

Towards principles and methods for good ontologies

Maria Keet

email: mkeet@cs.uct.ac.za

home: <http://www.meteck.org>

Department of Computer Science
University of Cape Town, South Africa

ISAO 2023, 10-14 July 2023, Sherbrooke, Canada

Outline

- 1 What is a 'good' ontology?
- 2 Methods
 - Theory
 - Techniques and tools
- 3 Summary

Quality of the ontology

- “**Bad ontologies** are (inter alia) those whose general terms lack the relation to corresponding universals in reality, and thereby also to corresponding instances.” $\Rightarrow$ need for grounding

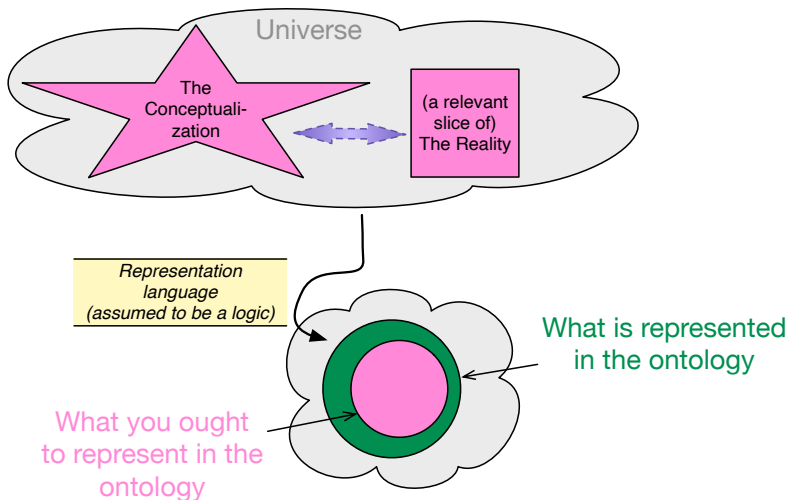
[Smith(2004)]

Quality of the ontology

- “**Bad ontologies** are (inter alia) those whose general terms lack the relation to corresponding universals in reality, and thereby also to corresponding instances.” $\Rightarrow$ need for grounding
- “**Good ontologies are reality representations**, and the fact that such representations are possible is shown by the fact that, as is documented in our scientific textbooks, very many of them have already been achieved, though of course always only at some specific level of **granularity** and to some specific degree of **precision, detail** and **completeness**.”

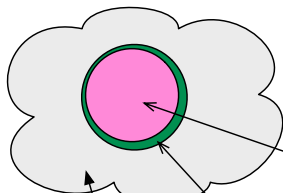
[Smith(2004)]

Quality of the ontology – the basic players

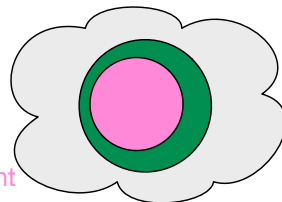


Quality of the ontology

Good



Less good

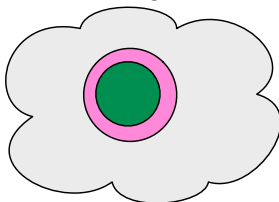


what you want or need to represent

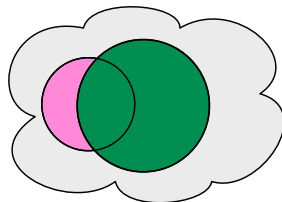
what you represented with the language

Universe

Probably bad



Worse



Why the difference between the pink and green circles (1/3)

- Example: 'each car has exactly 4 wheels' (except for 3-wheelers)
- OWL DL: $Car \sqsubseteq = 4 \text{ hasPart}. \top$ and/or $Car \sqsubseteq \exists \text{ hasPart}. Wheel$

Why the difference between the pink and green circles (1/3)

- Example: 'each car has exactly 4 wheels' (except for 3-wheelers)
- OWL DL: $Car \sqsubseteq = 4 \text{ hasPart. } \top$ and/or $Car \sqsubseteq \exists \text{ hasPart. } Wheel$

each car has exactly 4 things as part (no, it has more parts),
of which at least one wheel (yes, but actually 4, not possibly
also just 2 wheels)

Why the difference between the pink and green circles

(1/3)

- Example: 'each car has exactly 4 wheels' (except for 3-wheelers)
- OWL DL: $Car \sqsubseteq = 4 \text{ hasPart.} \top$ and/or $Car \sqsubseteq \exists \text{ hasPart.} Wheel$
each car has exactly 4 things as part (no, it has more parts), of which at least one wheel (yes, but actually 4, not possibly also just 2 wheels)
- OWL 2 DL: $Car \sqsubseteq = 4 \text{ hasPart.} Wheel$
- FOL: $\forall x (Car(x) \rightarrow \exists y^{=4} (hasPart(x, y) \wedge Wheel(y)))$

Why the difference between the pink and green circles

(1/3)

