

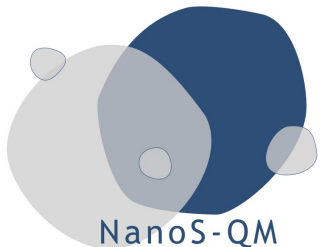


FAIR Research Data Management with Knowledge Graphs and Ontologies

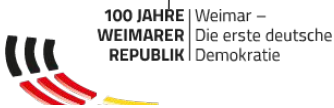
Prof. Dr. Harald Sack
InnoMatSafety
25.06.2021

Knowledge Graphs @ FIZ Karlsruhe

Projects, Services & NFDIs



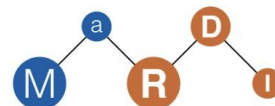
MATERIALDIGITAL



2019



2020



2021

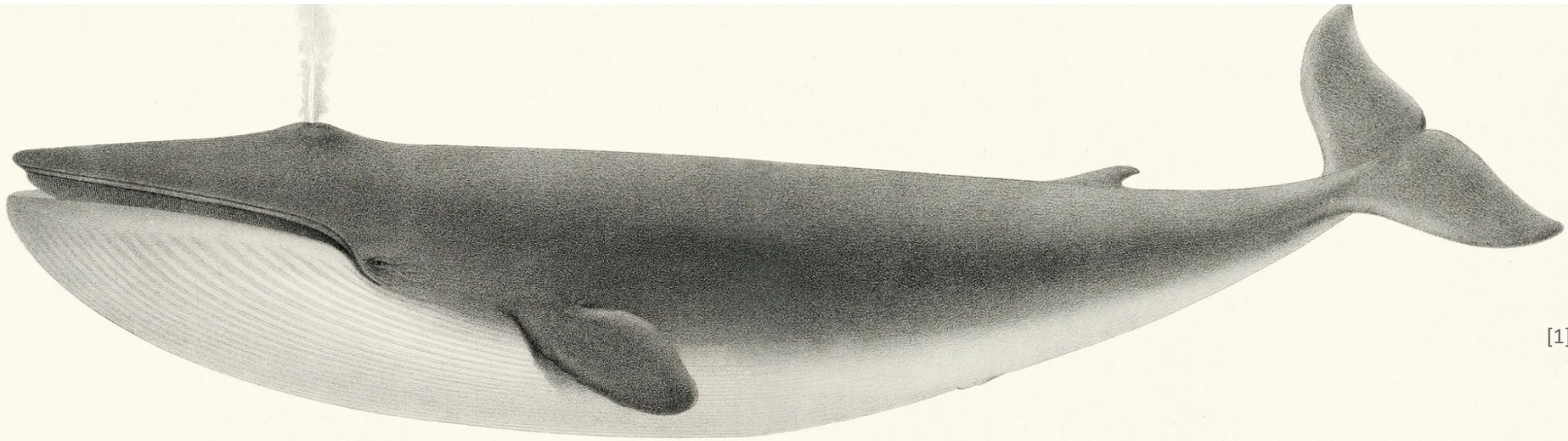


33.6

33.6 m

33.6m

33.6m^{[1] (1922)}



[1]

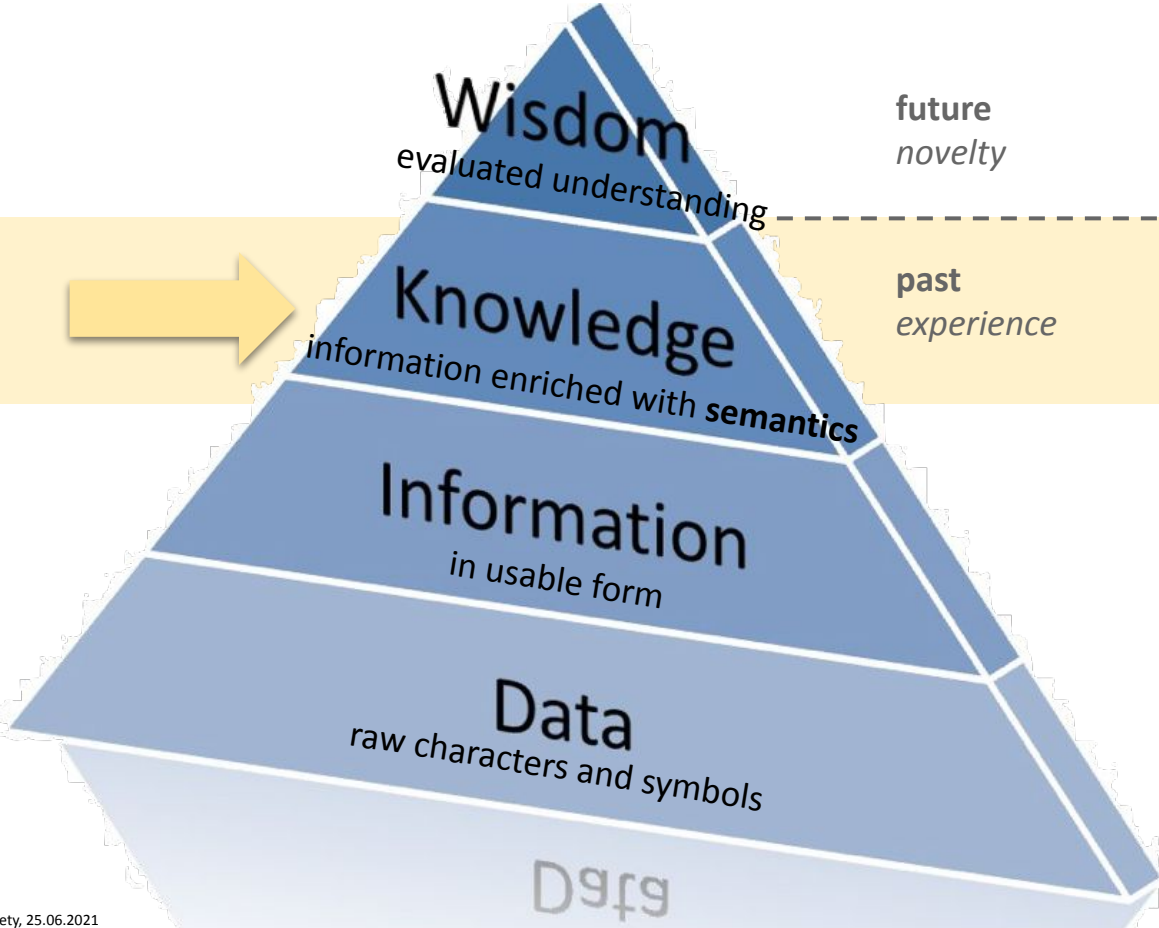
[1] S. G. Brown: *Balaenoptera musculus* (Linnaeus 1758) – *Blauwal*, in Jochen Niethammer, Franz Krapp (Hrsg.): Handbuch der Säugetiere Europas. Band 6: Meeressäuger, Teil I Wale und Delphine – Cetacea, Teil IB: Ziphiidae, Kogiidae, Physeteridae, Balaenidae, Balaenopteridae. Aula-Verlag Wiesbaden 1995

