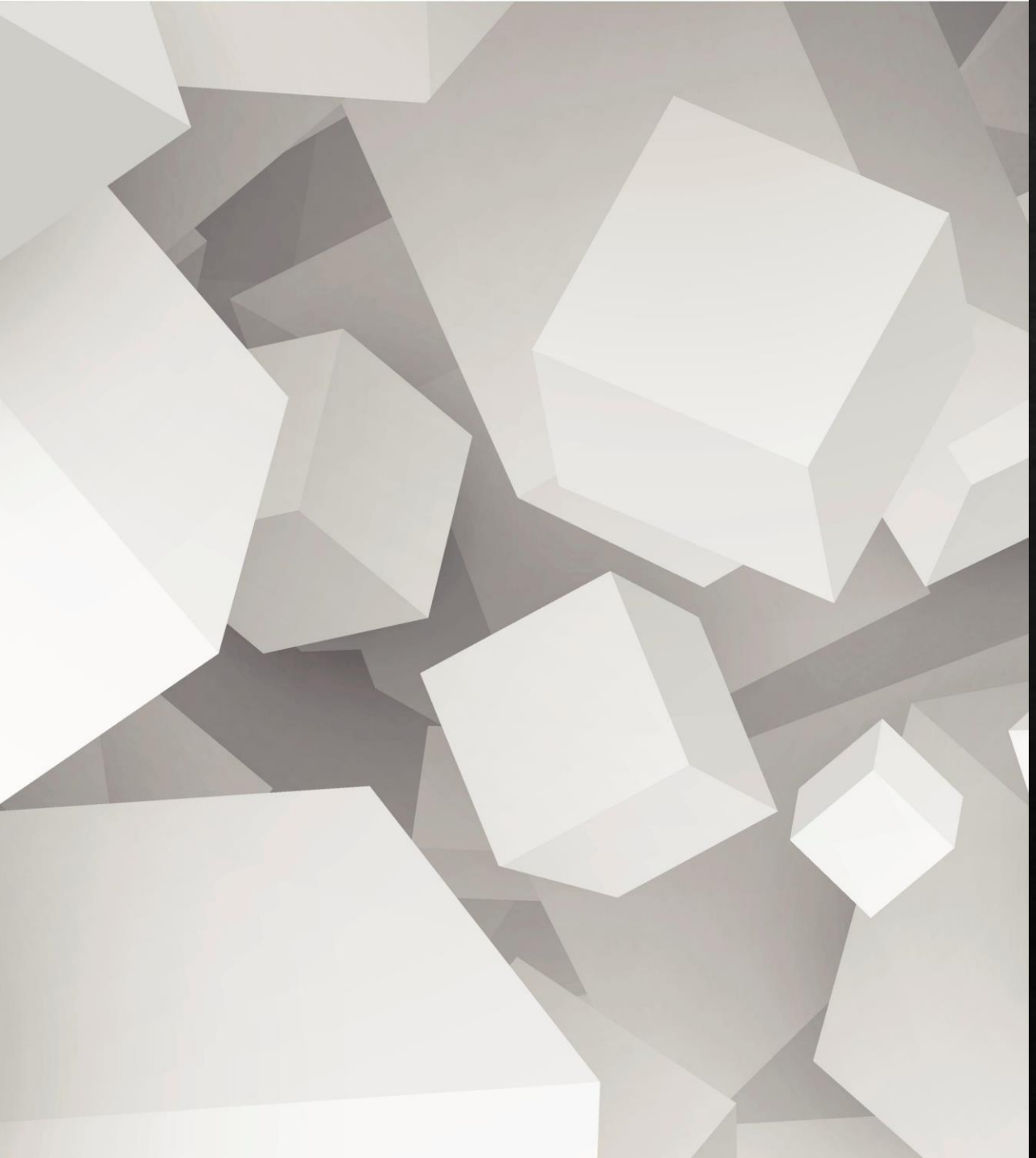
KNOWLEDGE GRAPHS

- Yash Varshney

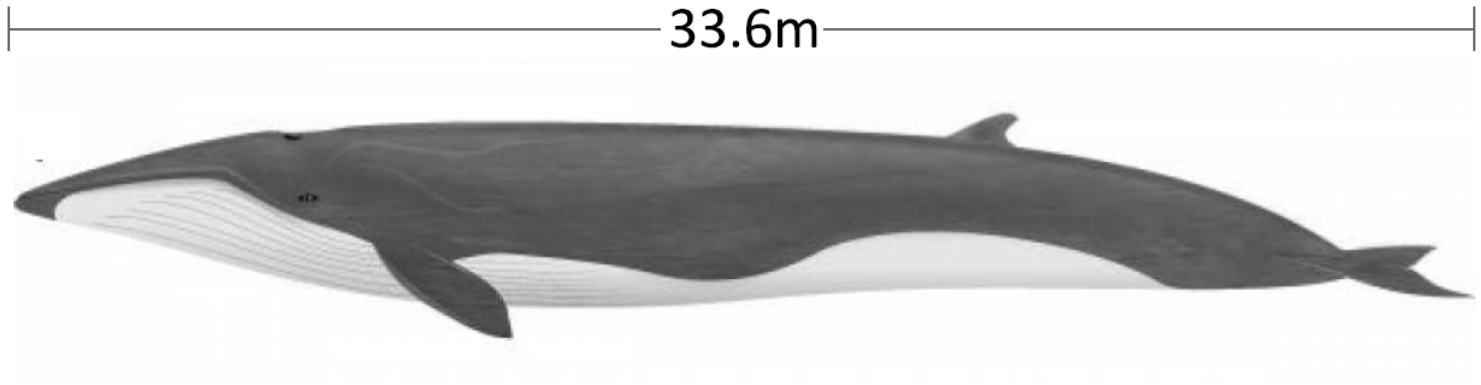


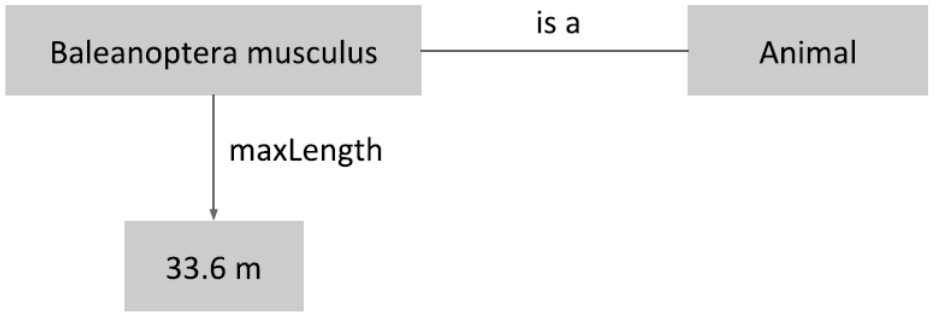
Index

- Data | Information | Knowledge
- How to represent Knowledge?
- Formal Knowledge Representation