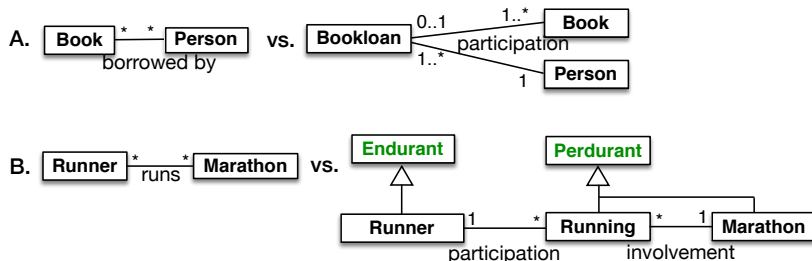
- Example: 'each car has exactly 4 wheels' (except for 3-wheelers)
- OWL DL: $Car \sqsubseteq = 4 \text{ hasPart}. \top$ and/or $Car \sqsubseteq \exists \text{ hasPart}. Wheel$

each car has exactly 4 things as part (no, it has more parts),
of which at least one wheel (yes, but actually 4, not possibly
also just 2 wheels)

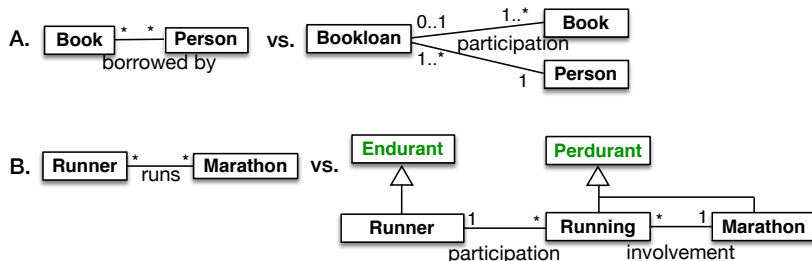
- OWL 2 DL: $Car \sqsubseteq = 4 \text{ hasPart}. Wheel$
- FOL: $\forall x (Car(x) \rightarrow \exists y^{=4} (hasPart(x, y) \wedge Wheel(y)))$

⇒ The representation language you choose matters

Why the difference between the pink and green circles (2/3)



Why the difference between the pink and green circles (2/3)



⇒ modelling style and granularity & precision ontologically (also when same language for both) [Fillottrani and Keet(2017)]

Why the difference between the pink and green circles (3/3)

The screenshot displays an ontology editor interface with two main panels. The left panel, titled 'Class hierarchy: EXM_000000000c', shows a tree structure of classes. The right panel, titled 'Annotations: EXM_000000000309', shows the annotations for the selected class.

Class hierarchy: EXM_000000000c

- owl:Thing
 - entity
 - continuant
 - generically dependent continuant
 - independent continuant
 - immaterial entity
 - material entity
 - EXM_000000000097
 - EXM_000000000099
 - EXM_000000000100
 - EXM_000000000219
 - EXM_000000000309** (highlighted in blue)
 - EXM_000000000342
 - EXM_000000000355
 - fiat object part
 - object
 - object aggregate
 - specifically dependent continuant
 - quality
 - realizable entity
 - occurrent

Annotations: EXM_000000000309

Annotations:

- rdfs:label** [language: zh]
exercise intervention
- hasDbXref** [type: xsd:string]
NCI: C62739

Description: EXM_000000000309

Equivalent To:

- SubClass Of: 'material entity'

General class axioms:

- SubClass Of (Anonymous Ancestor):
 - 'part of' **only** 'independent continuant'
 - 'part of' **only** continuant

⇒ **mistakes**. e.g., in FO alignment [Bernabé et al.(2023)], disjointness & intersection etc. etc.

Outline

1 What is a 'good' ontology?

2 Methods

- Theory
- Techniques and tools

3 Summary

How to develop good ontologies?

- Avoid common mistakes, with, e.g., “TIPS” (Typical Pitfall Prevention Scheme) [Keet et al.(2015)]
- Choose your language wisely (or create one) [Fillottrani and Keet(2020)]
- Use the automated reasoner
 - Common ‘culprits’ it detects: due to disjointness, cardinality constraints, domain & range axioms, violations of the logic and data types
- Reuse known working modelling solutions: a foundational ontology, one or more core ontologies, maybe also a relation ontology, ODPs
- Use modelling assistance, guidelines, and tools where available

Ontological, extra-logical, and logical principles that *explain*

- Different categories that the kinds of things are based on philosophical properties
 - e.g., rigidity and what follows from it (sortal, role etc); see OntoClean, OntoUML etc
- The meaning of class (/entity type/concept/universal) subsumption with property inheritance
- The meaning of class (/entity type/concept/universal) vs instance
- The meaning of property subsumption

Classes and instances

- An instance/individual is that thing that cannot be instantiated (any further)
- (notwithstanding the a representation language may let you 'play' with it, recast, and pun)

Classes and instances

- An instance/individual is that thing that cannot be instantiated (any further)
- (notwithstanding the a representation language may let you 'play' with it, recast, and pun)
- A subclass has all the properties of its superclass, and either or both:
 - at least one more property
 - at least one of the properties it inherited has its range more constrained

RBoxes: Questions and Problems

- What does subsumption mean for properties?
- What are the features of a 'good' RBox w.r.t. object property expressions?
- Modelling flaws in the RBox show up as unexpected or undesirable deductions regarding classes in the TBox, but current explanation algorithms mostly do not point to the actual flaw in the RBox
- How to guide the modeller how to revise the ontology once a flaw is found?