Data



Information

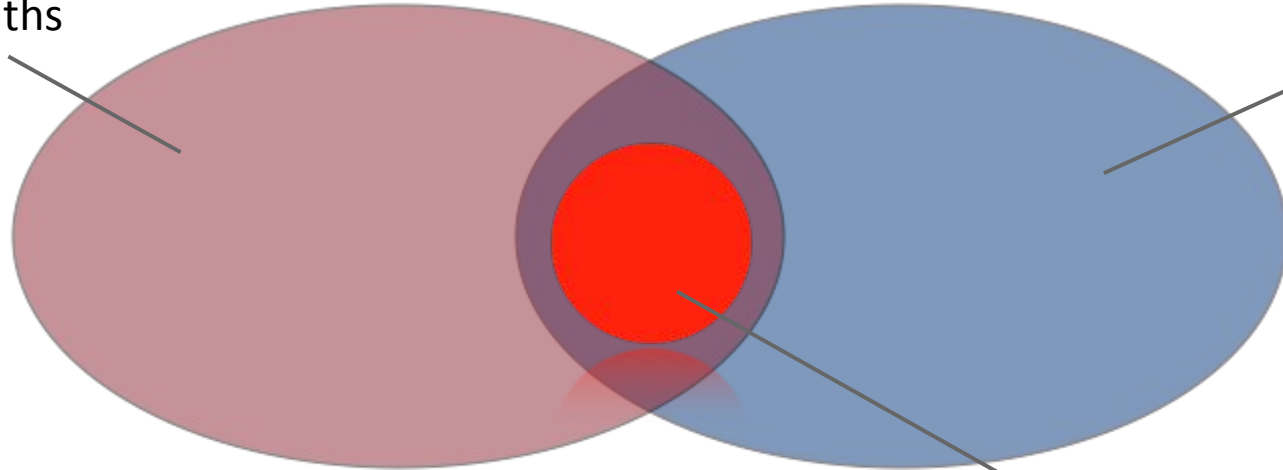
From Data to Knowledge



DIKW Pyramid, Ackoff 1989

What is Knowledge?

Truths



Beliefs

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?

The background of the slide is a reproduction of Michelangelo's famous fresco, 'The Creation of Adam'. It depicts Adam on the left, reclining and reaching out with his right hand towards God on the right. God is shown as an elderly man with a long white beard, wearing a pinkish-red robe, reclining and pointing his right index finger towards Adam. The two hands are just inches apart, creating a sense of tension and divine spark. Other figures, including the Virgin Mary and other angels, are visible in the background on the right side.

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

Philosophy 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.*

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.*

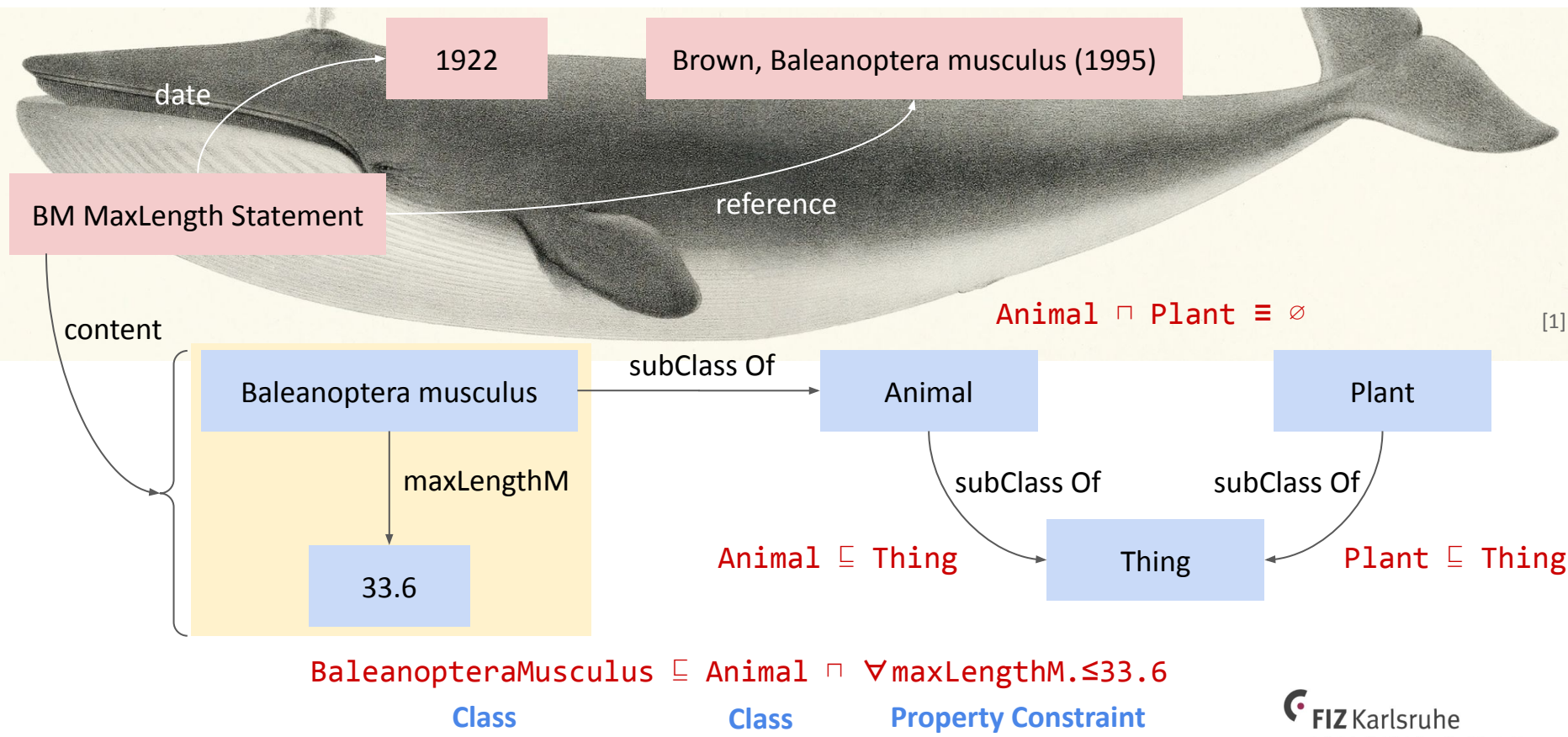
conceptualization:	abstract model (domain, identified relevant concepts, relations)
explicit:	meaning of all concepts must be defined
formal:	machine understandable
shared:	consensus about ontology