- Understanding the Meaning
- Semantic Web – Web of Data
- Data Access Traditional and Semantic Web

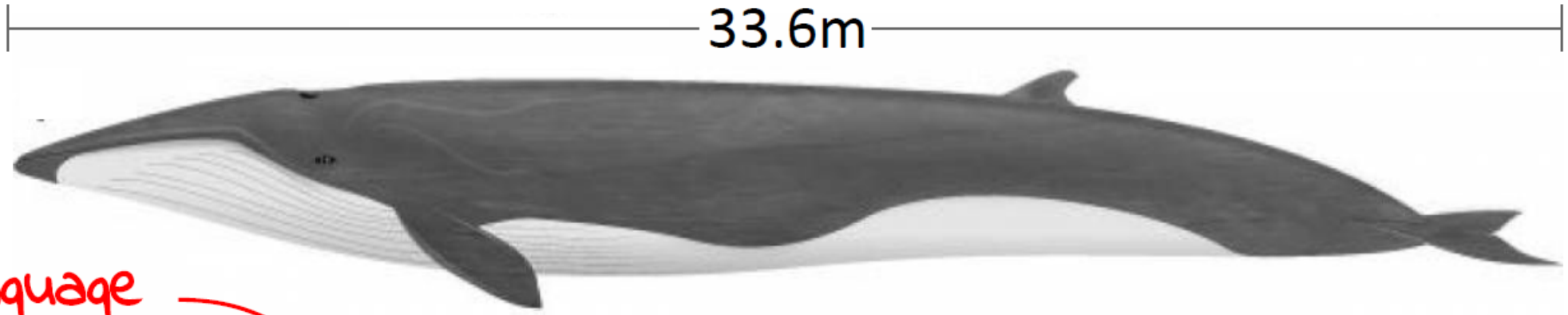
33.6

33.6 m





$\texttt{BaleanopteraMusculus} \sqsubseteq \texttt{Animal} \sqcap \forall \texttt{maxLength}.\leq 33.6$