RBoxes: Questions and Problems

- What does subbumption mean for properties?
 - What are the features of a 'good' RBox w.r.t. object property expressions?
 - Modelling flaws in the RBox show up as unexpected or undesirable deductions regarding classes in the TBox, but current explanation algorithms mostly do not point to the actual flaw in the RBox
 - How to guide the modeller how to revise the ontology once a flaw is found?
- ⇒ With Sub-Property compatibility Service (*SubProS* and *ProChainS*) [Keet(2012)]

Preliminaries (1/2)—OWL 2/*SR*OIQ

- “basic form” for sub-properties, i.e., $S \sqsubseteq R$,
- “complex form” with property chains
- $R \sqsubseteq C_1 \times C_2$ as shortcut for domain and range axioms
 $\exists R \sqsubseteq C_1$ and $\exists R^- \sqsubseteq C_2$ where C_1 and C_2 are generic classes;
 ObjectPropertyDomain(OPE CE) and
 ObjectPropertyRange(OPE CE) in OWL.
- $R \sqsubseteq \top \times \top$ when no domain and range axiom has been declared

Definition (User-defined Domain and Range Classes)

Let R be an OWL object property and $R \sqsubseteq C_1 \times C_2$ its associated domain and range axiom. Then, with the symbol D_R we indicate the *User-defined Domain* of R —i.e., $D_R = C_1$ —and with the symbol R_R we indicate the *User-defined Range* of R —i.e., $R_R = C_2$.

Object sub-properties

- Given $S \sqsubseteq R$, then all individuals in the property assertions involving property S must also be related to each other through property R .
- Subsumption for OWL object properties (DL roles) holds if the subsumed property is more constrained such that in every model, the set of individual property assertions is a subset of those of its parent property

Object sub-properties

- Given $S \sqsubseteq R$, then all individuals in the property assertions involving property S must also be related to each other through property R .
- Subsumption for OWL object properties (DL roles) holds if the subsumed property is more constrained such that in every model, the set of individual property assertions is a subset of those of its parent property
- Two ways to constrain a property, and either one suffices:
 - By specifying its domain or range
 - By declaring the property's characteristics

Constraining a property

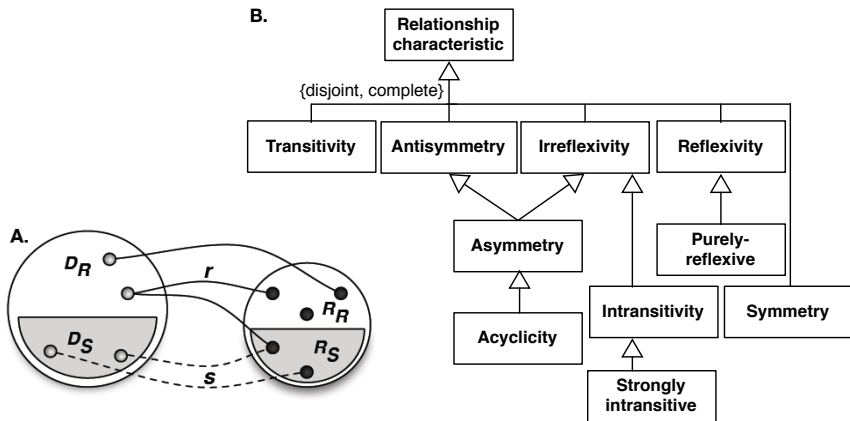


Figure: A: Example, alike the so-called 'subsetting' idea in UML; B: hierarchy of property characteristics (Based on Halpin 2001)

Constraining a property

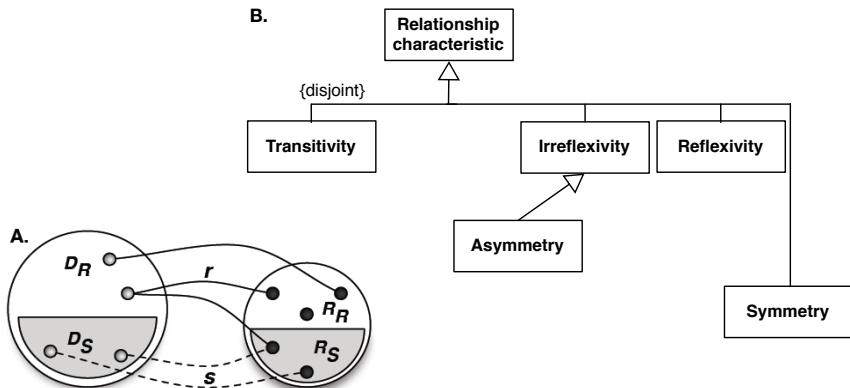


Figure: A: Example, alike the so-called 'subsetting' idea in UML; B: hierarchy of property characteristics relevant for OWL 2.