P A R E N T A L

ADVISORY

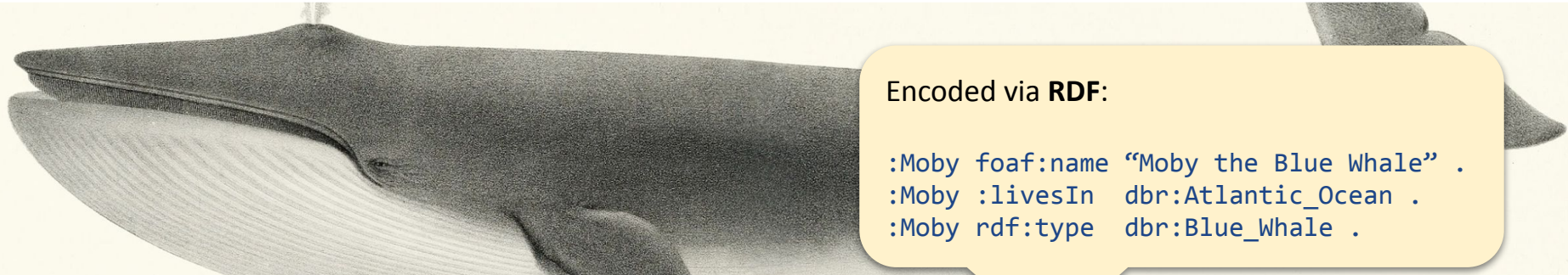
EXPLICIT SEMANTICS

(Mini) Example Ontology + Knowledge Graph



(Mini) Example Ontology + Knowledge Graph

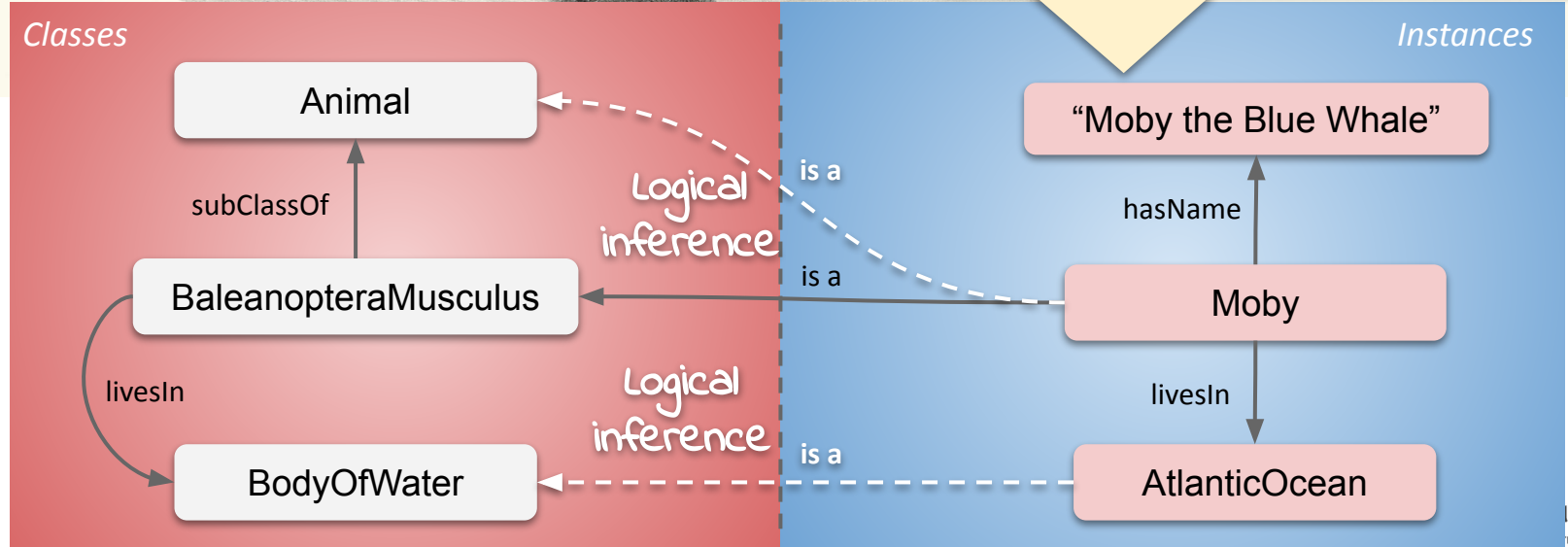
Logical Inference and Encoding



Encoded via **RDF**:

```

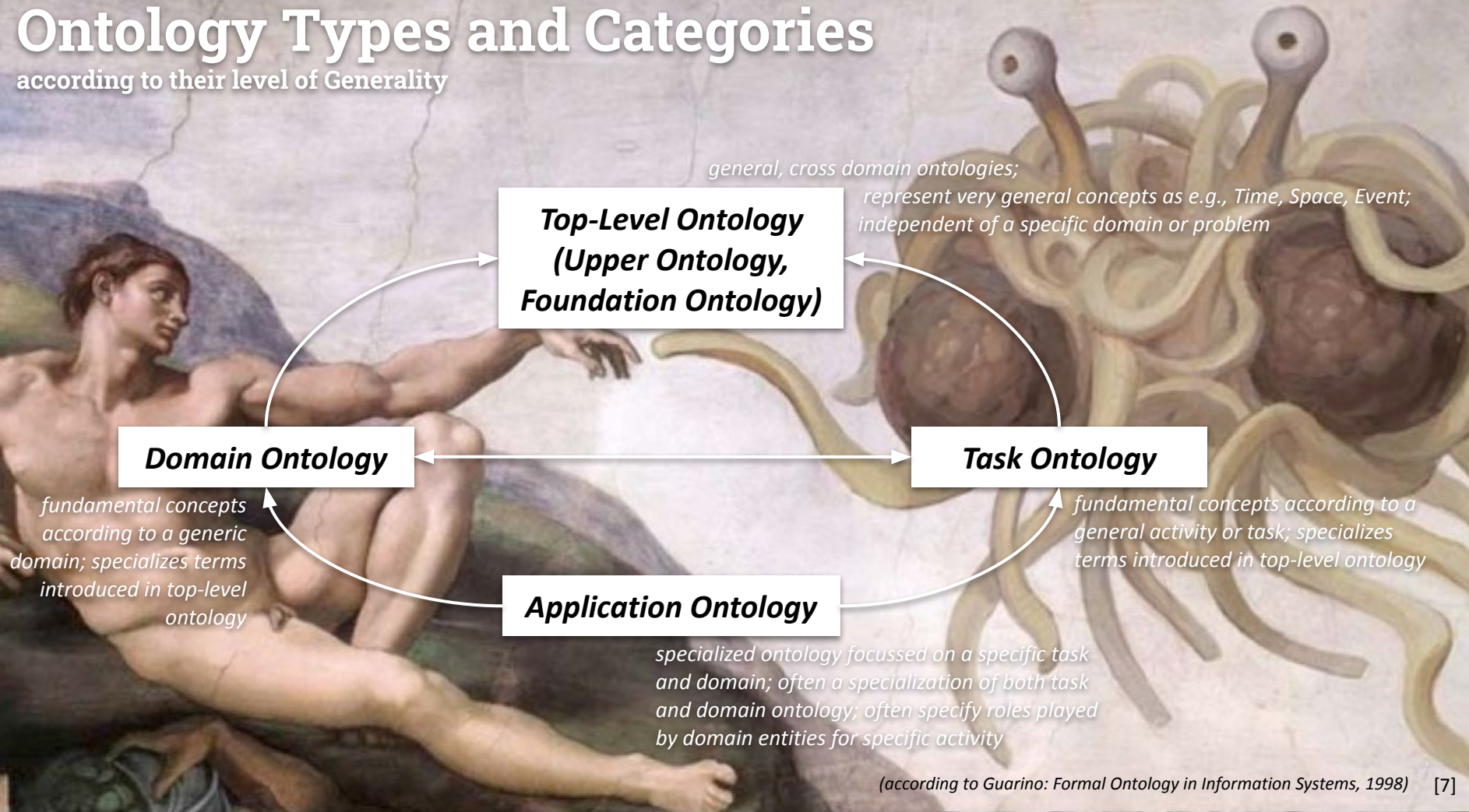
:Moby foaf:name "Moby the Blue Whale" .
:Moby :livesIn dbr:Atlantic_Ocean .
:Moby rdf:type dbr:Blue_Whale .
    
```



