Ontology and Knowledge Representation

Introductory lecture
Seminar, July 10-14, 2023

Stefan Schulz

Medical University of Graz (Austria)

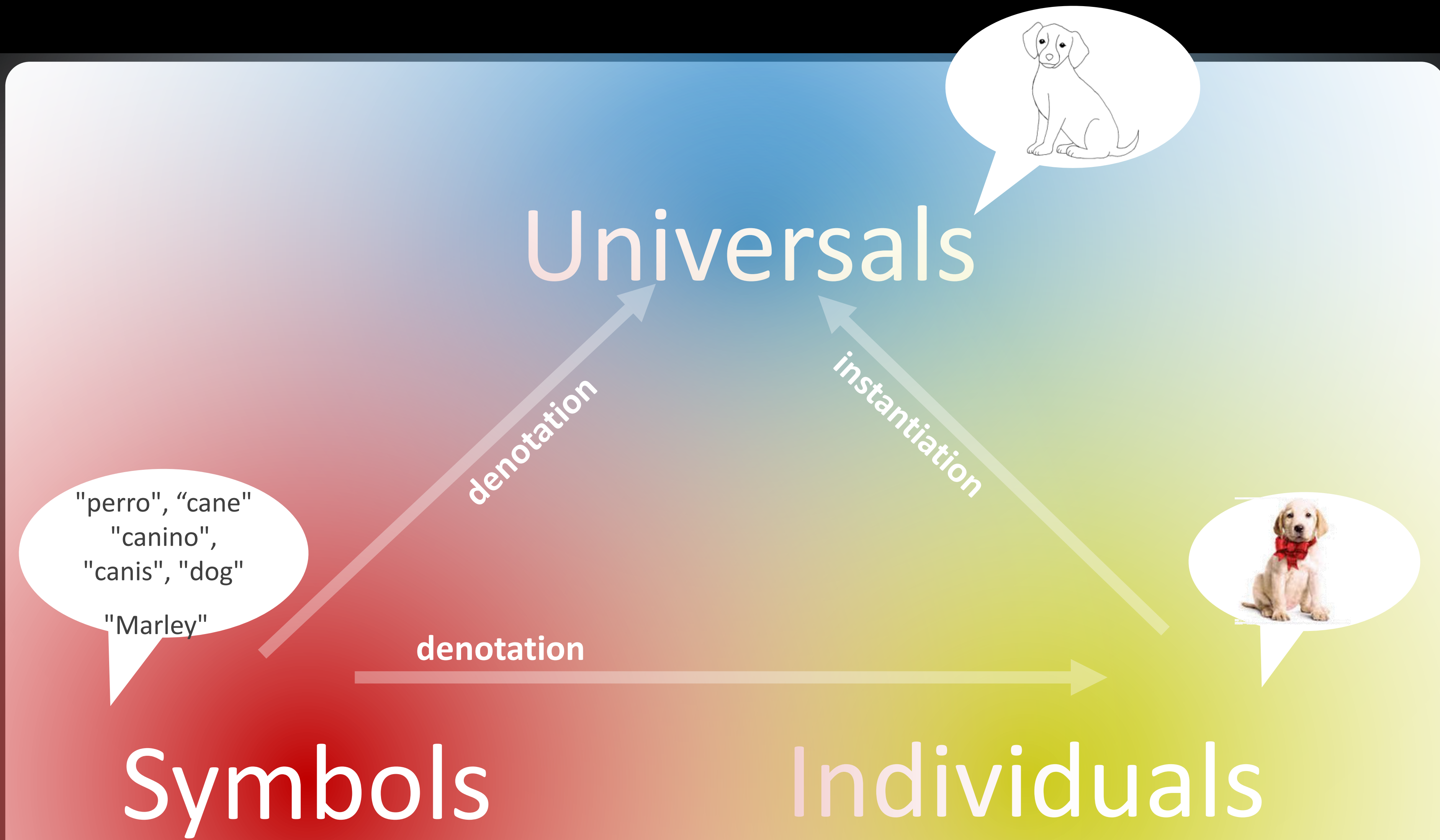
Averbis GmbH, Freiburg (Germany)

stefan.schulz@medunigraz.at

Topics

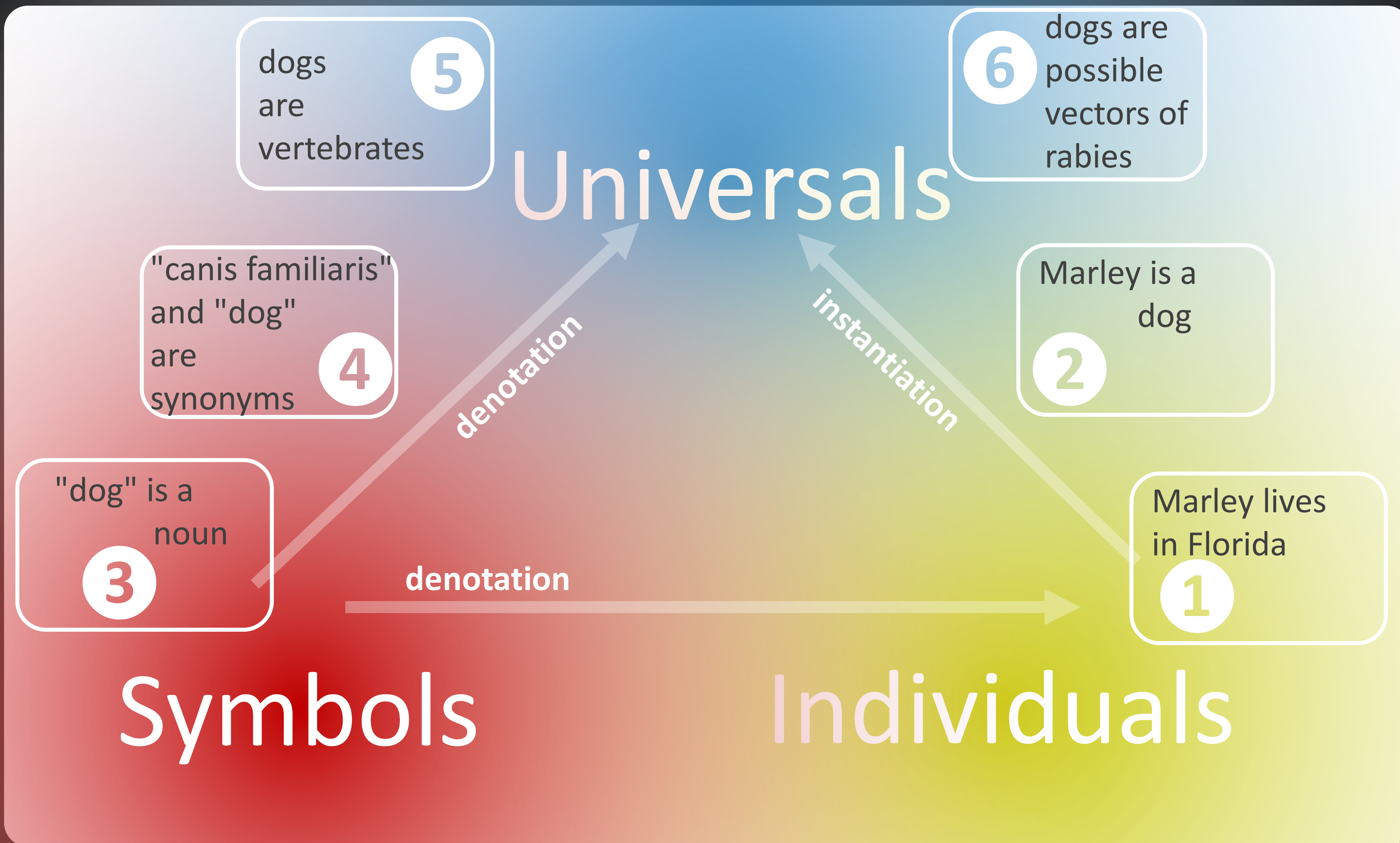
- Context: representation of knowledge in natural sciences / engineering
  - What are the types of knowledge to distinguish
  - How are they connected
  - What are the most important resources?
- Fundamental theories:
  - Ontology: theory of being
  - Epistemology: theory of knowledge
  - Semiotics: theory of signs
  - Semantics: theory of linguistic signs

Knowledge map



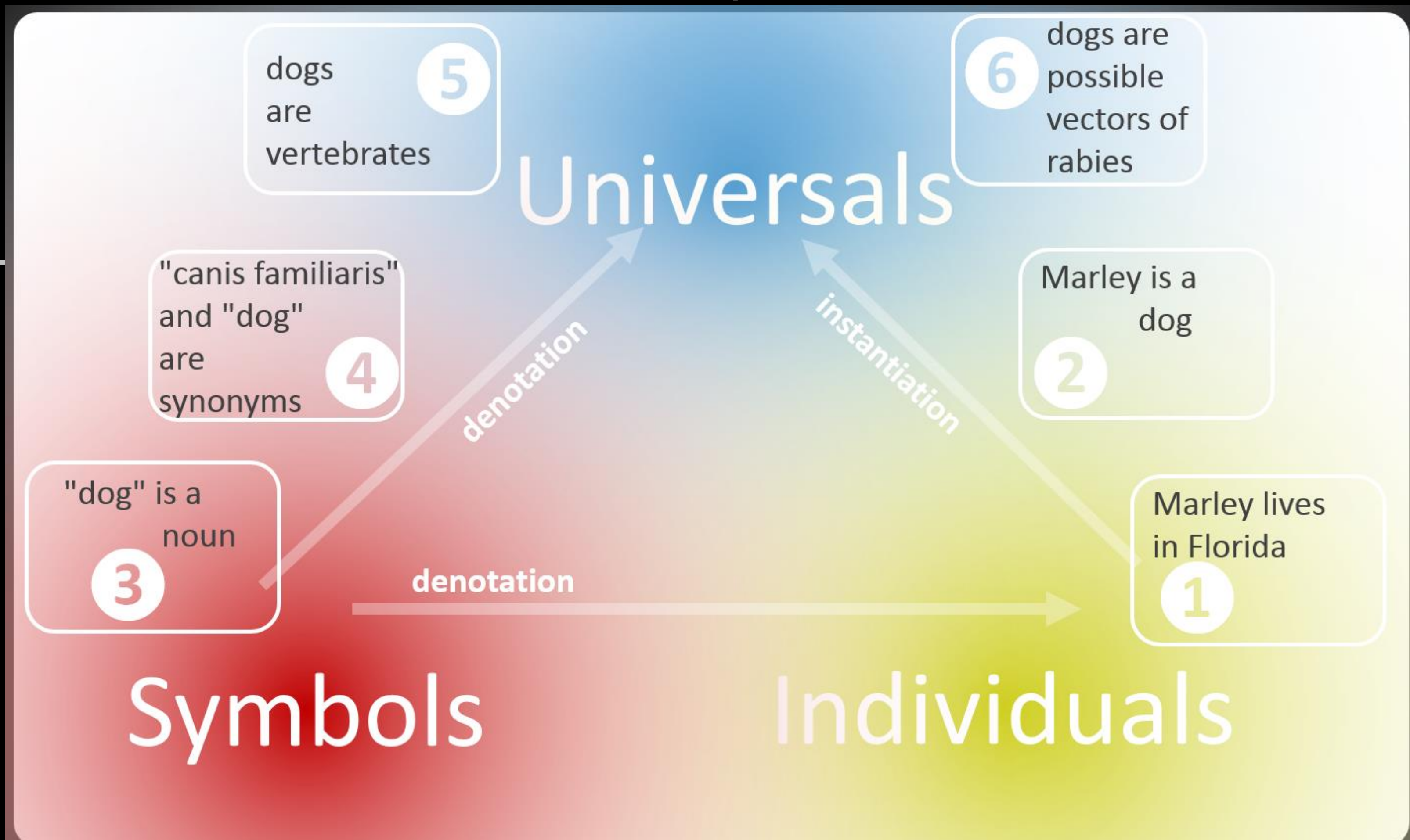
C. K. Ogden and I. A. Richards (1923) The Meaning of Meaning

