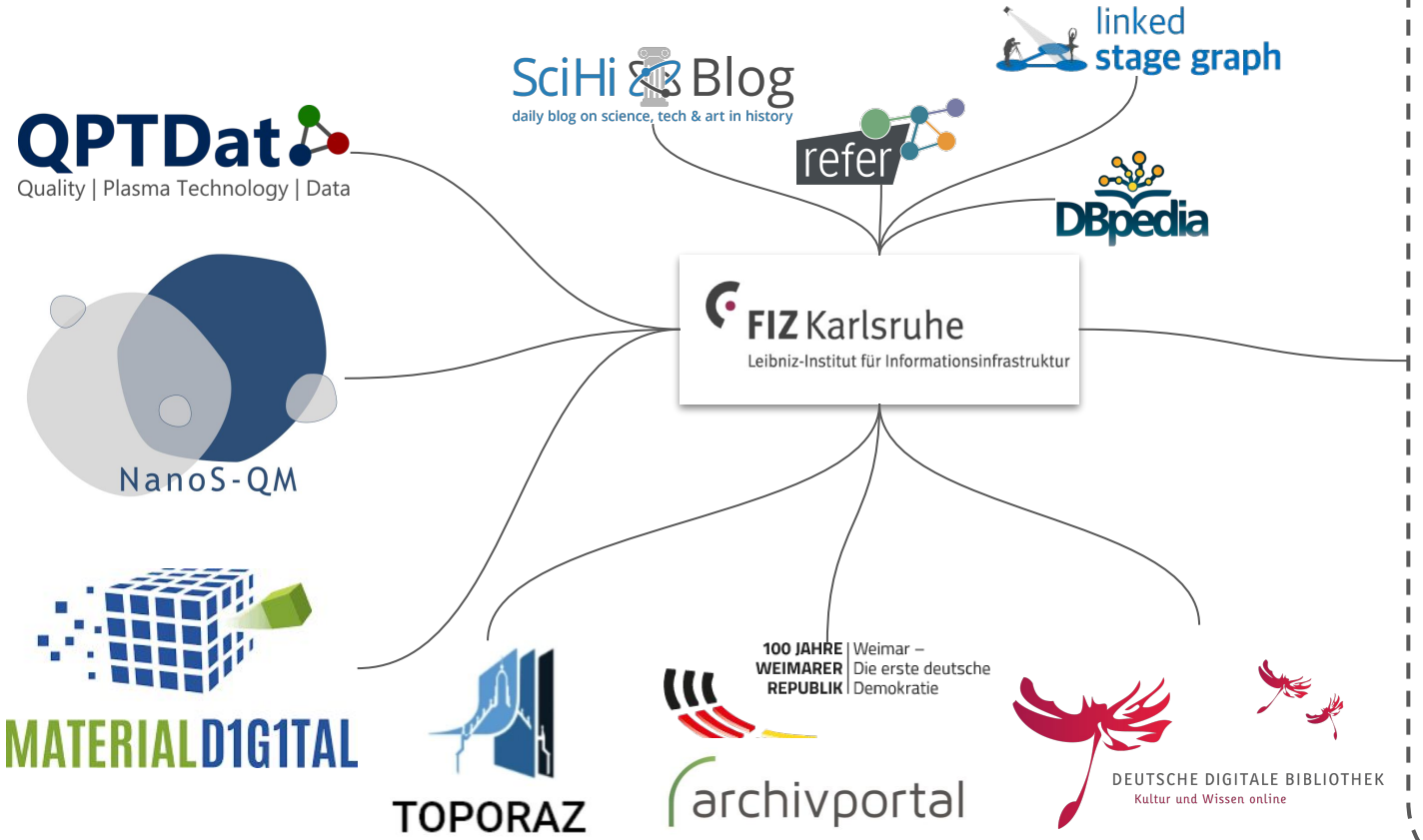
FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

# FAIR Research Data Management with Knowledge Graphs and Ontologies

Prof. Dr. Harald Sack
InnoMatSafety
25.06.2021

# Knowledge Graphs @ FIZ Karlsruhe

## Projects, Services & NFDIs

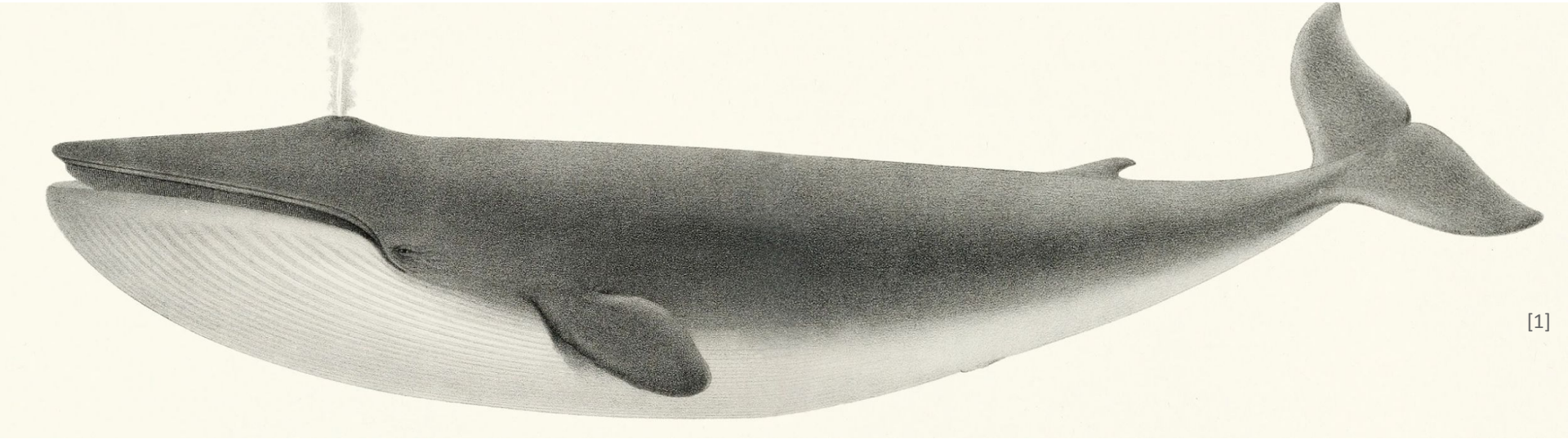


# 33.6

# 33.6 m


33.6m

33.6m [1] (1922)

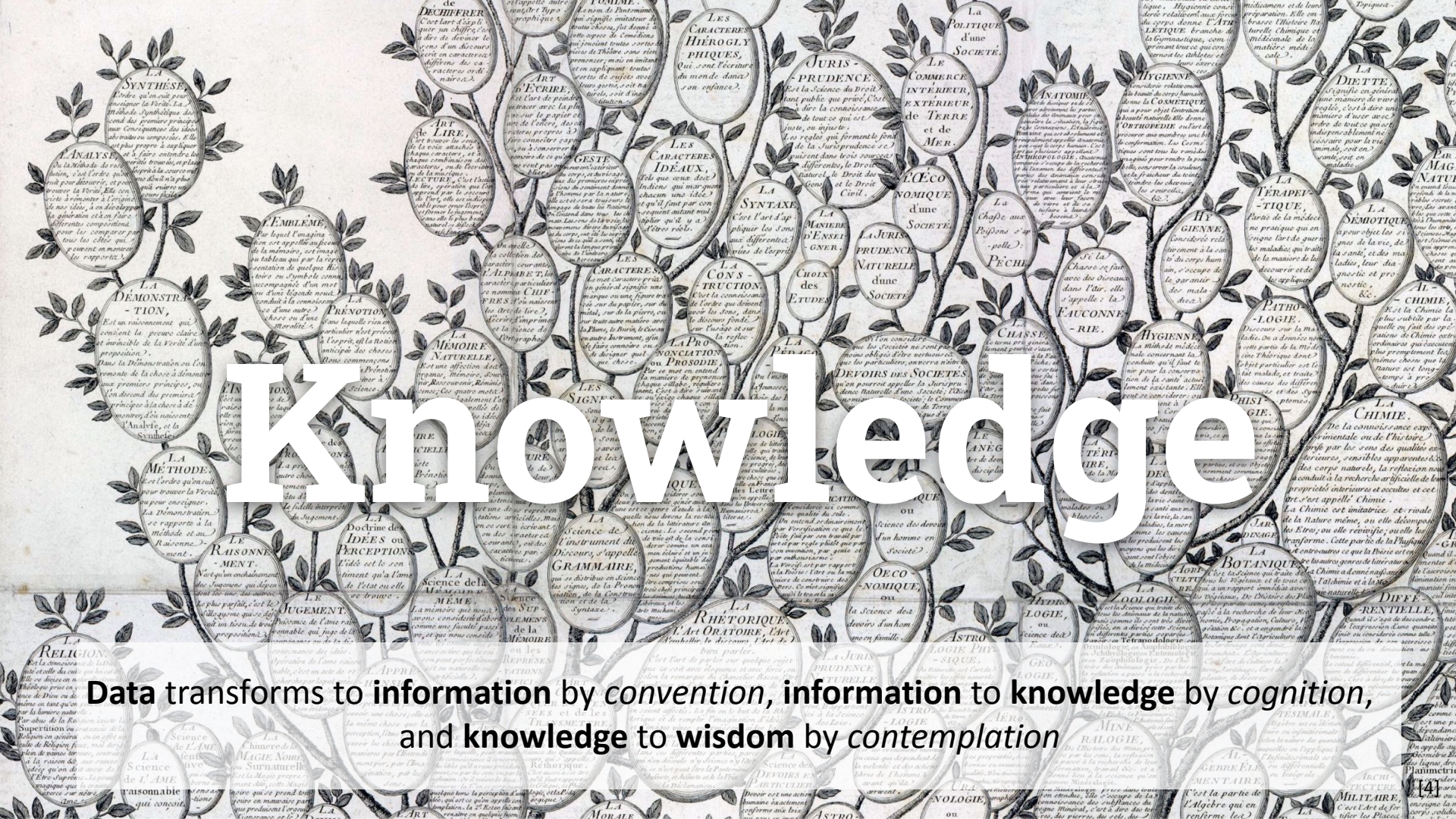


[1]

[1] S. G. Brown: *Balaenoptera musculus (Linnaeus 1758) – Blauwal*, in Jochen Niethammer, Franz Krapp (Hrsg.): Handbuch der Säugetiere Europas. Band 6: Meeressäuger, Teil I Wale und Delphine – Cetacea, Teil IB: Ziphidae, Kogiidae, Physeteridae, Balaenidae, Balaenopteridae. Aula-Verlag Wiesbaden 1995