[1]

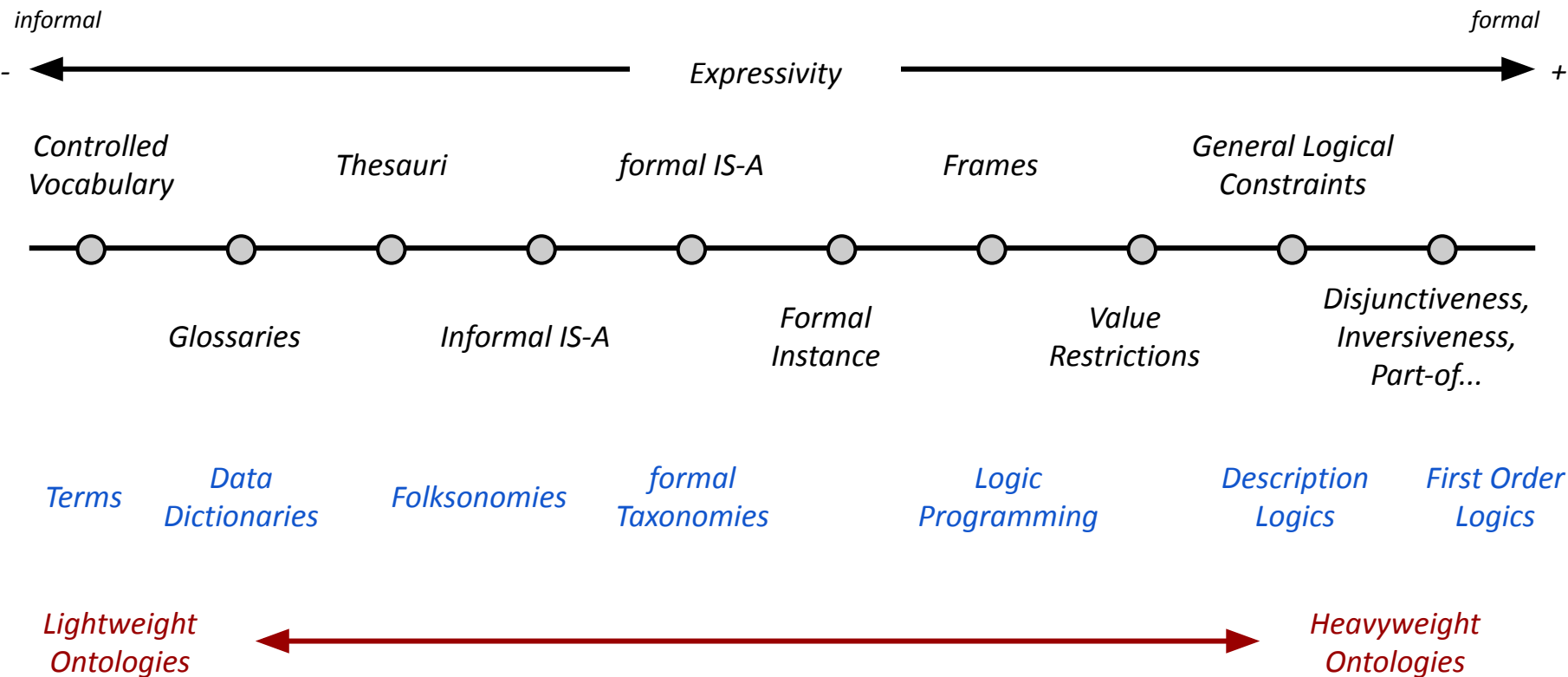
Ontology Types and Categories

according to their level of Generality



Ontology Types and Categories

according to their level of Semantic Expressivity



(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)

A woodcut-style illustration of a sea with a face. The sea is depicted with blue and green waves. In the upper right, a face with a beard and curly hair is breathing a stream of fire or light across the water. Two sailing ships are on the left, and a dragon's head is visible in the bottom left corner. The background shows a coastline with buildings and a cloudy sky.

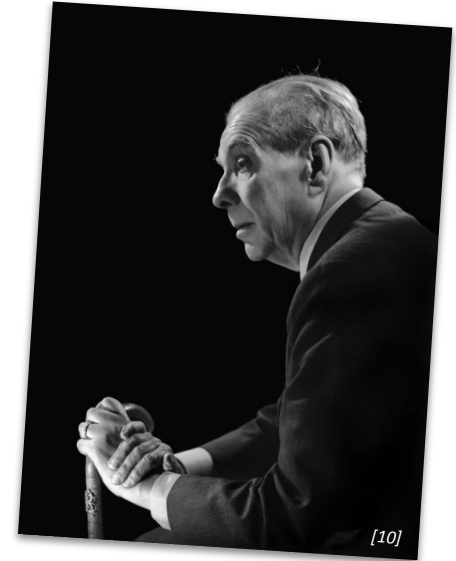
**“It does not do to leave a live dragon out of your calculations,
if you live near him.”**

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)?**

Geisteswissenschaften

Naturwissenschaften

Lebenswissenschaften

Sozialwissenschaften

Ingenieurwissenschaften

Forschungsdaten



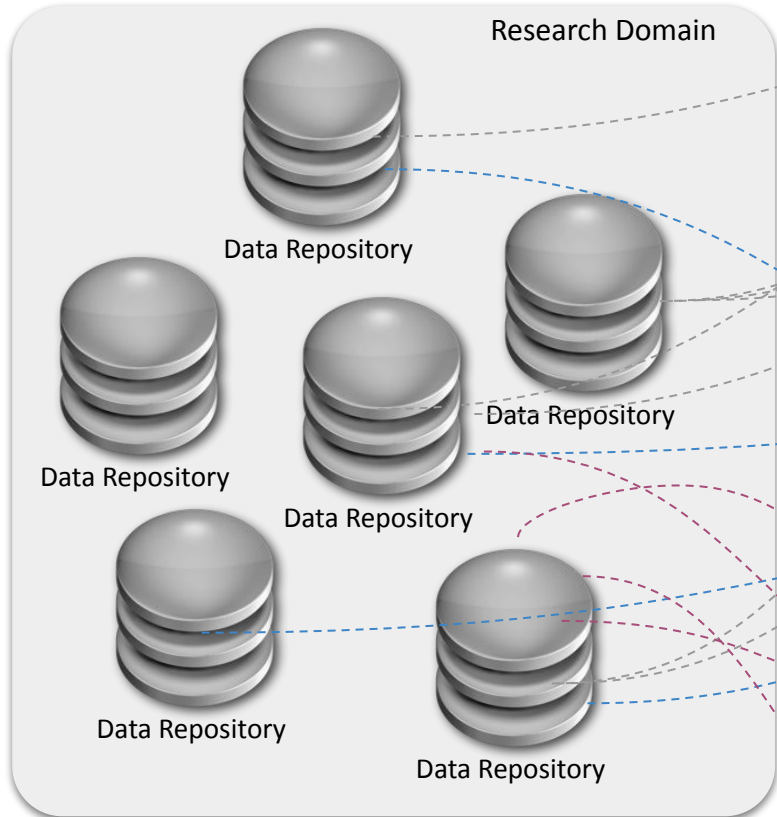
Forschungs- datensilos

- 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

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



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...?