Knowledge map



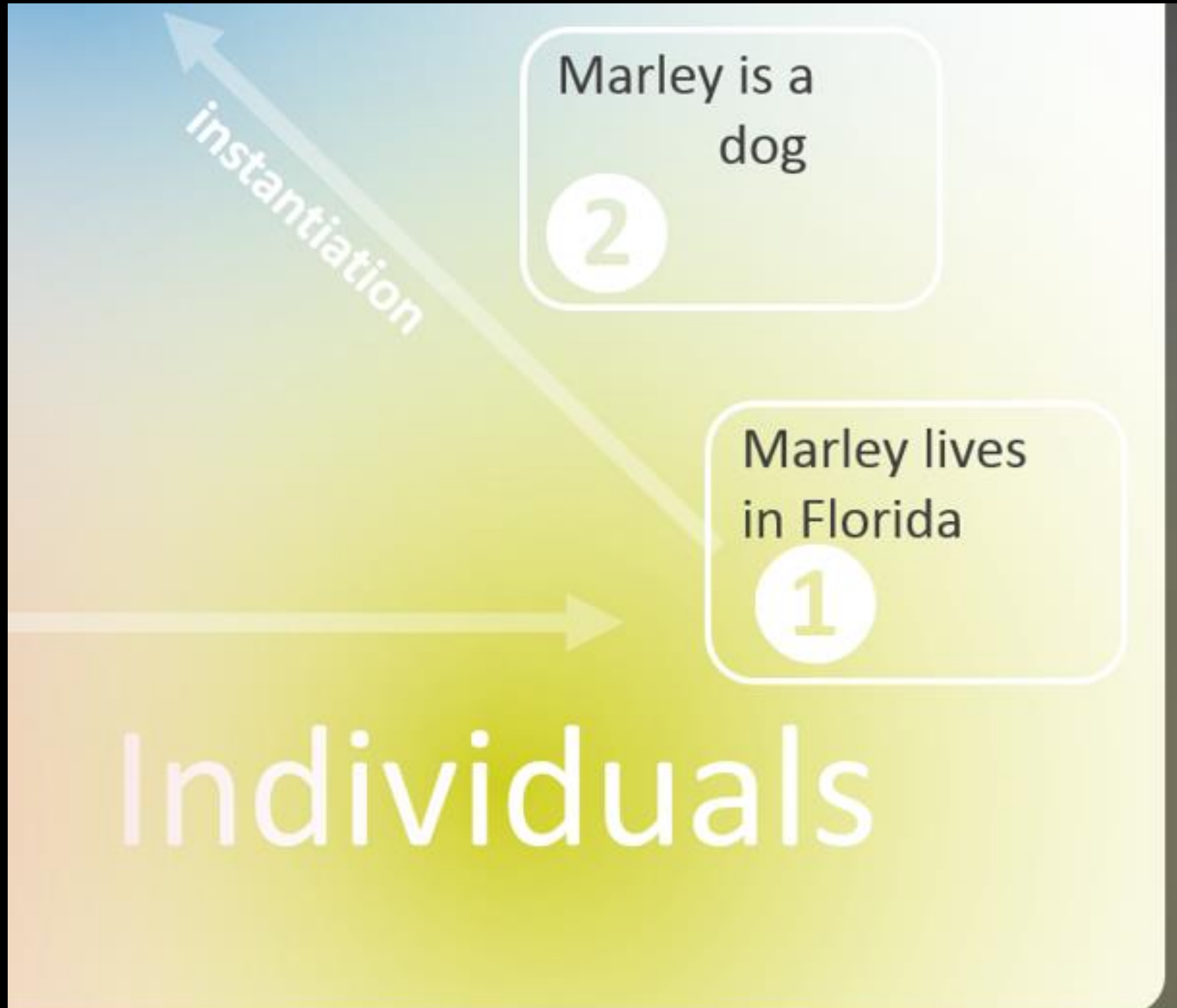
Ontological knowledge:
Axioms that are universally true

Contingent knowledge:
typical, likely, possible



Linguistic knowledge:
Statements about properties and meaning of signs of language

Factual knowledge:
Statements about concrete entities and their relationships



Factual knowledge:
Statements about concrete entities and their relationships

Statements at instance level

- RDF triples: data structures in RDF (Resource Description Framework) in the form of subject, predicate, object.
  - Subject: the entity to which the triple refers,
  - Predicate: relationship between subject and object
  - Object represents the value associated with this relationship.
  - Different syntaxes, e.g. TURTLE: https://www.w3.org/TR/turtle/ more readable
- Example:

Subject (Resource)	Predicate (Property)	Object (Value)
ex:Marley	ex:livesIn	ex:Montréal
ex:Marley	ex:isOwnedBy	ex:Pierre
ex:Marley	ex:owner	ex:Pierre
ex:Marley	ex:bornOn	ex:2019-02-03(xsd:date)
ex:Marley	ex:bornIn	ex:Brussels
ex:Montréal	ex:capitalOf	ex:Québec
ex:Québec	ex:partOf	ex:Canada
ex:Brussels	ex:capitalOf	ex:Belgium

Exercice

- DBpedia is an online knowledge base that extracts structured information from Wikipedia
- The extracted data includes information about people, places, events, works, etc.
- Choose a term from Dbpedia, e.g. ex. https://dbpedia.org/page/Montreal
- Identify Triples
- A-Box : representation of particulars
- T-Box: representation of universals

Statements between particulars and universals

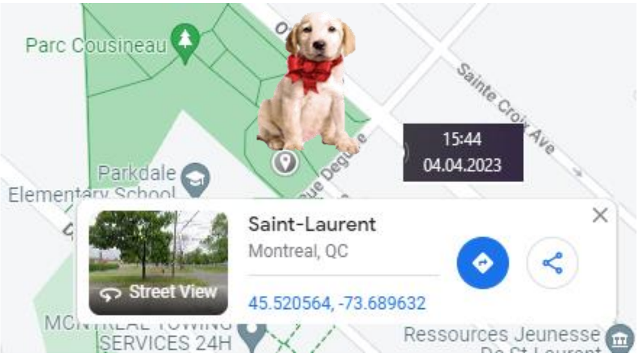
- Instantiation relationship "instance-of"
- Synonyms: "rdf:type", "a"
- Examples:

Subject (Resource)	Predicate (Property)	Object (Value)
ex:Marley	**a**	**ex:Dog**
ex:Pierre	**a**	**ex:Human**
ex:Canada	**a**	**ex:Country**
ex:Montréal	**a**	**ex:City**

- Dbpedia: https://dbpedia.org/page/Brussels rdf:Type https://dbpedia.org/page/Country

Statements between universals

- Type ≈ Universal ≈ Class ≈ Concept (≈ Kind ≈ Category):
  - which can be instantiated (⍰ 500 M dogs…)
  - undefined in place and time
- Particular: that which cannot be instantiated
  - There is only one Marley
  - Specified in terms of place and time

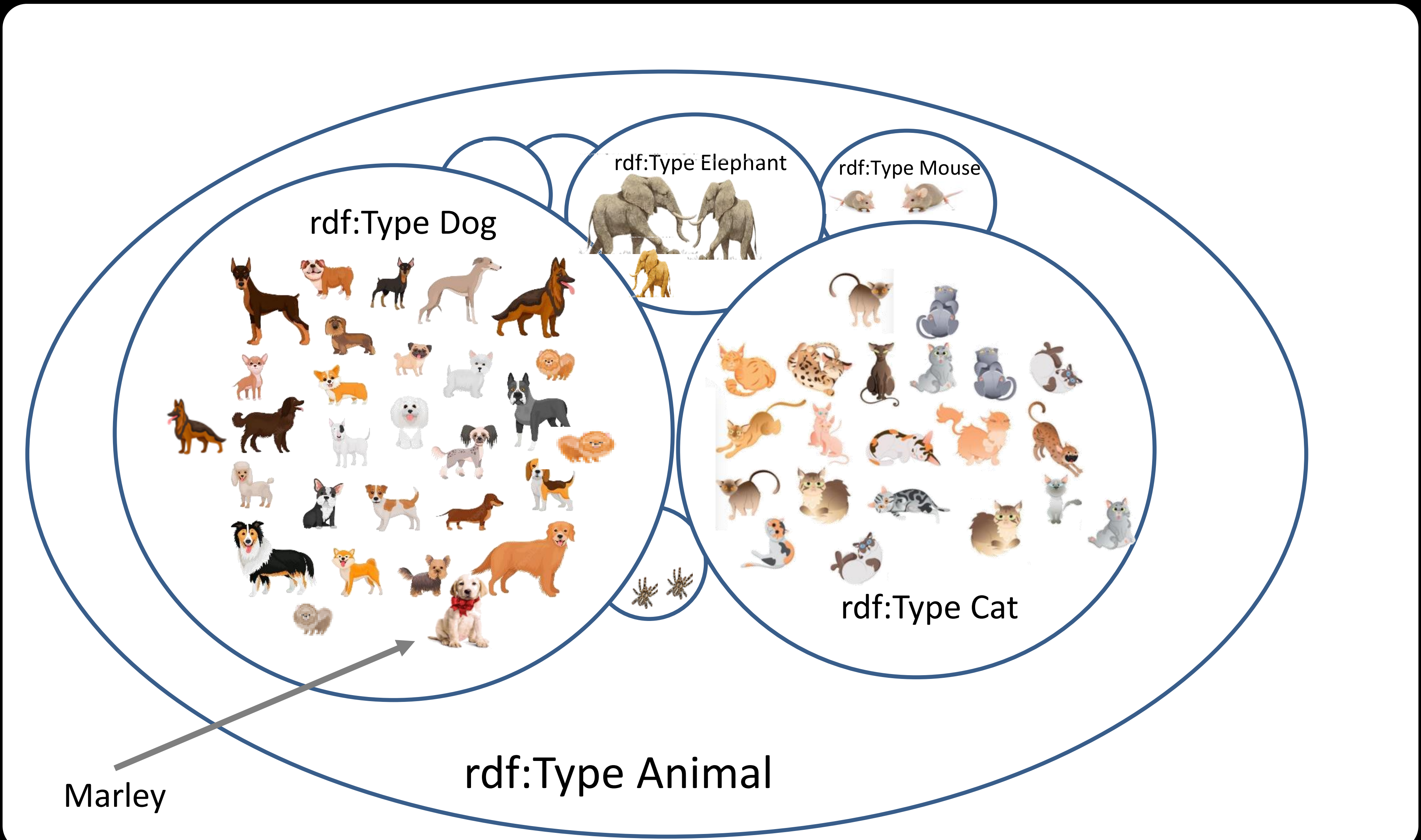


- Examples:

Subject (Resource)	Predicate (Property)	Object (Value)
ex:Dog	**rdfs:subClassOf:**	**ex:Vertebrate**
ex:Vertebrate	**rdfs:subClassOf:**	**ex:Animal**
ex:Capital	**rdfs:subClassOf:**	**ex:City**

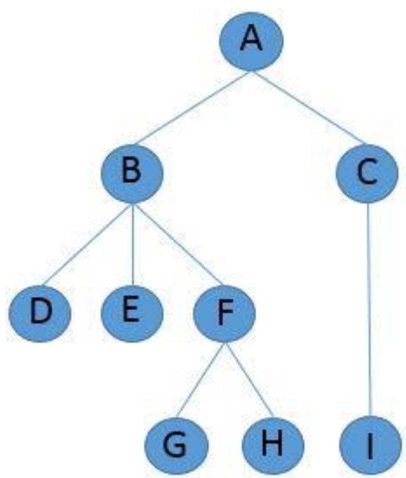
- How are instances and classes related?

