16th Annual International Conference on Intelligent Systems for Molecular Biology

The Ontology of Biological Taxa

Stefan Schulz
Holger Stenzhorn
Martin Boeker

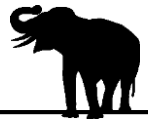
University Medical Center
Freiburg (Germany)
Institute of Medical Biometry
and Medical Informatics

Biological Taxa: Definition

Biological Taxa: Definition

- Taxa (singular taxon): Hierarchically structured labels or ranks used for biological classification
- Taxa can be attributed to organisms, populations, tissues, cells, cell components, and biological macromolecules
- Most biological discourse is related to some taxa
- Clarification taxa vs. species

Examples for Taxa

Taxon (Rank)	**Chimpanzee**	**Asian Elephant** 	**Drosophila**
Kingdom	Animalia	Animalia	Animalia
Phylum	Chordata	Chordata	Arthropoda
Subphylum	Vertebrata	Vertebrata	
Class	Mammalia	Mammalia	Insecta
Order	Primates	Proboscidea	Diptera
Superfamily		Elephantoidea	
Family	Hominides	Elephantidae	Drosophilidae
Subfamily			Drosophilinae
Genus	Pan	Elephas	Drosophila
Species	Simia troglodytes	Elephas maximus	Drosophila melanogaster

Taxa and biomedial vocabularies

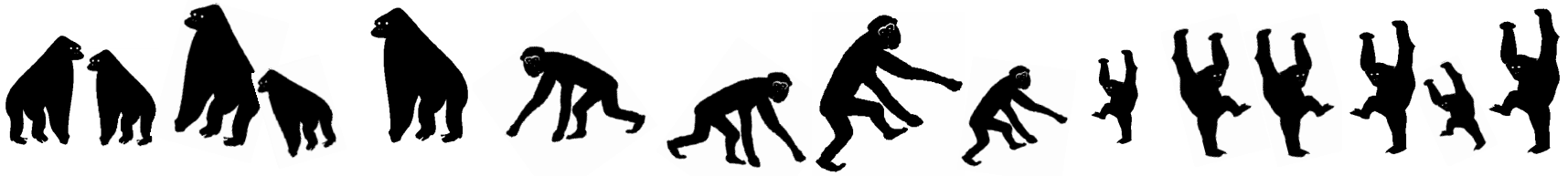
- MeSH: 3,497 entries
- Catalogue of Life: 1.75 M species by 2011
- NCBI taxonomy: 500,000 entries
- UNIPROT: 17,467 entries
- SNOMED CT: 27,400 entries
- OBO: 30 out of 66 ontologies are taxon-specific
- Nonspecific OBO ontologies:
  - GO : "*spore wall assembly (**sensu** Fungi)*"
    "*male tail morphogenesis (**sensu** Nematoda)*"
  - CL: "*non-visual cell (**sensu** Vertebrata)*"
    "*chemotactic amoeboid cell (**sensu** Mycetozoa)*"

The difficult concept of Species

- No agreement on proper definition of the term “species” and its ontological status
- 22 different conceptualizations of species
- Popular: „group of organisms that can interbreed and produce fertile offspring (Mayr, 1969) “
- Theoretical sound, difficult to apply, not generally valid
- Our approach: biological taxa need to be accounted for in biomedical ontologies, let alone whether they exist in nature or are merely (fiat) attributions by biologists

Basic stipulations on ontologies

Domain
(Particulars)

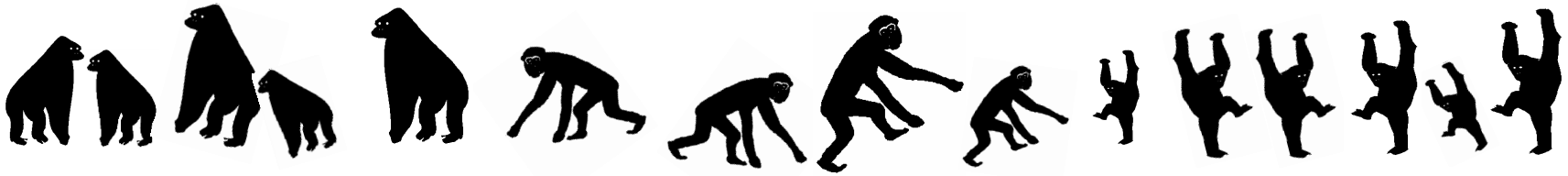


Basic stipulations on ontologies

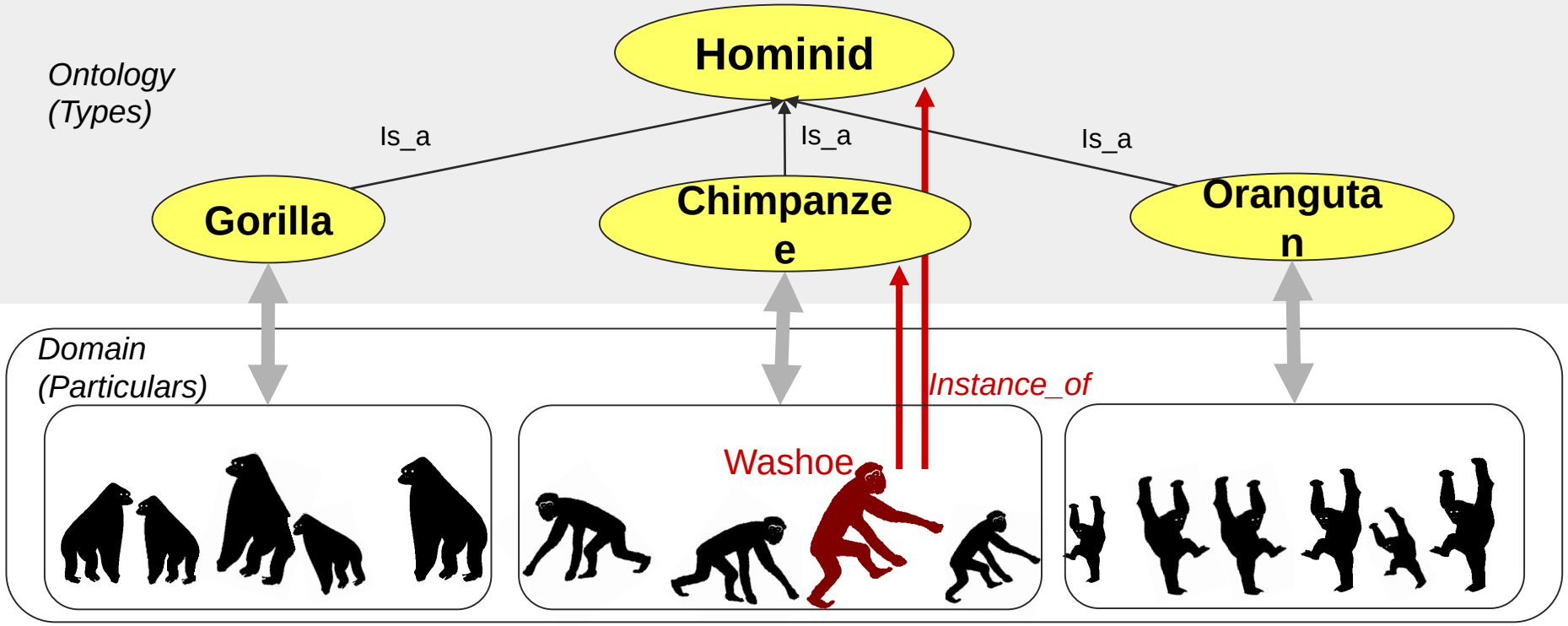
Ontology
(Types)