Outline Sub-Property compatibility Service

- First part extends the basic notions from the *RBox compatibility* (defined for *ALCQI*) [Keet and Artale(2008)]
- Checks the 'compatibility' of domain and range axioms w.r.t the object property hierarchy and the class hierarchy

Outline Sub-Property compatibility Service

- First part extends the basic notions from the *RBox compatibility* (defined for *ALCQI*) [Keet and Artale(2008)]
- Checks the 'compatibility' of domain and range axioms w.r.t the object property hierarchy and the class hierarchy
- Then checks whether the object property characteristic(s) conform to specification

Definition (Sub-Property compatibility Service (*SubProS*))

For each pair of object properties, $R, S \in \mathcal{O}$ such that $\mathcal{O} \models S \sqsubseteq R$, and $\mathcal{O}$ an OWL ontology adhering to the syntax and semantics as specified in OWL 2 Standard, check whether:

Test 1. $\mathcal{O} \models D_S \sqsubseteq D_R$ and $\mathcal{O} \models R_S \sqsubseteq R_R$;

Test 2. $\mathcal{O} \not\models D_R \sqsubseteq D_S$;

Test 3. $\mathcal{O} \not\models R_R \sqsubseteq R_S$;

Test 4. If $\mathcal{O} \models \text{Asym}(R)$ then $\mathcal{O} \models \text{Asym}(S)$;

Test 5. If $\mathcal{O} \models \text{Sym}(R)$ then $\mathcal{O} \models \text{Sym}(S)$ or $\mathcal{O} \models \text{Asym}(S)$;

...

Test 11. If $\mathcal{O} \models \text{Trans}(R)$ then $\mathcal{O} \not\models \text{Irr}(R)$, $\mathcal{O} \not\models \text{Asym}(R)$, $\mathcal{O} \not\models \text{Irr}(S)$, and $\mathcal{O} \not\models \text{Asym}(S)$;

An OWL object property hierarchy is said to be compatible iff

- *Test 1* and (2 or 3) hold for all pairs of property-subproperty in $\mathcal{O}$, and
- *Tests 4-11* hold for all pairs of property-subproperty in $\mathcal{O}$.

What to do if not compatible

- Guidelines for fixing a flaw, with one or more options for revision
 - “raising a warning” denotes that it is not a logical error but an ontological one
 - “forcing” a revision indicates there is a logical error that must be fixed in order to have a consistent ontology with satisfiable classes
 - “propose” indicates suggestions how the flaw can be best revised

Revisions (selection)

- A. If Test 1 fails, raise a warning “domain and range restrictions of either R or S are in conflict with the property hierarchy”, and propose to
- Change the object property hierarchy, i.e., either remove $S \sqsubseteq R$ and add $R \sqsubseteq S$ or add $S \equiv R$ to $\mathcal{O}$, or
 - Change domain and range restrictions of R and/or S , or
 - If the test on the domains fails, then propose a new axiom $R \sqsubseteq D'_R \times R_R$, where $D'_R \equiv D_R \sqcap D_S$ (and similarly when Test 1 fails on the range).
- B. ...
- C. Run *SubProS* again if any changes have been made in steps A or B, and record changes in the hierarchy (to be used in step I).

BioTop's inconsistent 'has process role'

'has process role' in BioTop (v. June 17, 2010) is inconsistent.

Relevant axioms are:

'has process role' $\sqsubseteq$ 'temporally related to' (E.1)

'has process role' $\sqsubseteq$ 'processual entity' $\times$ role (E.2)

'temporally related to' $\sqsubseteq$

'processual entity' $\sqcup$ quality $\times$

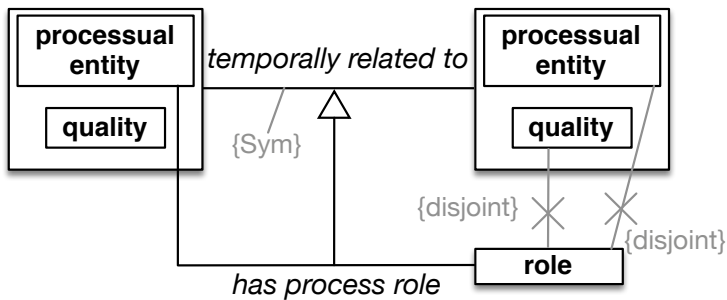
'processual entity' $\sqcup$ quality (E.3)

role $\sqsubseteq \neg$ quality (E.4)