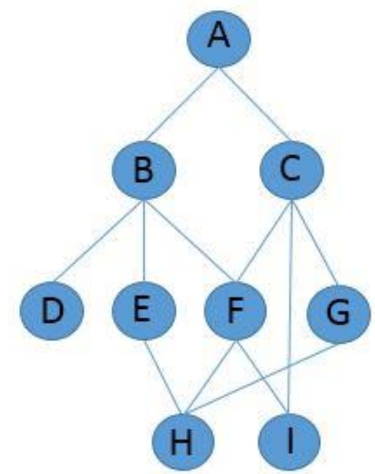
Universals / particulars
subclass / instance



Taxonomies

- if ***A*** is a subclass of ***B*** and ***i*** an instance de ***A***, then ***i*** is also an instance of ***B***
- If ***A*** is a subclass of ***B*** and ***B*** is a subclass of ***C***, then ***A*** is a subclass of ***C***
- Taxonomy



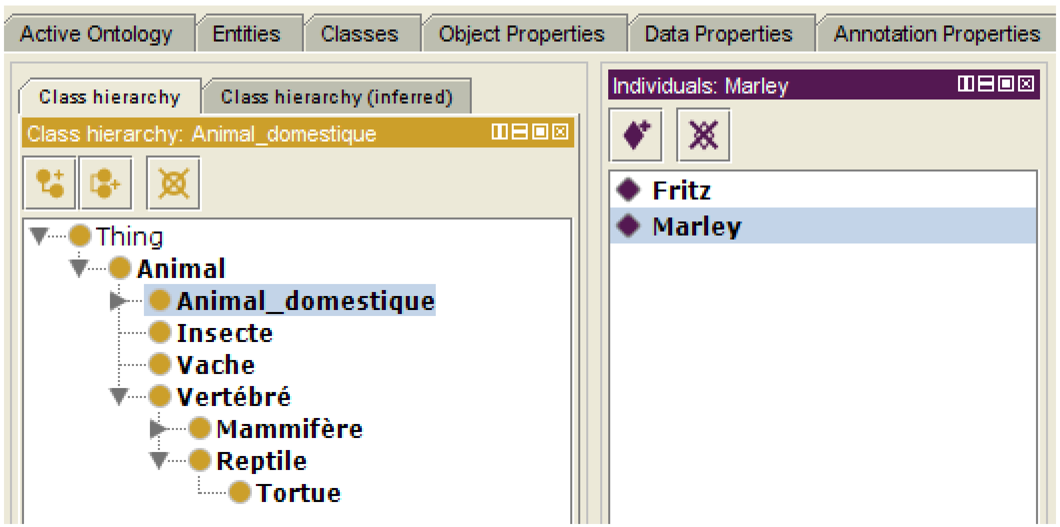


Explore:
https://www.onezoom.org/life

simple hierarchy / multiple hierarchy

Let's build classes and instances in Protégé

- https://protege.stanford.edu/

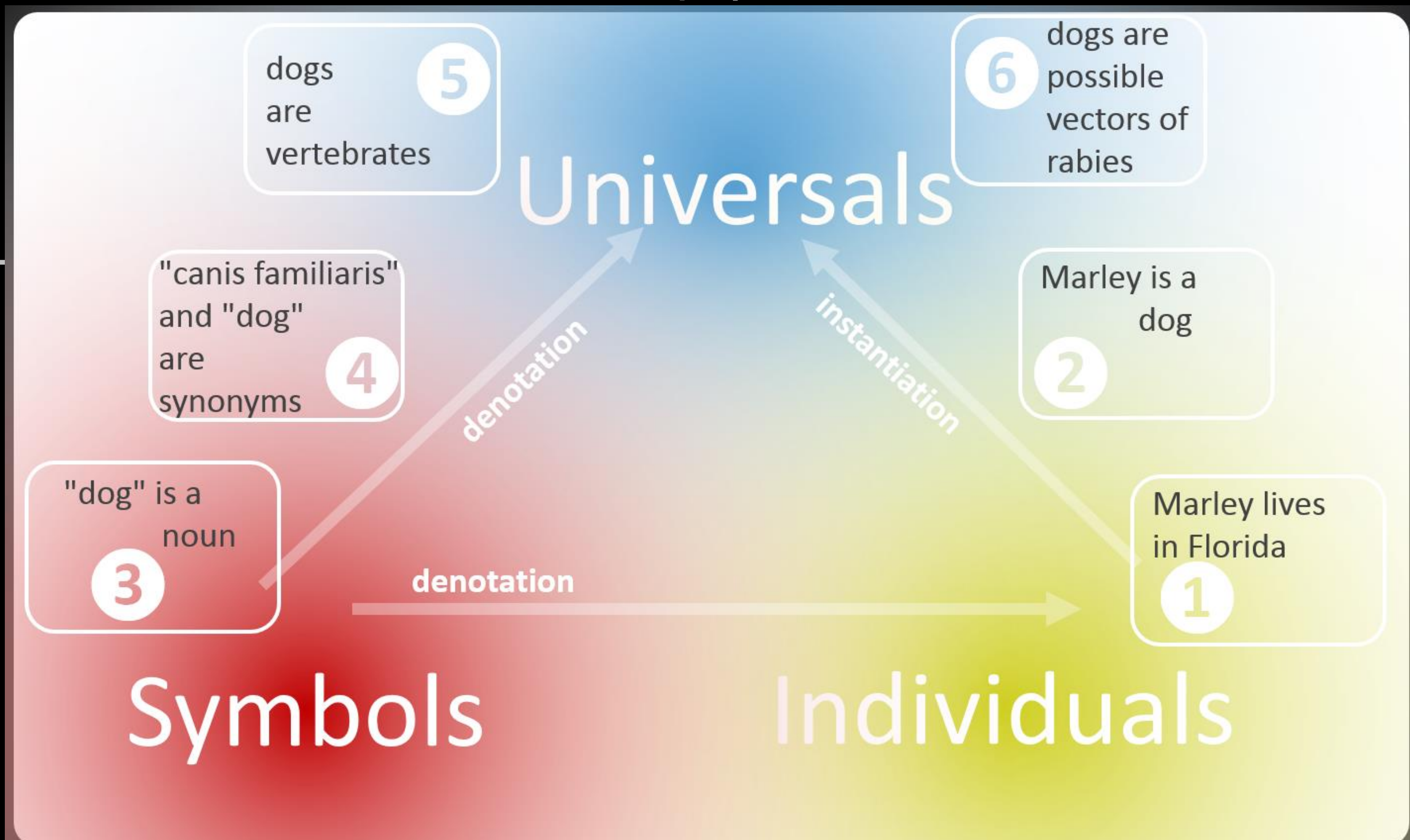


Protégé

- Taxonomy (with its formal semantics): "backbone" of ontologies
- Formal ontologies: knowledge organization systems (KOS) which are based on logic and describe classes of entities of reality (objects, places, processes, qualities, …) by the entity properties they have in commmon

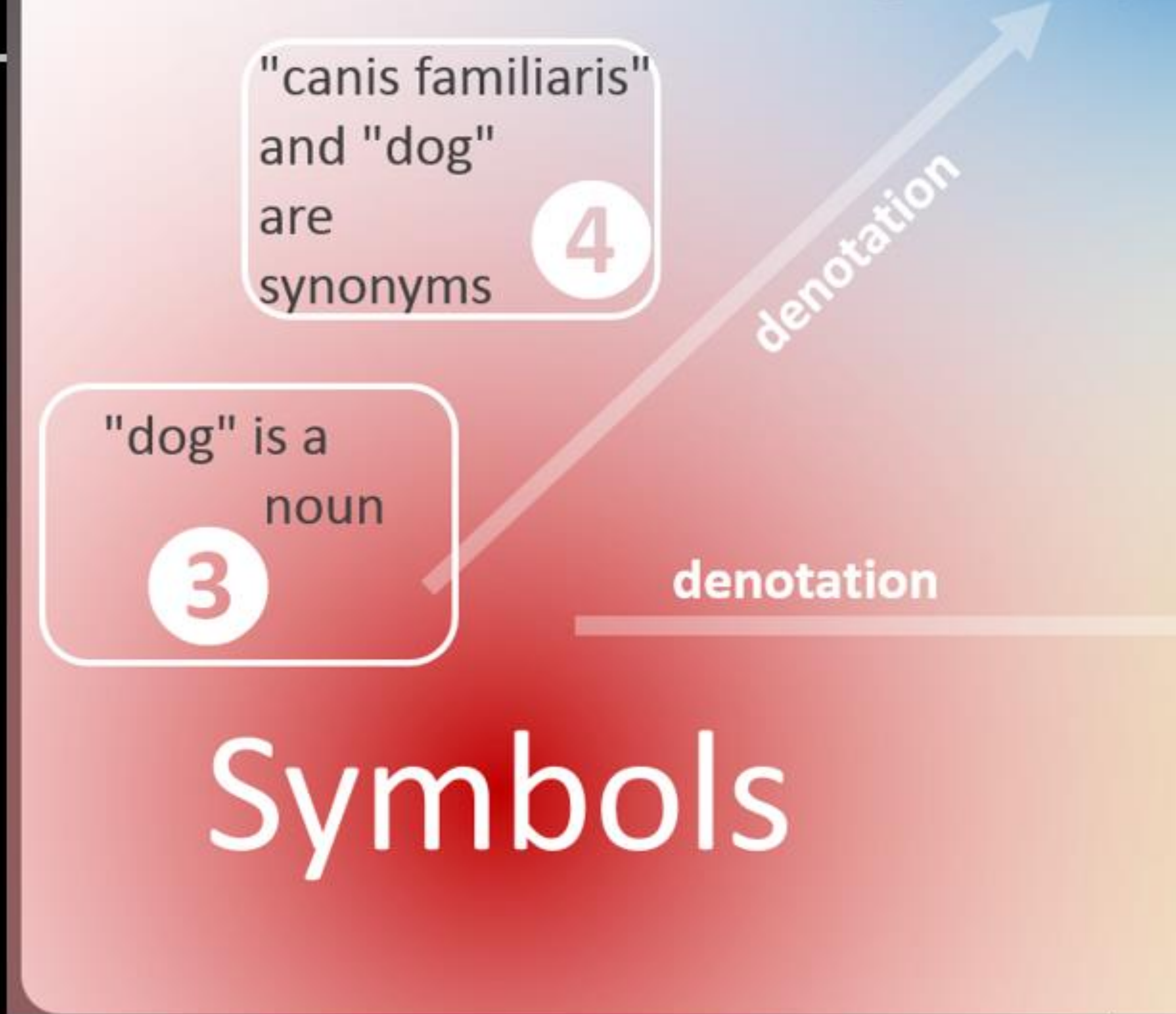
Ontological knowledge:
Axioms that are universally true

Contingent knowledge:
typical, likely, possible



Linguistic knowledge:
Statements about properties and meaning of signs of language

Factual knowledge:
Statements about concrete entities and their relationships



Linguistic knowledge:
Statements about properties and meaning of signs of language

SKOS / Linked Data representations

Subject (Resource)	Predicate (Property)	Object (Value)
`ex:Dog`	`rdf:type`	`skos:Concept`
`ex:Dog`	`skos:prefLabel`	`"dog"@en;`
`ex:Dog`	`skos:prefLabel`	`"chien"@fr;`
`ex:Animal`	`rdf:type`	`skos:Concept`
`ex:Animal`	`skos:broader`	`ex:Dog`
`wr:dog`	`lemon:sense`	`wr:dog-English-Noun-1`
`wr:dog`	`lemon:sense`	`wr:dog-English-Verb-1`
`wr:dog-English-Noun-1`	`wt:hasPoS`	`wt:Noun`