Language

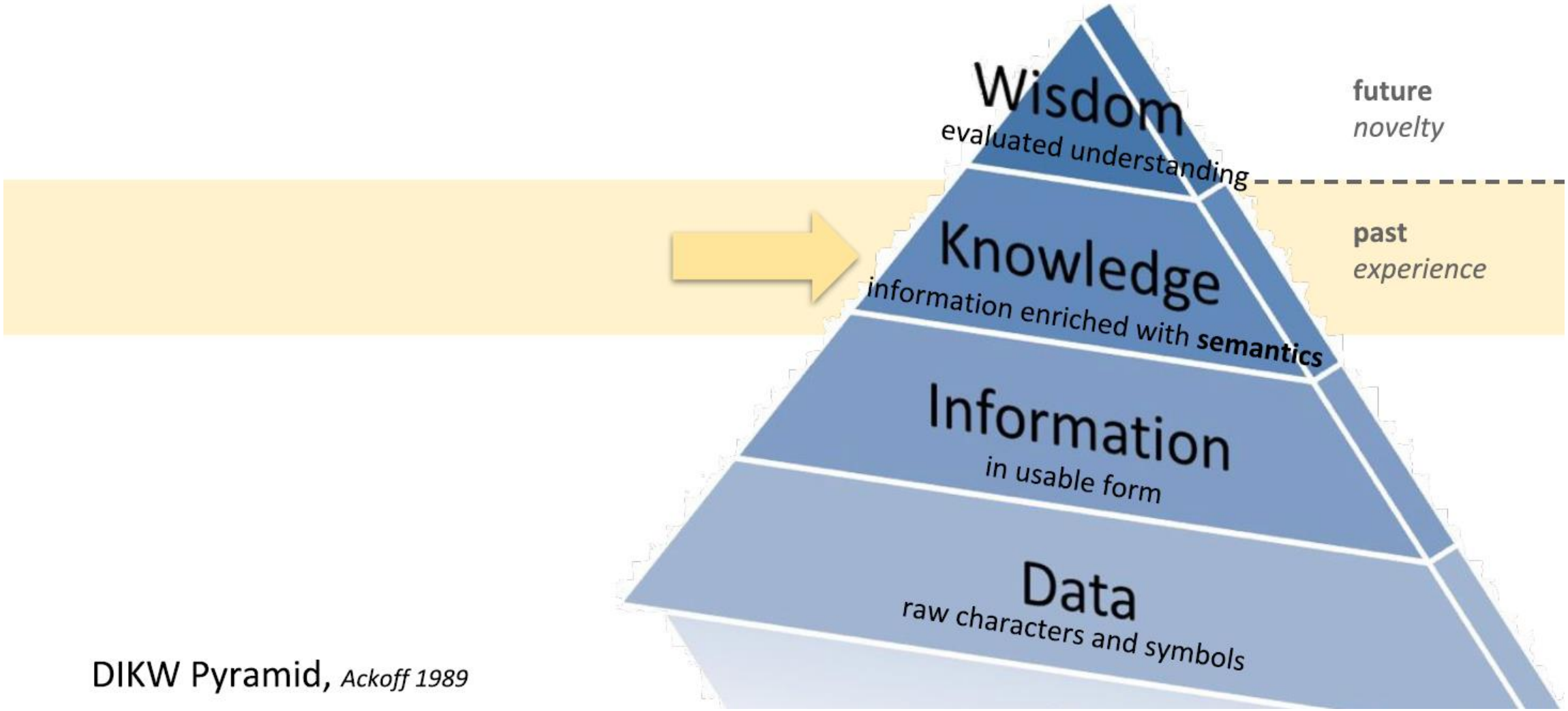
- **We want to express more:**
  - The blue whale is a whale. A whale is a mammal. A mammal is an animal.
  - The whale lives in the oceans. An ocean is a body of water.
  - This is ‘Moby’, a specific blue whale. He lives in the Atlantic Ocean.
  - The longest ever measured blue whale had a length of 33.6 m
  - This means that - up to now and unless we may find a longer one - the largest blue whale measures 33.6 m, or no blue whale is longer than 33.6 m.
  - Moby is not longer than 33.6 m.
  - If you happen to find a longer whale, then either it is no blue whale or we have to change our previous assumptions.

Data | Information | Knowledge

- Data is raw
- Its simply exists and has no significance
- It can be or not be useful

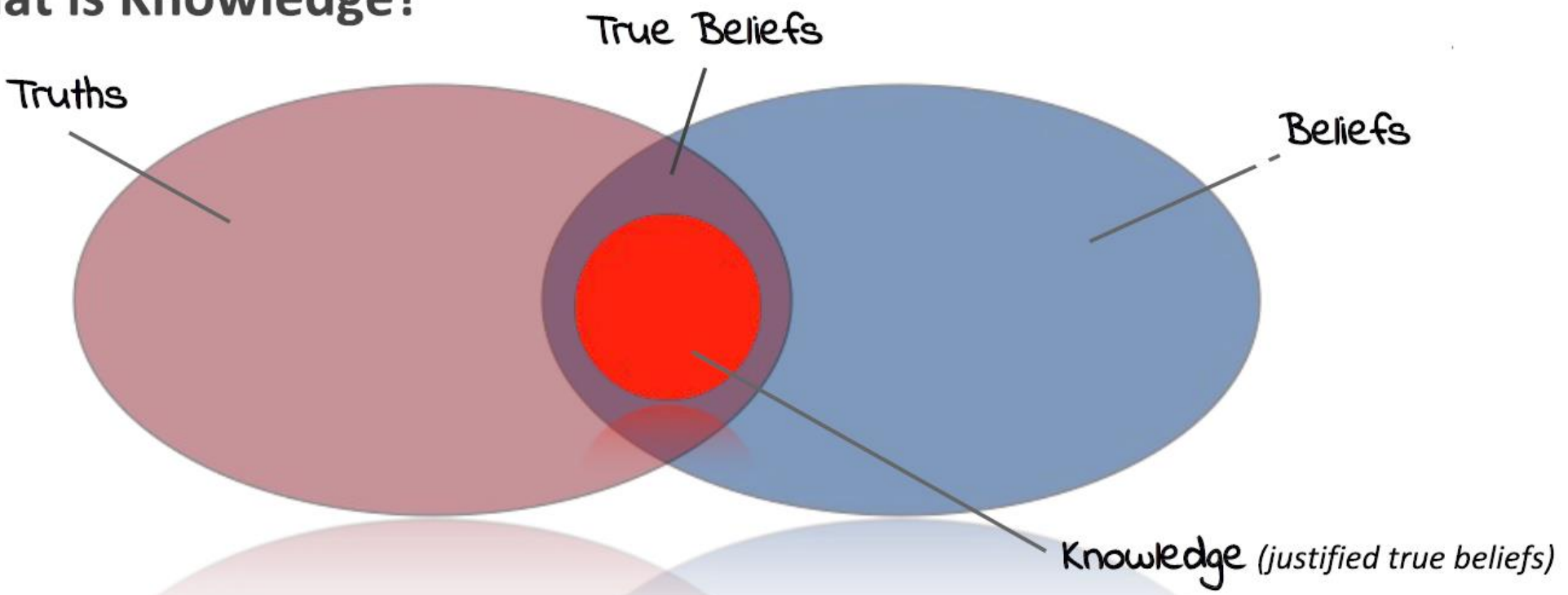
- Information is the data that has meaning
- Again, information can be useful or not be.