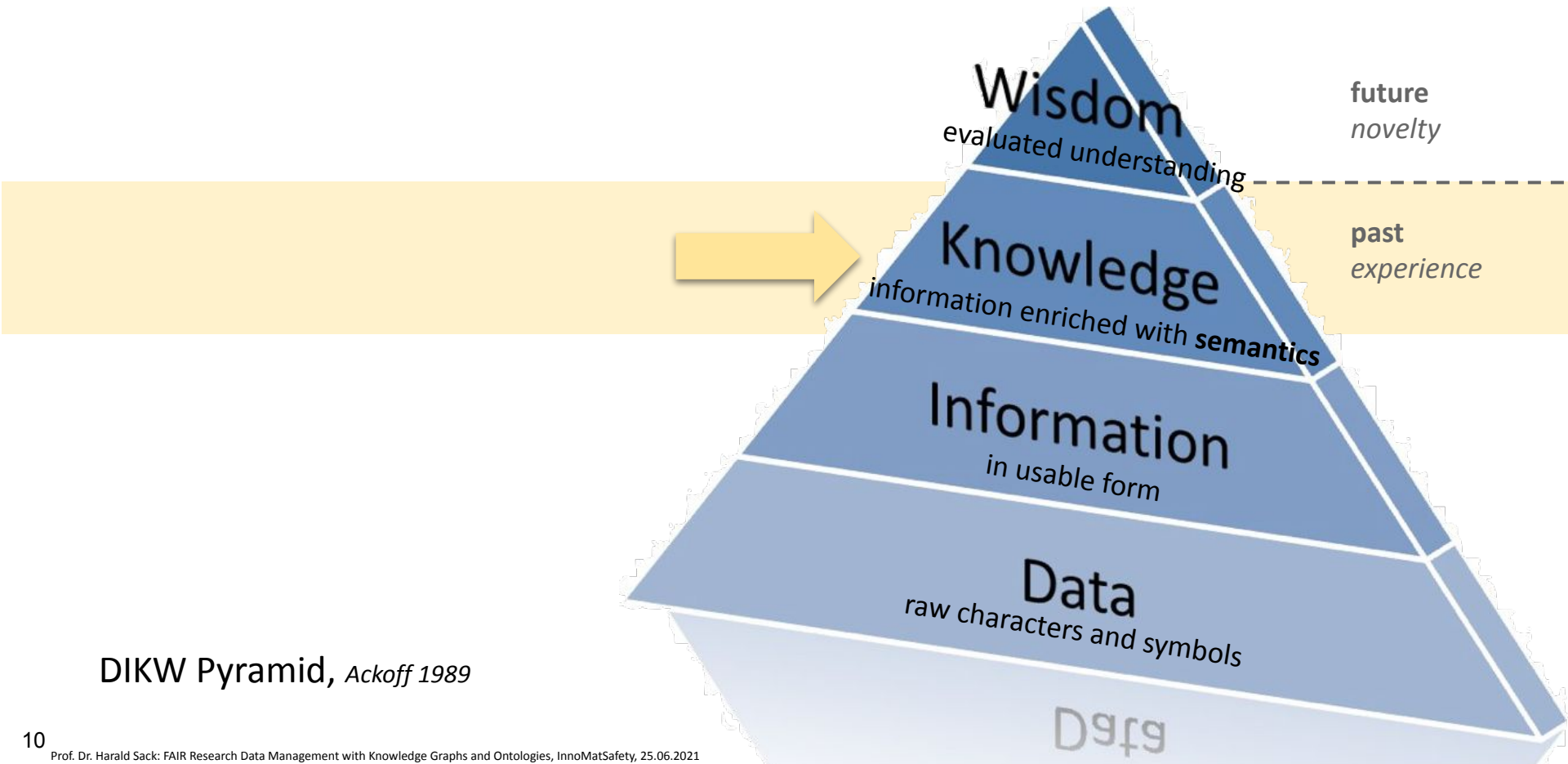

[2]


[3]



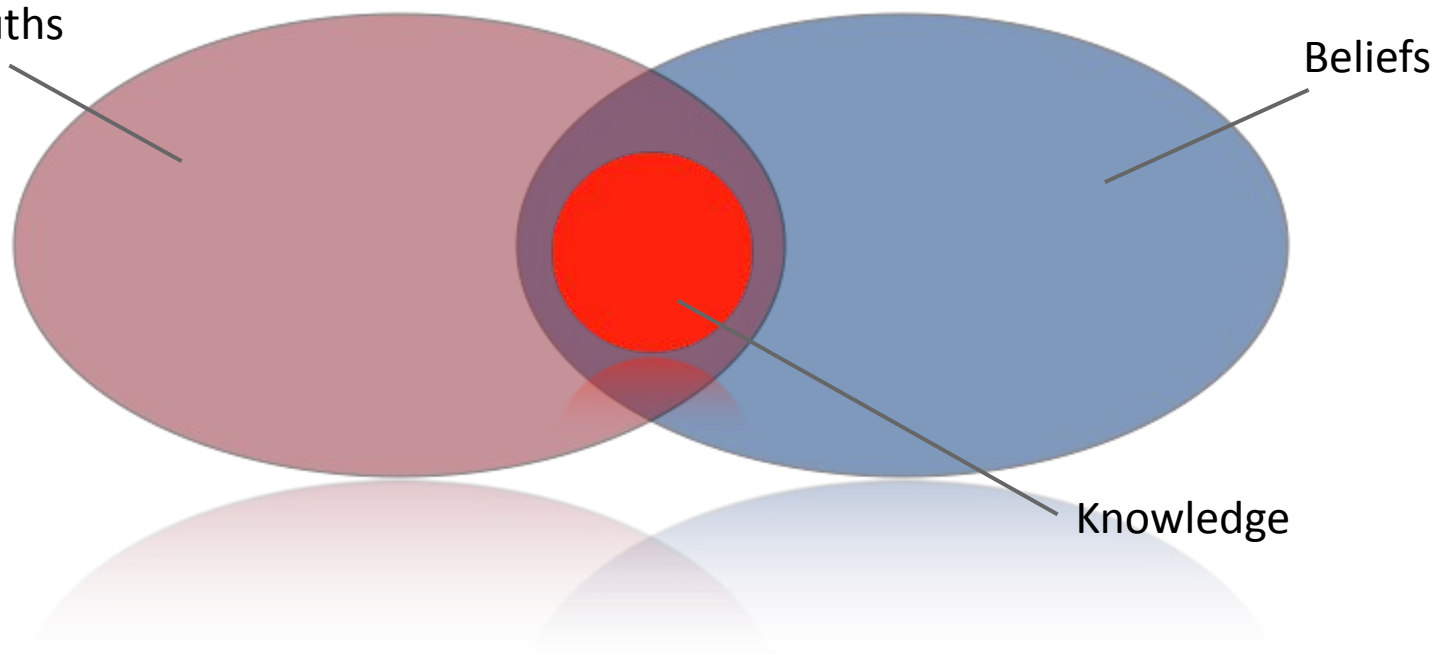

**Data** transforms to **information** by *convention*, **information** to **knowledge** by *cognition*, and **knowledge** to **wisdom** by *contemplation*

# From Data to Knowledge




DIKW Pyramid, *Ackoff 1989*

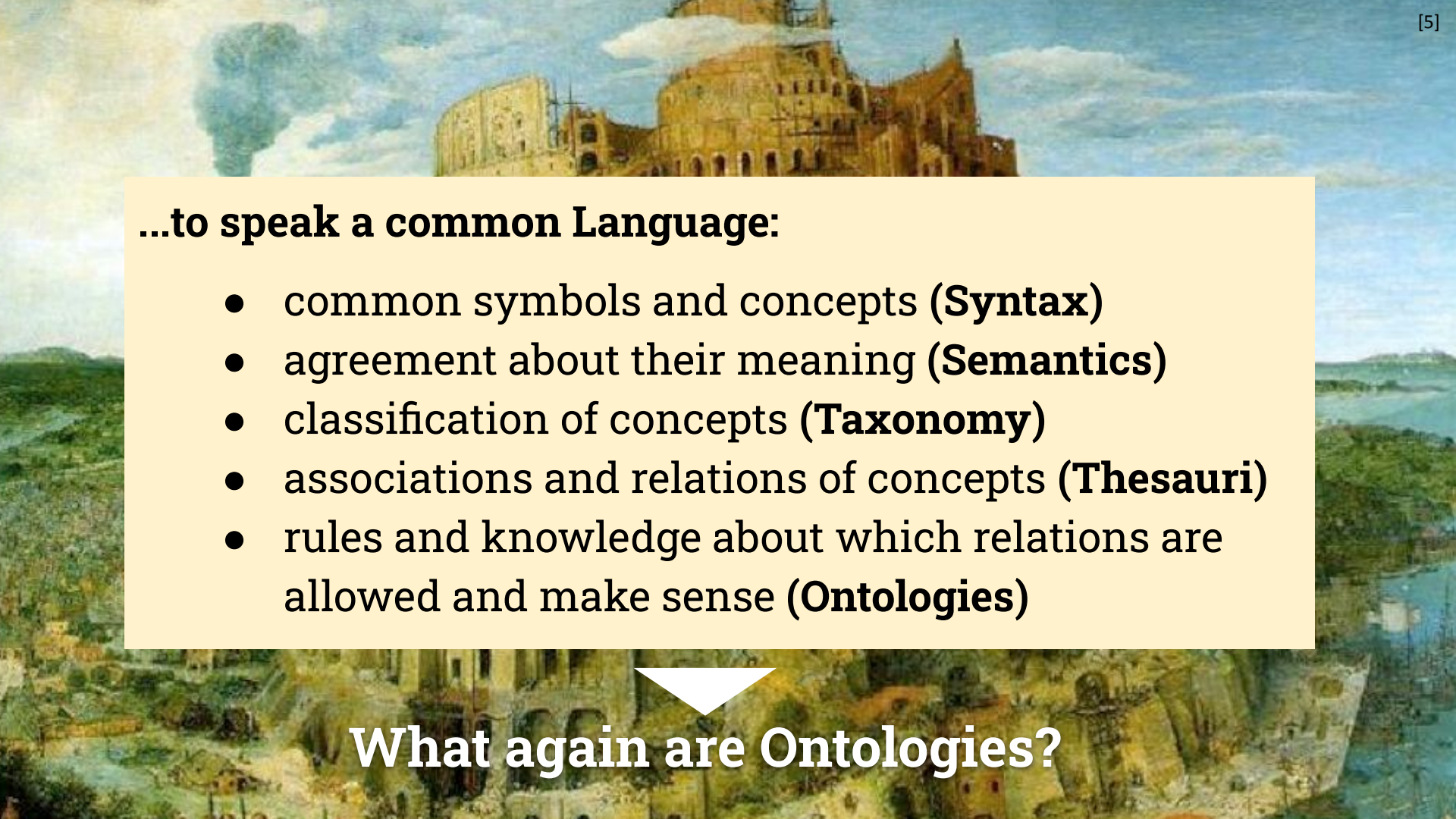
# What is Knowledge?