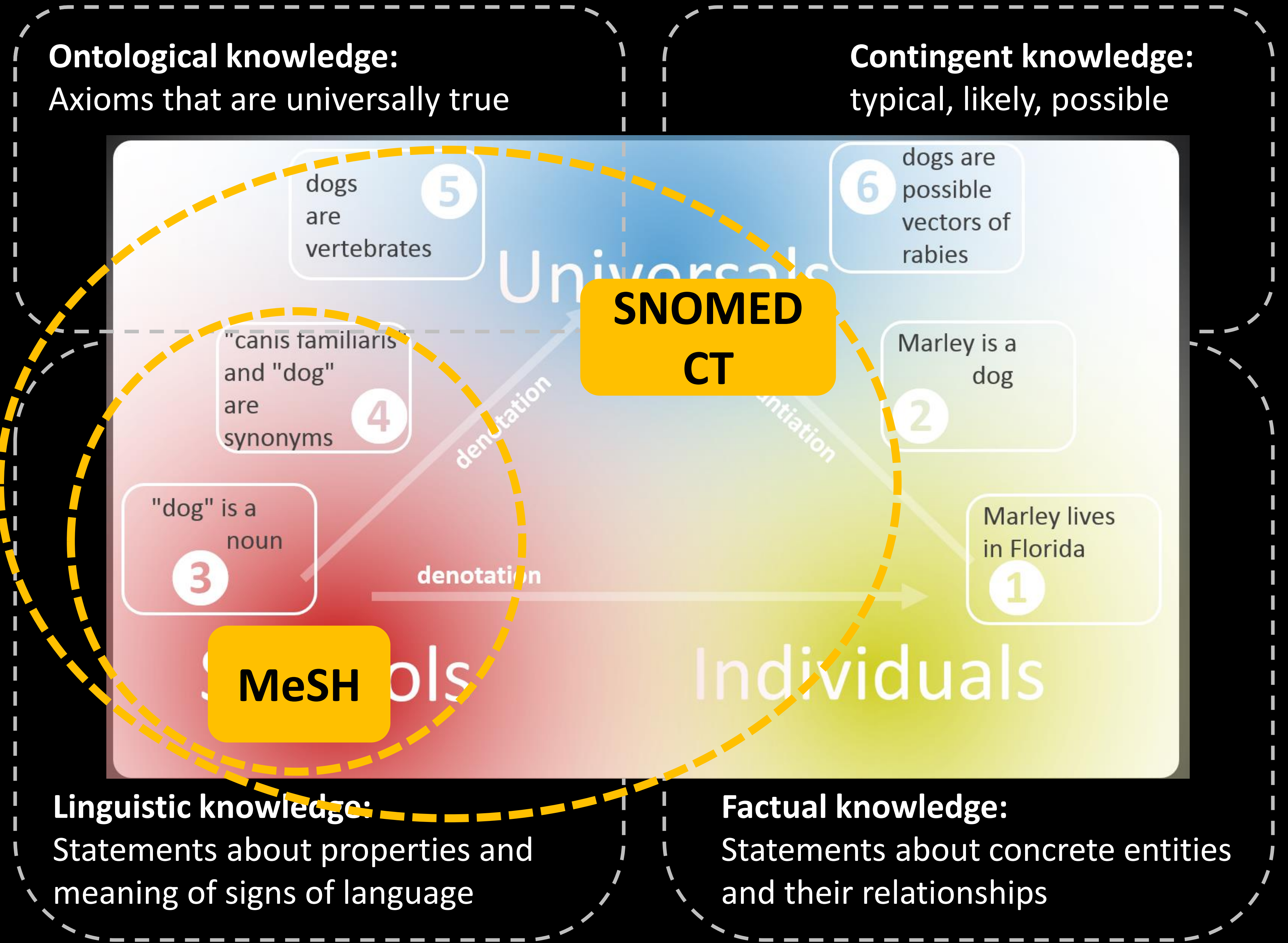
Syntax TURTLE : https://www.w3.org/TR/turtle/
Wiktionary: http://wiki.dbpedia.org/wiktionary-rdf-extraction

What matters when representing linguistic/cognitive entities?

- Different Symbols
  - Names like "Marley" or "Montreal" denote individuals
  - Terms like "dog" or "city" denote universals
- Horizontal relations between symbols:
  - Synonymy: the same symbol denotes different things
  - Homonymy: the same thing is denoted by different symbols
  - Translation: synonymy between expressions in two natural languages
  - Synonymy is often "near-synonymy"
- Vertical (hierarchical) relationships between symbols :
  - symbol A has a broader meaning than symbol B
  - Symbol B has a narrower meaning than symbol A
- Many of the relations describe grammatical and morphological features, e.g.
  - symbol A has a broader meaning than symbol B
  - Symbol B has a narrower meaning than symbol A

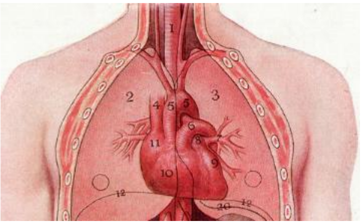
Which resources describe relations?

- Dictionary: contains words and terms with definitions and other characteristics (grammar, morphology, translations, synonyms)
- Hierarchical thesaurus / semantic dictionary: dictionary with richer relations, particularly with hierarchical relations (broader… narrower)
- Medical Thesaurus: MeSH (Medical Subject Headings)
- General Thesaurus of the English Language: WordNet
- Important: dictionaries and thesauri do not provision formal descriptions like ontologies.
- Comparison: SNOMED CT - MeSH

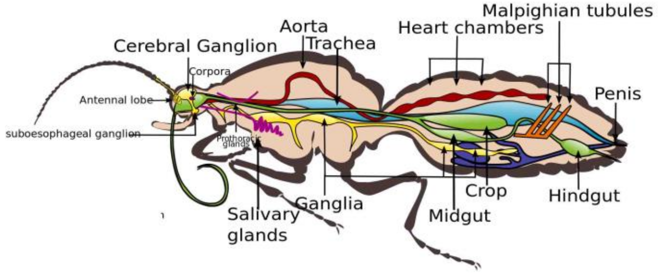

Ontological knowledge:
Axioms that are universally true
Contingent knowledge:
typical, likely, possible
dogs are vertebrates
5
dogs are possible vectors of rabies
6
Universals
SNOMED CT
"canis familiaris" and "dog" are synonyms
4
Marley is a dog
2
denotation
"dog" is a noun
3
denotation
Marley lives in Florida
1
MeSH
Individuals
Linguistic knowledge:
Statements about properties and meaning of signs of language
Factual knowledge:
Statements about concrete entities and their relationships

The importance of labels

- Anatomic Structure, System, or Substance
  - Organ
    - Heart

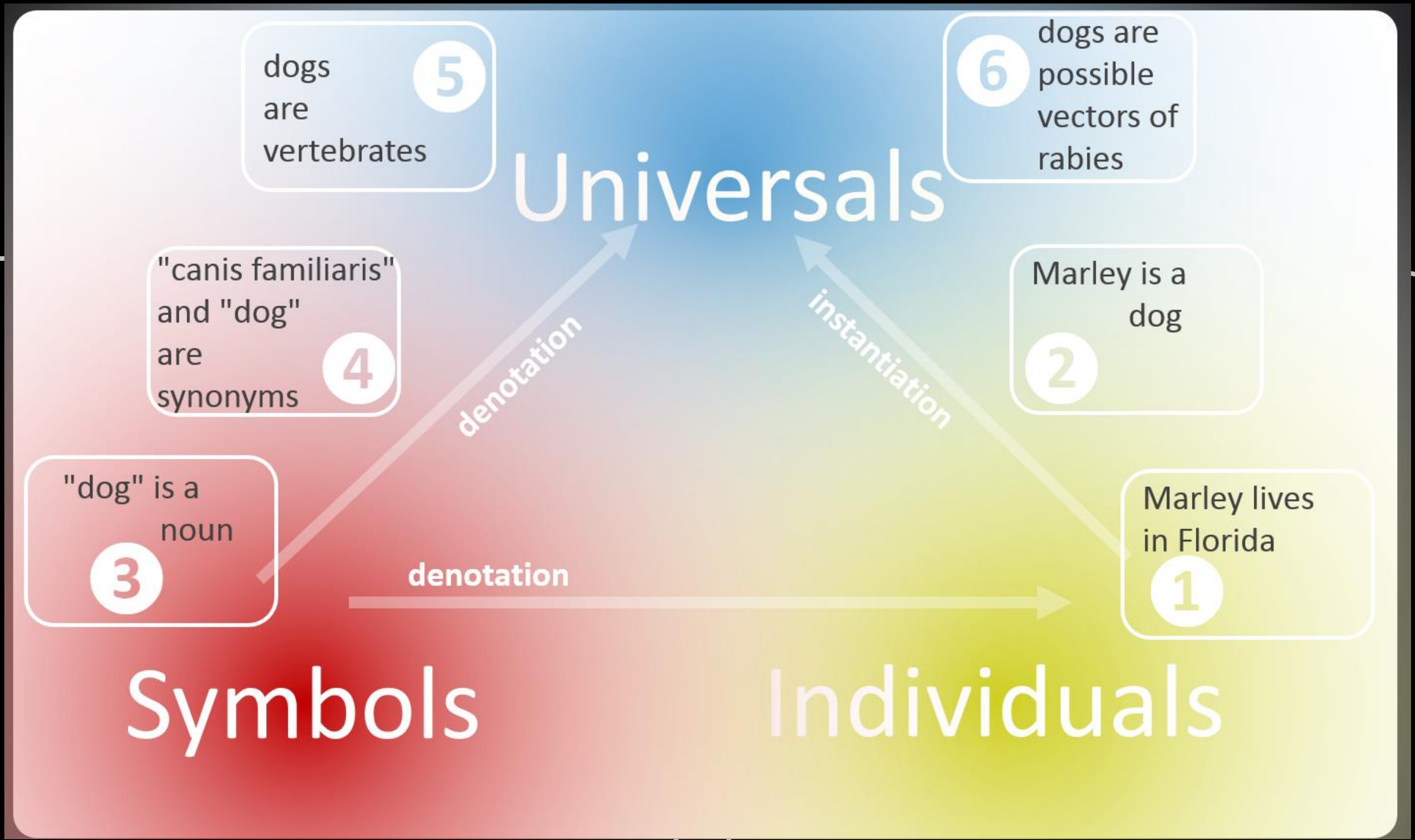


- anatomical entity
  - vessel
    - dorsal vessel
      - heart



Ontological knowledge:
Axioms that are universally true

Contingent knowledge:
typical, likely, possible



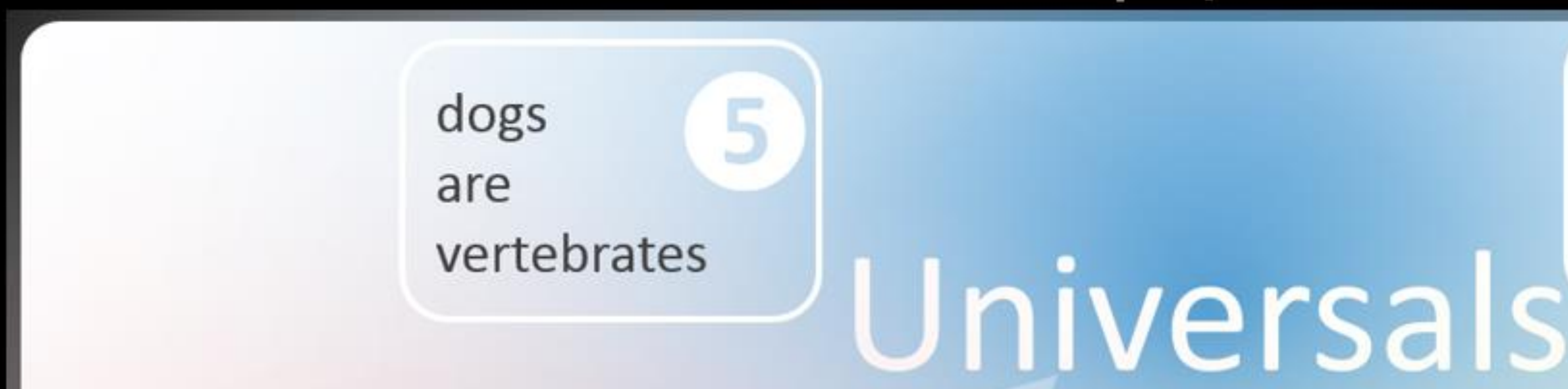
Linguistic knowledge:
Statements about properties and meaning of signs of language

Factual knowledge:
Statements about concrete entities and their relationships

6

Ontological knowledge:

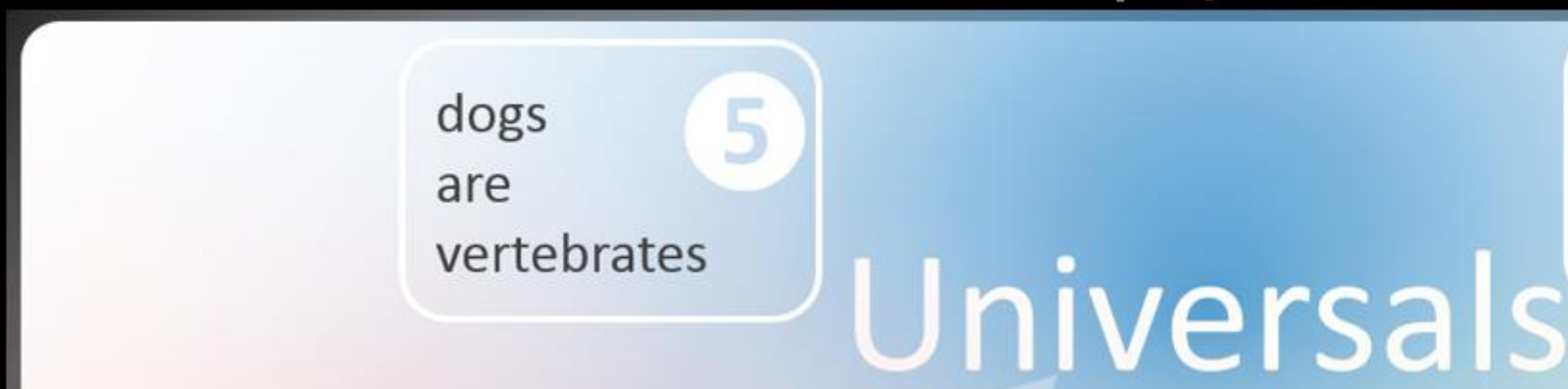
Axioms that are universally true



6

Ontological knowledge:

Axioms that are universally true



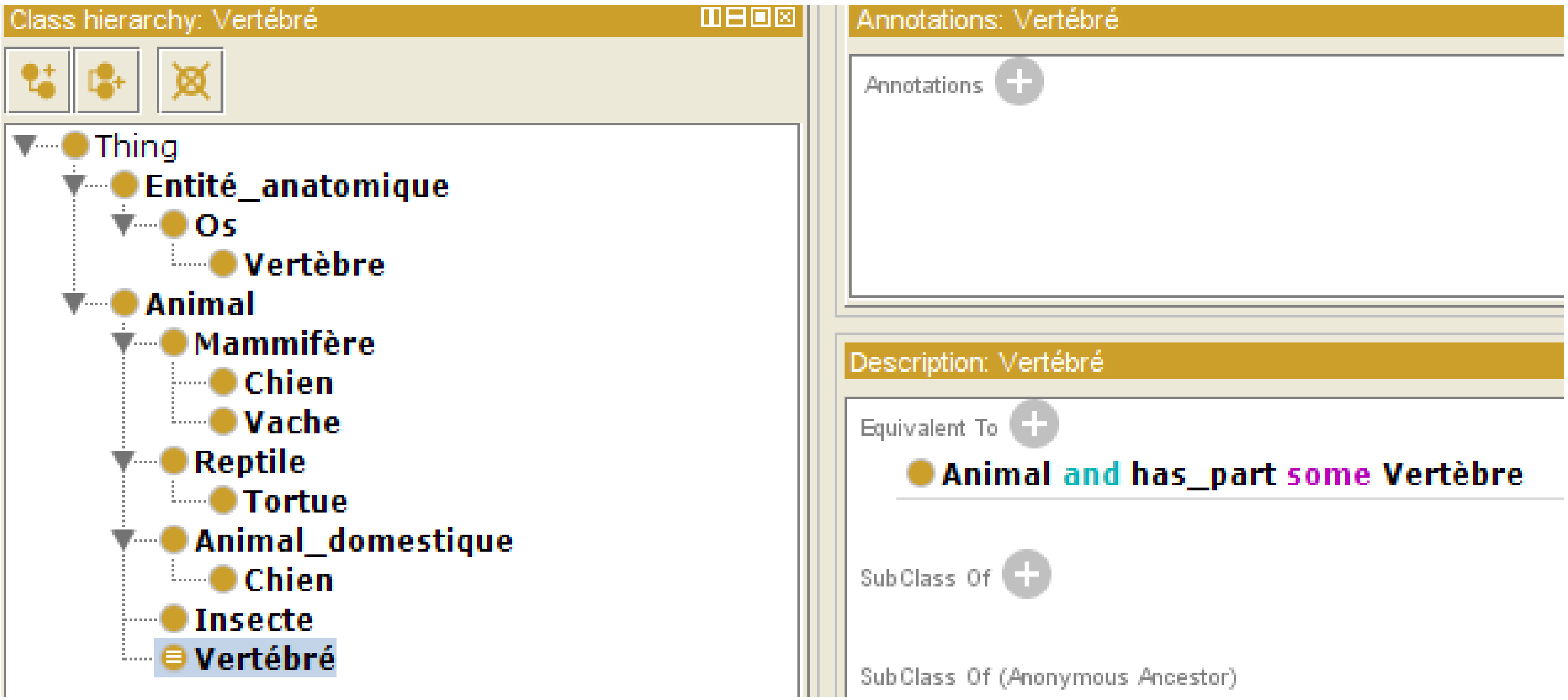
Representation in OWL

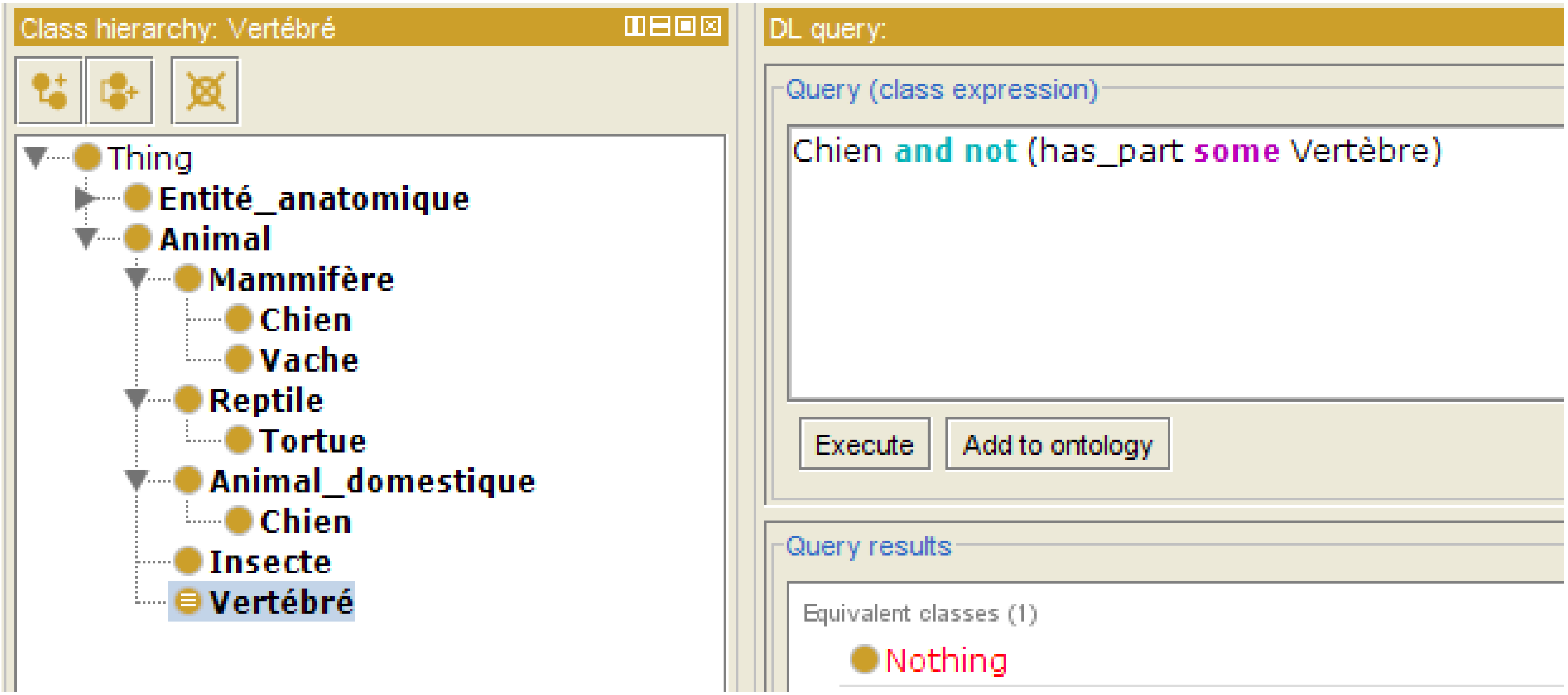
```
Dog subclassOf Vertebrate
Vertebrate subclassOf Animal
Vertebra subclassOf Bone
Vertebrate equivalentTo Animal and
          has-part some Vertebra
```

computable inference
(e.g. HermiT or Fact++
OWL reasoner)

There is no dog that has no bones

OWL Manchester Syntax: https://www.w3.org/TR/owl2-manchester-syntax/
HermiT reasoner: http://www.hermit-reasoner.com/
Fact++ reasoner: http://owl.man.ac.uk/factplusplus/


Class hierarchy: Vertébré
Thing
Entité_anatomique
Os
Vertèbre
Animal
Mammifère
Chien
Vache
Reptile
Tortue
Animal_domestique
Chien
Insecte
Vertébré
Annotations: Vertébré
Annotations
Description: Vertébré
Equivalent To
Animal and has_part some Vertèbre
SubClass Of
SubClass Of (Anonymous Ancestor)


Class hierarchy: Vertébré
Thing
Entité_anatomique
Animal
Mammifère
Chien
Vache
Reptile
Tortue
Animal_domestique
Chien
Insecte
Vertébré
DL query:
Query (class expression)
Chien and not (has_part some Vertèbre)
Execute
Add to ontology
Query results
Equivalent classes (1)
Nothing

Characteristics of formal ontologies

- A class definition or description in an ontology always indicates what must necessarily be true for all instances of that class
- Equivalent statements:
  - All cells have a nucleus
  - There is no cell without a nucleus
- Strong restriction of what can be expressed by ontologies. However, it is the basis of logic-based reasoning

e

Contingent knowledge:
typical, likely, possible

Contingent knowledge

Alan Rector (2008):

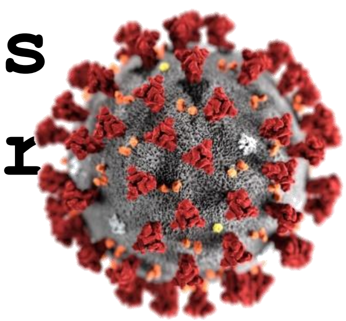
" very few interesting items of knowledge that are truly ontological…". Much current work on informatics ontologies is aimed at integrating probabilistic and typical reasoning with universal “ontological” reasoning effectively. Hence, the background information for a clinical system often goes well beyond the ontology, in this strict sense. Brachman introduced the notion of the ontology as a “conceptual coat rack” on which other information is held."

Rector A (2008): Semantichealth Deliverable 6.1

Representation of contingent knowledge with triples

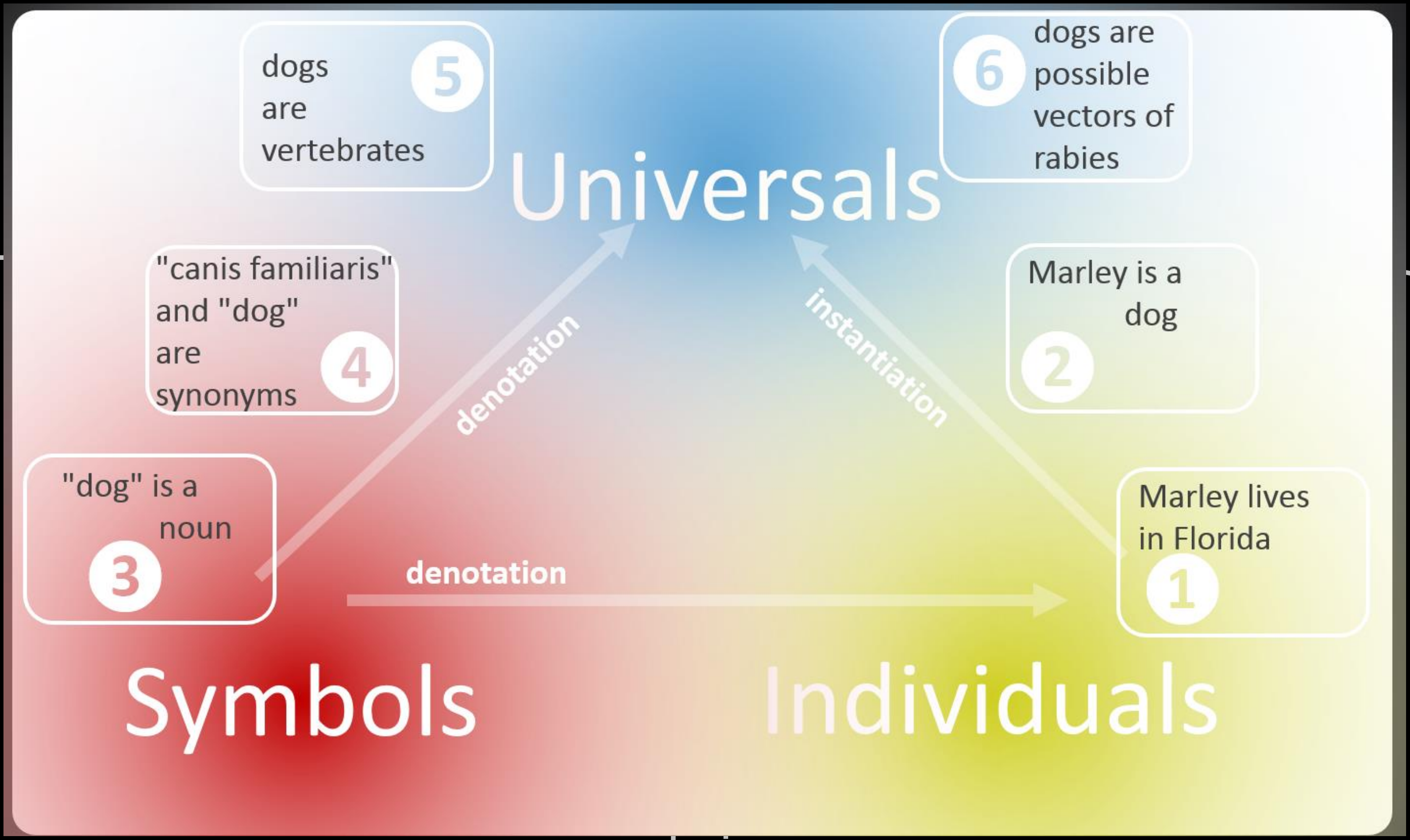
- Possible, but no formal semantics!
- Complex and context-dependent interpretations

<Subject>	<Predicate>	<Object>
:Dog	**:vector-of**	**:Rabies**
:Tobacco	**:causes**	**:Cancer**
:Aspirin	**:treats**	**:Pain**
:Bird	**:capable-of**	**:Flying**
:Fever	**:suggests**	**:Influenza**



Ontological knowledge:
Axioms that are universally true

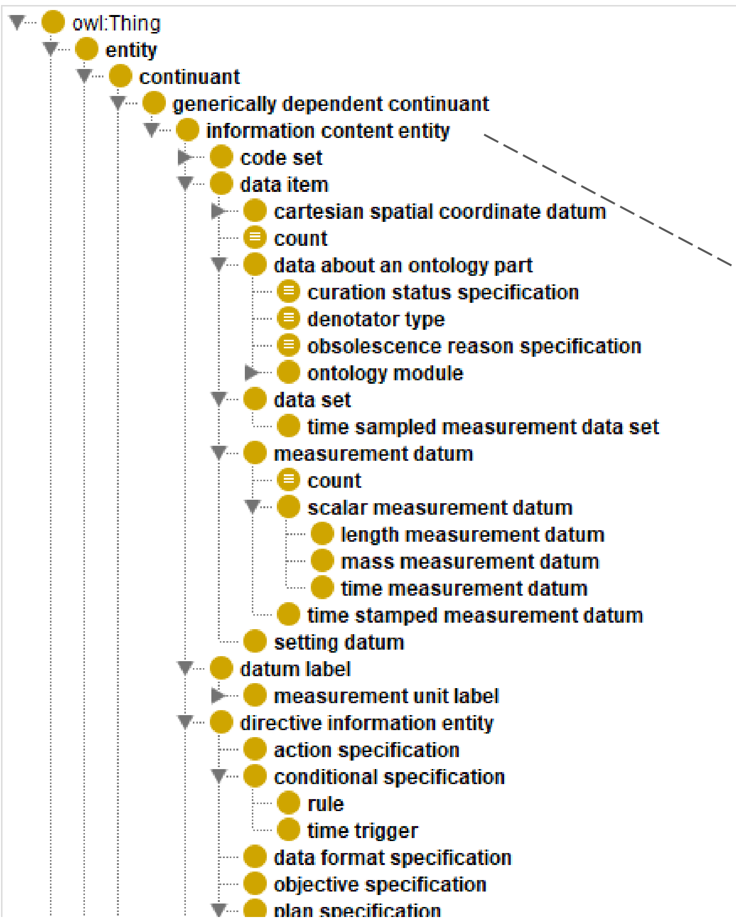
Contingent knowledge:
typical, likely, possible



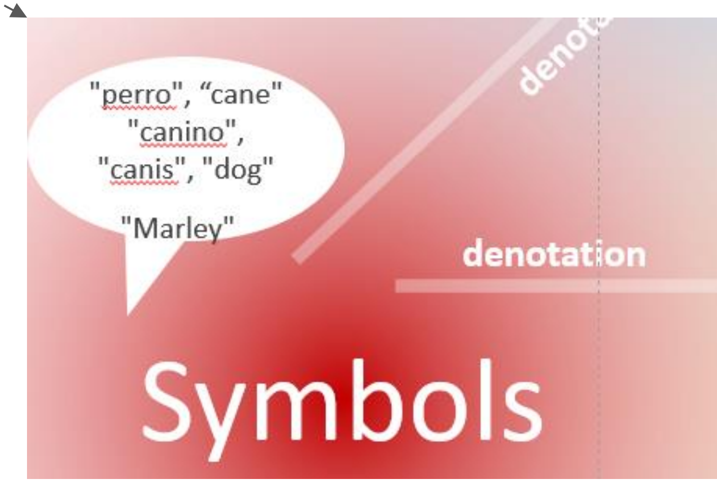
Linguistic knowledge:
Statements about properties and meaning of signs of language

Factual knowledge:
Statements about concrete entities and their relationships

Ontologies of information



- Information content entity, e.g. in IAO (information artefact ontology): „A generically dependent continuant that is about some thing"
- Ontology class for all symbolic entities



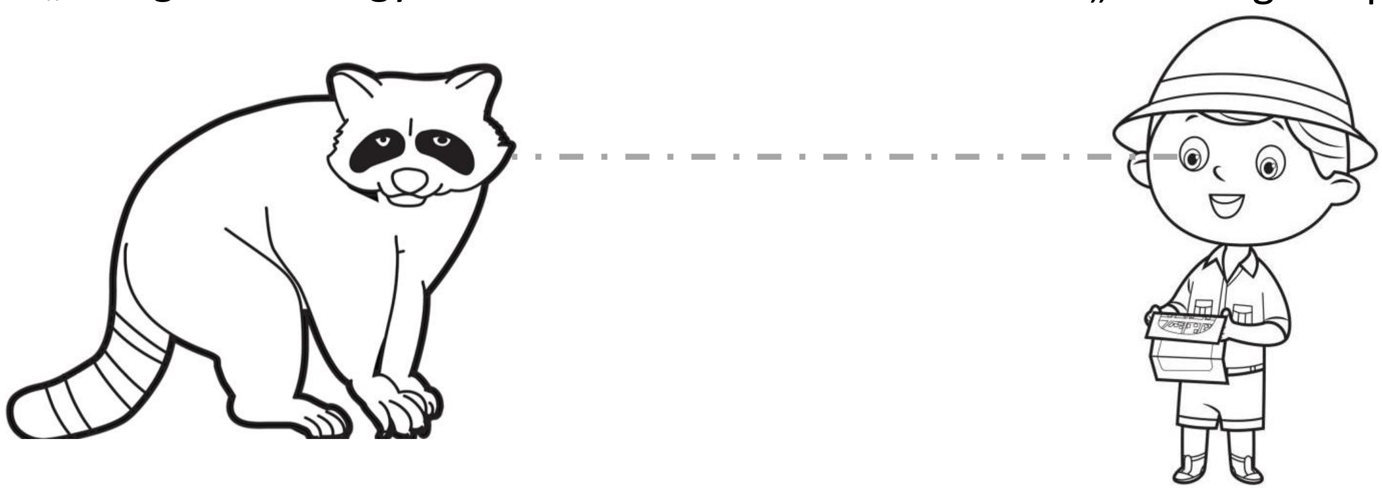
What‘s the difference

- biological_sex
  - non_binary
  - binary
    - female
    - male
      - male_racoon_12345

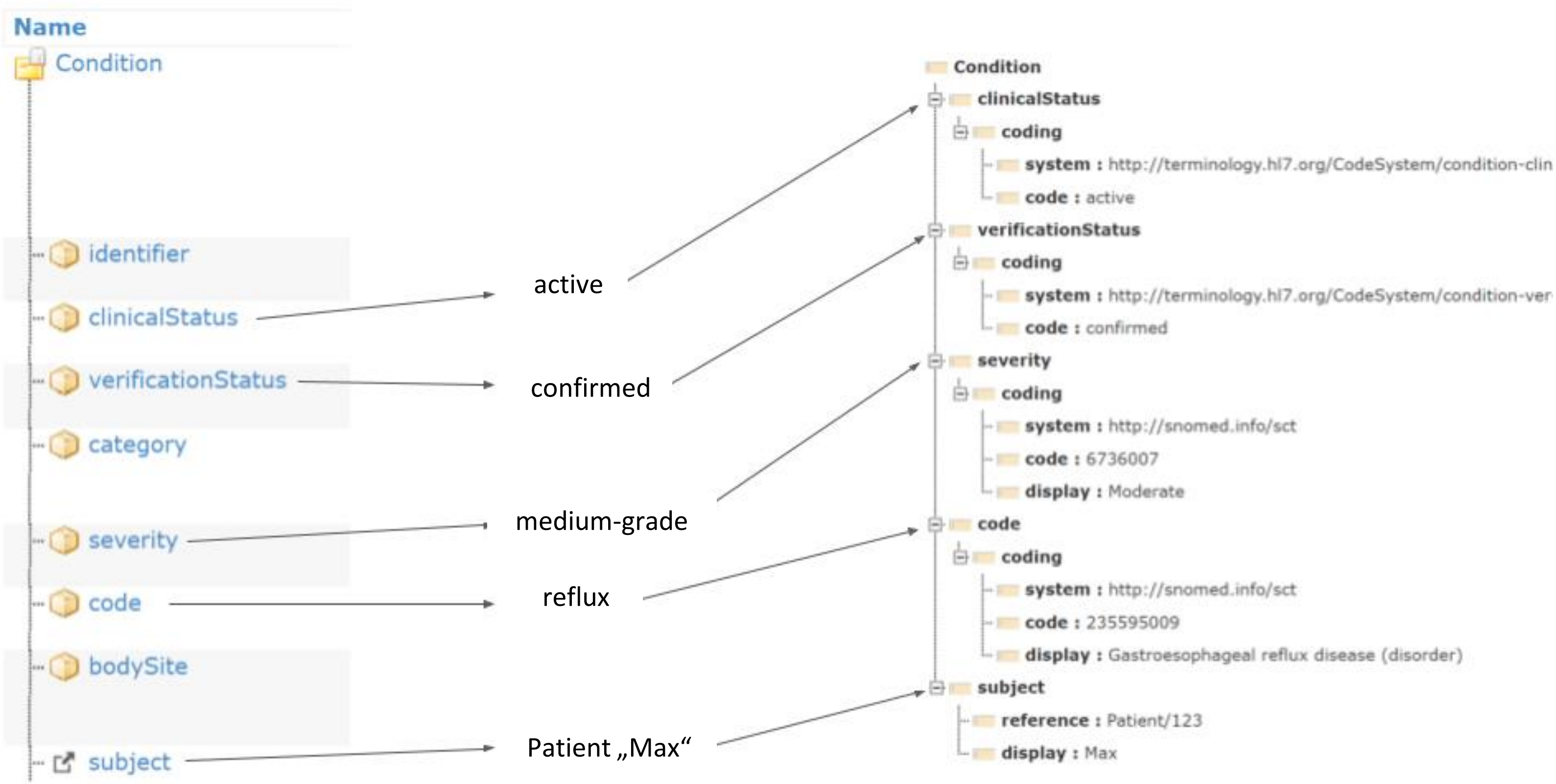
- biological_sex_data
  - female
  - male
  - other
  - unknown
  - unspecified
    - sex_data_racoon_12345

„being“ – ontology

„knowing“ – epistemology



Example: FHIR



OWL in a nutshell

- Description logics:
  - Subset of First-Order Logics
  - Axiom-based, not graph-based (can be expressed as graphs)
  - Different syntaxes: RDF/XML, OWL/XML, **Manchester OWL**, …
- OWL profiles
  - Different expressiveness: OWL Full > OWL DL > OWL EL
  - The more expressive, the more intractable
- OWL reasoners (Pellet, HermIT, Fact++, ELK)
  - Compute satisfiability, inferences

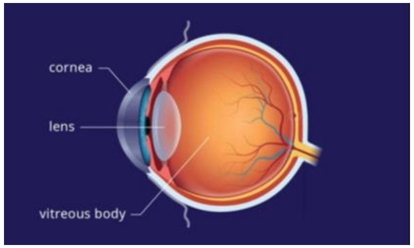
stefan.schulz@medunigraz.at

OWL EL

- NExpTime-complete (good scalability)
- Constructors (Manchester syntax):
  - Class equivalence, subsumption, disjointness: **`equivalentTo`**, **`subclassOf`**, **`DisjointClasses`**
    `Primate subclassOf Animal`
    `DisjointClasses (Plant, Animal)`
  - Relationships (object properties) with algebraic properties: transitivity, reflexivity, domain / range restrictions
  - Object intersection **`and`** / objectSomeValuesFrom **`some`**
    `Vertebrate equivalentTo Animal and has-part some Vertebra`
  - Assertions about individuals (subject predicate object triples)
    `part-of(Udine, Italy)`
- Cannot express negation (only via … **`subclassOf Nothing`**)

TBox – Abox distinction

- Compare:



  - Lens is part of Eye

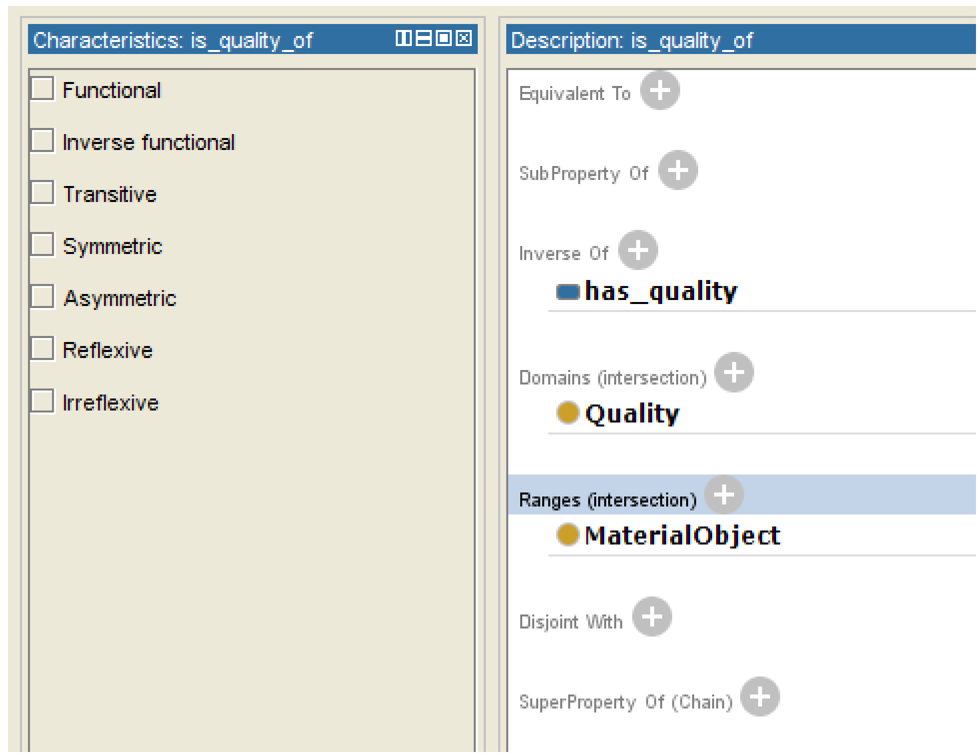
Italy is part of Europe

Lens
  subclass-of
    part-of some Eye

Part-of(Italy, Europe)

Binary relations (object properties)

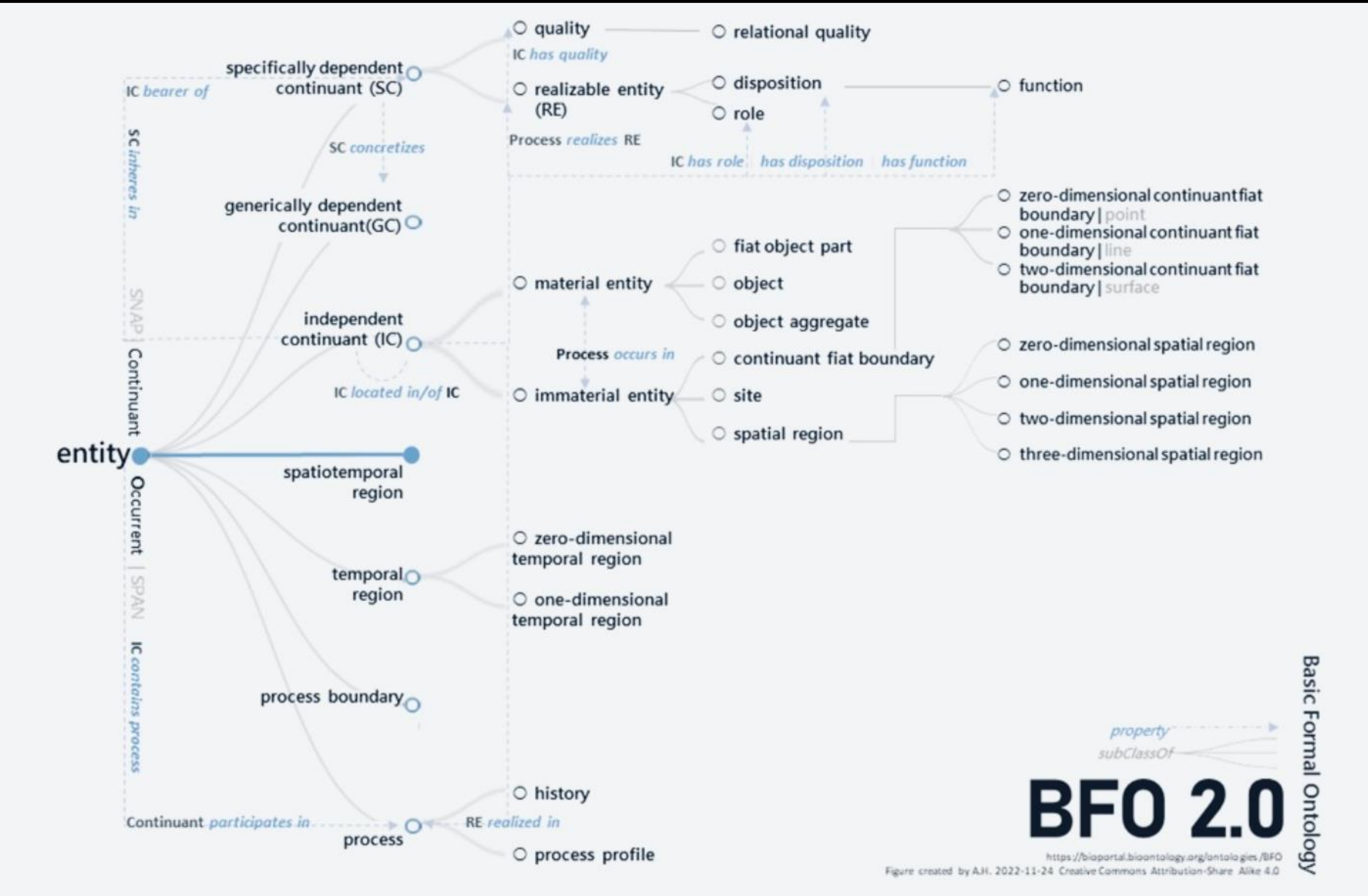
- Domain (left side)
- Range (right side)
- Inverse relation
- Algebraic characteristics:
  - Transitivity:
    r (a, b)
    r (b, c)
    r (a, c)
  - Symmetry:
    r (a, b) ←→ r (b, a)
  - Reflexivity:
    r(a, a)



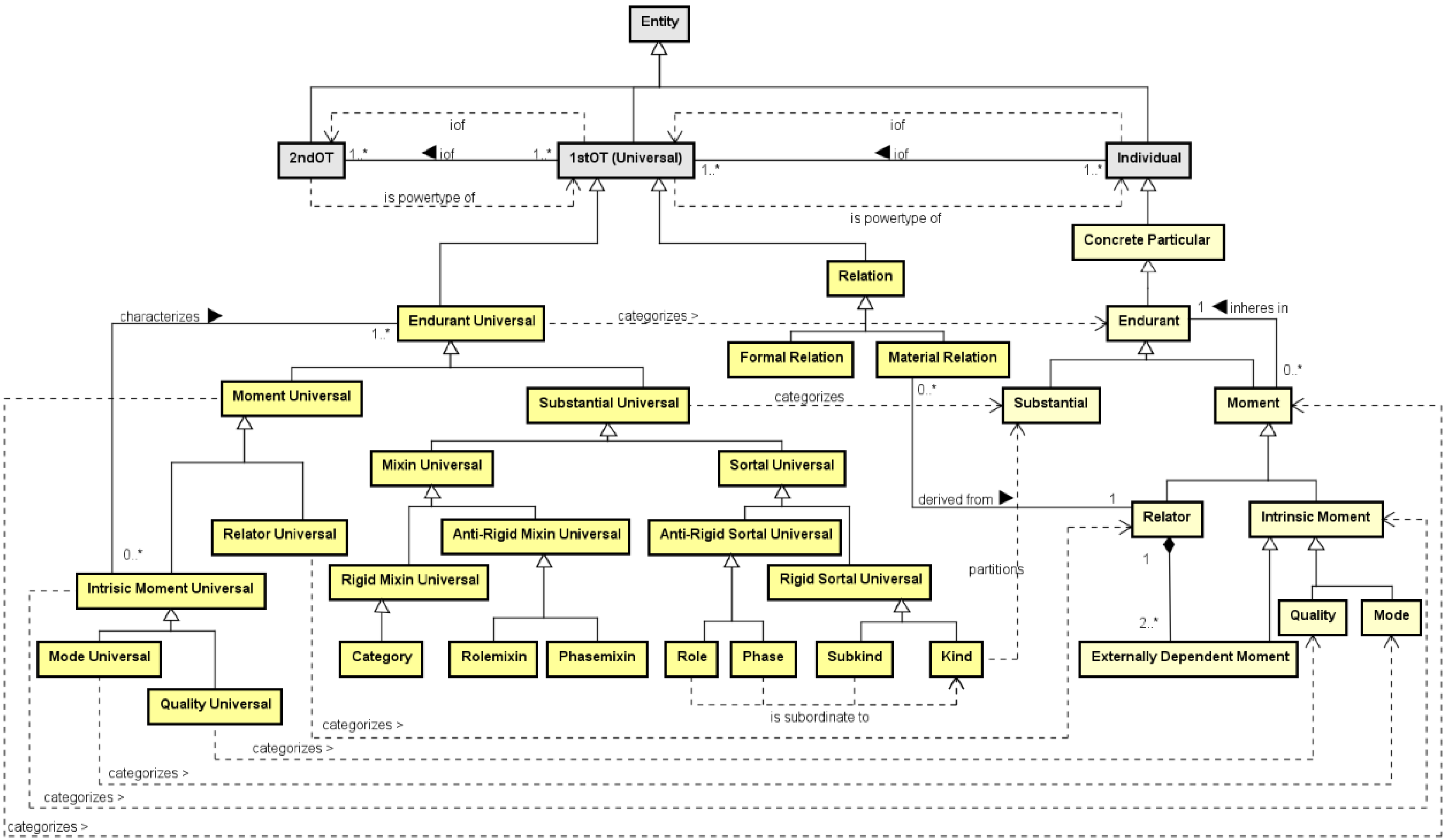
Upper level ontologies

- Also: foundational ontologies
- Typically provide:
  - Logical language (description logics, FOL)
  - Set of upper-level classes
  - Set of relations (object properties)
  - Constraining axioms (disjointness, domain/range restrictions)
- Purpose:
  - Facilitate domain ontology modelling
  - Supporting interoperability
  - Expressing particular point of view