- Knowledge is the collection of information
- It has significance and needs to be useful.



DIKW Pyramid, *Ackoff 1989*

What is Knowledge?



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

How to represent Knowledge?

- **(Natural) Language can be a way to represent knowledge**
- **What is Language?**
  - **Language** is a system of conventional **spoken, manual, or written symbols** that combine to **convey meaning**, and by means of which human beings, as members of a social group and participants in its culture, **express** themselves.
  - One of the most important functions of language is **communication**.

Natural Language ??

I am a Linguist.

I love language more than most people.

1. Paraphrasing
2. Ambiguity

Formal Knowledge Representation

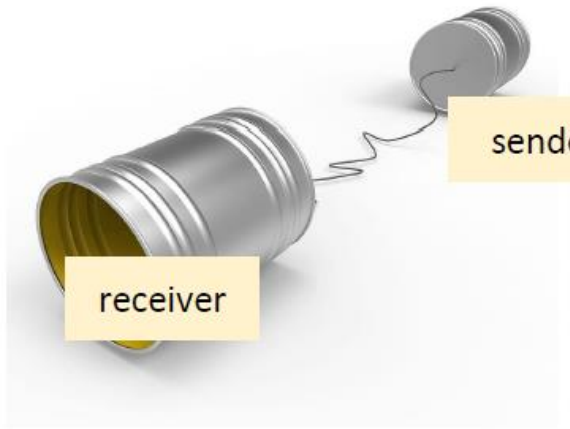
- **Formal Knowledge Representation**
  - is a field of **artificial intelligence (AI),**
  - which (unambiguously) captures the **semantics (meaning)** of **concepts, properties, relationships,** and **entities**
  - of specific **knowledge domains,** i.e., fields of interest or areas of concern,
  - as **structured data.**
- **Machines (computers)** must be able to **understand** formal knowledge representations.
- To "**understand**" a knowledge representation, the machine must be able to **interpret it correctly**.

What does it mean to “understand”?

- **Understanding** (in general) is the ability to grasp the meaning of information.
- **Information** is conveyed in a **message** using a specific **language** from a sender to a receiver.

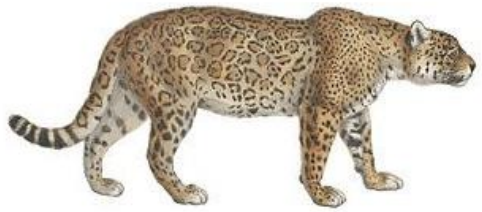
- **Information is understood** by the receiver of a message, if the receiver **interprets** the information **correctly**.

[2,3,4,5]



Experience

Concept



Context

symbolizes

refers to

Experience

Symbol

Object

stands for

"Jaguar"

Pragmatics

Understanding The Meaning

- **Correct Interpretation** depends on
  - Syntax,
  - Semantics,
  - Context,
  - Pragmatics, and
  - Experience.

SYNTAX

- =[greek] Arrangement, Ordering
- In **grammatics** syntax denotes the study of the **principles** and processes **by which sentences are constructed** in particular languages.
- In **formal languages**, syntax is just a set of rules, by which well formed expressions can be created from a fundamental set of symbols (alphabet).
- In **computer science**, syntax defines the normative structure of data.

SEMANTICS

- =[greek] pertains to the character, the study of meaning
- is part of the linguistics which focuses on **Sense and Meaning** of language or symbols of language.
- is the **study of interpretation of signs or symbols** as used by agents or communities within particular circumstances and **contexts**.
- Semantics asks, how sense and meaning of complex concepts can be derived from simple concepts based on the **rules of syntax**.
- The semantics of a message depends on **context** and **pragmatics**.

CONTEXT

- [lat.] contextus = interweaved
- denotes the **surrounding of a symbol** (concept) in an expression resp. its **relationship with surrounding expressions** (concepts) and further related elements,
- Context denotes **all elements of any sort of communication that define the interpretation of the communicated content**.

PRAGMATICS

- =[greek] action
- reflects the **intention by which the language is used** to communicate a message.
- In linguistics, pragmatics denotes the **study of applying language in different situations**.
- It also **denotes the intended purpose** of the speaker.
- Pragmatics studies **the ways in which context contributes to meaning**.

EXPERIENCE

- **Experience** considers all information that you have learned and put in context with the world you are living in.
- Experience in this sense is often referred to as **common sense knowledge** or **world knowledge**.

