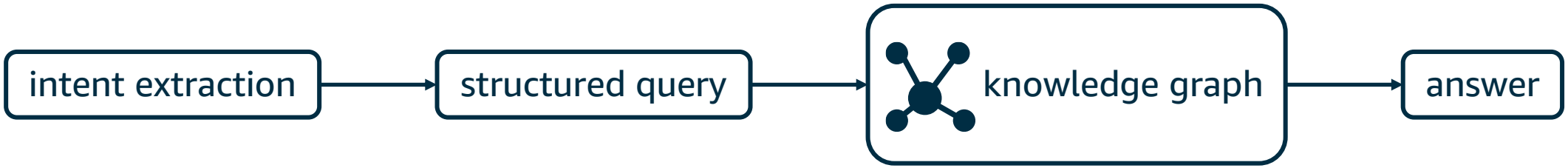
2021:

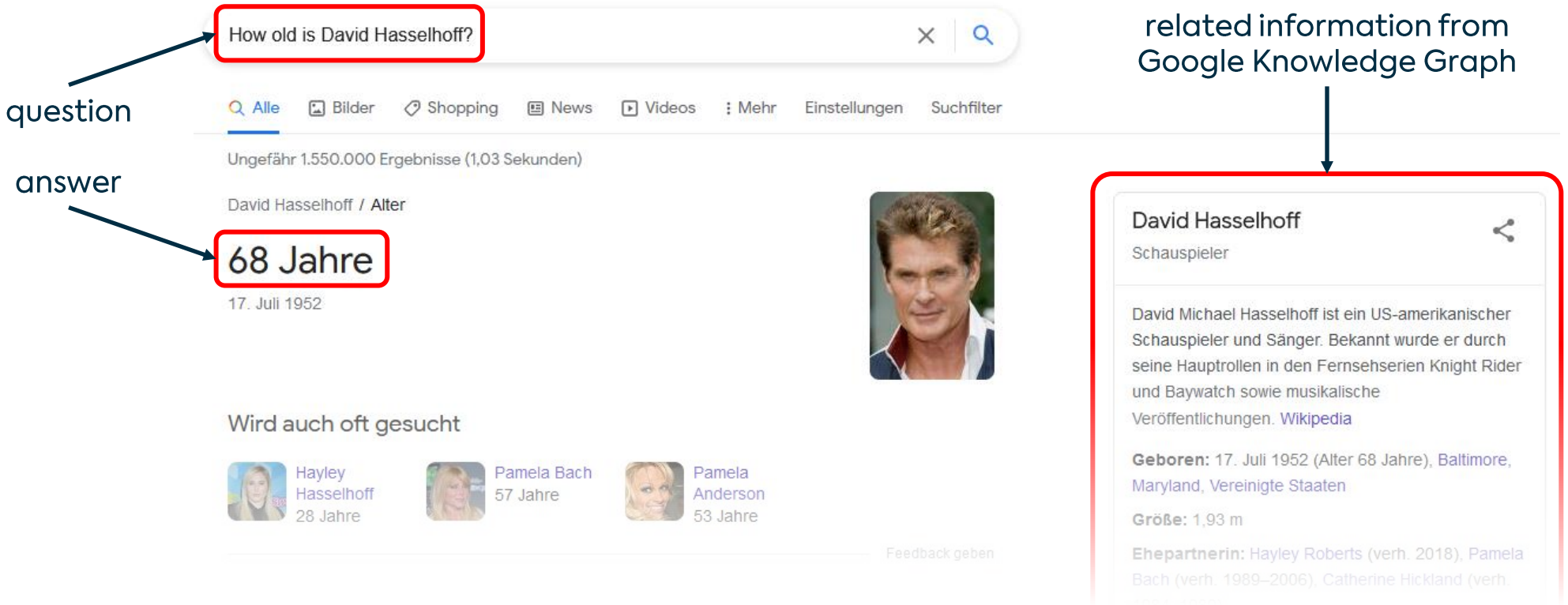
- >1300 data collections
- >16000 links



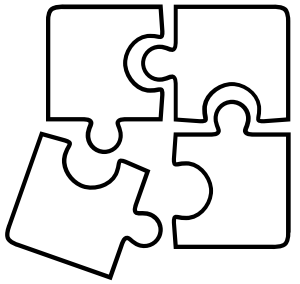
Barry Smith. Ontologies for Space and Ground Systems (Tutorial), Ground System Architectures Workshop (2020)

Use Case: Internet Search





Wrapup: Ontologies in the Semantic Web



- **The Semantic Web** extends the WWW by machine-readable information
- **Ontologies** help share meaning between humans and computers
  - Linked Open Data allows standardized structured queries
  - Semantic interoperability remains a major challenge
- **Use cases** revolve around inference and (derived) meaning, e.g.:
  - Information consumption (search, discovery, recommendation, etc.)
  - Semantic data integration & governance
  - Virtual assistants like Siri or Google Assistant