Ontologies and Knowledge Graphs for FAIR Research Data Management

Prof. Dr. Harald Sack
FAIR Research Data in Plasma Medicine
28 October 2020
650 cm
650 cm
650
$650 \times 10^9 \text{ly}^3$
$650 \times 10^9 \text{ ly}^3$
$650 \times 10^9 \text{ly}^3$
Information
Knowledge
From Data to Knowledge

DIKW Pyramid, Ackoff 1989
“People can’t share knowledge if they don’t speak a common language”

Thomas Davenport (1997)
What is Knowledge?

Traditional Definition: „Knowledge is a subset of all true beliefs“
...to speak a common Language:

- common symbols and concepts (**Syntax**)
- agreement about their meaning (**Semantics**)
- classification of concepts (**Taxonomy**)
- associations and relations of concepts (**Thesauri**)
- rules and knowledge about which relations are allowed and make sense (**Ontologies**)

But what exactly are Ontologies?
 „A theory of being, which tries to explain the being itself, by developing a system of universal categories and their intrinsic relationships...“

Philosophy Definition
An ontology is an explicit, formal specification of a shared conceptualization.

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


**conceptualization:** abstract model
(domain, identified relevant concepts, relations)

**explicit:** meaning of all concepts must be defined

**formal:** machine understandable

**shared:** consensus about ontology
PARENTAL ADVISORY
EXPLICIT SEMANTICS
Golden Altar ⊑ High Altar ⊓ ∃widthCM.=650

650 cm
From Ontology to Knowledge Graphs

Golden Altar

High Altar

Altar

{GoldenAltar} ⊆ HighAltar ∩ ∃ widthCM. = 650
HighAltar ⊆ Altar

isLocated(HighAltar, Apse)

isPartOf(Apse, Church)

SecularBuilding ⊆ Building

Church ⊆ Building

SecularBuilding ⊆ Building

isPartOf(Apse, Church)

Church ⊆ SecularBuilding ∩ SecularBuilding ≡ ∅
From Ontology to Knowledge Graphs

**Classes**

- **Altar**
- **HighAltar**
- **Church**

**Instances**

- **“Golden Altar”**
- **GoldenAltar**
- **ChurchOfOurLady**

Terminological Knowledge

Assertional Knowledge

Prof. Dr. Harald Sack: Ontologies and Knowledge Graphs for FAIR Research Data Management, FAIR Research Data in Plasma Medicine, 28 October 2020
From Ontology to Knowledge Graphs

Encoded via RDF:

`:GoldenAltar foaf:name “Golden Altar” .
`:GoldenAltar :locatedIn dbr:ChurchOfOurLady .

Terminological Knowledge

Classes

Altar

HighAltar

Church

Instances

“Golden Altar”

GoldenAltar

:locatedIn

:ChurchOfOurLady

Logical inference

Logical inference

Is a

Is a

Is a

Assertional Knowledge
Research Data Management
without Ontologies and Knowledge Graphs

- Research Data is locked up in small data islands
- Access only via proprietary APIs
- Without prior knowledge specific Research Data is difficult to find
- Cross connections between Data Repositories are next to impossible
- FAIR principles are only hard to implement
Research Data Management

with Ontologies and Knowledge Graphs

Data Repository

Knowledge Graph 1

Knowledge Graph 2

Knowledge Graph 3

Research Domain
FAIR Research Data Management

with Ontologies and Knowledge Graphs

Knowledge Graphs

Implement all 4 FAIR Principles

- Findability
- Accessibility
- Interoperability
- Reproducibility

for Research Data Management
But how to Begin...?
Ontology Types and Categories

according to their level of Generality

**Top-Level Ontology**
(Upper Ontology, Foundation Ontology)

- general, cross domain ontologies;
- represent very general concepts as e.g., Time, Space, Event;
- independent of a specific domain or problem

**Domain Ontology**

- fundamental concepts according to a generic domain; specializes terms introduced in top-level ontology

**Application Ontology**

- specialized ontology focussed on a specific task and domain; often a specialization of both task and domain ontology; often specify roles played by domain entities for specific activity