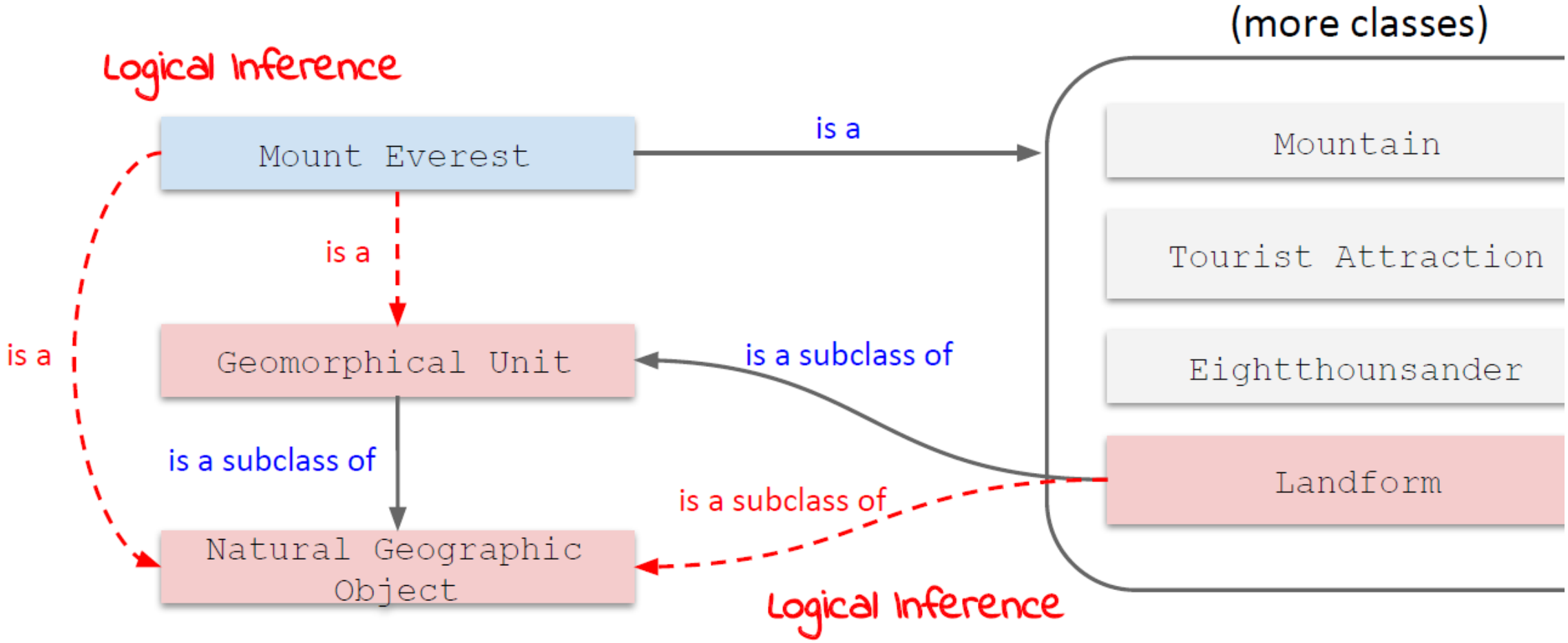
Successful Communication

- For **successful communication**,
  - information has to be correctly transmitted (**Syntax**)
  - the meaning (**Semantics**) of the transmitted information must be interpreted correctly (= **understanding**)
- **Understanding** depends on
  - the **context** of both sender and receiver and
  - the **pragmatics** of the sender
- (Personal) **experience** determines
  - how sender and receiver **interpret the semantics, context,** and **pragmatics** of a message, and thus its intended meaning