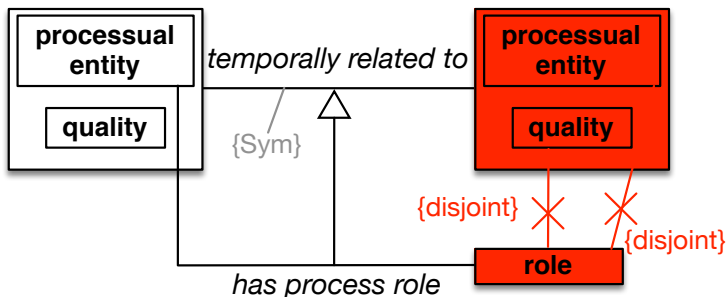
role $\sqsubseteq \neg$ 'processual entity' (E.5)

Sym('temporally related to') (E.6)

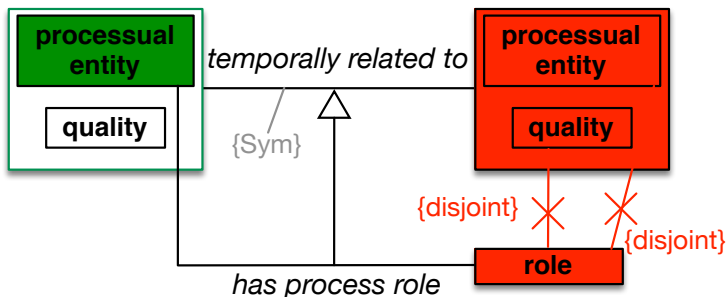
Diagrammatically



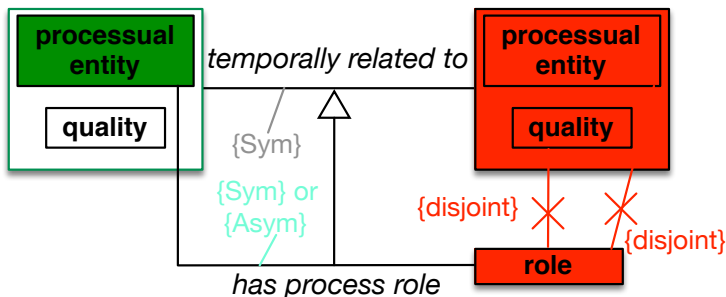
test 1 fails, test 2 passes, test 5 fails



test 1 fails, test 2 passes, test 5 fails



test 1 fails, test 2 passes, test 5 fails

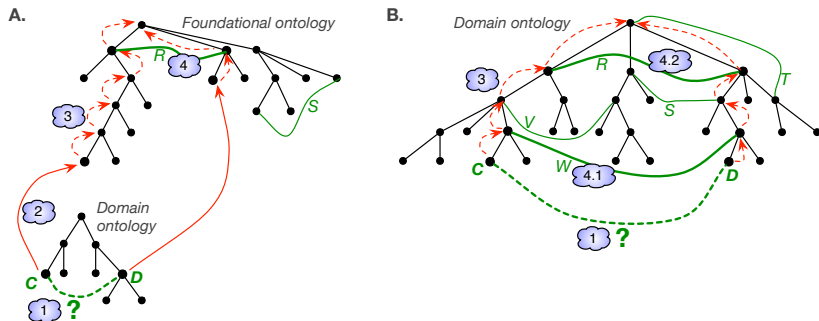


BioTop's inconsistent 'has process role'

Use *SubProS* to isolate the flaw:

- Test 1: fail, because $R_{\text{hasprocessrole}} \sqsubseteq R_{\text{temporallyrelatedto}}$ is false, as the ranges (see E.2 cf. E.3) are disjoint (see E.4, E.5) and therewith 'has process role' is inconsistent;
- Test 2 and 3: pass.
- Test 4: not applicable.
- Test 5: fail, because $\mathcal{O}$ does not contain $\text{Sym}(\text{'has process role'})$.
- Test 6–11: not applicable.

Avoid it: GENERATOR, the Guided ENTity reuse and class Expression geneRATOR

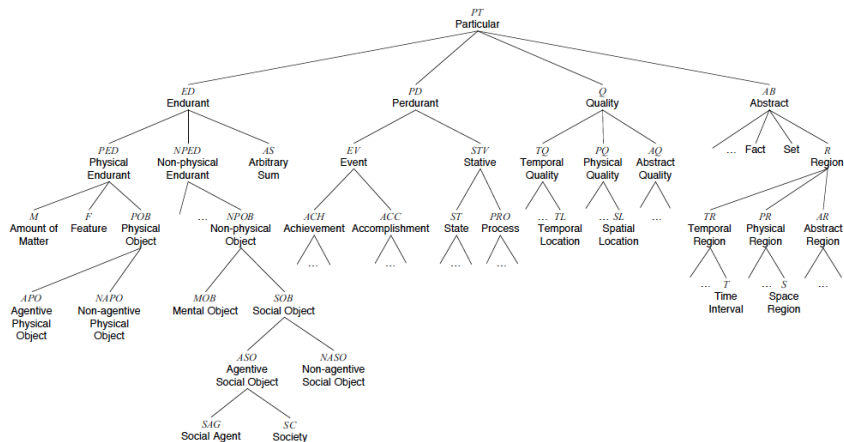


Option A (with a reasoner, a taxonomy of part-whole relations and DOLCE) is instantiated as FORZA (Foundational Ontology and Reasoner-enhanced axiomatization) [Keet et al.(2013)]