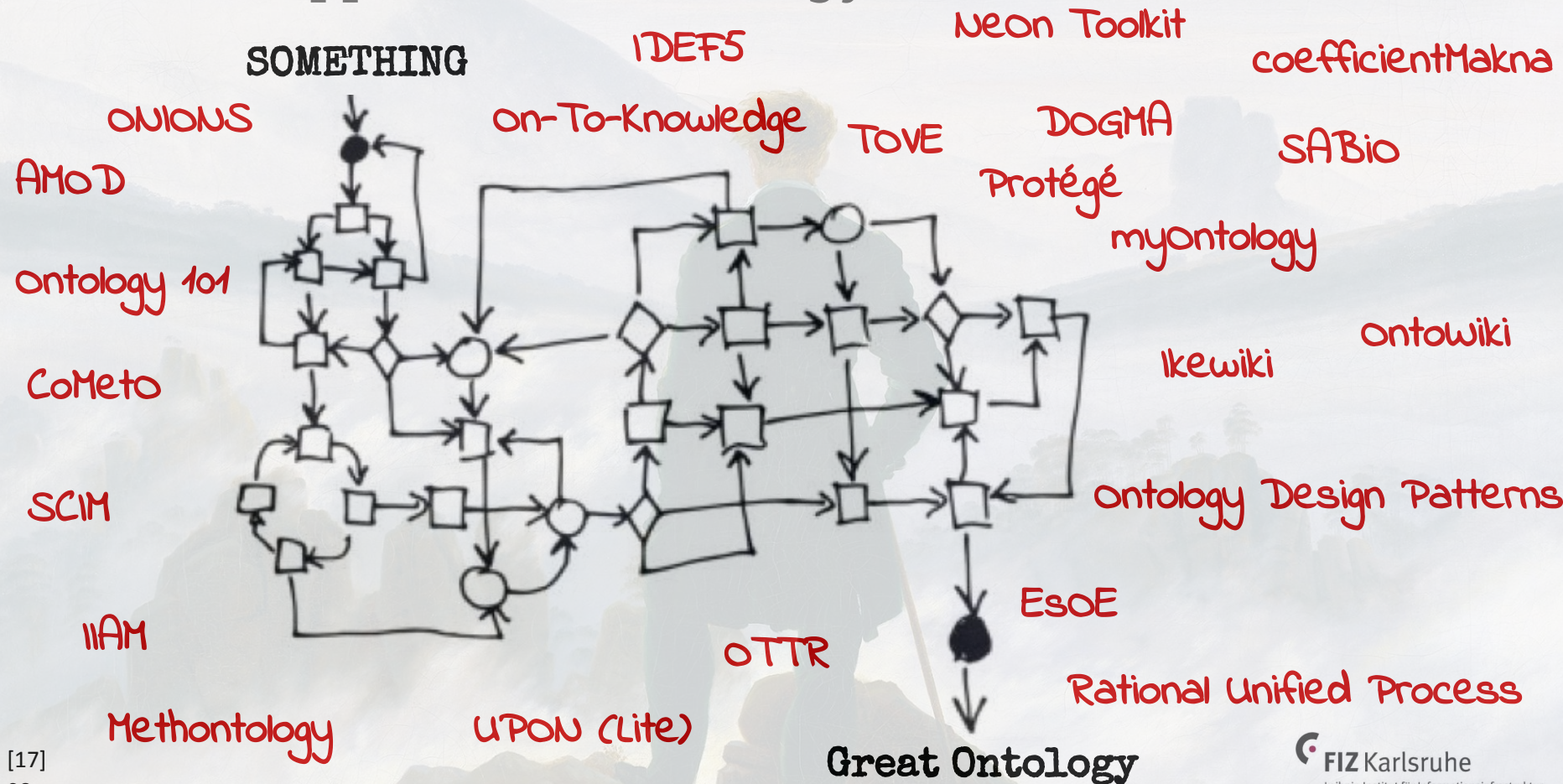
Follow an Approved Methodology

SOMETHING



Great Ontology

Follow an Approved Methodology



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}]
0.000000e+00	2.1538249e-01	2.2127591e-01
1.000000e-02	2.2320410e-01	2.2851489e-01
2.000000e-02	2.3078706e-01	2.3700471e-01
3.000000e-02	2.3957809e-01	2.4612475e-01
4.000000e-02	2.4898703e-01	2.5569295e-01
5.000000e-02	2.5889461e-01	2.6656408e-01
6.000000e-02	2.7120663e-01	2.7901766e-01
7.000000e-02	2.8447237e-01	2.9209201e-01
8.000000e-02	2.9853002e-01	3.0861118e-01
9.000000e-02	3.1697947e-01	3.2641678e-01
1.000000e-01	3.3656863e-01	3.4837557e-01
1.100000e-01	3.6049250e-01	3.7427430e-01
1.200000e-01	3.8862354e-01	4.0343478e-01
1.300000e-01	4.2297845e-01	4.3891770e-01
1.400000e-01	4.6555629e-01	4.8310615e-01
1.500000e-01	5.1581989e-01	5.3864561e-01
1.600000e-01	5.7837521e-01	6.0616555e-01
1.700000e-01	6.4984874e-01	6.8350098e-01
1.800000e-01	7.3012722e-01	7.6446633e-01
1.900000e-01	8.1671138e-01	8.5748202e-01
2.000000e-01	9.0275181e-01	9.4726775e-01

Ontologies and Knowledge Graphs for Research Data Management

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

Fields +

- z/d [1] string »
- Ion density (PIC-ITAP) [10¹⁵ m⁻³] string »
- Ion density (PIC-INP) [10¹⁵ m⁻³] string »
- Ion density (Fluid-DDAn) [10¹⁵ m⁻³] string »
- Ion density (Fluid-DDA53) [10¹⁵ m⁻³] string »

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

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. Kählert, Hanno Sun, Anbang Loffhagen, Detlef
Release Date	2019-06-14
Resources	Benchmark data for CCRF discharge plasmas - time averaged ion density (argon, 20 Pa) Benchmark data for CCRF discharge plasmas - time averaged ion density (argon, 40 Pa) Benchmark data for CCRF discharge plasmas - time averaged ion density (argon, 80 Pa) Benchmark data for CCRF discharge plasmas - time averaged ion density (helium, 10 Pa) Benchmark data for CCRF discharge plasmas - time averaged ion density (helium, 20 Pa) Benchmark data for CCRF discharge plasmas - time averaged ion density (helium, 40 Pa) Raw benchmark data for CCRF discharge plasmas - time averaged ion density (helium, 80 Pa) Show more
Identifier	60dbccd4-8be4-4f41-896c-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 high frequency low pressure 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

Ontologies and Knowledge Graphs for Research Data Management

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

semantic information



Item Discussion

plasma (Q10251)

state of matter consisting of ionized gas
materia plasmática | gas ionizado

~ In more languages
Configure

Language	Label	Description	Also known as
English	plasma	state of matter consisting of ionized gas	materia plasmática gas ionizado
German	Plasma	Gas, dessen Bestandteile teilweise oder vollständig als Ionen und Elektronen vorliegen	
French	plasma	état de la matière où sont mélangés des électrons, des ions et des noyaux atomiques	
Bavarian	No label defined	No description defined	

All entered languages

Statements

instance of

- fundamental state of matter
 - sourcing circumstances: disputed
 - < 0 references
 - + add reference
- classical state of matter
 - < 0 references
 - + add reference
 - + add value

subclass of

- gas
 - sourcing circumstances: disputed

Properties and parameters

Definition

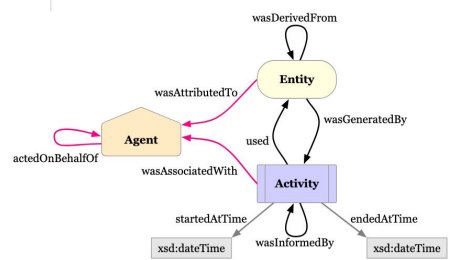
Plasma is a state of matter in which an ionized gaseous substance becomes highly electrically conductive to the point that long-range electric and magnetic fields dominate the behaviour of the matter.^{[21][22]} The plasma state can be contrasted with the other states: solid, liquid, and gas.