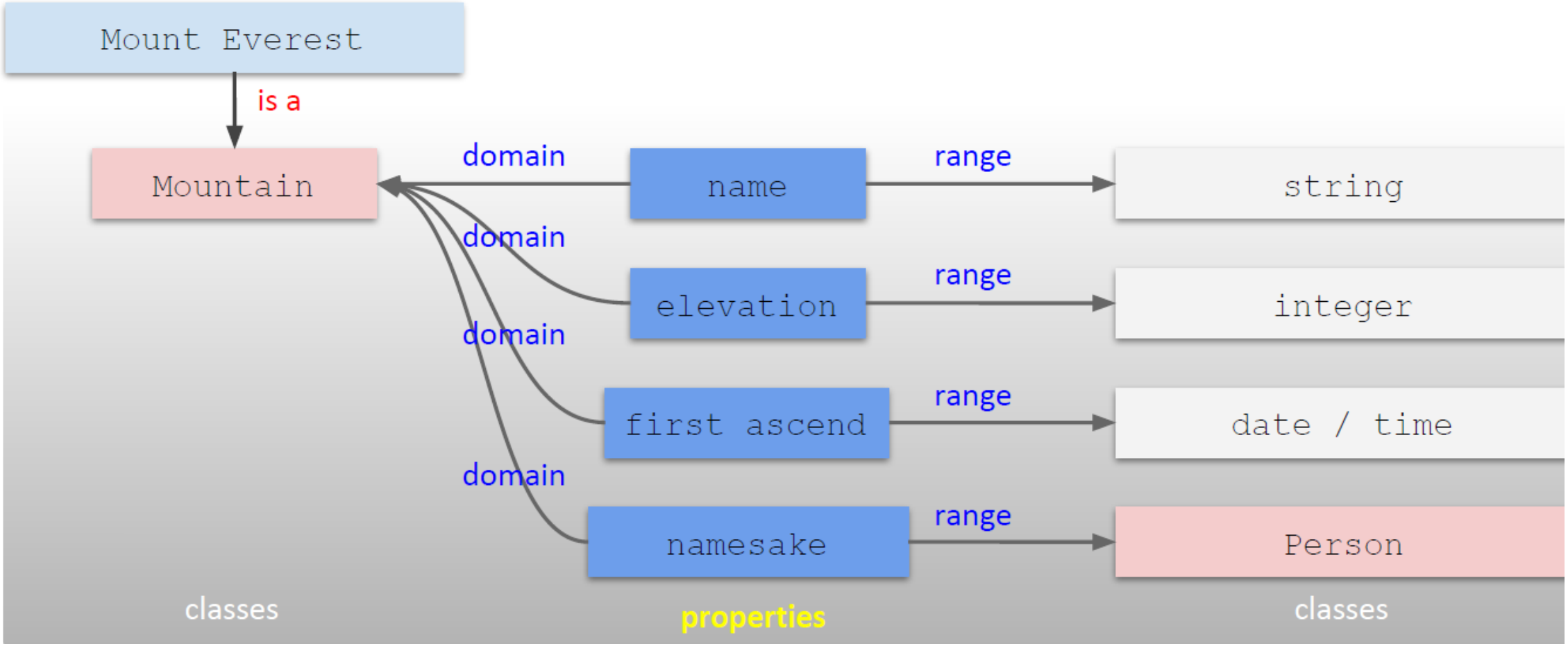
Semantic Web = A web of Data

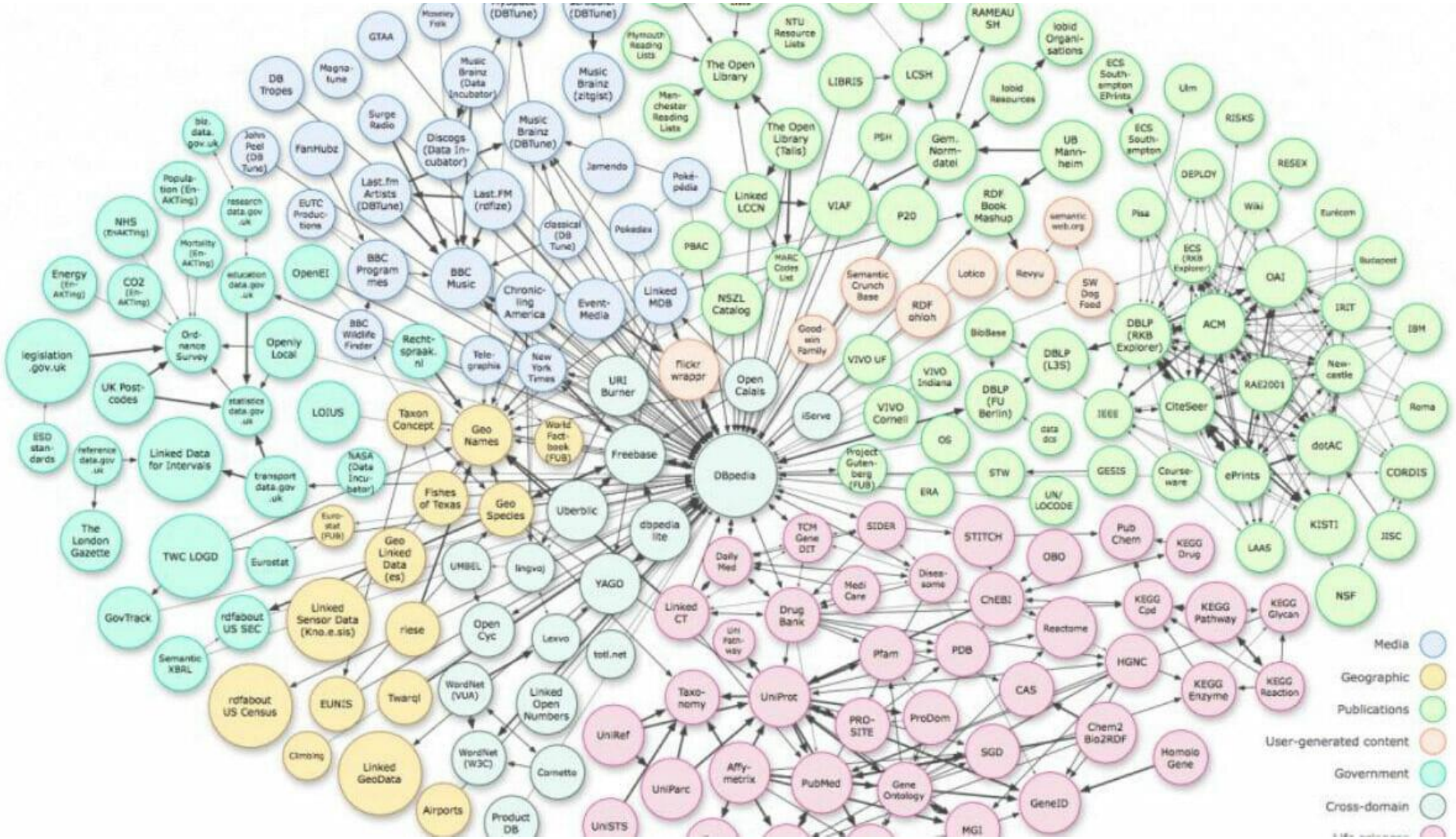
- The Semantic Web is an **Extension of the current Web.**
- The meaning of information (Semantics) is made explicit by **formal (structured) and standardized knowledge representations (Ontologies)**.
- Thereby it will be possible,
  - to **process** the meaning of information automatically,
  - to **relate** and **integrate** heterogeneous data,
  - to **deduce** implicit (not evident) information from existing (evident) information in an automated way.
- The Semantic Web is kind of a **global database** that contains a **universal network of semantic propositions**.

- The Meaning (Semantics) of information is expressed with the help of knowledge representations (**Ontologies**)



- The Meaning (Semantics) is expressed with the help of knowledge representations (**Ontologies**)




DBpedia
The Open Library
Linked MDB
NSZL Catalog
BBC Music
BBC Programmes
Music Brainz (DBTune)
Jamendo
Last.FM (rdfize)
Discogs (Data Incubator)
Chronicling America
Event-Media
Open Calais
Freebase
Geo Names
Taxon Concept
LOIUS
URI Burner
flickr wrappr
Uberblic
dbpedia lite
YAGO
Linked Sensor Data (Kno.e.sis)
Linked GeoData
Linked Open Numbers
WordNet (VUA)
WordNet (W3C)
Open Cyc
Lexvo
totl.net
Twarql
riese
EUNIS
rdfabout US Census
rdfabout US SEC
Semantic XBRL
GovTrack
TWC LOGD
The London Gazette
Eurostat
Eurostat (FUB)
Geo Linked Data (es)
UMBEL
lingvoj
Fishes of Texas
Geo Species
NASA (Data Incubator)
transport data.gov.uk
Linked Data for Intervals
reference data.gov.uk
ESD standards
legislation.gov.uk
UK Post-codes
Ordnance Survey
statistics data.gov.uk
Openly Local
education data.gov.uk
research data.gov.uk
OpenEI
BBC Wildlife Finder
Recht-spraak.nl
Tele-graphis
New York Times
Population (EnAKTing)
Mortality (EnAKTing)
Energy (EnAKTing)
CO2 (EnAKTing)
NHS (EnAKTing)
biz.data.gov.uk
John Peel (DB Tune)
FanHubz
Magna-tune
DB Tropes
Surge Radio
GTAA
Last.fm Artists (DBTune)
EUTC Productions
Pokedex
Poképedia
classical (DB Tune)
PBAC
Linked LCCN
VIAF
P20
RDF Book Mashup
Gem. Norm-datei
PSH
LCSH
LIBRIS
RAMEAU SH
Iobid Organisations
Iobid Resources
UB Mannheim
The Open Library (Talis)
NTU Resource Lists
Plymouth Reading Lists
Manchester Reading Lists
Musicbrainz (zitgist)
Music Brainz (Data Incubator)
Moseley Folk
(DBTune)
(DBTune)
Lotico
Revyu
SW Dog Food
semantic web.org
Semantic Crunch Base
RDF ohloh
Good-win Family
BioBase
DBLP (L3S)
DBLP (FU Berlin)
VIVO UF
VIVO Indiana
VIVO Cornell
iServe
OS
ERA
STW
GESIS
Course-ware
ePrints
dotAC
CORDIS
KISTI
JISC
NSF
LAAS
IEEE
CiteSeer
RAE2001
ACM
DBLP (RKB Explorer)
OAI
IRIT
IBM
Roma
Newcastle
Budapest
Eurécom
Wiki
DEPLOY
RISKS
RESEX
Pisa
ECS (RKB Explorer)
ECS Southampton
ECS Southampton EPrints
Ulm
data dcs
UN/LOCODE
Project Gutenberg (FUB)
SIDER
STITCH
OBO
Pub Chem
KEGG Drug
KEGG Cpd
KEGG Pathway
KEGG Glycan
KEGG Enzyme
KEGG Reaction
Reactome
ChEBI
HGNC
CAS
Chem2 Bio2RDF
Homolo Gene
GeneID
MGI
Gene Ontology
PubMed
SGD
PDB
Pfam
UniProt
Taxo-nomy
UniRef
UniParc
UniSTS
Affy-metrix
PRO-SITE
ProDom
Drug Bank
Linked CT
Medi Care
Disea-some
Daily Med
TCM Gene DIT
UH Path-way
Media
Geographic
Publications
User-generated content
Government
Cross-domain

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

Most apps use only a subset of the stack

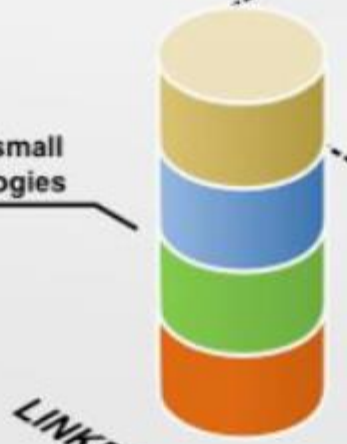
Querying allows fine-grained data access

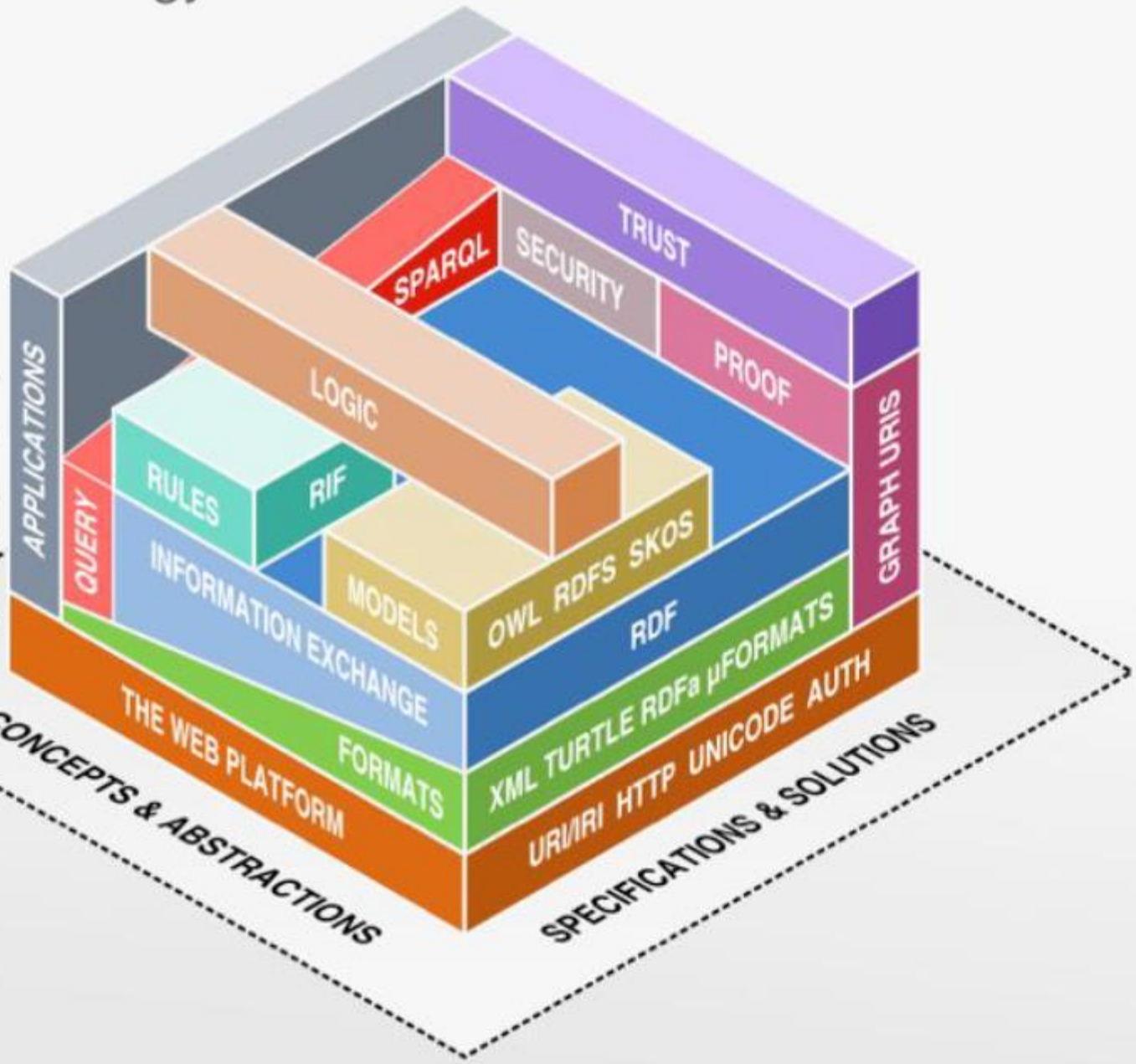
Standardized information exchange is key

Formats are necessary, but not too important

The Semantic Web is based on the Web

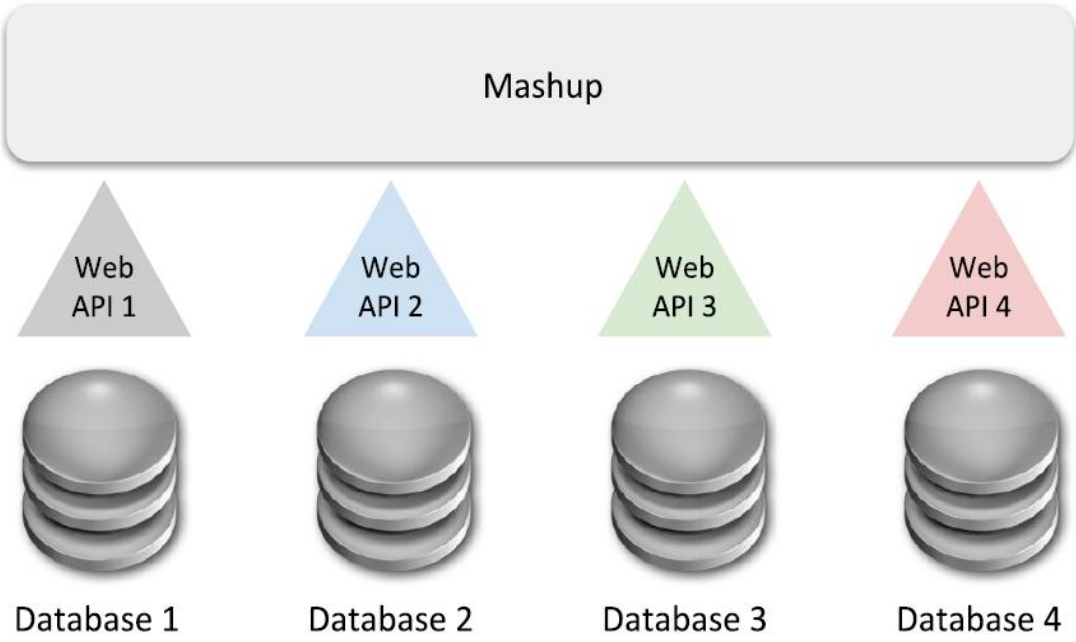
Linked Data uses a small selection of technologies





Data Access in Traditional Web

- There is a number of different (proprietary) **Web APIs**, data exchange formats, and **Mashups** on top of that.



Data Access in Semantic Web

- **Apply Linked Data technology**
  - to publish (structured) data on the Web
  - to draw connections from one data source to data from other data sources