**Task Ontology**

- fundamental concepts according to a general activity or task; specializes terms introduced in top-level ontology

(according to Guarino: Formal Ontology in Information Systems, 1998)
Ontology Types and Categories

according to their level of Generality

**Top-Level Ontology**
*Upper Ontology, Foundation Ontology*

- general, cross domain ontologies;
- represent very general concepts as e.g., Time, Space, Event;
- independent of a specific domain or problem

**Domain Ontology**

- fundamental concepts according to a generic domain; specializes terms introduced in top-level ontology

**Application Ontology**

- specialized ontologies and domain; often specializes by domain entities

(according to Guarino: Formal Ontology in Information Systems, 1998)
Ontology Types and Categories
according to their level of Generality

Domain Ontology
fundamental concepts according to a generic domain; specializes terms introduced in top-level ontology

Top-level Ontology (Universal, Foundation)

Application Ontology

(according to Guarino: Formal Ontology in Information Systems, 1998)
Ontology Types and Categories

according to their level of Generality

- **General, cross domain ontologies:**
  - represent very general concepts as e.g., Time, Space, Event;
  - independent of a specific domain or problem

- **Task Ontology**
  - fundamental concepts according to a general activity or task;
  - specializes terms introduced in top-level ontology

(according to Guarino: Formal Ontology in Information Systems, 1998)
Ontology Types and Categories according to their level of Generality

- **Top-Level Ontology** (Upper Ontology, Foundation Ontology)
  - general, cross-domain

- **Domain Ontology**
  - fundamental concepts according to a generic domain; specializes terms introduced in top-level ontology

- **Application Ontology**
  - specialized ontology focused on a specific task and domain; often a specialization of both task and domain ontology; often specify roles played by domain entities for specific activity

(according to Guarino: Formal Ontology in Information Systems, 1998)

Prof. Dr. Harald Sack: Ontologies and Knowledge Graphs for FAIR Research Data Management, FAIR Research Data in Plasma Medicine, 28 October 2020
Ontology Types and Categories

according to their level of Semantic Expressivity

Informal

- Controller Vocabulary

Thesauri

formal IS-A

Frames

General Logical Constraints

formal

Expressivity

Glossaries

informal IS-A

formal Instance

Value Restrictions

Disjunctiveness, Inversiveness, Part-of...

Terms

Data Dictionaries

Folksonomies

formal Taxonomies

Logic Programming

Description Logics

First Order Logics

Lightweight Ontologies

Heavyweight Ontologies

(according to Guarino: Formal Ontology in Information Systems, 1998)

(according to Lassila and McGuinness: The Role of Frame-Based Representation on the Semantic Web, 2001)
“It does not do to leave a live dragon out of your calculations, if you live near him.”

J.R.R. Tolkien, *The Hobbit or There and Back again* (1937)
**Ontologies as Interpretations of Reality**

*Various categories of animals from "a certain Chinese encyclopedia" according to Jorge Luis Borges:*

- Those that belong to the emperor
- Embalmed ones
- Those that are trained
- Suckling pigs
- Mermaids (or Sirens)
- Fabulous ones
- Stray dogs
- Those that are included in this classification
- Those that tremble as if they were mad
- Innumerable ones
- Those drawn with a very fine camel hair brush
- Et cetera
- Those that have just broken the flower vase
- Those that, at a distance, resemble flies

---

*Jorge Luis Borges: The Analytical Language of John Wilkins (1942)*

[15]

Jorge Luis Borges
(1899-1986)
Follow an Approved Methodology
Follow an Approved Methodology
Follow an Approved Methodology

1. **Natural Language Description**
2. **Terminology & Glossary**
3. **Competency Questions**
4. **Concepts**
5. **Classes & Relations**
6. **Restrictions & Constraints**
7. **Instances and Statements**
8. **Formalization**