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 and is affected by the fields created by the other charges. In turn this governs collective behaviour with many degrees of variation.^{[1][21][23]} Three factors define a plasma:^{[24][25]}

1. **The plasma approximation:** The plasma approximation applies when the plasma parameter, Λ ,^[26] 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.^{[21][22]}
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.^[27]
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.^[28]

unstructured information

semantic information



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** [show]
- Scientists** [show]

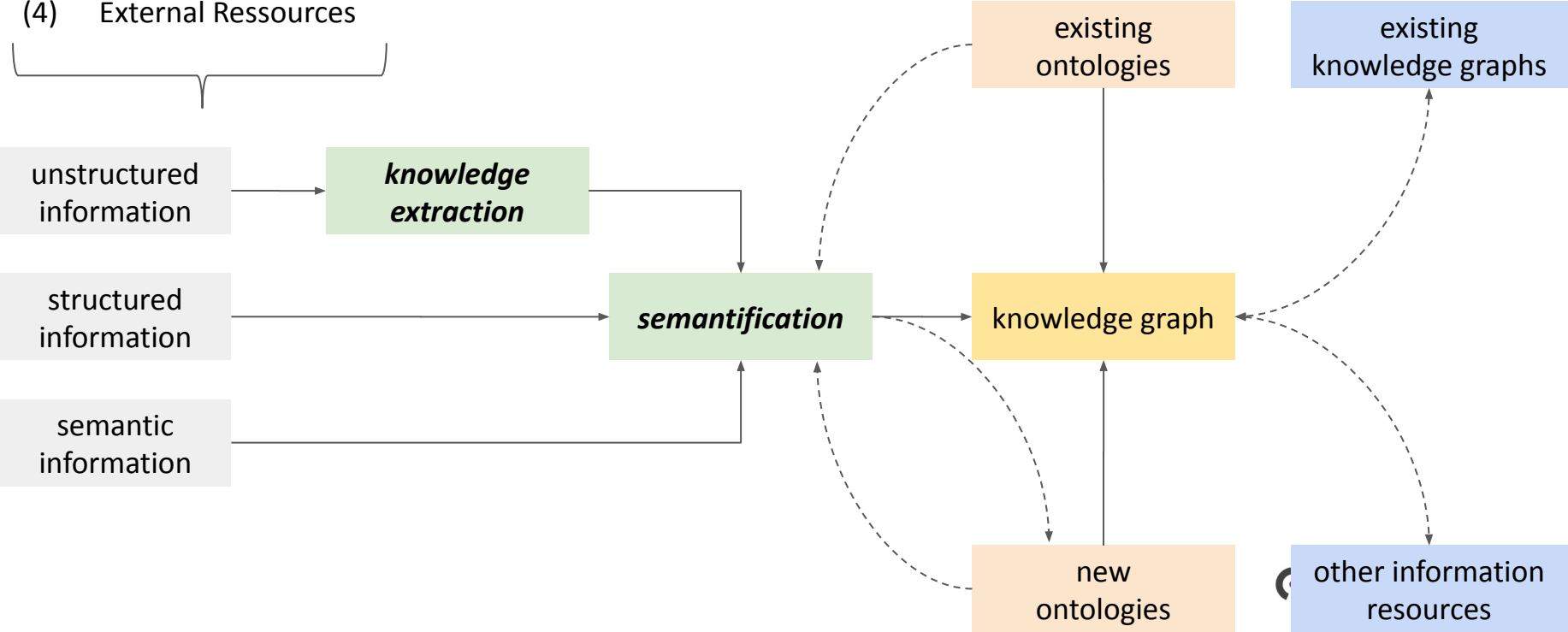
V · T · E

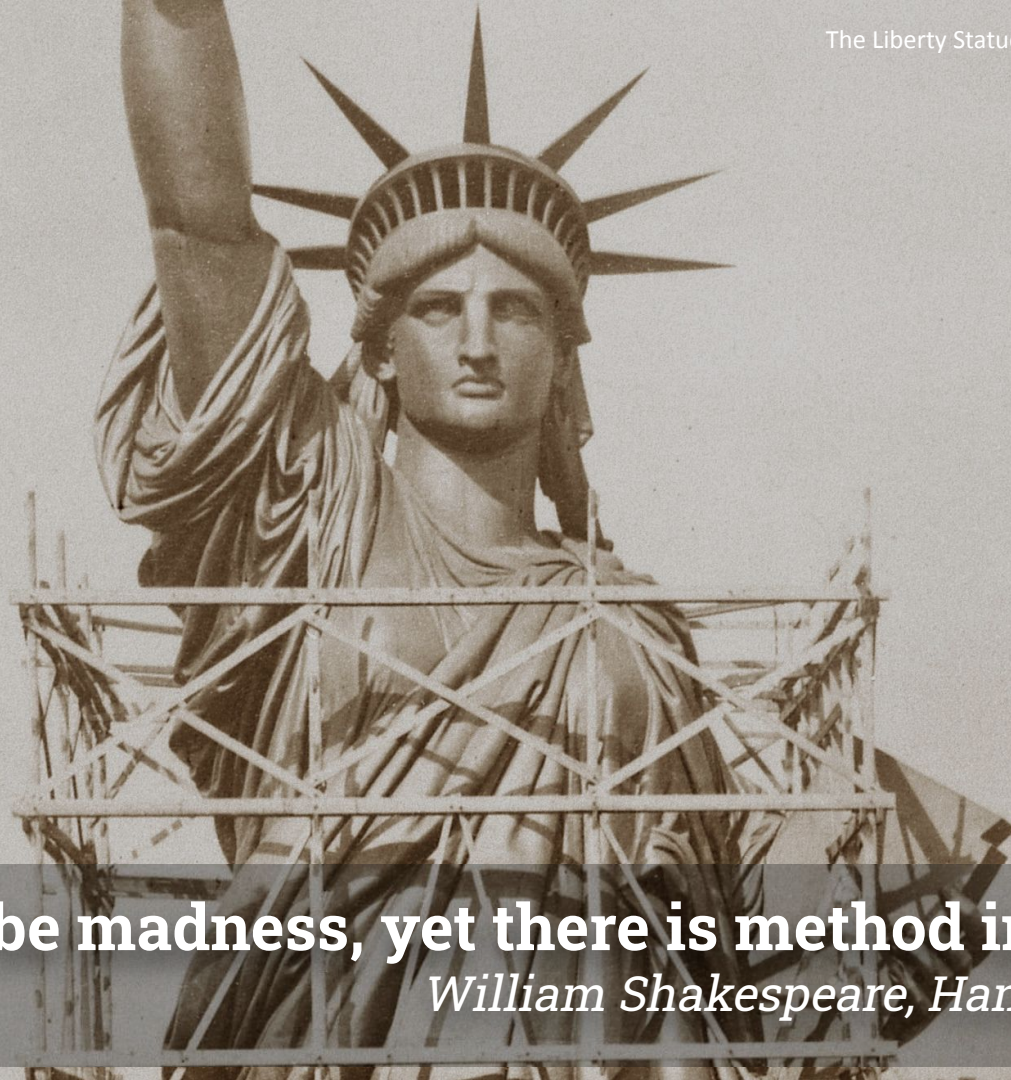
structured information



Ontologies and Knowledge Graphs for Research Data Management

- (1) (Raw) Research Data
- (2) Schema Information
- (3) Metadata
- (4) External Ressources