Hominid

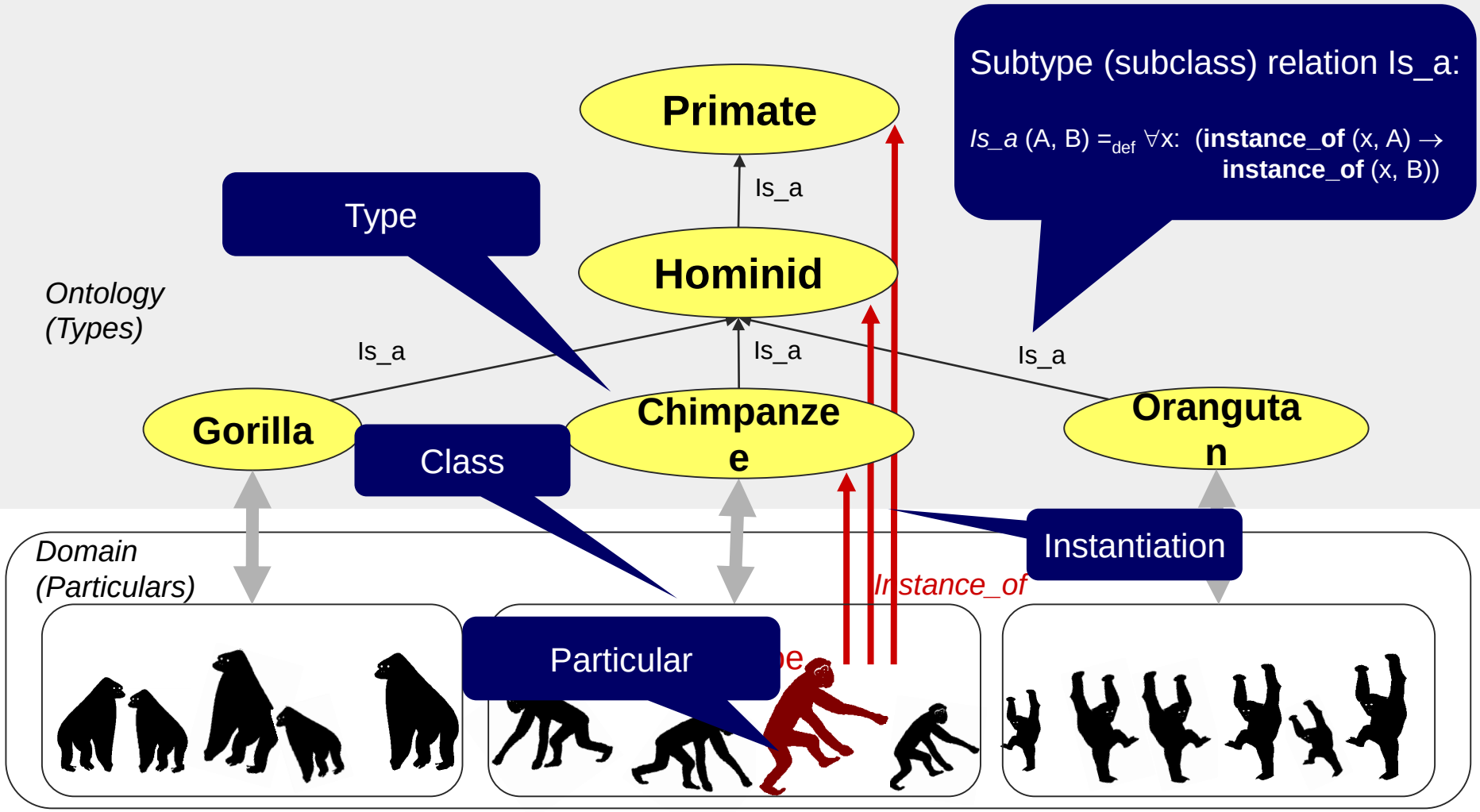
Domain
(Particulars)



Basic stipulations on ontologies



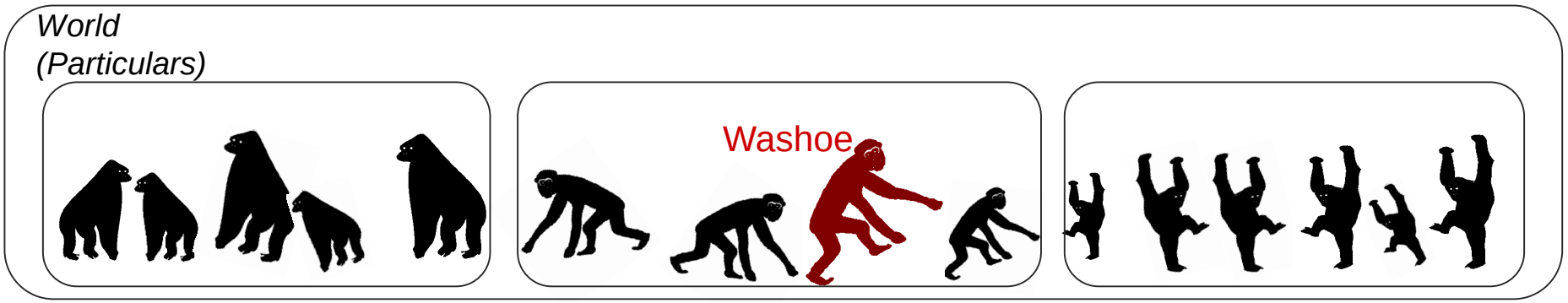
Basic stipulations on ontologies

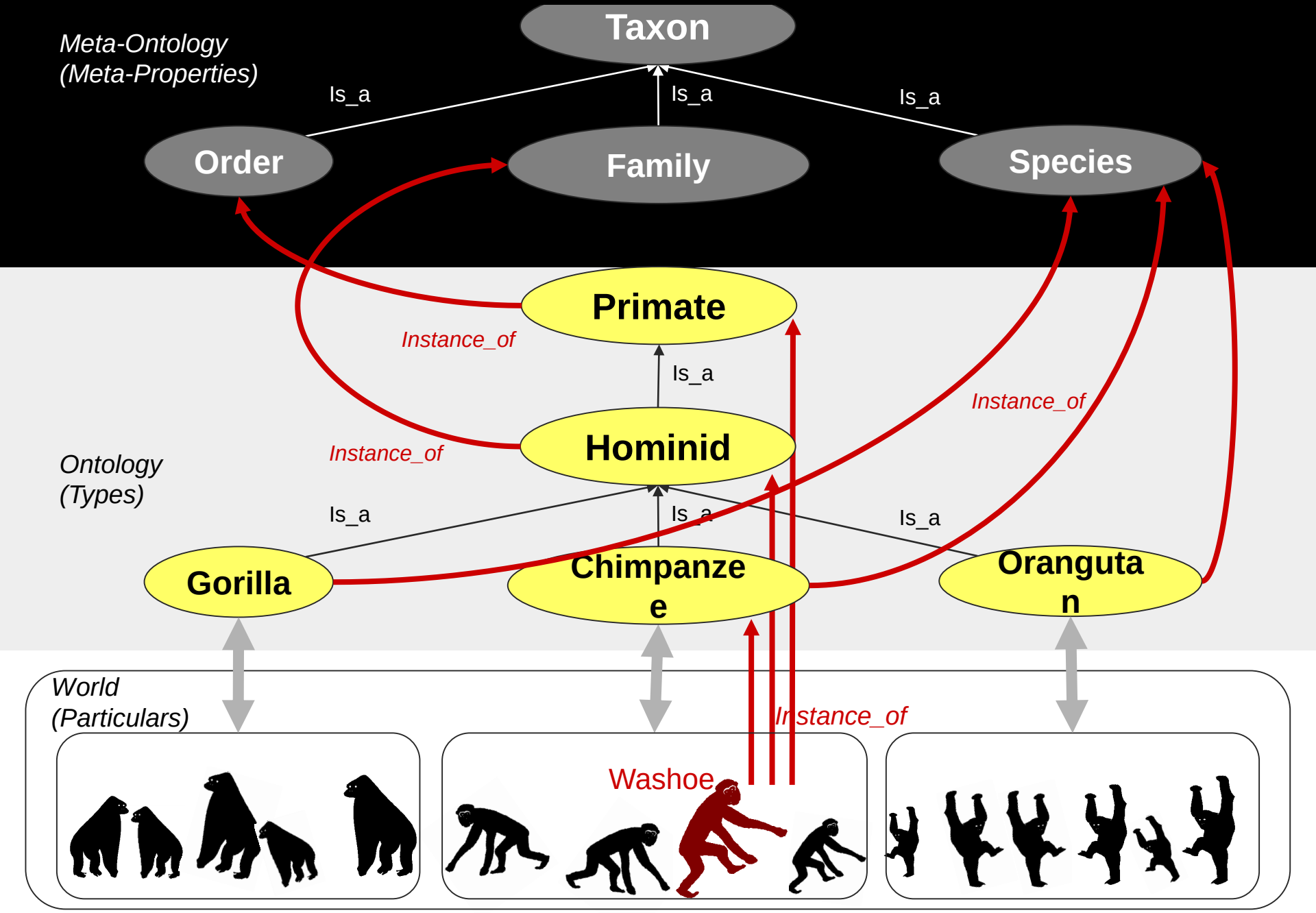


How to represent biological taxa?

How to represent biological taxa?

1. **Meta-Properties**


World
(Particulars)
Washoe

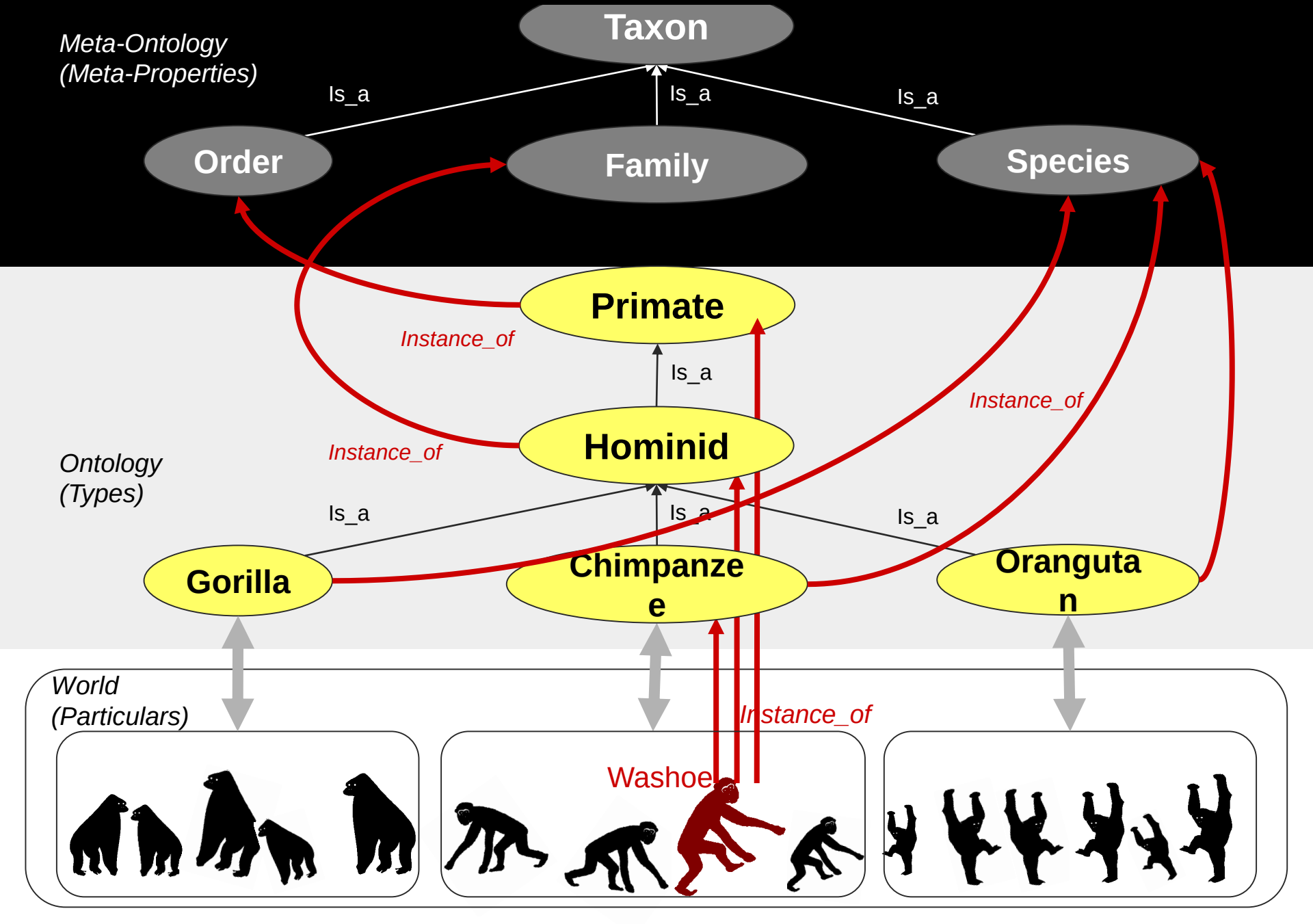

Meta-Ontology
(Meta-Properties)
Taxon
Is_a
Is_a
Is_a
Order
Family
Species
Primate
Instance_of
Is_a
Instance_of
Instance_of
Hominid
Ontology
(Types)
Is_a
Is_a
Is_a
Gorilla
Chimpanze
e
Oranguta
n
World
(Particulars)
Instance_of
Washoe

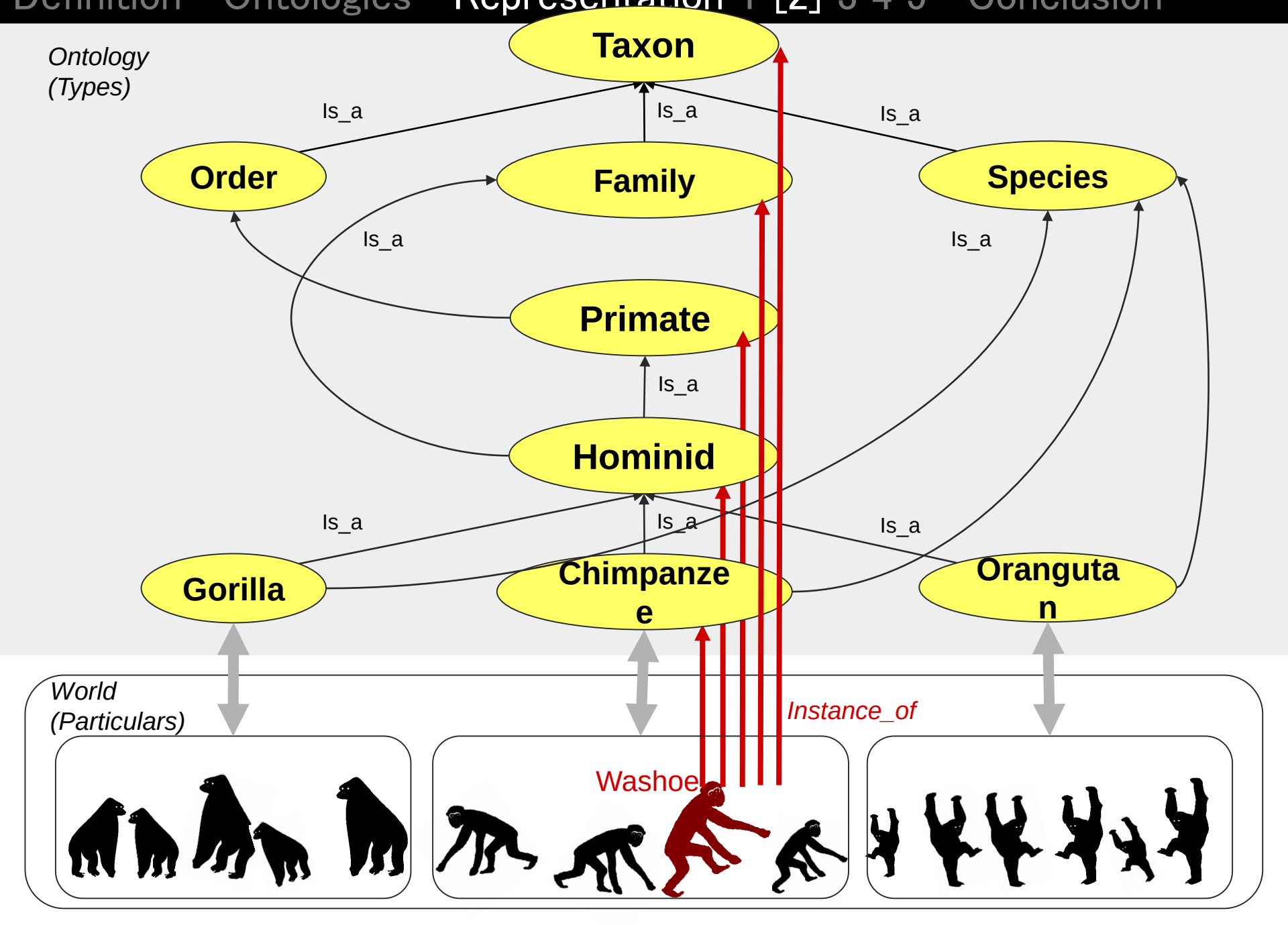
How to represent biological taxa?

1. **Meta-Properties**
   - \+ Intuitive
   - \- Instances of instances cannot be expressed by common representational formalisms (OWL, OBO)

How to represent biological taxa?

1. **Meta-Properties**
   + Intuitive
   - Instances of instances cannot be expressed by common representational formalisms (OWL, OBO)
2. **Supertypes:**


Meta-Ontology
(Meta-Properties)
Taxon
Is_a
Is_a
Is_a
Order
Family
Species
Primate
Instance_of
Is_a
Instance_of
Instance_of
Hominid
Ontology
(Types)
Is_a
Is_a
Is_a
Gorilla
Chimpanzee
Orangutan
World
(Particulars)
Instance_of
Washoe

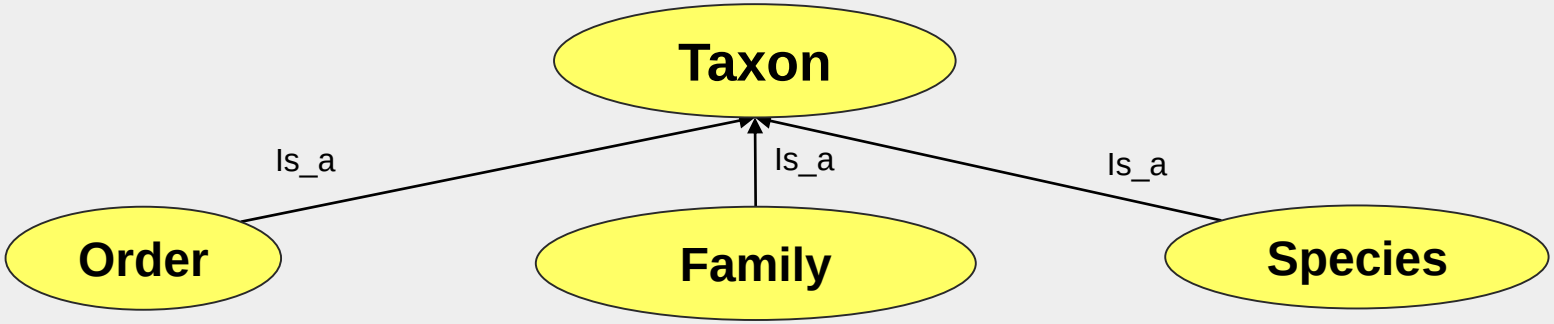

Ontology
(Types)
Taxon
Is_a
Is_a
Is_a
Order
Family
Species
Is_a
Is_a
Primate
Is_a
Hominid
Is_a
Is_a
Is_a
Gorilla
Chimpanze
e
Oranguta
n
World
(Particulars)
Instance_of
Washoe

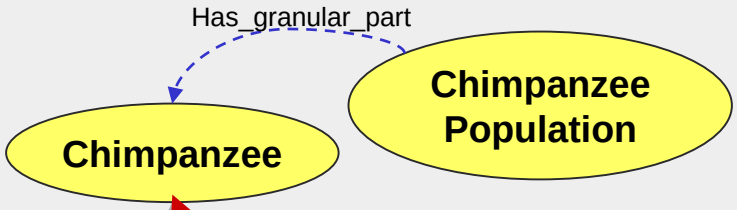
How to represent biological taxa?

1. **Meta-Properties**
   + Intuitive
   - Instances of instances cannot be expressed by common representational formalisms (OWL, OBO)
2. **Supertypes:**
   + Intuitive
   - Unintended inferences

How to represent biological taxa?

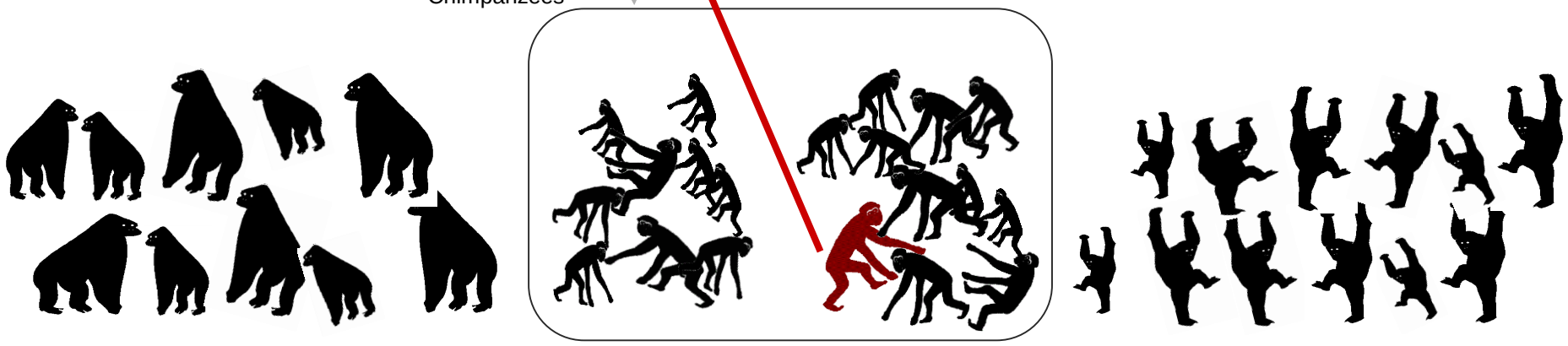
1. **Meta-Properties**
   + Intuitive
   - Instances of instances cannot be expressed by common representational formalisms (OWL, OBO)
2. **Supertypes:**
   + Intuitive
   - Unintended inferences
3. **Population instances**

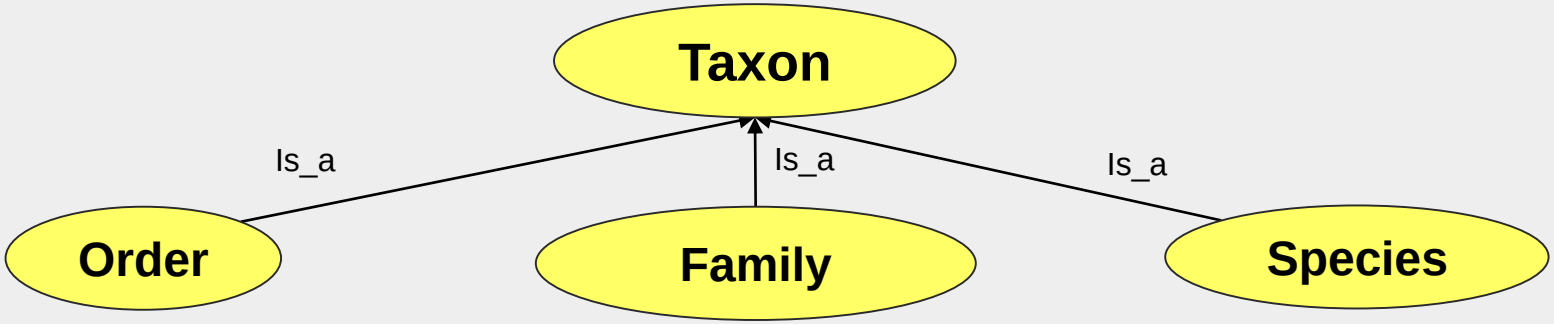


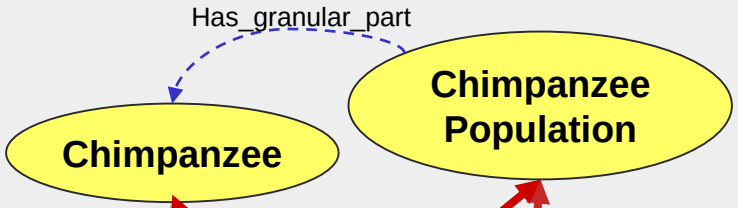


World (Particulars)

Class of Chimpanzees

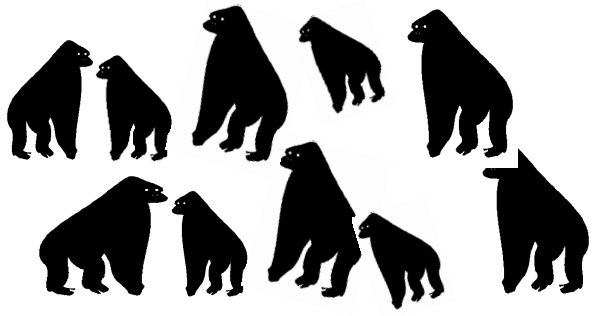


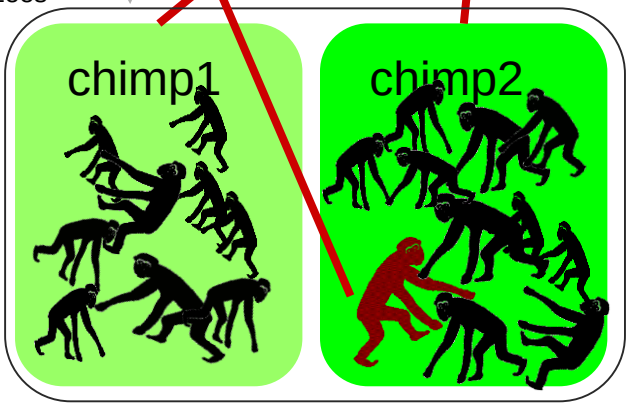


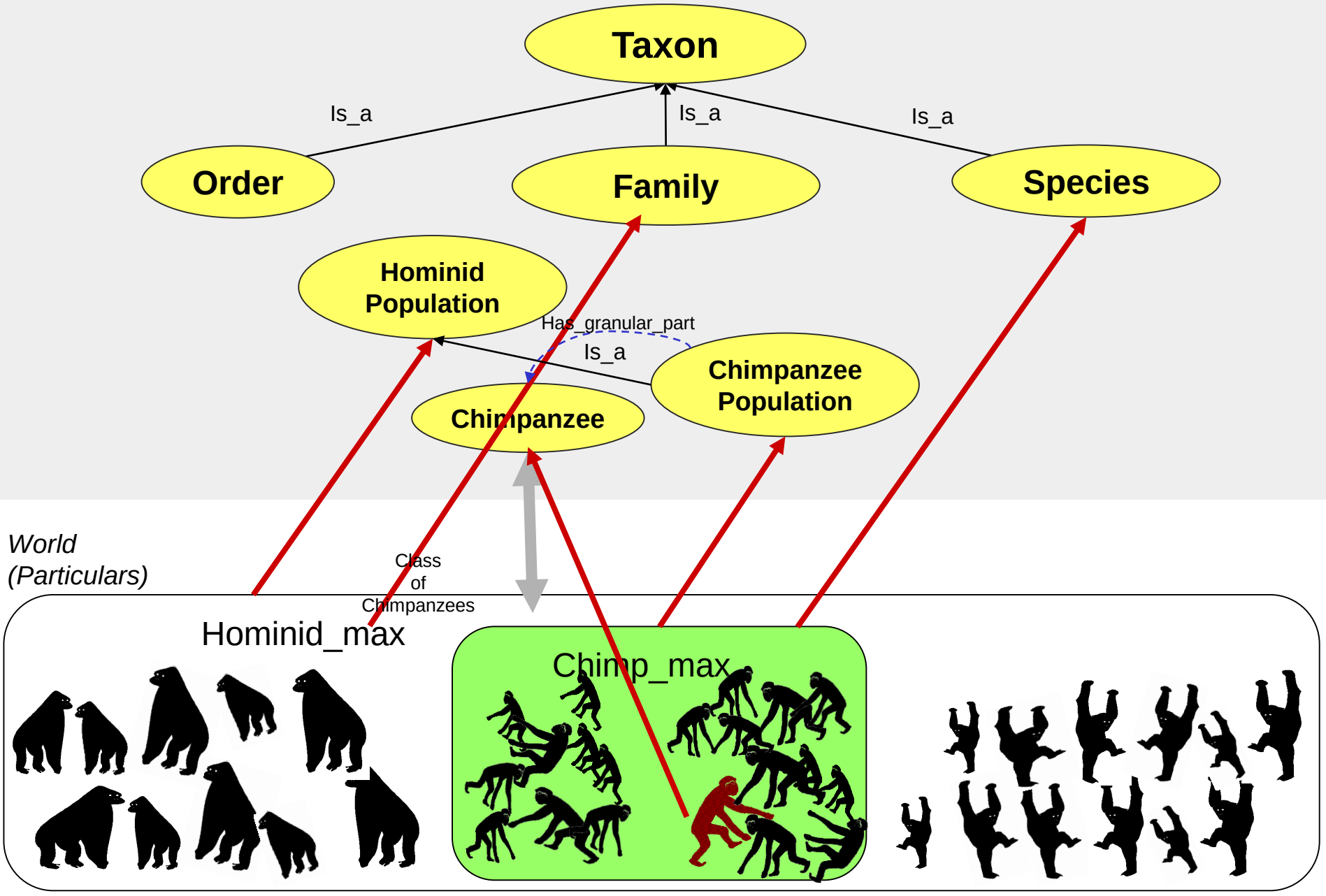


World (Particulars)

Class of Chimpanzees






Taxon
Is_a
Is_a
Is_a
Order
Family
Species
Hominid Population
Has_granular_part
Is_a
Chimpanzee Population
Chimpanzee
World (Particulars)
Class of Chimpanzees
Hominid_max
Chimp_max

How to represent biological taxa?

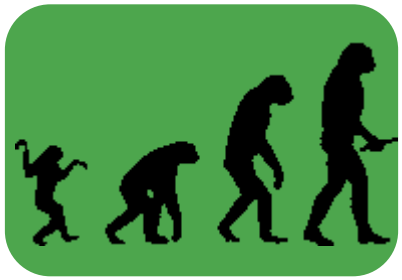
1. **Meta-Properties**
   + Intuitive
   - Instances of instances cannot be expressed by common representational formalisms (OWL, OBO)
2. **Supertypes:**
   + Intuitive
   - Unintended inferences
3. **Population instances**
   + Allows instantiation of abstract nodes (taxon, species, family…)
   - Requires A-Box extension (not expressible in OBO),
   - Only representation of organisms.

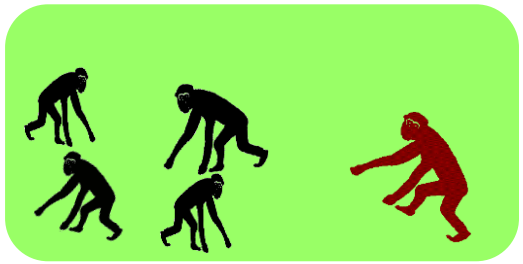
How to represent biological taxa?

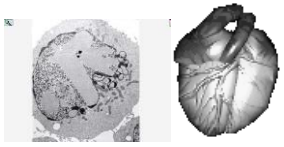
1. **Meta-Properties**
   + Intuitive
   - Instances of instances cannot be expressed by common representational formalisms (OWL, OBO)
2. **Supertypes:**
   + Intuitive
   - Unintended inferences
3. **Population instances**
   + Allows instantiation of abstract nodes (taxon, species, family…)
   - Requires A-Box extension (not expressible in OBO),
   - Only representation of organisms.
4. **Qualities**

Qualities in Upper Ontologies

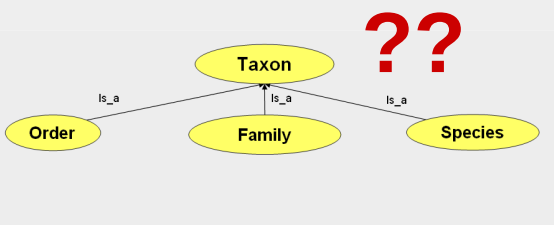
- BFO : "*A dependent continuant that is exhibited if it inheres in an entity or categorical property. Examples: the color of a tomato, the ambient temperature of air, the circumference shape of a nose, the mass of a piece of gold, the weight of a chimpanzee*"
- DOLCE : "*…the basic entities we can perceive or measure: shapes, colors, sizes, sounds, smells, as well as weights, lengths, electric charges*"
- The relation *inheres_in* links qualities to their bearers







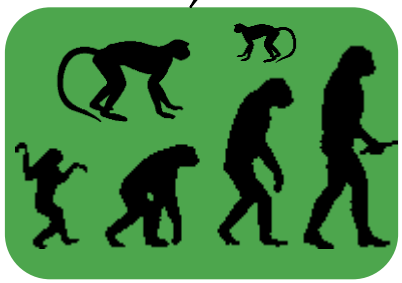
Hela Cell

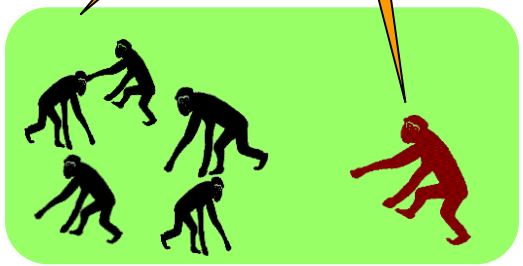


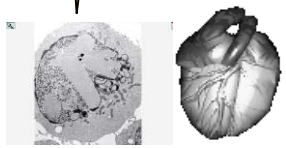
Quality
Is_a
Taxon_Quality
Is_a
Kingdom_Animalia_ Quality
Is_a
Phylum_Chordata_ Quality
Is_a
Class_Mammalia Quality
Is_a
Order_Primatae Quality
Is_a
Family_Hominides Quality
Is_a
Species_Simia T. Quality
Species Homo S. Quality

World (Particulars)

q3 q2 q1 q7 q6 q5 q4







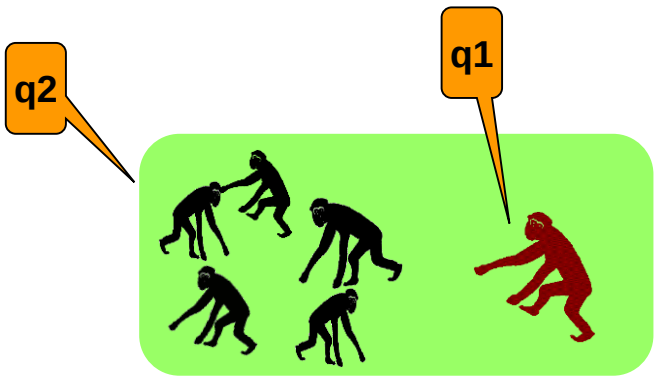
Hela Cell

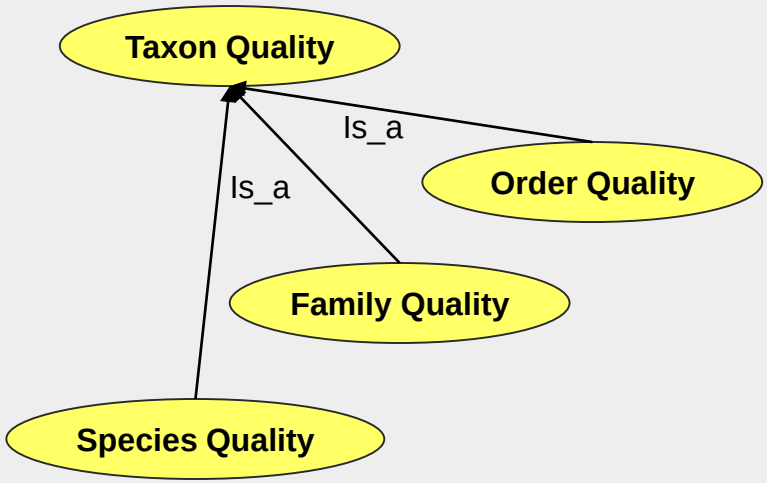
How to represent biological taxa?

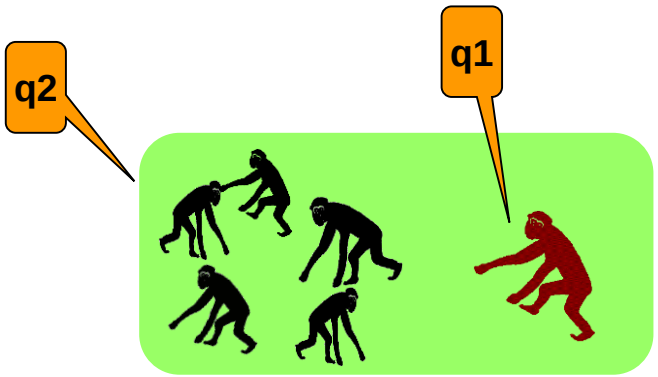
1. **Meta-Properties**
   + Intuitive
   - Instances of instances cannot be expressed by common representational formalisms (OWL, OBO)
2. **Supertypes:**
   + Intuitive
   - Unintended inferences
3. **Population instances**
   + Allows instantiation of abstract nodes (taxon, species, family…)
   - Requires A-Box extension (not expressible in OBO),
   - Only representation of organisms.
4. **Qualities**
   + Flexible and intuitive
   - No place for abstract nodes

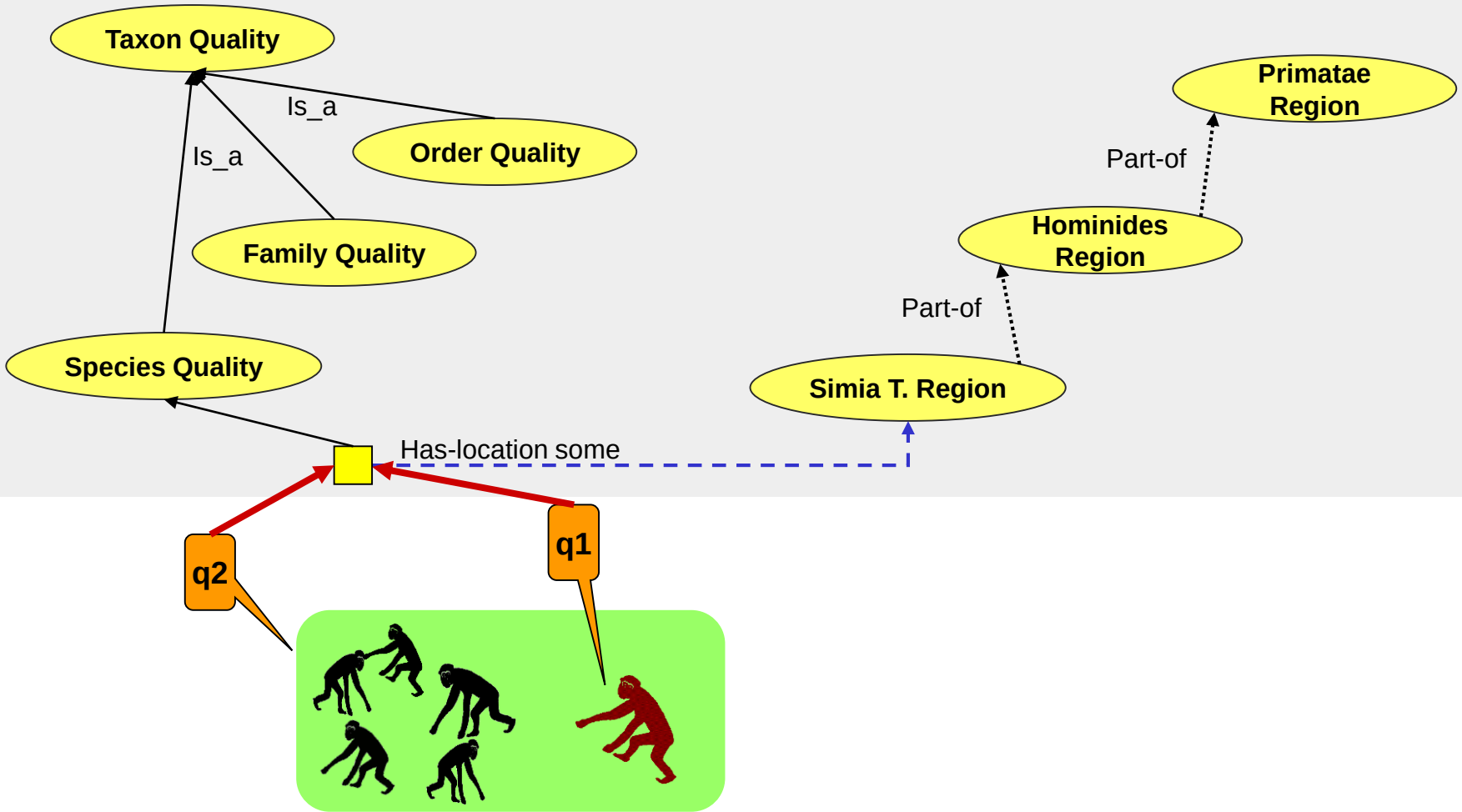
How to represent biological taxa?

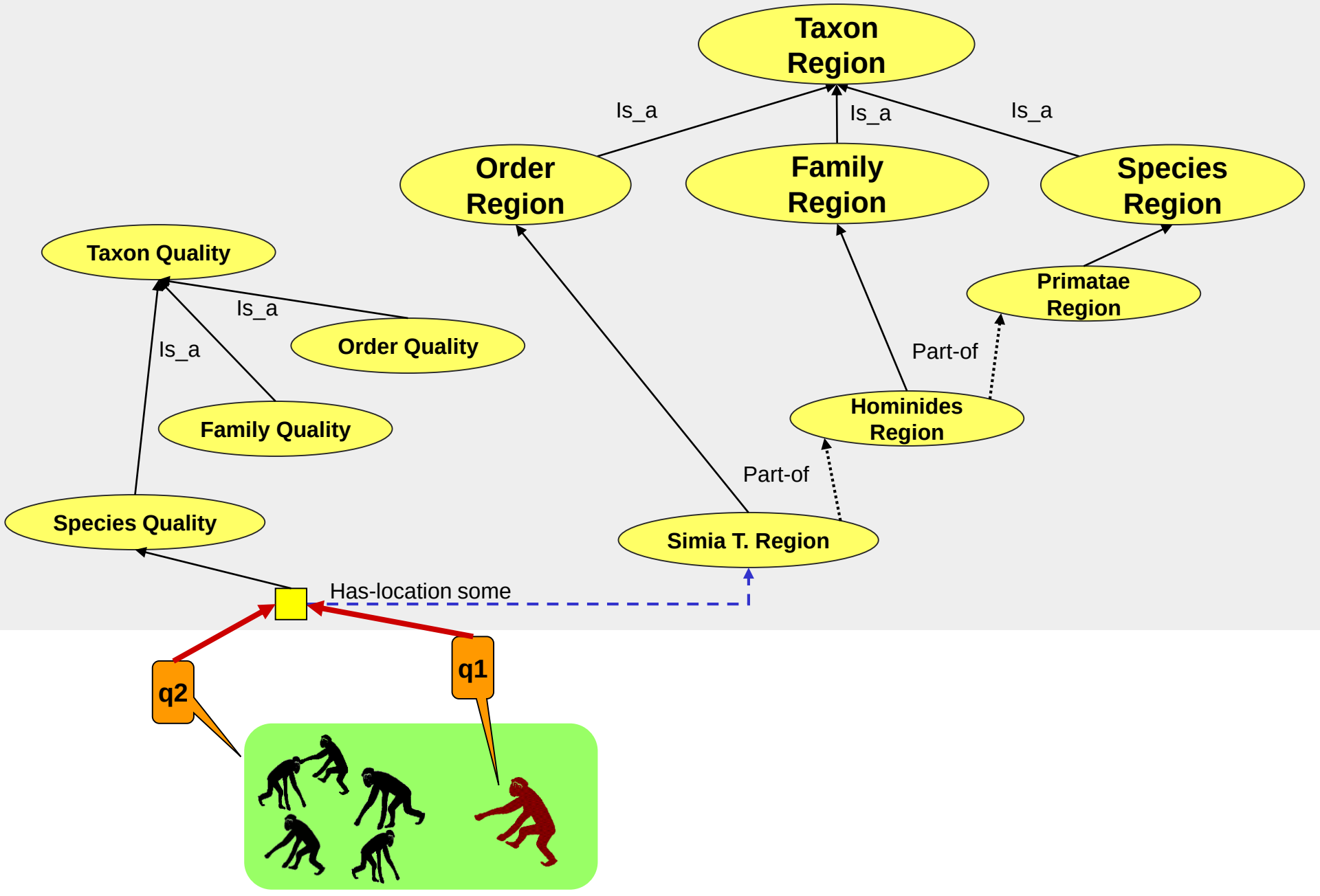
1. **Meta-Properties**
   + Intuitive
   - Instances of instances cannot be expressed by common representational formalisms (OWL, OBO)
2. **Supertypes:**
   + Intuitive
   - Unintended inferences
3. **Population instances**
   + Allows instantiation of abstract nodes (taxon, species, family…)
   - Requires A-Box extension (not expressible in OBO),
   - Only representation of organisms.
4. **Qualities**
   + Flexible and intuitive
   - No place for abstract nodes
5. **Qualia**


q2
q1


Taxon Quality
Is_a
Order Quality
Is_a
Family Quality
Species Quality


q2
q1


Taxon Quality
Is_a
Order Quality
Is_a
Family Quality
Species Quality
Primatae Region
Part-of
Hominides Region
Part-of
Simia T. Region
Has-location some
q2
q1


Taxon Region
Is_a
Is_a
Is_a
Order Region
Family Region
Species Region
Taxon Quality
Primatae Region
Is_a
Order Quality
Is_a
Part-of
Family Quality
Hominides Region
Part-of
Species Quality
Simia T. Region
Has-location some
q2
q1

How to represent biological taxa?

1. **Meta-Properties**
   + Intuitive
   - Instances of instances cannot be expressed by common representational formalisms (OWL, OBO)
2. **Supertypes:**
   + Intuitive
   - Unintended inferences
3. **Population instances**
   + Allows instantiation of abstract nodes (taxon, species, family…)
   - Requires A-Box extension (not expressible in OBO),
   - Only representation of organisms.
4. **Qualities**
   + Flexible and intuitive
   - No place for abstract nodes
5. **Qualia**
   + Abstract nodes can be represented as quality regions
   - Complex representation

Preferred Representations

Preferred Representations

1. **Meta-Properties**
   + Intuitive
   - Instances of instances cannot be expressed by common representational formalisms (OWL, OBO)
2. **Supertypes:**
   + Intuitive
   - Unintended inferences

Preferred Representations

1. **Meta-Properties**
   + Intuitive
   - Instances of instances cannot be expressed by common representational formalisms (OWL, OBO)
2. **Supertypes:**
   + Intuitive
   - Unintended inferences
3. **Population instances**
   + Allows instantiation of abstract nodes (taxon, species, family…)
   - Requires A-Box extension (not expressible in OBO),
   - Only representation of organisms.
4. **Qualities**
   + Flexible and intuitive
   - No place for abstract nodes
5. **Qualia**
   + Abstract nodes can be represented as quality regions
   - Complex representation

Preferred Representations

1. **Meta-Properties**
   + Intuitive
   - Instances of instances cannot be expressed by common representational formalisms (OWL, OBO)
2. **Supertypes:**
   + Intuitive
   - Unintended inferences

"BFO"

3. **Population instances**
   + Allows instantiation of abstract nodes (taxon, species, family…)
   - Requires A-Box extension (not expressible in OBO),
   - Only representation of organisms.
4. **Qualities**
   + Flexible and intuitive
   - No place for abstract nodes
5. **Qualia**
   + Abstract nodes can be represented as quality regions
   - Complex representation

Preferred Representations

1. **Meta-Properties**
   + Intuitive
   - Instances of instances cannot be expressed by common representational formalisms (OWL, OBO)
2. **Supertypes:**
   + Intuitive
   - Unintended inferences

"DOLCE"

3. **Population instances**
   + Allows instantiation of abstract nodes (taxon, species, family…)
   - Requires A-Box extension (not expressible in OBO),
   - Only representation of organisms.
4. **Qualities**
   + Flexible and intuitive
   - No place for abstract nodes
5. **Qualia**
   + Abstract nodes can be represented as quality regions
   - Complex representation

Summary

- Favored representation: Taxa are qualities
- Two flavors:
  - Subtype hierarchies of qualities
  - Inclusion hierarchies of quality regions
- Compatible with OBO / OWL-DL
- Qualities can inhere in populations, organisms, body parts, biomolecules (*"sensu"*)
- Compatible with Mayr's concept of species as populations: each taxon quality corresponds to exactly one group of organisms

Use cases

- Embedded in top-level ontology BioTop
- Demonstration taxon quality hierarchy “taxdemo” in http://purl.org/biotop
- NCBI taxonomy converted into OWL-DL taxon quality hierarchy (Dumontier Lab, Carleton University, Canada)
- Suggested formalism for the organism hierarchy redesign of SNOMED CT

The Ontology of Biological Taxa

Stefan Schulz
Holger Stenzhorn
Martin Boeker

Acknowledgements:

Alan Rector, Manchester
Michel Dumontier, Ottawa
anonymous reviewers