Basic Formal Ontology (BFO)



Unified Formal Ontology (UFO)



Domain upper-level ontologies

- Examples
  - BioTop (BTL2)
  - SemanticScience integrated ontology (SIO)
  - SNOMED CT top hierarchies and concept model
- Characteristics
  - Focus on a particular domain
  - Usage does not require the „philosophical overhead“ of foundational ontologies
  - Ideally link to some foundational level ontology

BTL2
(BioTopLite v2)

- thing
  - Nothing
  - 'particular at some time'
    - condition
    - disposition
    - function
    - 'immaterial object'
      - 'immaterial three dimensional physical entity'
      - 'one dimensional physical entity'
      - 'two dimensional physical entity'
      - 'zero dimensional physical entity'
    - 'information object'
      - plan
    - 'material object'
      - 'collective material entity'
        - 'amount of pure substance'
        - 'plurality of organisms'
      - compound
        - atom
        - 'compound of collective material entities'
        - 'mono molecular entity'
        - 'poly molecular composite entity'
          - 'structured biological entity'
            - cell
            - 'cellular component'
            - organism
            - 'organism part'
          - universe
      - 'subatomic particle'
      - wave

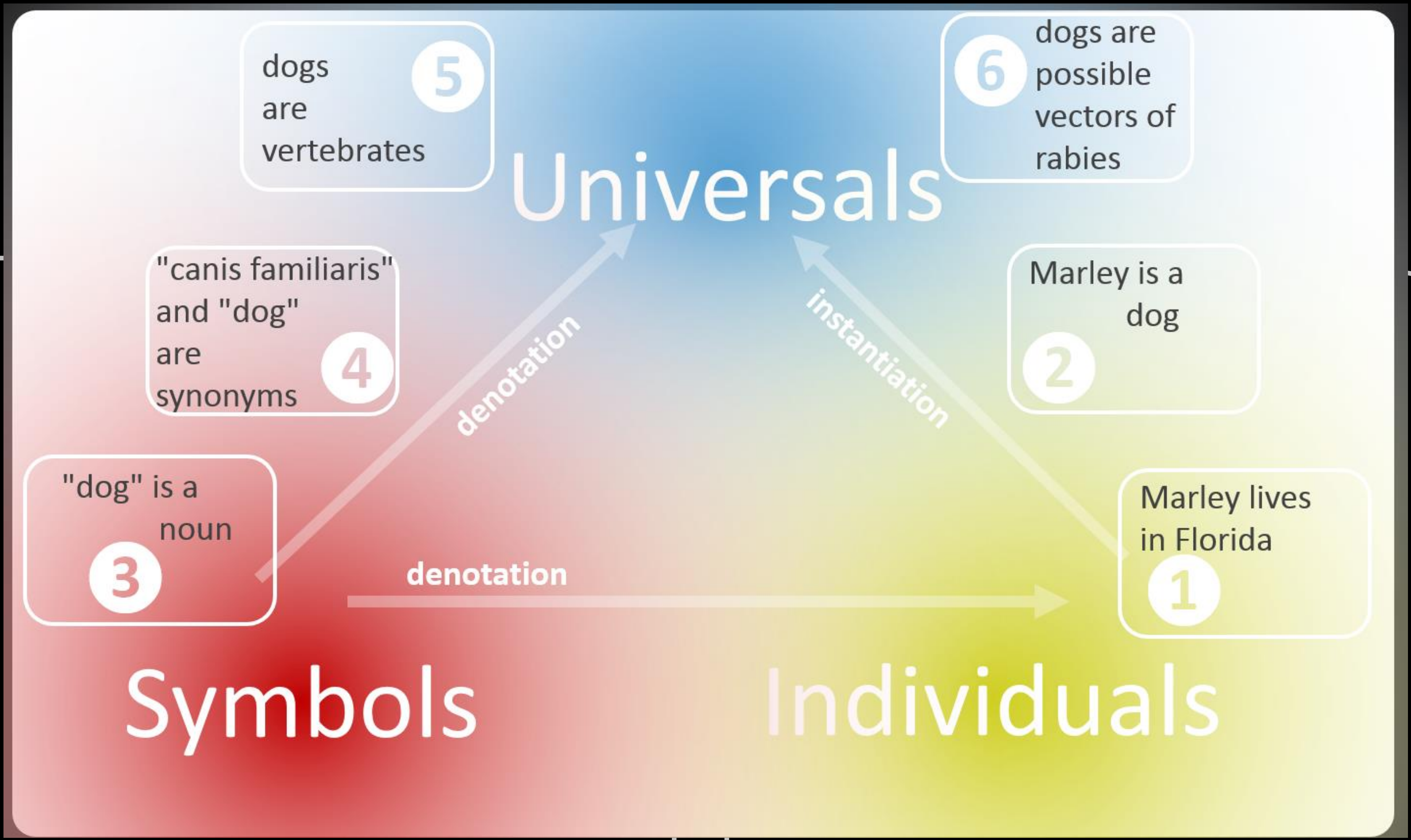
- process
  - action
    - 'plan execution'
  - 'collective process'
  - 'instantaneous process'
  - life
  - situation
- quality
  - canonicity
  - 'object quality'
    - 'physical length'
    - 'physical mass'
    - 'physical volume'
    - 'taxon quality'
  - 'process quality'
    - 'physical force'
- role
- 'temporal region'
  - 'point in time'
  - 'time interval'
- 'value region'
  - 'canonicity value region'
    - 'canonical value region'
    - 'noncanonical value region'
  - 'taxon value region'

- 'top object property'
  - 'at some time'
  - causes
    - 'has realization'
    - 'is agent in'
  - 'has condition'
  - 'has participant'
    - 'has agent'
    - 'has outcome'
    - 'has patient'
    - 'is life of'
  - includes
    - 'has part'
      - 'has boundary'
      - 'has component part'
      - 'has granular part'
    - 'is bearer of'
  - 'is caused by'
    - 'has agent'
    - 'is realization of'
  - 'is condition of'
  - 'is included in'
    - 'inheres in'
    - 'is part of'
      - 'is boundary of'
      - 'is component part of'
      - 'is granular part of'
  - 'is participant in'
    - 'has life'
    - 'is agent in'
    - 'is outcome of'
    - 'is patient in'
  - 'is preceded by'
  - 'is projection of'
  - 'is referred to at time'
  - 'is represented by'
  - precedes
  - 'projects onto'
  - represents

https://biotopontology.github.io/

Ontological knowledge:
Axioms that are universally true

Contingent knowledge:
typical, likely, possible

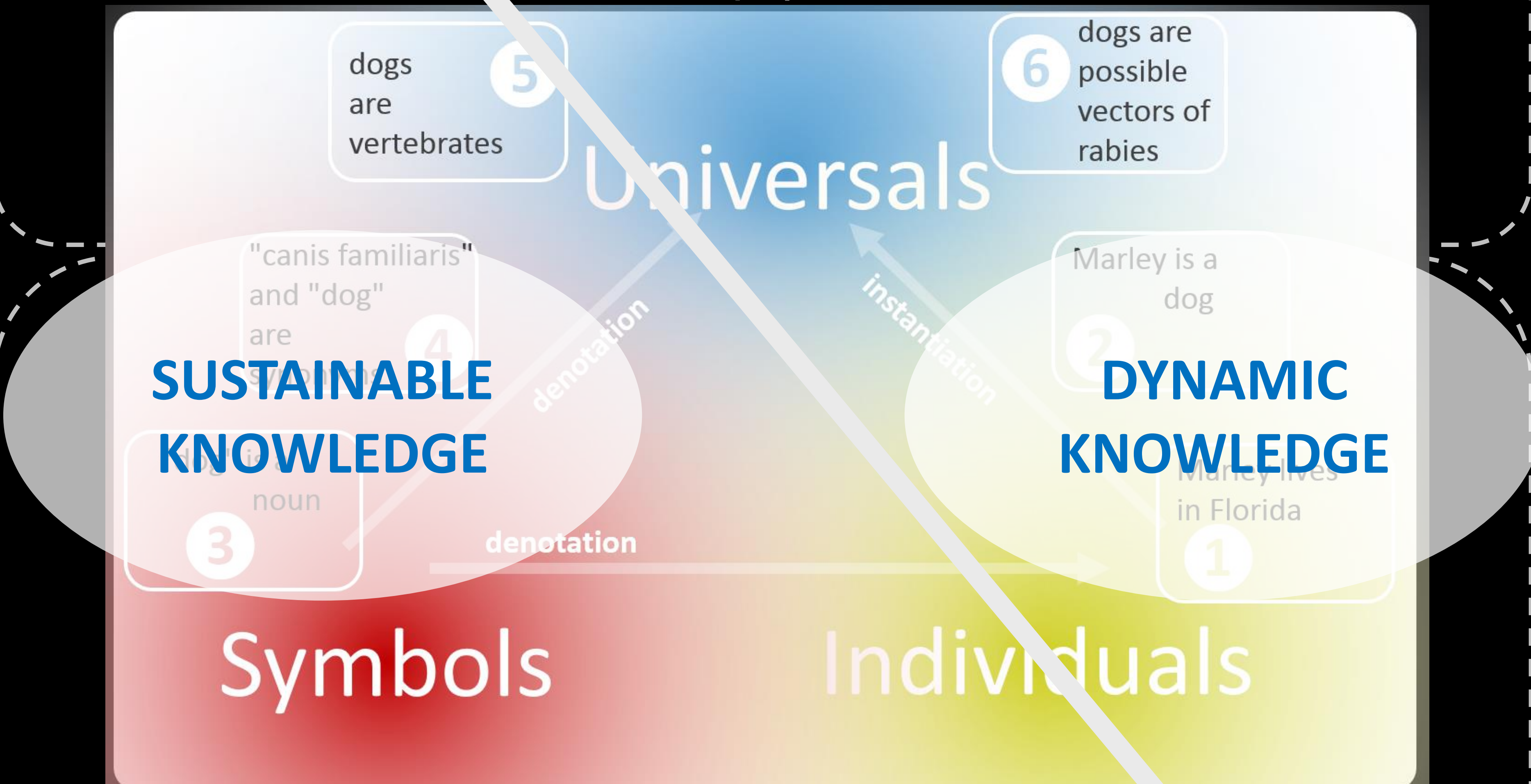


Linguistic knowledge:
Statements about properties and meaning of signs of language

Factual knowledge:
Statements about concrete entities and their relationships

Ontological knowledge:
Axioms that are universally true

Contingent knowledge:
typical, likely, possible



Linguistic knowledge:
Statements about properties and meaning of signs of language

Factual knowledge:
Statements about concrete entities and their relationships