“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

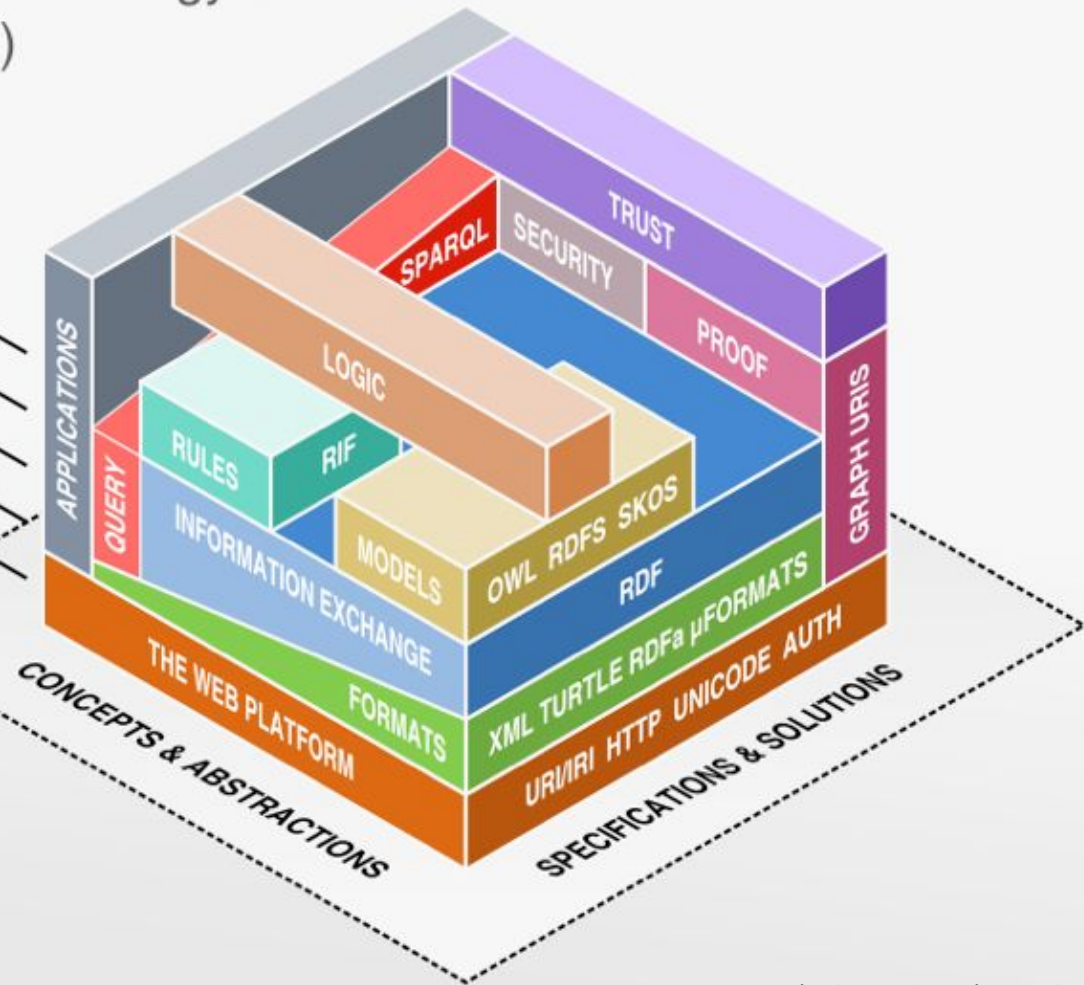
Standardized information exchange is key

Formats are necessary, but not too important

The Semantic Web is based on the Web

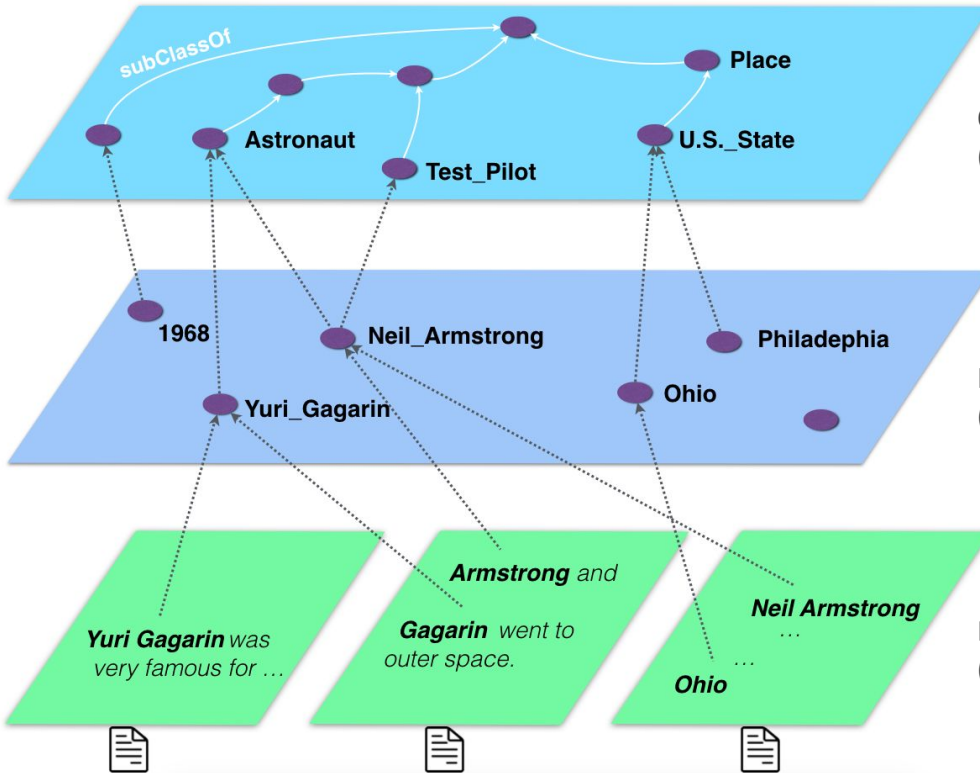
Linked Data uses a small selection of technologies