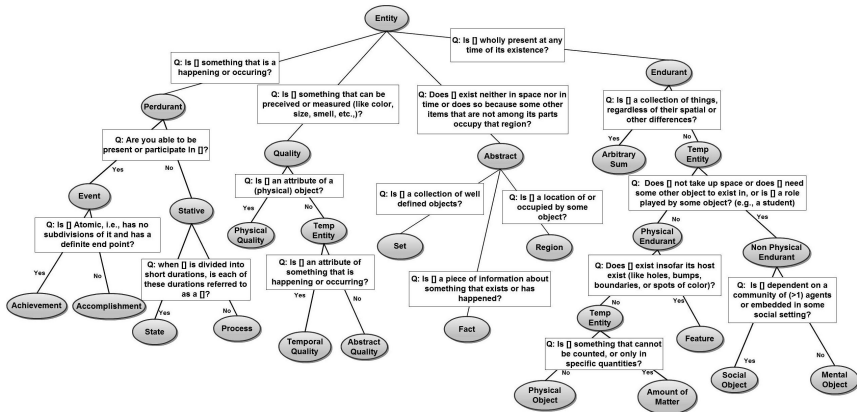
So you want to use a foundational ontology?

- Find the right one for you:
 - Mostly: a foundational ontology or module thereof is better than none
 - Ontological and non-ontological parameters that determine the outcome (e.g., ONSET [Khan and Keet(2012)])
- Use it:
 - Typically: import, don't extend
 - Align your top entities to FO entities (D3 [Keet et al.(2013)], BFO classifier [Bernabé et al.(2023)]) and try to reuse or refine properties as much as possible

Outline of DOLCE categories



D3



The African Wildlife Ontology and DOLCE

- Where does Plant fit in DOLCE?
- Giraffes drink Water: where should we put Water?
- Impalas run (fast). Where should we put Running?
- Lions eat impalas, and in the process, the impalas die. Where should we put Death?
- Generic examples of DOLCE's 'leaf' categories: see Table 1, p21 in the D18.pdf

The African Wildlife Ontology and DOLCE

- Where does Plant fit in DOLCE?
 - as a subtype of Non-Agentive Physical Object
- Giraffes drink Water: where should we put Water?
- Impalas run (fast). Where should we put Running?
- Lions eat impalas, and in the process, the impalas die. Where should we put Death?
- Generic examples of DOLCE's 'leaf' categories: see Table 1, p21 in the D18.pdf

The African Wildlife Ontology and DOLCE

- Where does Plant fit in DOLCE?
 - as a subtype of Non-Agentive Physical Object
- Giraffes drink Water: where should we put Water?
 - as a subtype of Amount of Matter
- Impalas run (fast). Where should we put Running?
- Lions eat impalas, and in the process, the impalas die. Where should we put Death?
- Generic examples of DOLCE's 'leaf' categories: see Table 1, p21 in the D18.pdf

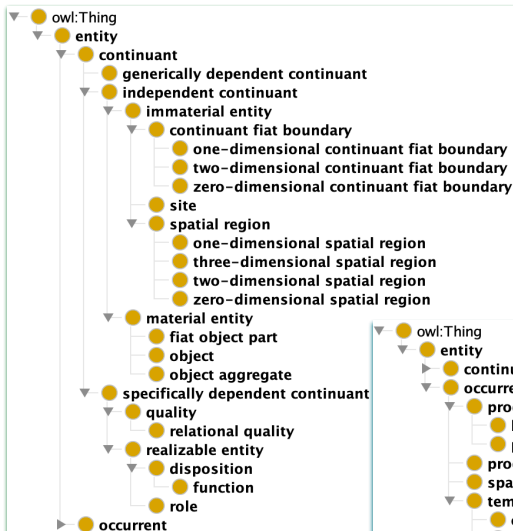
The African Wildlife Ontology and DOLCE

- Where does Plant fit in DOLCE?
 - as a subtype of Non-Agentive Physical Object
- Giraffes drink Water: where should we put Water?
 - as a subtype of Amount of Matter
- Impalas run (fast). Where should we put Running?
 - as a subtype of Process
- Lions eat impalas, and in the process, the impalas die. Where should we put Death?
- Generic examples of DOLCE's 'leaf' categories: see Table 1, p21 in the D18.pdf