Traditional Definition: „Knowledge is a subset of all true beliefs“

„People can‘t share knowledge if they don‘t speak a common language“

*Thomas Davenport (1997)*

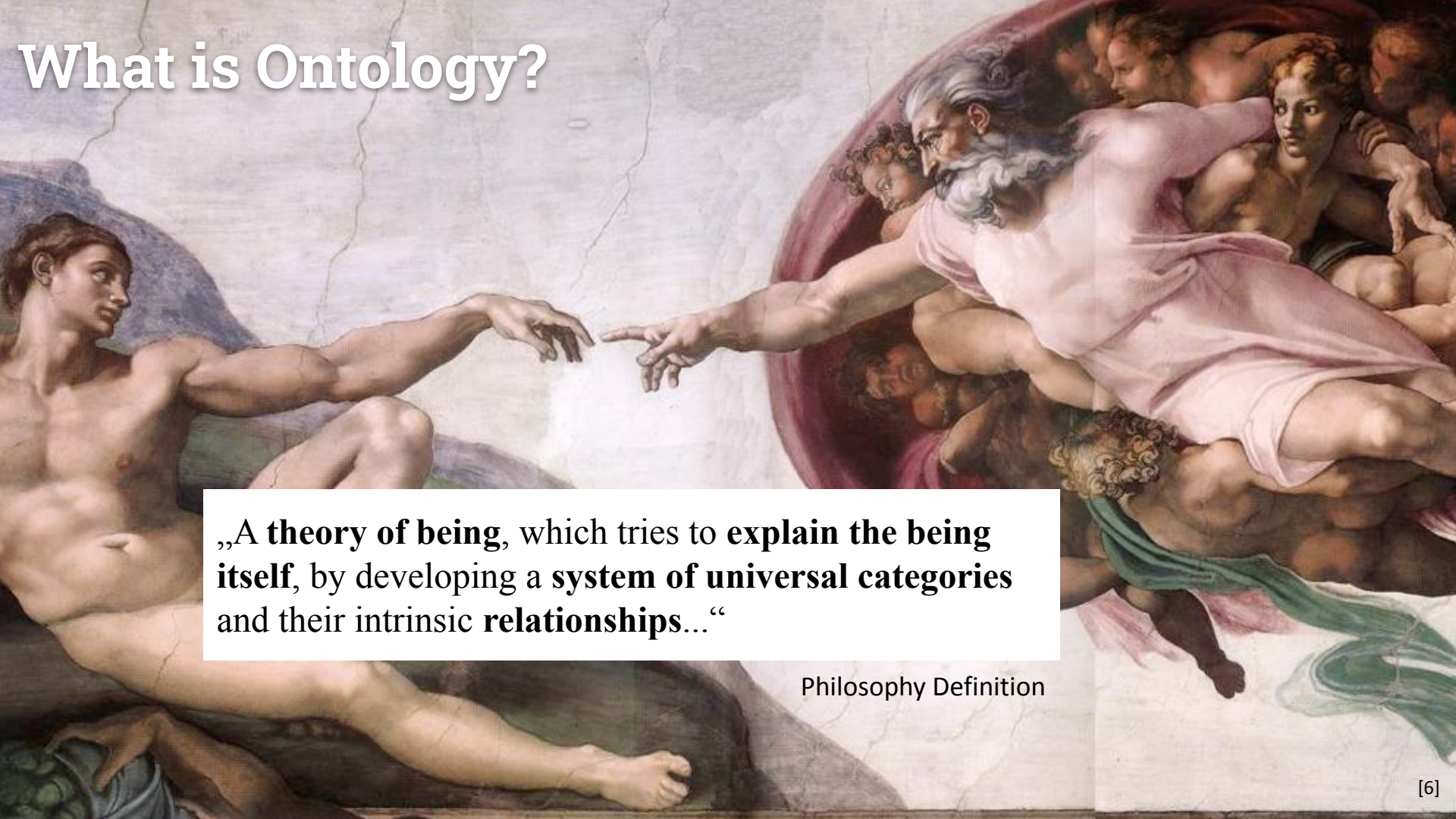


**...to speak a common Language:**

- 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)**

## What again are Ontologies?

# What is Ontology?

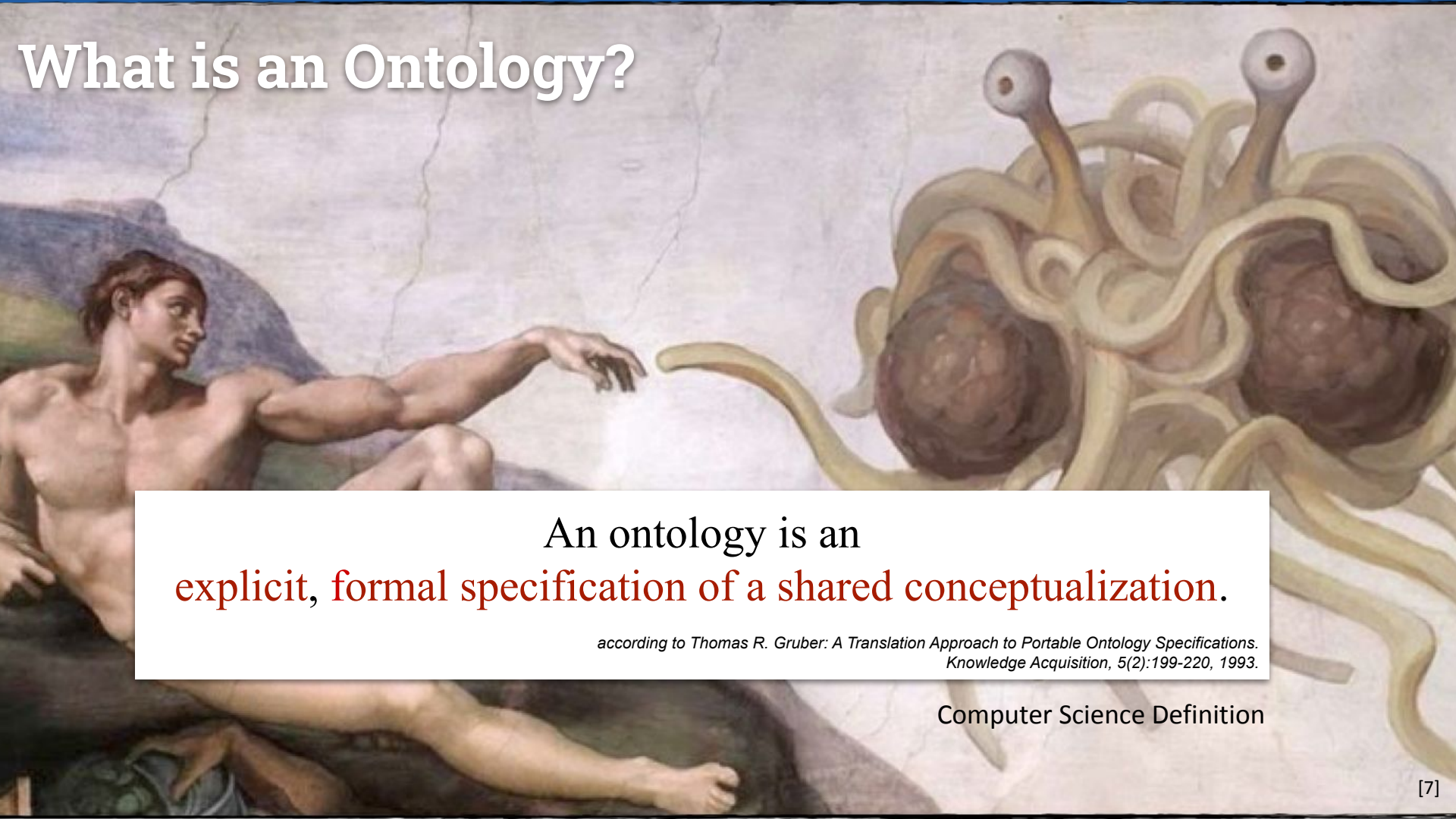


„A **theory of being**, which tries to **explain the being itself**, by developing a **system of universal categories** and their intrinsic **relationships**...“

Philosophy Definition

[6]

# What is an Ontology?



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

*according to Thomas R. Gruber: A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2):199-220, 1993.*

Computer Science Definition

# What is an Ontology?

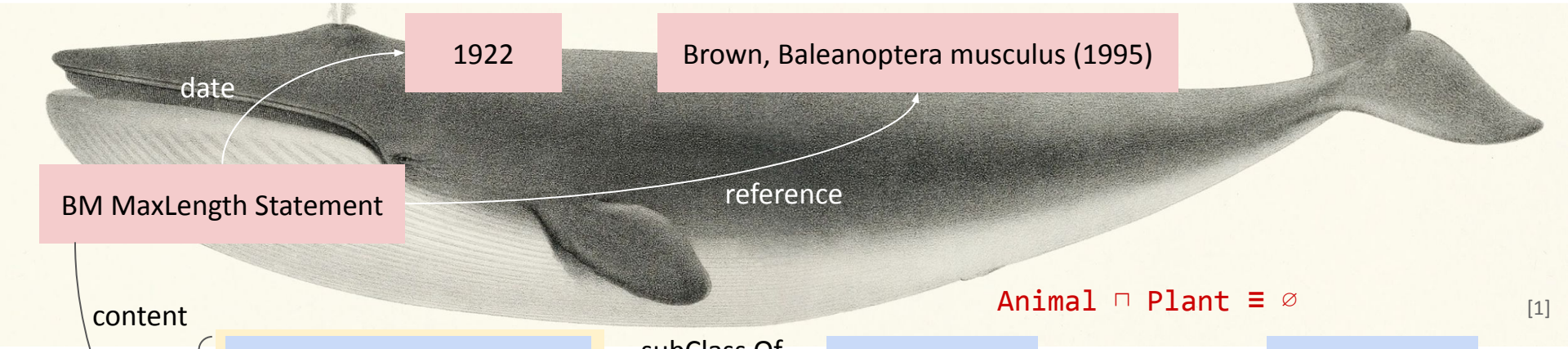
An ontology is an
explicit, formal specification of a shared conceptualization.

*according to Thomas R. Gruber: A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2):199-220, 1993.*

| | |
|---|---|
| **conceptualization**: | abstract model<br>(domain, identified relevant concepts, relations) |
| **explicit**: | meaning of all concepts must be defined |
| **formal**: | machine understandable |
| **shared**: | consensus about ontology |


[8]

# (Mini) Example Ontology + Knowledge Graph




$Animal \sqcap Plant \equiv \varnothing$ [1]

$Animal \sqsubseteq Thing$

$Plant \sqsubseteq Thing$

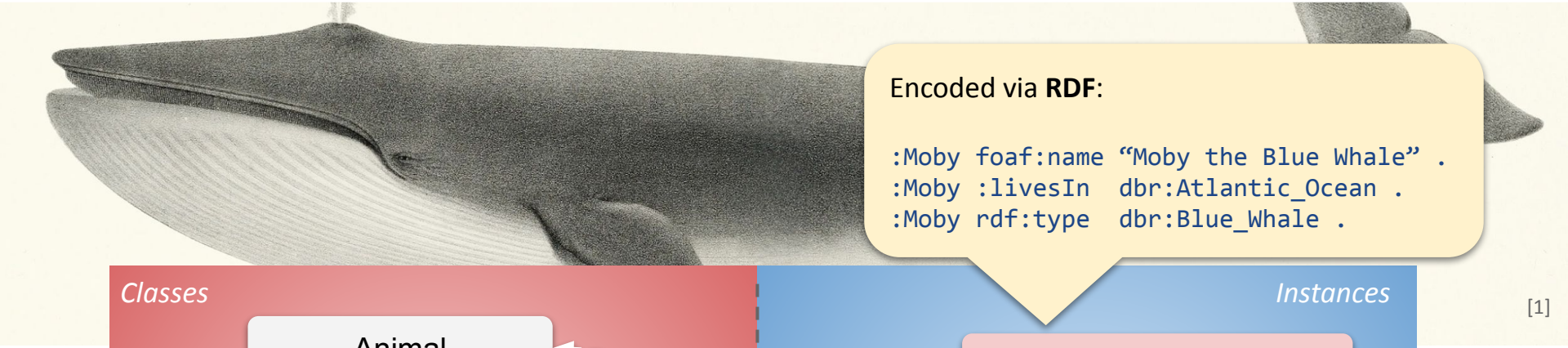
$BaleanopteraMusculus \sqsubseteq Animal \sqcap \forall maxLengthM.\leq 33.6$

Class Class Property Constraint

# (Mini) Example Ontology + Knowledge Graph

## Logical Inference and Encoding

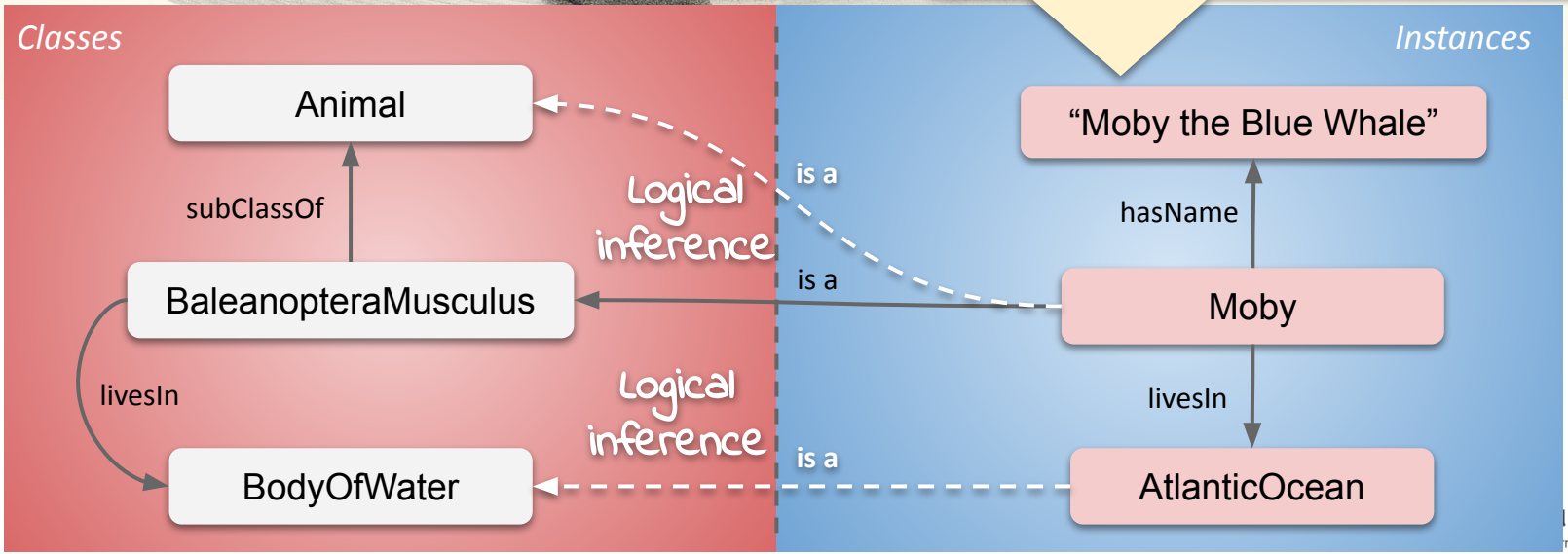
FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur



Encoded via **RDF**:

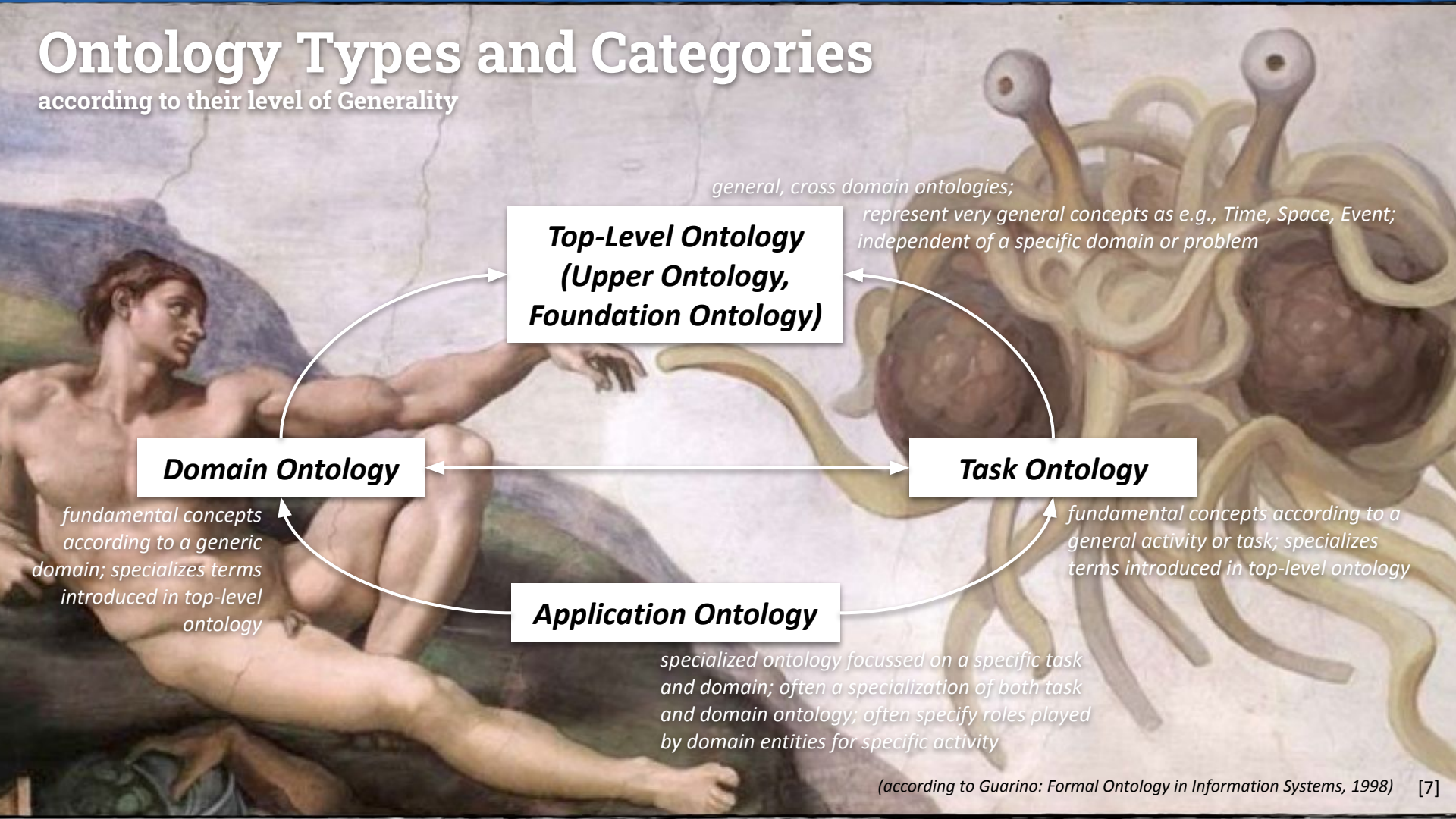
```
:Moby foaf:name “Moby the Blue Whale” .
:Moby :livesIn  dbr:Atlantic_Ocean .
:Moby rdf:type  dbr:Blue_Whale .
```

[1]



# Ontology Types and Categories

**according to their level of Generality**

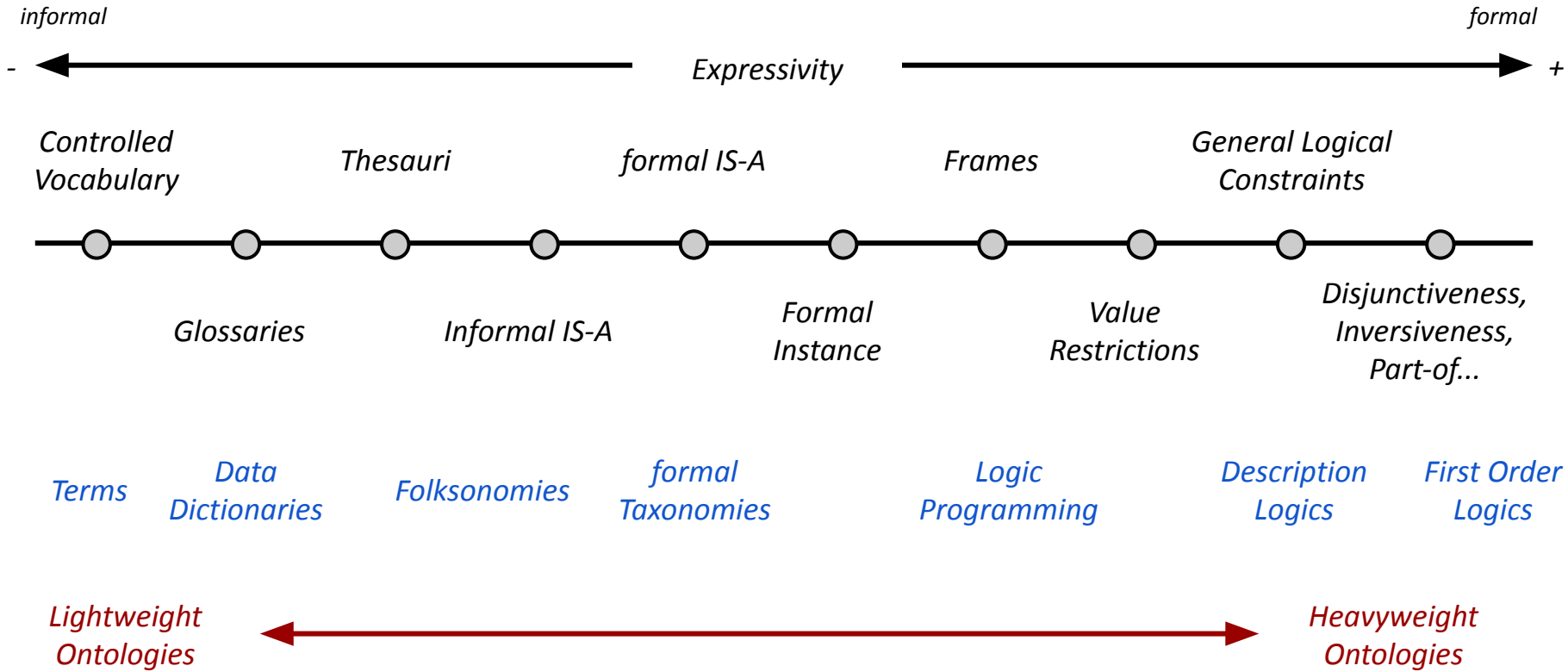



*(according to Guarino: Formal Ontology in Information Systems, 1998)*

# Ontology Types and Categories

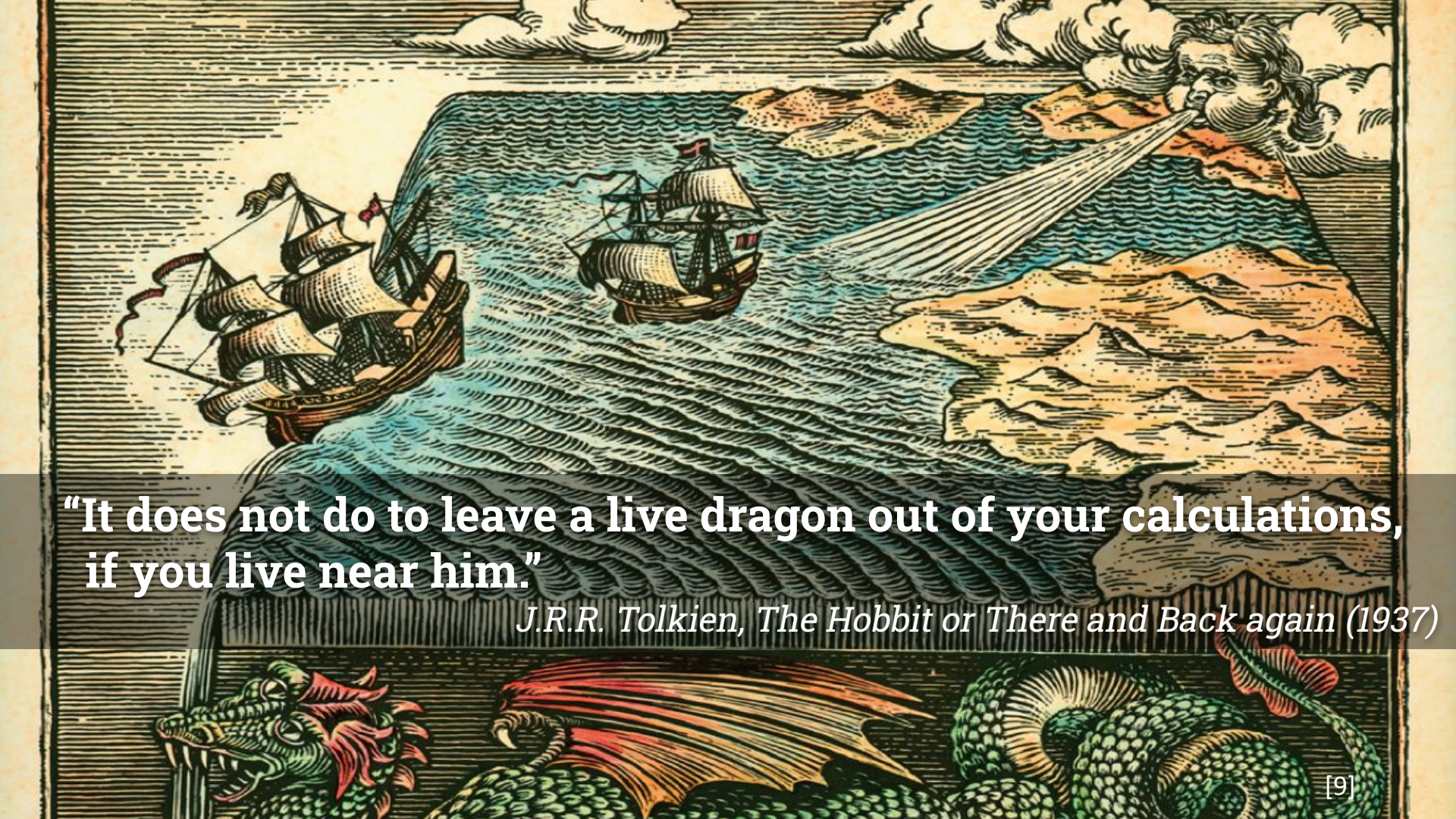
**according to their level of Semantic Expressivity**

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur




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

*(according to Lassila and McGuiness:The Role of Frame-Based Representation on the Semantic Web, 2001*



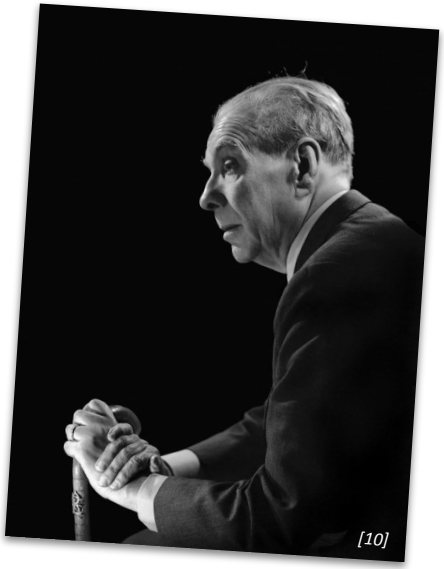
**“It does not do to leave a live dragon out of your calculations, if you live near him.”**

*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:

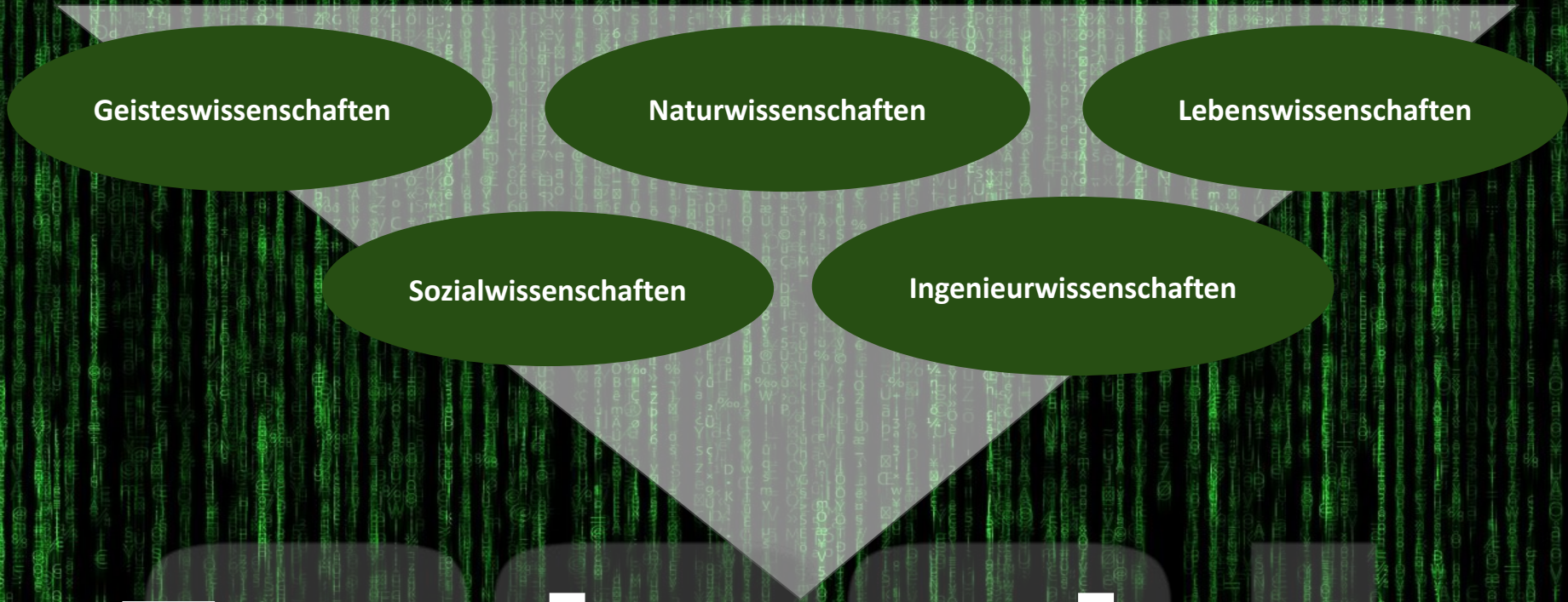
- Those that belong to the emperor
- Embalmed ones
- Those that are trained
- Suckling pigs
- Mermaids (or Sirens)
- Fabulous ones
- Stray dogs
- Those that are included in this classification
- Those that tremble as if they were mad
- Innumerable ones
- Those drawn with a very fine camel hair brush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




Jorge Luis Borges
(1899-1986)

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

# What is the Purpose?
# What will be the Application(s)?




# Forschungsdaten


# Forschungs-datensilos

[12]

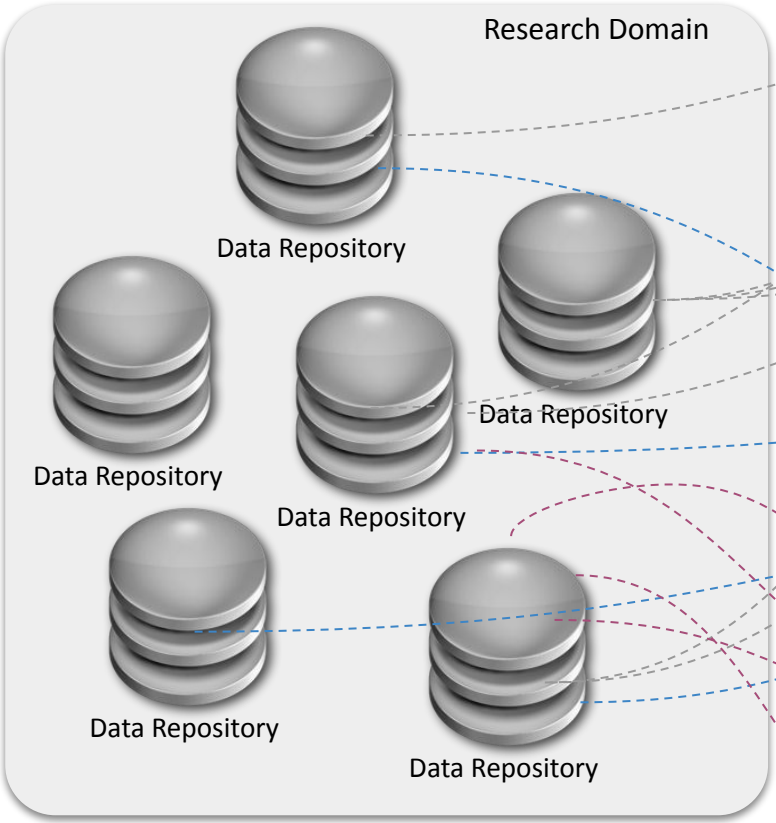
- Forschungsdaten abgeschlossen in **lokalen Datensilos**
- Zugriff nur via **proprietäre APIs**
- Ohne **spezielles Vorwissen** können Forschungsdaten kaum gefunden werden
- **Quervernetzung** zwischen Datenrepositorien **nahezu unmöglich**
- **FAIR Prinzipien** sind oft nur **unzureichend** umsetzbar

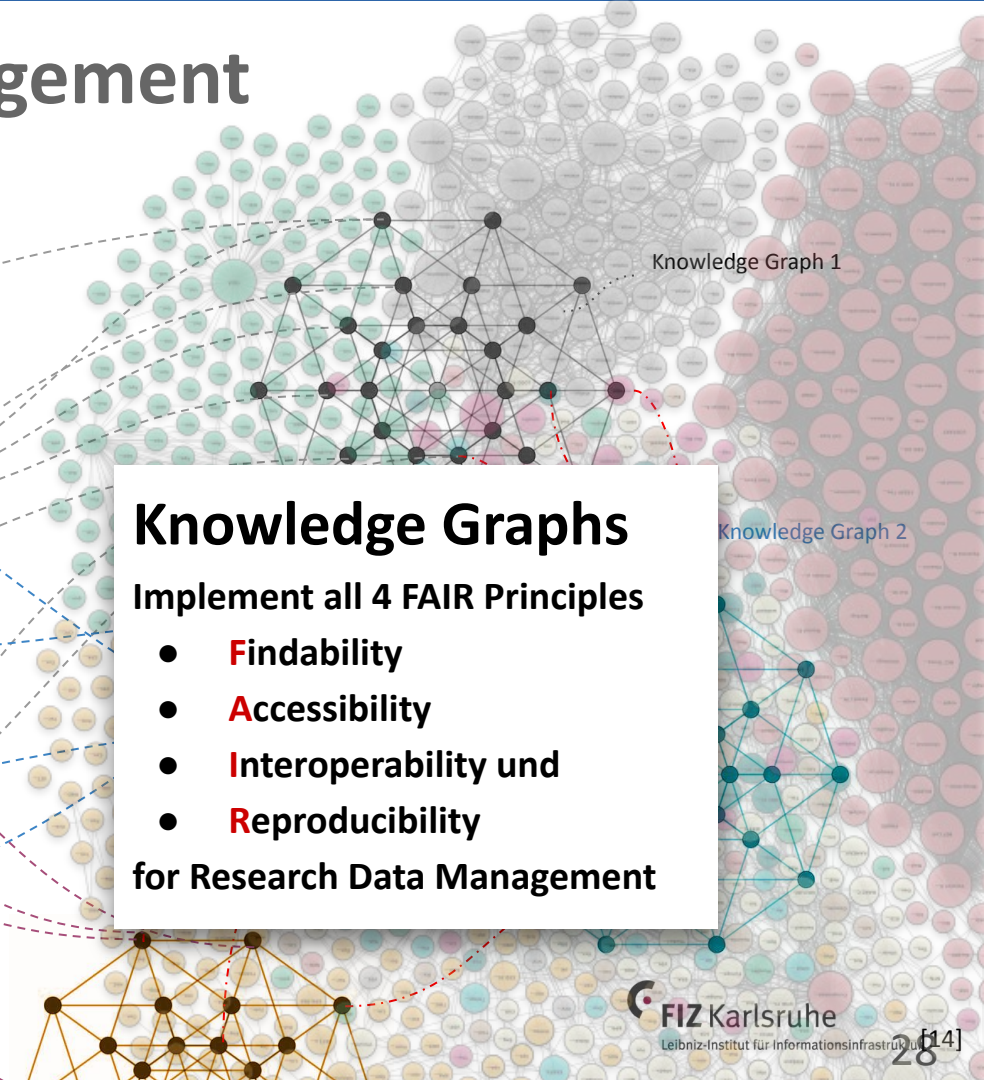
# vernetzte Forschungsdaten

https://www.publicdomainpictures.net/pictures/380000/velka/globale-digitalisierung-201105.jpg

# FAIR Research Data Management

## with Ontologies and Knowledge Graphs




## Knowledge Graphs

**Implement all 4 FAIR Principles**

- **F**indability
- **A**ccessibility
- **I**nteroperability und
- **R**eproducibility

**for Research Data Management**

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

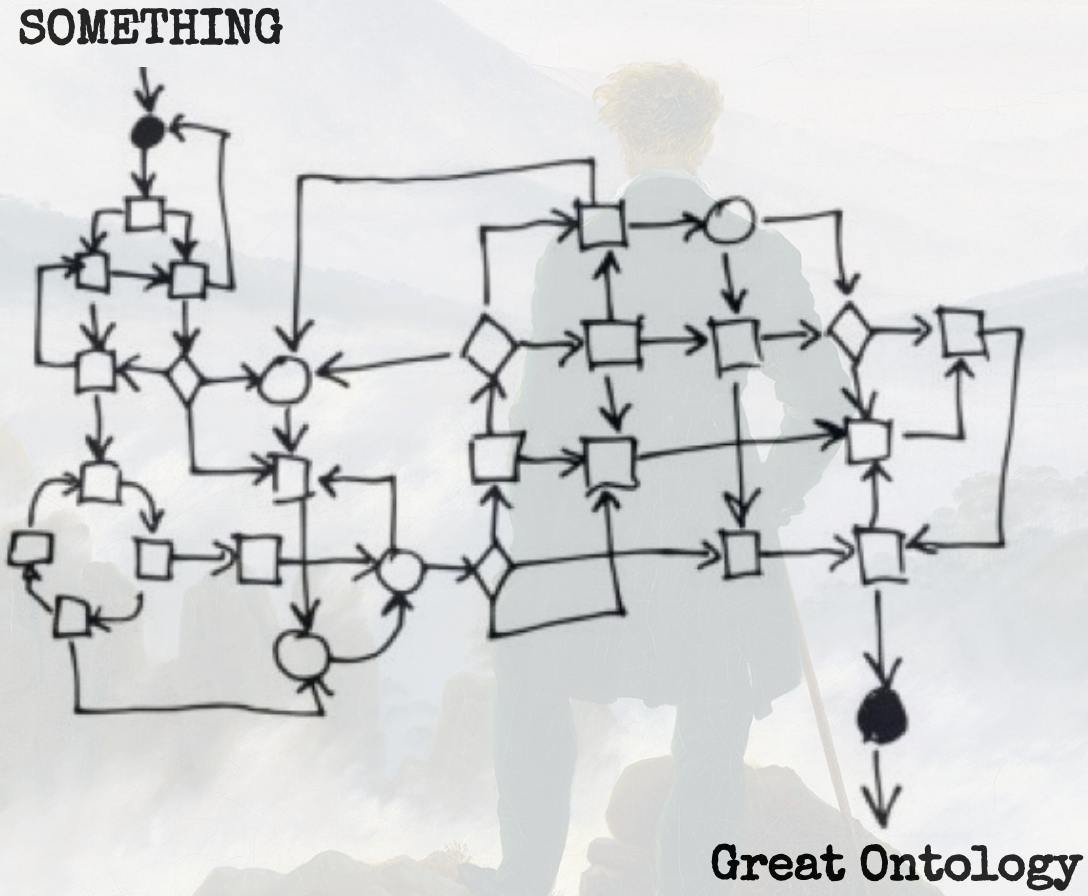
# Vernetzte Forschungsdaten

- **Standardisierung** von Metadaten
- Nutzung von **Normdaten**
- **Ontologien** und semantische Technologien
- **Wissensgraphen**
- Umsetzung der **FAIR Prinzipien**
  - **F**indability
  - **A**ccessibility
  - **I**nteroperability
  - **R**eusable

# But how to get there…?

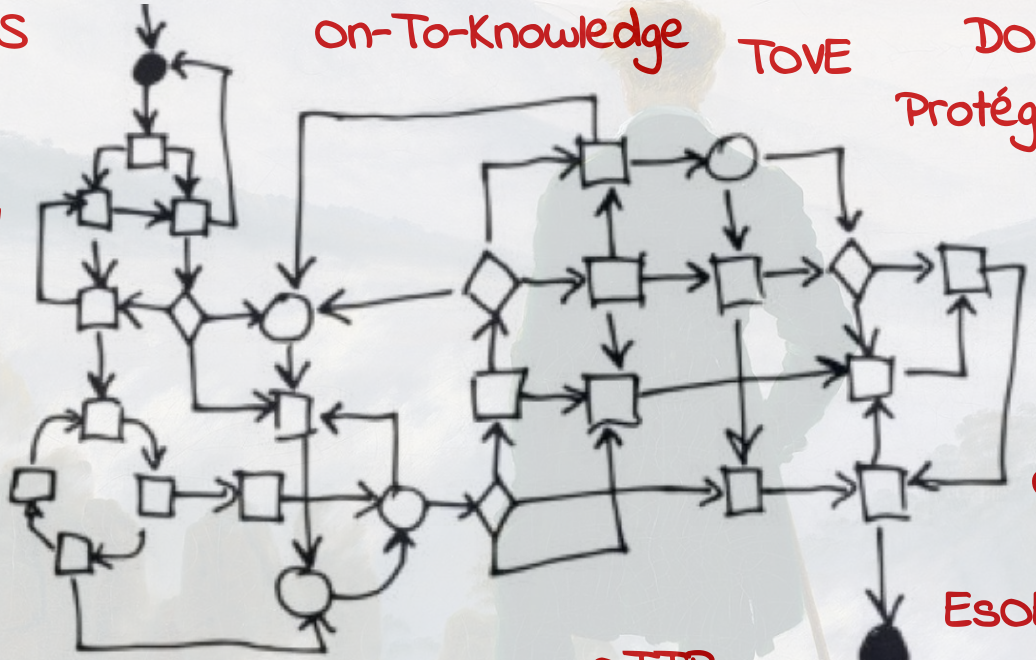
*Caspar David Friedrich, Wanderer über dem Nebelmeer, 1818 [10]*

# Follow an Approved Methodology




[16]

# Follow an Approved Methodology




Neon Toolkit
IDEF5
coefficientMakna
ONIONS
On-To-Knowledge
TOVE
DOGMA
SABio
AMOD
Protégé
myontology
Ontology 101
Ontowiki
Ikewiki
CoMeto
SCIM
Ontology Design Patterns
EsoE
IIAM
OTTR
Rational Unified Process
Methontology
UPON (Lite)

[17]



FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

# Ontologies and Knowledge Graphs for Research Data Management

## (1) (Raw) Research Data

| z/d [1] | Ion density (PIC-ITAP) [10^15 m^(-3)] | Ion density (PIC-INP) [10^15 m^(-3)] |
|---|---|---|
| 0.0000000e+00 | 2.1538249e-01 | 2.2127591e-01 |
| 1.0000000e-02 | 2.2320410e-01 | 2.2851489e-01 |
| 2.0000000e-02 | 2.3078706e-01 | 2.3700471e-01 |
| 3.0000000e-02 | 2.3957809e-01 | 2.4612475e-01 |
| 4.0000000e-02 | 2.4898703e-01 | 2.5569295e-01 |
| 5.0000000e-02 | 2.5889461e-01 | 2.6656408e-01 |
| 6.0000000e-02 | 2.7120663e-01 | 2.7901766e-01 |
| 7.0000000e-02 | 2.8447237e-01 | 2.9209201e-01 |
| 8.0000000e-02 | 2.9853002e-01 | 3.0861118e-01 |
| 9.0000000e-02 | 3.1697947e-01 | 3.2641678e-01 |
| 1.0000000e-01 | 3.3656863e-01 | 3.4837557e-01 |
| 1.1000000e-01 | 3.6049250e-01 | 3.7427430e-01 |
| 1.2000000e-01 | 3.8862354e-01 | 4.0343478e-01 |
| 1.3000000e-01 | 4.2297845e-01 | 4.3891770e-01 |
| 1.4000000e-01 | 4.6555629e-01 | 4.8310615e-01 |
| 1.5000000e-01 | 5.1581989e-01 | 5.3864561e-01 |
| 1.6000000e-01 | 5.7837521e-01 | 6.0616555e-01 |
| 1.7000000e-01 | 6.4984874e-01 | 6.8350098e-01 |
| 1.8000000e-01 | 7.3012722e-01 | 7.6446633e-01 |
| 1.9000000e-01 | 8.1671138e-01 | 8.5748202e-01 |
| 2.0000000e-01 | 9.0375181e-01 | 9.4726775e-01 |

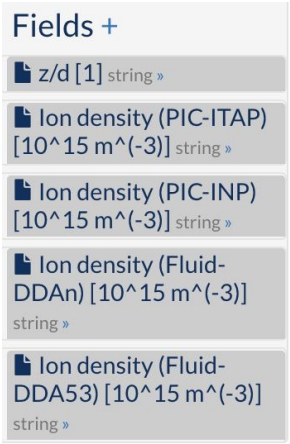
QPTDat
Quality | Plasma Technology | Data

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

# Ontologies and Knowledge Graphs for Research Data Management

(1) (Raw) Research Data

**(2) Schema Information**




**structured 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

**unstructured information**

QPTDat
Quality | Plasma Technology | Data

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

# Ontologies and Knowledge Graphs for Research Data Management

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

**structured**
**+**
**unstructured**
**Information**

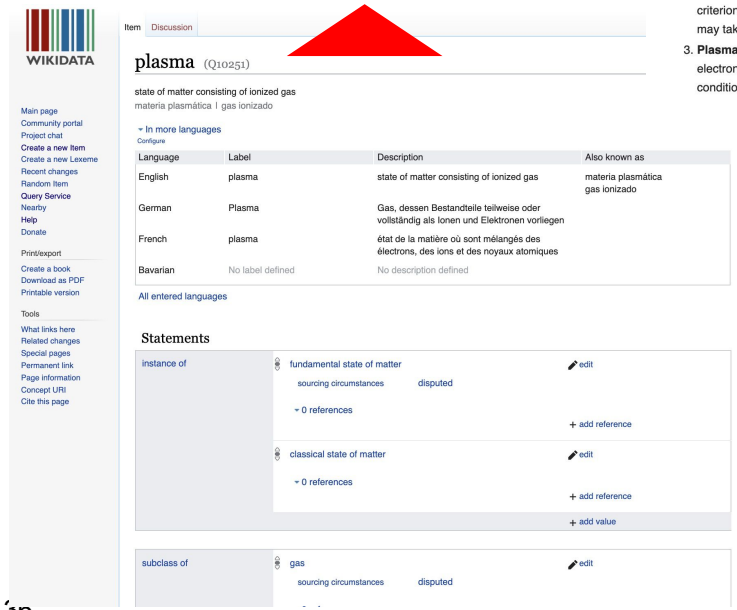
| Field | Value |
|---|---|
| Group | Plasma Modelling |
| Authors | Becker, Markus M.<br>Kählert, Hanno<br>Sun, Anbang<br>Loffhagen, Detlef |
| Release Date | 2019-06-14 |
| Resources | Benchmark data for CCRF discharge plasmas - time averaged ion density (argon, 20 Pa)<br>Benchmark data for CCRF discharge plasmas - time averaged ion density (argon, 40 Pa)<br>Benchmark data for CCRF discharge plasmas - time averaged ion density (argon, 80 Pa)<br>Benchmark data for CCRF discharge plasmas - time averaged ion density (helium, 10 Pa)<br>Benchmark data for CCRF discharge plasmas - time averaged ion density (helium, 20 Pa)<br>Show more |
| Identifier | 60dbcdd4-8be4-4f41-896c-e725bdb37fe2 |
| Permanent Identifier (DOI) | doi:10.34711/inptdat.72 |
| Permanent Identifier (URI) | https://www.inptdat.de/node/72 |
| Is supplementing | M. M. Becker et al., Plasma Sources Sci. Technol. 26 (2017) 044001 |
| Plasma Source Name | CCP |
| Plasma Source Specification | AC<br>high frequency<br>low pressure<br>non-thermal |
| Plasma Source Properties | Low-pressure RF plasma between plane electrodes separated by the distance d, driven by a sinusoidal voltage with amplitude V0 and frequency f; d = 2.5 cm (argon) resp. 6.7 cm (helium); V0 = 50-250 V; f = 13.56 MHz; Current density: 10 A/m^2 |

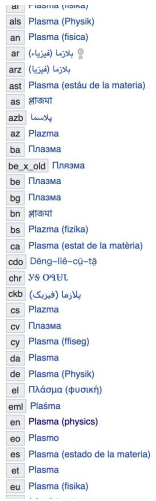
Leibniz-Institut für Informationsinfrastruktur

# Ontologies and Knowledge Graphs for Research Data Management

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

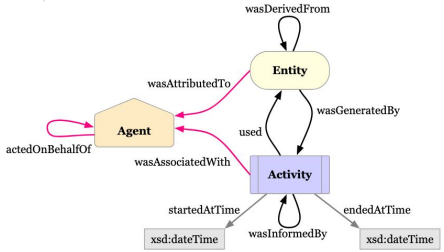
**semantic information**

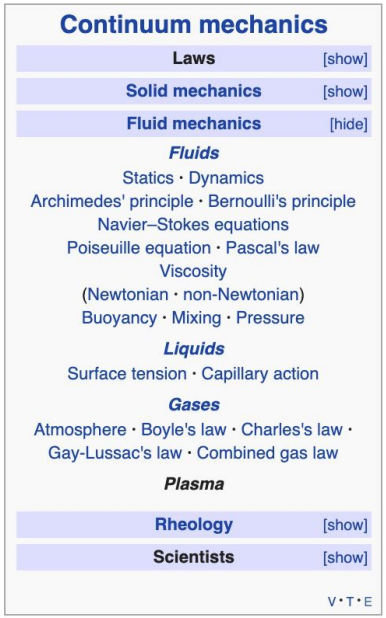



**unstructured information**

**semantic information**



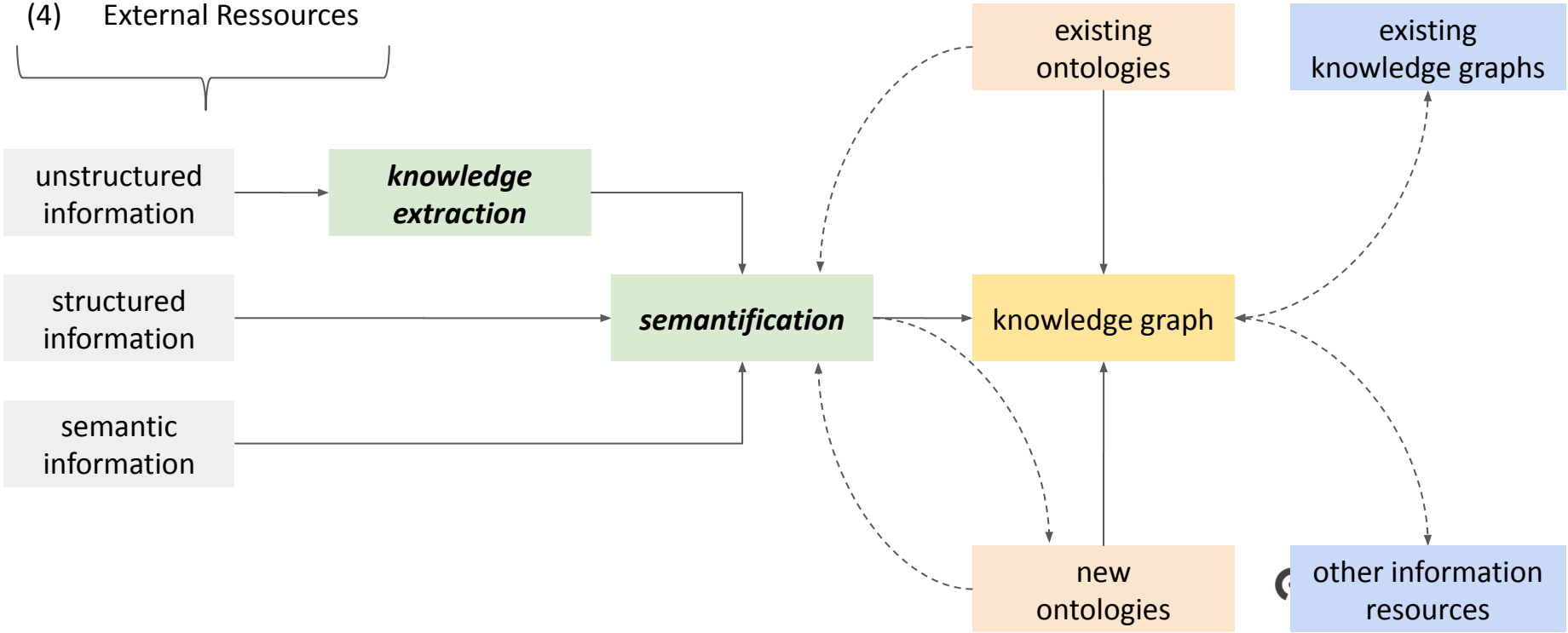

**structured information**

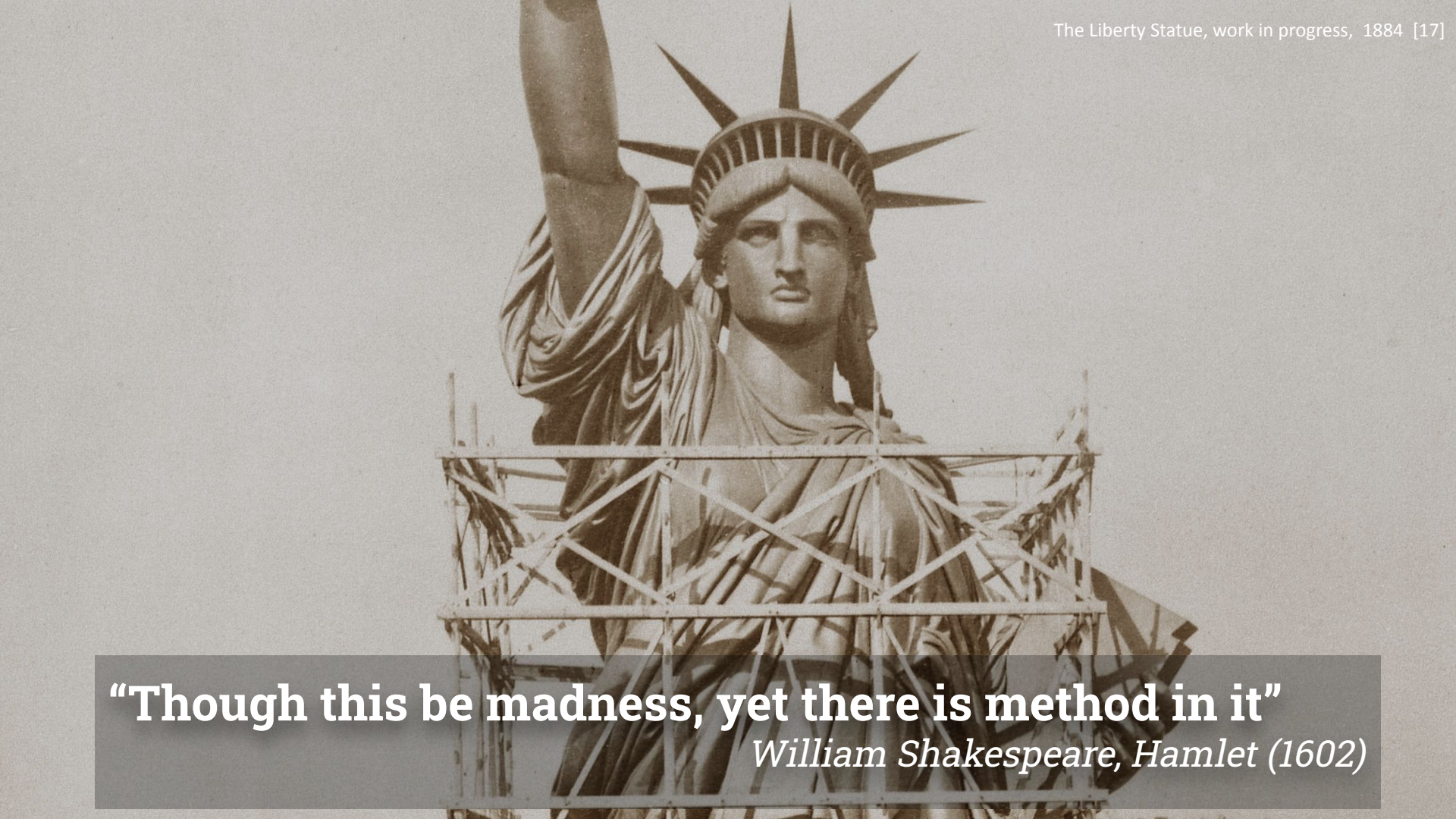
QPTDat
Quality | Plasma Technology | Data

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 Ressources



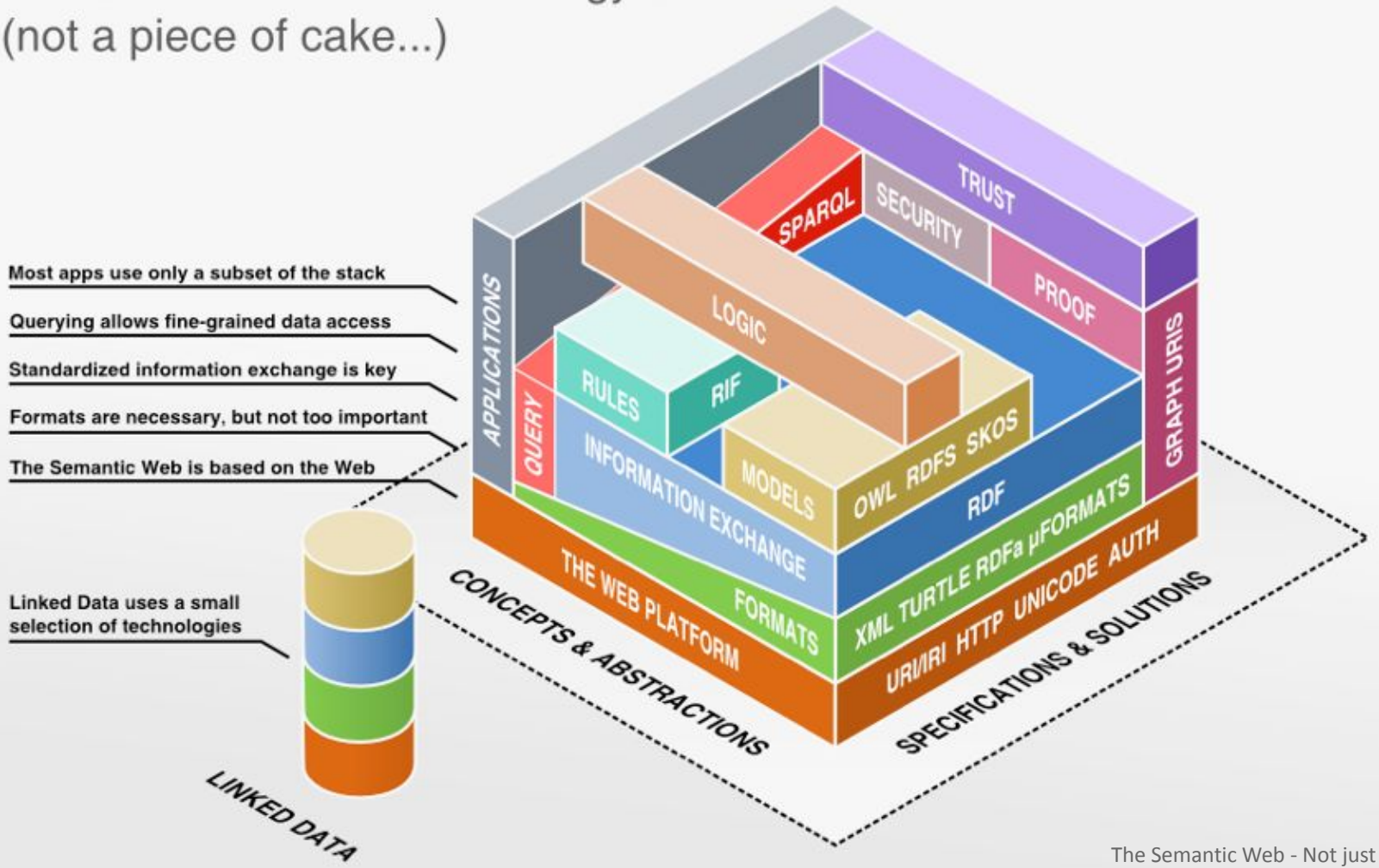


The Liberty Statue, work in progress, 1884 [17]

**“Though this be madness, yet there is method in it”**

*William Shakespeare, Hamlet (1602)*

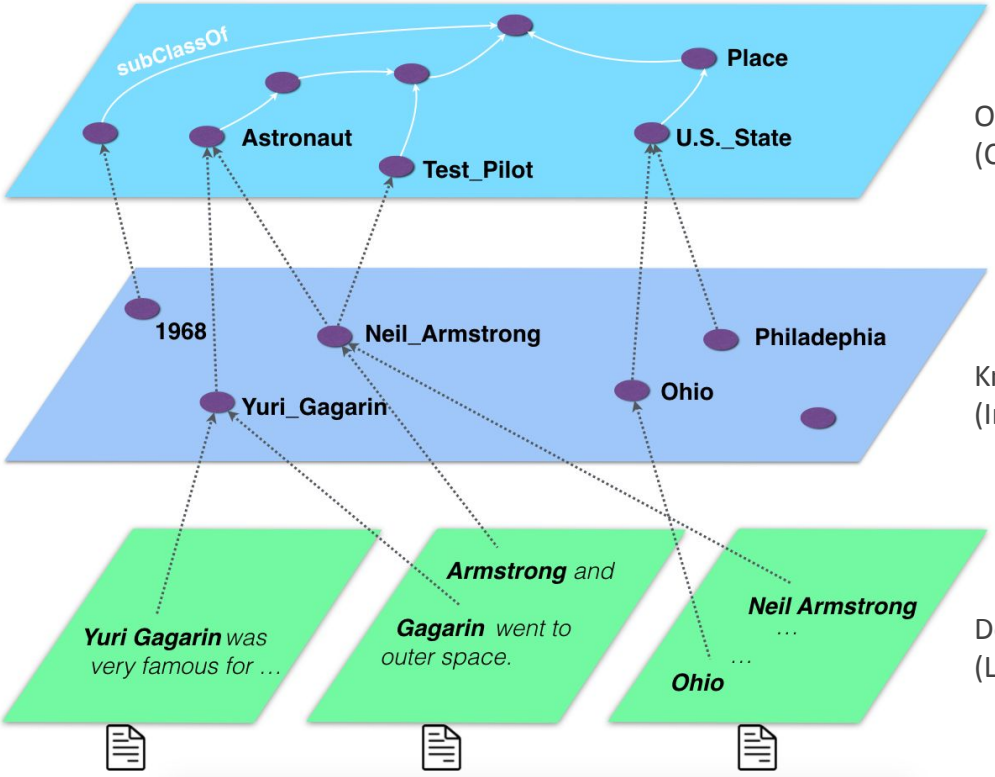
# The Semantic Web Technology Stack
(not a piece of cake...)




The Semantic Web - Not just a piece of cake, [18]

# Semantic Search & Retrieval

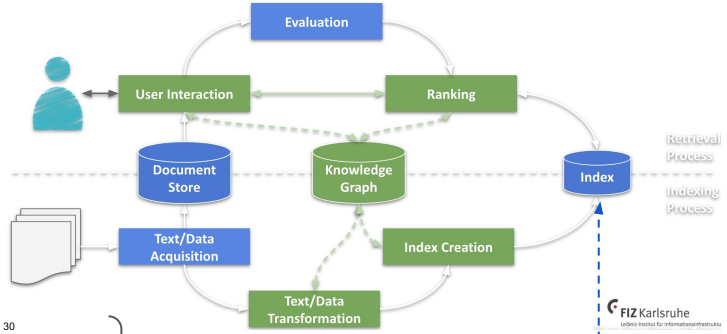
## FAIR Research Data Management - **Findability & Accessibility**




Ontology (Classes & Relations)

Knowledge Graph (Instances)

Documents (Literals)

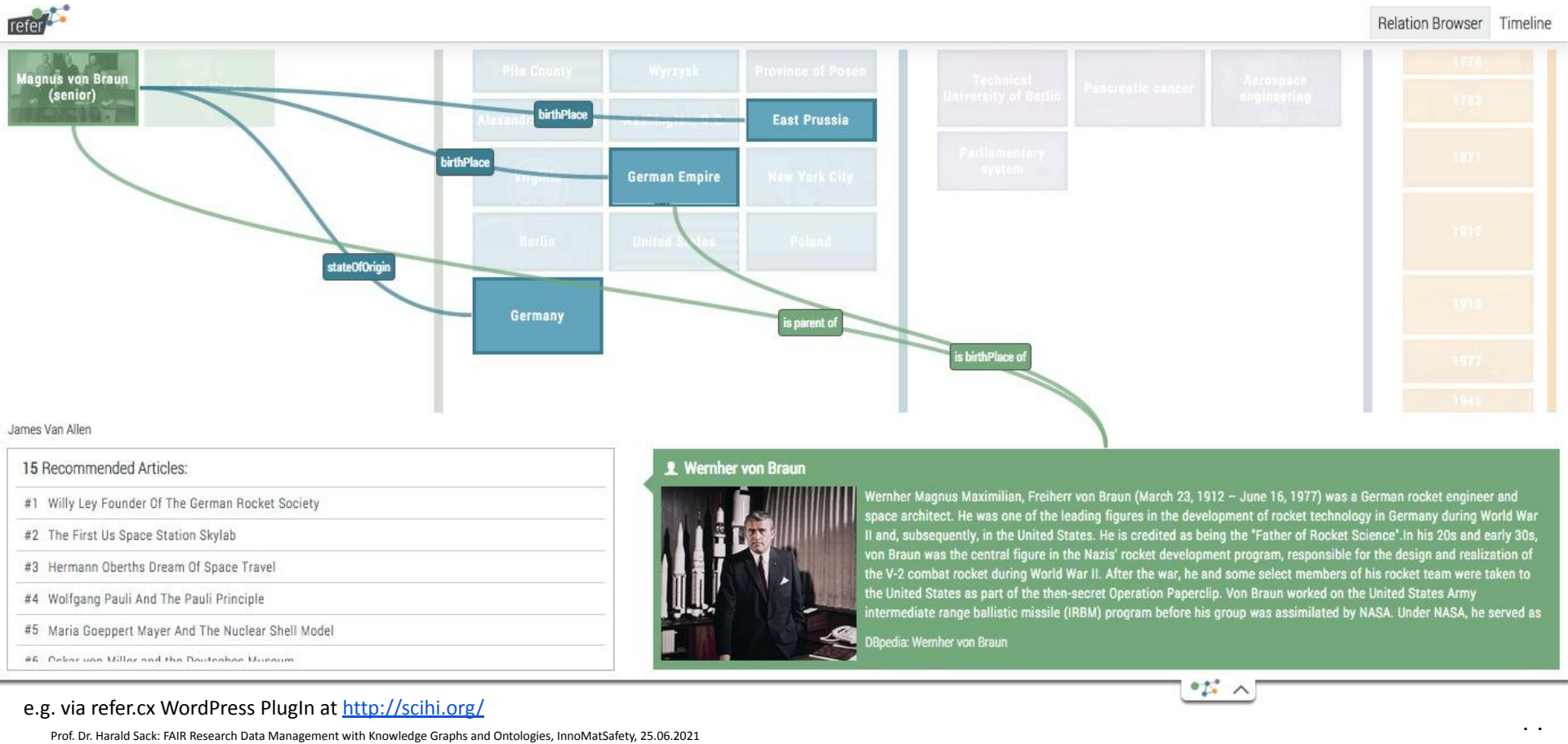



Jörg Waitelonis, Claudia Exeler, and Harald Sack. **Linked Data enabled Generalized Vector Space Model to improve document retrieval.** In Proc. of NLP & DBpedia 2015 workshop in conjunction with 14th International Semantic Web Conference (ISWC2015), CEUR Workshop Proceedings, Vol1581, pp 33-44, 2015.

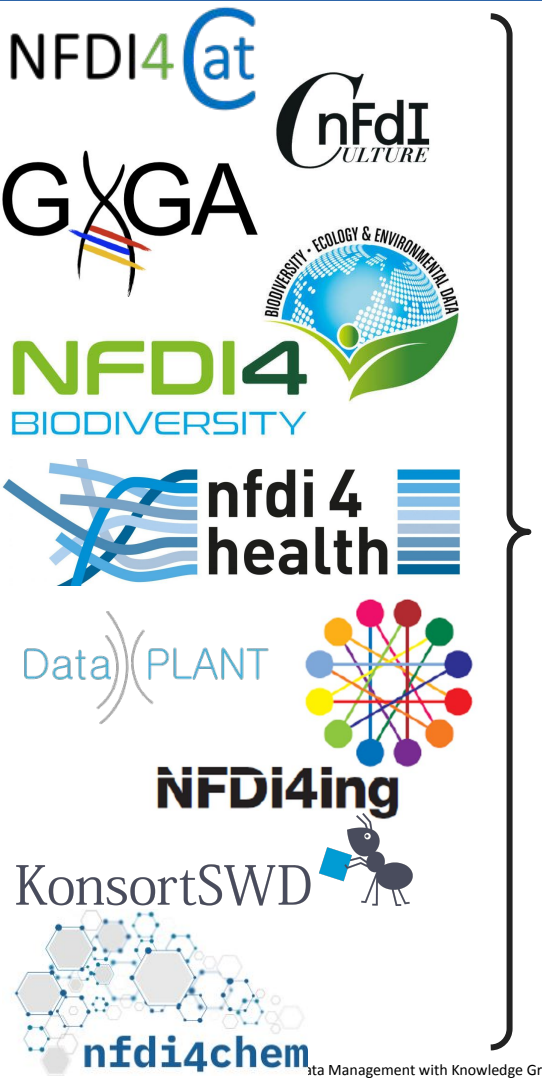
# Exploration & Recommendation

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

## FAIR Research Data Management - **Findability & Accessibility & Reusability**




e.g. via refer.cx WordPress PlugIn at http://scihi.org/




nfdi

# Innovative Informationssysteme

- Verbessertes Retrieval
- Föderierte Suche
- Semantische Suche
- Explorative Suche
- Ähnlichkeitsbasierte Suche
- Intelligente Empfehlungen
- Question Answering
- Explainable AI

# Community-basierte Kuratierung und Management


- **Probleme mit Wikibase:**
  - Eingeschränkte Vernetzbarkeit
  - kein W3C konformes Vokabular (RDF, RDFS, OWL) nutzbar
  - keine **explizite Semantik** (damit auch kein **Reasoning** möglich)
  - kein dezidiertes **Rechte- und Zugriffsmanagement**
  - komplexe Architektur erfordert **aufwändige Anpassungen**
  - Abhängigkeit von Wikimedia Foundation


**Prof. Dr. Harald Sack**

*Knowledge Graphs for Research Data Management*

harald.sack@fiz-karlsruhe.de
twitter: lysander07

InnoMatSafety
25.06.2021

# Take Home Messages:

- Die **Vernetzung von Daten**, **Informationen und Wissen** wird immer wichtiger.
- **Ontologien und Wissensgraphen** helfen bei der effektiven Umsetzung der **FAIR Prinzipien** im Forschungsdatenmanagement.
- Die **NFDI** bietet die Chance der großflächigen (intelligenten) Vernetzung von Forschungsdaten.
- Wissensgraphen ermöglichen **großflächige Vernetzung und Integration von Forschungsdaten** innerhalb der NFDI und darüber hinaus.

**Image References:**


[1] The Sulphurbottom (Sibbaldius sulfureus) from Natural history of the cetaceans and other marine mammals of the western coast of North America (1872) by Charles Melville Scammon (1825-1911). [Public Domain], https://www.rawpixel.com/board/328227/charles-melville-scammons-marine-mammals

[2] Matrix Computer Screen, [Public Domain] https://pixabay.com/illustrations/matrix-code-computer-pc-data-356024/

[3] UBC Library Card Catalog, Paul Joseph, [cc-by-2.0], https://commons.wikimedia.org/wiki/File:2009_3544505541_card_catalog.jpg

[4] Tree of knowledge based on the French Encyclopedie from 1780, [Public Domain] https://commons.wikimedia.org/wiki/File:Essai_d%27une_distribution_g%C3%A9n%C3%A9alogique_des_sciences_et_des_arts_principaux,_1780.jpg

[5] Pieter Bruegel the Elder, The Tower of Babel, 1563, [Public Domain] https://commons.wikimedia.org/wiki/File:Pieter_Bruegel_the_Elder_-_The_Tower_of_Babel_(Vienna)_-_Google_Art_Project_-_edited.jpg

[6] Michelangelo Buonarotti, Creazione di Adamo, c. 1512, [Public Domain] https://en.wikipedia.org/wiki/The_Creation_of_Adam#/media/File:Michelangelo_-_Creation_of_Adam_(cropped).jpg

[7] Niklas Jansson, Touched by His Noodly Appendage, Niklas Janson, [Public Domain] https://commons.wikimedia.org/wiki/File:Touched_by_His_Noodly_Appendage_HD.jpg

[8] Parental Advisory logo, [Public Domain] https://commons.wikimedia.org/wiki/File:Parental_Advisory_label.svg

[9] A fantasy map of a flat earth. Photograph: Antar Dayal/Getty Images/Illustration Works <link>

[10] Jorge Luis Borges by Annemarie Heinrich, 1967, [public domain] https://commons.wikimedia.org/wiki/File:Jorge_Luis_Borges_by_Annemarie_Heinrich,_1967.jpg

[11] Albrecht Dürer, Melancholia I, 1514, [public domain], https://commons.wikimedia.org/wiki/File:D%C3%BCrer_Melancholia_I.jpg

[12] Silos, CC0, https://pxhere.com/en/photo/773866

[13] Globale Digitalisierung, public domain, https://www.publicdomainpictures.net/pictures/380000/velka/globale-digitalisierung-201105.jpg

[14] The Linked Data Cloud, 2019, [cc-by] https://lod-cloud.net

[15] Caspar David Friedrich, Wanderer über dem Nebelmeer, 1818, public domain, https://upload.wikimedia.org/wikipedia/commons/b/b9/Caspar_David_Friedrich_-_Wanderer_above_the_sea_of_fog.jpg

[16] The Software Deveopment Process, Geek & Poke, http://geekandpoke.typepad.com/geekandpoke/2012/01/simply-explained-dp.html

[17] Liberty Statue, work in progress, 1884, https://commons.wikimedia.org/wiki/File:Statue_de_la_Libert%C3%A9,_en_construction.jpg

[18] The Semantic Web, Not just a piece of cake, http://bnode.org/blog/2009/07/08/the-semantic-web-not-a-piece-of-cake