**KG Population**
## (1) (Raw) Research Data

<table>
<thead>
<tr>
<th>z/d [1]</th>
<th>Ion density (PIC-ITAP) [10^15 m^(-3)]</th>
<th>Ion density (PIC-INP) [10^15 m^(-3)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00000000e+00</td>
<td>2.1538249e-01</td>
<td>2.2127591e-01</td>
</tr>
<tr>
<td>1.00000000e-02</td>
<td>2.2320410e-01</td>
<td>2.2851489e-01</td>
</tr>
<tr>
<td>2.00000000e-02</td>
<td>2.3078706e-01</td>
<td>2.3700471e-01</td>
</tr>
<tr>
<td>3.00000000e-02</td>
<td>2.3957809e-01</td>
<td>2.4612475e-01</td>
</tr>
<tr>
<td>4.00000000e-02</td>
<td>2.4898703e-01</td>
<td>2.5569295e-01</td>
</tr>
<tr>
<td>5.00000000e-02</td>
<td>2.5889461e-01</td>
<td>2.6656408e-01</td>
</tr>
<tr>
<td>6.00000000e-02</td>
<td>2.7120663e-01</td>
<td>2.7901766e-01</td>
</tr>
<tr>
<td>7.00000000e-02</td>
<td>2.8447237e-01</td>
<td>2.9209201e-01</td>
</tr>
<tr>
<td>8.00000000e-02</td>
<td>2.9853002e-01</td>
<td>3.0861118e-01</td>
</tr>
<tr>
<td>9.00000000e-02</td>
<td>3.1697947e-01</td>
<td>3.2641678e-01</td>
</tr>
<tr>
<td>1.00000000e-01</td>
<td>3.3656863e-01</td>
<td>3.4837557e-01</td>
</tr>
<tr>
<td>1.10000000e-01</td>
<td>3.6049250e-01</td>
<td>3.7427430e-01</td>
</tr>
<tr>
<td>1.20000000e-01</td>
<td>3.8862354e-01</td>
<td>4.0343478e-01</td>
</tr>
<tr>
<td>1.30000000e-01</td>
<td>4.2297845e-01</td>
<td>4.3891770e-01</td>
</tr>
<tr>
<td>1.40000000e-01</td>
<td>4.6555629e-01</td>
<td>4.8310615e-01</td>
</tr>
<tr>
<td>1.50000000e-01</td>
<td>5.1581989e-01</td>
<td>5.3864561e-01</td>
</tr>
<tr>
<td>1.60000000e-01</td>
<td>5.7837521e-01</td>
<td>6.0616555e-01</td>
</tr>
<tr>
<td>1.70000000e-01</td>
<td>6.4984874e-01</td>
<td>6.8350098e-01</td>
</tr>
<tr>
<td>1.80000000e-01</td>
<td>7.3012722e-01</td>
<td>7.6446633e-01</td>
</tr>
<tr>
<td>1.90000000e-01</td>
<td>8.1671138e-01</td>
<td>8.5748202e-01</td>
</tr>
<tr>
<td>2.00000000e-01</td>
<td>9.0278416e-01</td>
<td>9.4737795e-01</td>
</tr>
</tbody>
</table>
Ontologies and Knowledge Graphs for Research Data Management

(1) (Raw) Research Data
(2) Schema Information

Benchmark data for fluid modelling of low-pressure CCRF discharge plasmas

Plasma Chemical Processes

The dataset contains data from comparative studies of capacitively coupled radio-frequency (CCRF) discharges in helium and argon at pressures between 10 and 80 Pa applying two different fluid modeling approaches as well as two independently developed particle-in-cell Monte Carlo collision (PIC-MCC) codes. The dataset provides a test bed for future studies of simple ccrf discharge configurations in helium and argon at pressures ranging from 10 to 80 Pa.

- plasma modelling/simulation
- benchmark data

structured information  unstructured information
Ontologies and Knowledge Graphs for Research Data Management

(1) (Raw) Research Data
(2) Schema Information
(3) Metadata

structured + unstructured information
Ontologies and Knowledge Graphs for Research Data Management

1. (Raw) Research Data
2. Schema Information
3. Metadata
4. External Resources

Properties and parameters

Definition
Plasma is a state of matter in which an ionized gaseous substance becomes highly electrically conductive to the point that long-range electric and magnetic fields dominate the behaviour of the matter. The plasma state can be contrasted with the other states: solid, liquid, and gas. Plasma is an electrically neutral medium of unbound positive and negative particles (i.e., the overall charge of a plasma is roughly zero). Although these particles are unbound, they are not "free" in the sense of not experiencing forces. Moving charged particles generate an electric current within a magnetic field, and any movement of a charged plasma particle affects is also affected by the fields created by the other changes. In turn, this governs collective behaviour with many degrees of variation. Three factors define a plasma:

1. The plasma approximation: The plasma approximation applies when the plasma parameter, $A$ [2], representing the number of charge carriers within a sphere (called the Debye sphere whose radius is the Debye screening length) surrounding a given charged particle, is sufficiently high as to shield the electrostatic influence of the particle outside of the sphere.

2. Bulk interactions: The Debye screening length (defined above) is short compared to the physical size of the plasma. This criterion means that interactions in the bulk of the plasma are more important than those at its edges, where boundary effects may take place. When this criterion is satisfied, the plasma is quasineutral.

3. Plasma frequency: The electron plasma frequency (measuring plasma oscillations of the electrons) is large compared to the electron-neutral collision frequency (measuring frequency of collisions between electrons and neutral particles). When this condition is valid, electrostatic interactions dominate over the processes of ordinary gas kinetics.

unstructured information

structured information

semantic information

continuum mechanics

fluid mechanics

laminar flow

turbulent flow

statics

dynamics

archimedes’ principle

bernoulli’s principle

navier-stokes equations

poiseuille equation

pascal’s law

viscosity

(newtonian • non-newtonian)

buoyancy

mixing • pressure

liquids

surface tension • capillary action

gases

atmosphere • boyle’s law • charles’s law • gay-lussac’s law • combined gas law

Plasma

Rheology

Scientists

continuum mechanics

laws

show

Solid mechanics

show

fluid mechanics

hide

fluids

statics • dynamics

archimedes’ principle • bernoulli’s principle

navier-stokes equations

poiseuille equation • pascal’s law

viscosity

(newtonian • non-newtonian)

buoyancy • mixing • pressure

liquids

surface tension • capillary action

gases

atmosphere • boyle’s law • charles’s law • gay-lussac’s law • combined gas law

plasma

rheology

scientists

v • t • e

FIZ Karlsruhe

Leibniz-Institut für Informationsinfrastruktur
Ontologies and Knowledge Graphs for Research Data Management

1. (Raw) Research Data
2. Schema Information
3. Metadata
4. External Resources

unstructured information → knowledge extraction → semantification

existing ontologies → knowledge graph
new ontologies
existing knowledge graphs
other information resources
“Though this be madness, yet there is method in it”

William Shakespeare, Hamlet (1602)
The Semantic Web Technology Stack
(not a piece of cake...)

- Most apps use only a subset of the stack
- Querying allows fine-grained data access
- Standardized information exchange is key
- Formats are necessary, but not too important
- The Semantic Web is based on the Web
- Linked Data uses a small selection of technologies

The Semantic Web - Not just a piece of cake, [16]
Linked Data and Knowledge Graphs
Semantic Search & Retrieval
Ontology & Knowledge Graph Applications

User Interaction → Document Store → Knowledge Graph → Ranking → Evaluation → User Interaction

Text/Data Acquisition → Knowledge Graph → Index Creation → Index → Retrieval Process

Indexing Process

Messy Store Room with Boxes, [17]
Semantic Search & Retrieval

Ontology & Knowledge Graph Applications

Ontology
(Classes & Relations)

Knowledge Graph
(Instances)

Document Text
(Labels)

An Evolving Knowledge Graph

For Plasma Technology

Electrolytic Plasma Technology (EPT) (de:Elektrolytischer Plasmatronik)
“Technology presumes there's just one right way to do things and there never is”