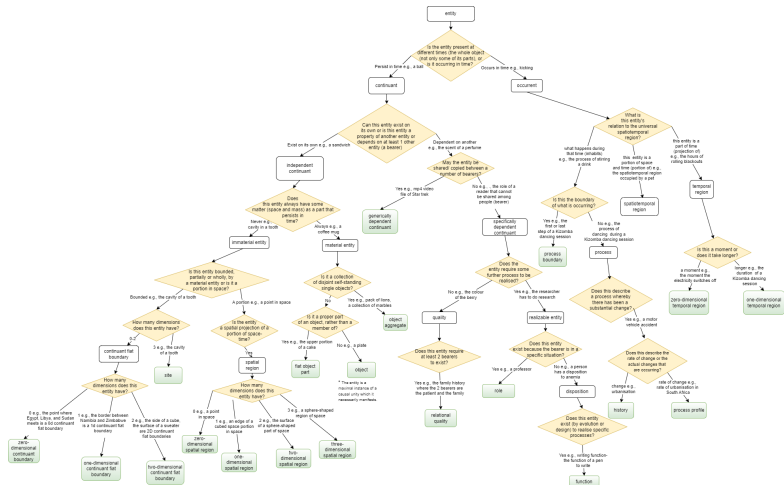
The African Wildlife Ontology and DOLCE

- Where does Plant fit in DOLCE?
 - as a subtype of Non-Agentive Physical Object
- Giraffes drink Water: where should we put Water?
 - as a subtype of Amount of Matter
- Impalas run (fast). Where should we put Running?
 - as a subtype of Process
- Lions eat impalas, and in the process, the impalas die. Where should we put Death?
 - as a subtype of Achievement...
- Generic examples of DOLCE's 'leaf' categories: see Table 1, p21 in the D18.pdf

BFO Taxonomy



A decision diagram



and a tool: <https://github.com/mkeet/BF02DecisionDiagram>

Parts and wholes: which relation?

- Is Sherbrooke a more specific instance of Québec Province, or a part of it?

Parts and wholes: which relation?

- Is Sherbrooke a more specific instance of Québec Province, or a part of it?
- Is a tunnel part of the mountain?

Parts and wholes: which relation?

- Is Sherbrooke a more specific instance of Québec Province, or a part of it?
- Is a tunnel part of the mountain?
- What is the difference, if any, between how Cell nucleus and Cell are related and how Receptor and Cell wall?

Parts and wholes: which relation?

- Is Sherbrooke a more specific instance of Québec Province, or a part of it?
- Is a tunnel part of the mountain?
- What is the difference, if any, between how Cell nucleus and Cell are related and how Receptor and Cell wall?
- All Canadian provinces share a border with something that is not Canada (e.g., Québec borders with the USA and the sea, Nunavut with the sea); Kansas is surrounded by only other USA states. How to represent this differences?

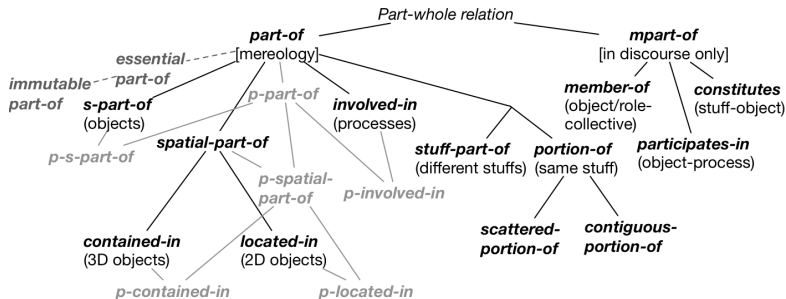
Parts and wholes: which relation?

- Is Sherbrooke a more specific instance of Québec Province, or a part of it?
- Is a tunnel part of the mountain?
- What is the difference, if any, between how Cell nucleus and Cell are related and how Receptor and Cell wall?
- All Canadian provinces share a border with something that is not Canada (e.g., Québec borders with the USA and the sea, Nunavut with the sea); Kansas is surrounded by only other USA states. How to represent this differences?
- Any difference between Brain part of Human and Hand part of Boxer? (assume boxers must have their own hands)

Parts and wholes: which relation?

- Is Sherbrooke a more specific instance of Québec Province, or a part of it?
- Is a tunnel part of the mountain?
- What is the difference, if any, between how Cell nucleus and Cell are related and how Receptor and Cell wall?
- All Canadian provinces share a border with something that is not Canada (e.g., Québec borders with the USA and the sea, Nunavut with the sea); Kansas is surrounded by only other USA states. How to represent this differences?
- Any difference between Brain part of Human and Hand part of Boxer? (assume boxers must have their own hands)
- Hand is part of musician, musician part of orchestra, but clearly, the musician's hands are not part of the orchestra. Is part-of then not transitive, or is there a problem with the example?

Try to be more precise with part-whole relations



pw hierarchy (aligned): <http://www.meteck.org/swdsont.html>

Sampling of range of methods and tools

- The automated reasoner & explanations (rudimentary baseline)

Sampling of range of methods and tools

- The automated reasoner & explanations (rudimentary baseline)
- **Logic-based:** Advocatus Diaboli, to add disjointness where there should be [Ferré and Rudolph(2012)], TDDonto2: test before you add the axiom [Davies et al.(2019)]
- **Ontology- and modelling-based:** OntoPartS, to select the right mereological or mereotopological relation [Keet et al.(2012)]
- **Heuristics-based:** OOPS! [Poveda-Villalón et al.(2012)]

References I



C. H. Bernabé, C. M. Keet, Z. C. Khan, and Z. Mahlaza.

A method to improve alignments between domain and foundational ontologies.