LINKED DATA



Semantic Search & Retrieval

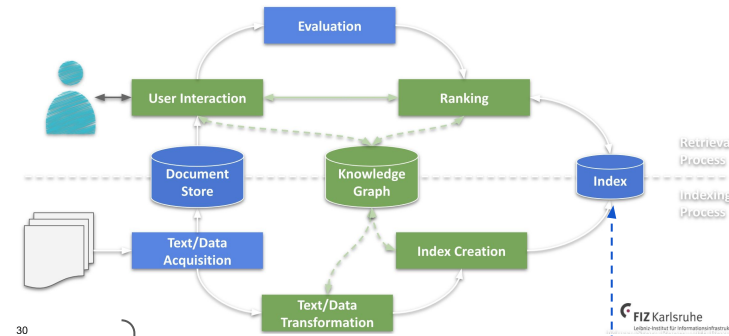
FAIR Research Data Management - **Findability & Accessibility**



Ontology
(Classes & Relations)

Knowledge Graph
(Instances)

Documents
(Literals)



30

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

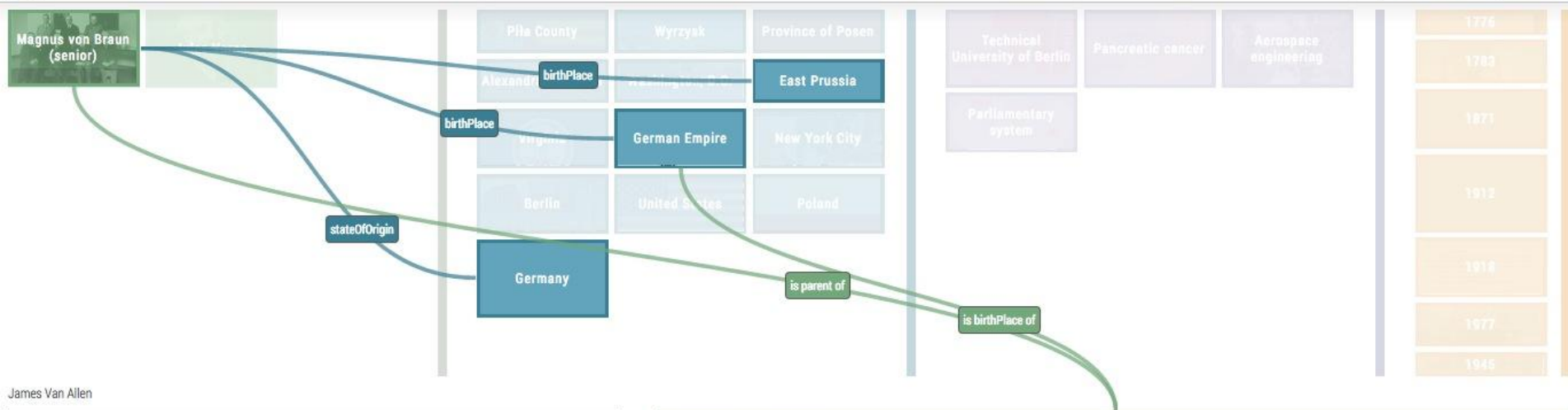
Jörg Waitelonis, Claudia Exeler, and Harald Sack. **Linked Data enabled Generalized Vector Space Model to improve document retrieval.** In Proc. of NLP & DBpedia 2015 workshop in conjunction with 14th International Semantic Web Conference (ISWC2015), CEUR Workshop Proceedings, Vol1581, pp 33-44, 2015.

Exploration & Recommendation

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



Relation Browser Timeline



James Van Allen

15 Recommended Articles:

- #1 Willy Ley Founder Of The German Rocket Society
- #2 The First Us Space Station Skylab
- #3 Hermann Oberths Dream Of Space Travel
- #4 Wolfgang Pauli And The Pauli Principle
- #5 Maria Goeppert Mayer And The Nuclear Shell Model
- #6 Oskar von Miller and the Deutsches Museum

Wernher von Braun



Wernher Magnus Maximilian, Freiherr von Braun (March 23, 1912 – June 16, 1977) was a German rocket engineer and space architect. He was one of the leading figures in the development of rocket technology in Germany during World War II and, subsequently, in the United States. He is credited as being the "Father of Rocket Science". In his 20s and early 30s, von Braun was the central figure in the Nazis' rocket development program, responsible for the design and realization of the V-2 combat rocket during World War II. After the war, he and some select members of his rocket team were taken to the United States as part of the then-secret Operation Paperclip. Von Braun worked on the United States Army intermediate range ballistic missile (IRBM) program before his group was assimilated by NASA. Under NASA, he served as

DBpedia: Wernher von Braun

e.g. via refer.cx WordPress Plugin at <http://scihi.org/>





Innovative Informationssysteme

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

Community-basierte Kuratierung und Management



species: marine mammal
B. musculus

Language	Label	Description	Also known as
English	Balaenoptera musculus	species of marine mammal	B. musculus blue whale
German	Blauwal	Art der Gattung Balaenoptera	Balaenoptera musculus Blauwale
French	baleine bleue	espèce de cétacés de la famille des Balaenopteridae	Balaenoptera musculus rorqual bleu baleine bleue
Bavarian	No label defined	No description defined	

All entered languages

Statements

instance of	taxon
	0 references

image



Bluewhale877.jpg
1,792 x 1,128; 1.36 MB

media legend

Rorqual blau adult a l'est de l'oceà Pacific (Catalan)

1 reference



Faroe stamp 402 blue whale (Balaenoptera musculus) crop.jpg
302 x 358; 93 KB


0 references

Wikidata: Balaenoptera musculus

ar	حوت أزرق
ast	Balaenoptera musculus
avk	Megenol (Balaenoptera musculus)
az	Göy balina
be_x_old	Блактны кіт
be	Блактны кіт
bg	Син кит
bn	সীত কিত
br	Balum glas
bs	Plavi kit
ca	Rorqual blau
ceb	Balaenoptera musculus
cs	Morvil obrovský
cy	Morvil Glas
da	Bålhval
de	Blauwal
el	Γαλάττα βαλινά
en	Blue whale
eo	Blua baleno
es	Balaenoptera musculus
et	Sinivaal
eu	Balea urdin
fa	نهنگ آبی
fi	Sinivalas
fr	Baleine bleue
fy	Blauwe finfisk
ga	Miol mór gorm
gl	Balea azul
he	לווייתן כחול
hi	नीली हूँस
hr	Plavetni kit
hu	Kék bálna
hy	Կապույտ կեն
id	Paus biru
io	Baleno blua
is	Steypireyður
it	Balaenoptera musculus
ja	シロナガスクジラ
jv	Paus biru
kab	Tizmekt tazegzawt
ka	ლოვითი კეზბაბი
kk	Кек кит
kl	Tunnullit
kn	ಹಿರಿಯ ಕೆಪುಟ
ko	대왕고래
kw	Morvil Glas
la	Balaenoptera musculus
li	Blauwe vinvès
lt	Mėlynasis banginis
lv	Zilais valis

- **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

2. NFDI4CULTUR
Wikibase Workshop
07.07.2021



“Technology presumes there's
just one right way to do things
and there never is.”

*Robert M. Pirsig, Zen and the Art of Motorcycle
Maintenance (1974)*

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](https://commons.wikimedia.org/wiki/File:Pieter_Bruegel_the_Elder_-_The_Tower_of_Babel_(Vienna)_-_Google_Art_Project_-_edited.jpg)
- [6] Michelangelo Buonarroti, Creazione di Adamo, c. 1512, [Public Domain] [https://en.wikipedia.org/wiki/The_Creation_of_Adam#/media/File:Michelangelo_-_Creation_of_Adam_\(cropped\).jpg](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 Development 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>