In *14th International Conference on Formal Ontology in Information Systems 2023 (FOIS'23)*, FAIA, page accepted. IOS Press, 2023.



Kieren Davies, C. Maria Keet, and Agnieszka Lawrynowicz.

More effective ontology authoring with test-driven development and the TDDonto2 tool.

International Journal on Artificial Intelligence Tools, 28(7):1950023, 2019.



Sebastien Ferré and Sebastian Rudolph.

Advocatus diaboli – exploratory enrichment of ontologies with negative constraints.

In A ten Teije et al., editors, *18th International Conference on Knowledge Engineering and Knowledge Management (EKAW'12)*, volume 7603 of *LNAI*, pages 42–56. Springer, 2012.

Oct 8-12, Galway, Ireland.



P. R. Fillottrani and C. M. Keet.

An analysis of commitments in ontology language design.

In B. Brodaric and F. Neuhaus, editors, *11th International Conference on Formal Ontology in Information Systems 2020 (FOIS'20)*, volume 330 of *FAIA*, pages 46–60. IOS Press, 2020.



Pablo R. Fillottrani and C. Maria Keet.

Patterns for heterogeneous tbox mappings to bridge different modelling decisions.

In E. Blomqvist et al., editors, *Proceeding of the 14th Extended Semantic Web Conference (ESWC'17)*, volume 10249 of *LNCS*, pages 371–386. Springer, 2017.

30 May - 1 June 2017, Portoroz, Slovenia.

References II



C. M. Keet, M. C. Suárez-Figueroa, and M. Poveda-Villalón.

Pitfalls in ontologies and tips to prevent them.

In A. Fred, J. L. G. Dietz, K. Liu, and J. Filipe, editors, *Knowledge Discovery, Knowledge Engineering and Knowledge Management: IC3K 2013 Selected papers*, volume 454 of *CCIS*, pages 115–131. Springer, Berlin, 2015.



C. Maria Keet.

Detecting and revising flaws in OWL object property expressions.

In A. ten Teije et al., editors, *18th International Conference on Knowledge Engineering and Knowledge Management (EKAW'12)*, volume 7603 of *LNAI*, pages 252–266. Springer, 2012.

URL <http://www.meteck.org/files/EKAW12subProsChains.pdf>.

Oct 8-12, Galway, Ireland.



C. Maria Keet and Alessandro Artale.

Representing and reasoning over a taxonomy of part-whole relations.

Applied Ontology – Special issue on Ontological Foundations for Conceptual Modeling, 3(1-2):91–110, 2008.

URL http://www.meteck.org/files/A007_pw_AK07.pdf.



C. Maria Keet, Francis C. Fernández-Reyes, and Annette Morales-González.

Representing mereotopological relations in OWL ontologies with ONTOPARTS.

In E. Simperl et al., editors, *Proceedings of the 9th Extended Semantic Web Conference (ESWC'12)*, volume 7295 of *LNCS*, pages 240–254. Springer, 2012.

URL <http://www.meteck.org/files/OnToPartSESWC12.pdf>.

29-31 May 2012, Heraklion, Crete, Greece.

References III



C. Maria Keet, M. Tahir Khan, and Chiara Ghidini.

Guided ENTITY reuse and class Expression geneRATOR.

In Richard Benjamins, Mathieu d'Aquin, Andrew Gordon, , and José Manuel Gómez-Pérez, editors, *Seventh International Conference on Knowledge Capture (K-CAP'13)*, page a26 (poster&demo). ACM, 2013. 23-26 June 2013, Banff, Canada.



Zubeida Khan and C. Maria Keet.

ONSET: Automated foundational ontology selection and explanation.

In A. ten Teije et al., editors, *18th International Conference on Knowledge Engineering and Knowledge Management (EKAW'12)*, volume 7603 of *LNAI*, pages 237–251. Springer, 2012. Oct 8-12, Galway, Ireland.



María Poveda-Villalón, Mari Carmen Suárez-Figueroa, and Asunción Gómez-Pérez.

Validating ontologies with OOPS!

In A. ten Teije et al., editors, *18th International Conference on Knowledge Engineering and Knowledge Management (EKAW'12)*, volume 7603 of *LNAI*, pages 267–281, Germany, 2012. Springer. Oct 8-12, Galway, Ireland.



B. Smith.

Beyond concepts, or: Ontology as reality representation.

In A. Varzi and L. Vieu, editors, *Formal Ontology and Information Systems. Proceedings of the Third International Conference (FOIS'04)*, pages 73–84. Amsterdam: IOS Press, 2004.

Thank you!

Questions?

- My textbook on ontology engineering (aimed at computer scientists)
- Free pdf (and slides and exercises) at <https://people.cs.uct.ac.za/~mkeet/OEbook/>
- Also available in paperback:

