Encoding biases' influences on development and use of ontologies in the life sciences

C. Maria Keet

Department of Computer Science, University of Cape Town, South Africa
mkeet@cs.uct.ac.za

Bio-ontologies COSI Keynote
Intelligent Systems for Molecular Biology Conference 2022
Madison, Wisconsin, USA, 10-14 July 2022

⇒ Different ways of representing the same meaningful unit

- level of detail desired or needed
- purpose of the ontology
- (and whether it is an ontology or actually another artefact)

⇒ Conflicting requirements can't be all met in the same (one type of) ontology

image source: https://www.flickr.com/photos/mentealterne/1224542289/

Outline

1. Context and motivation
2. Encoding biases
   - Ontology patterns
   - Examples of effects for ontology development
   - From patterns to styles
3. Effects in applications
4. Conclusions

Outline

Introduction

- Ontologies
  - For their own sake
  - For communication among humans
  - Used for many different ontology-driven information systems (database integration and linking, recommender systems, NLP, textbook annotation and search, question generation, Q&A systems, etc.)

Introduction

- Ontologies
  - For their own sake
  - For communication among humans
  - Used for many different ontology-driven information systems (database integration and linking, recommender systems, NLP, textbook annotation and search, question generation, Q&A systems, etc.)

$\Rightarrow$ Different reasons may affect how the knowledge is represented in the ontology

$\Rightarrow$ Historically called **encoding bias** [Uschold and Gruninger(1996)]

Historically, it's also separate from confounding factors

- Thus not about

  "Is it a concept? Is a universal?"

  "Is it subsumption? Is it parthood?"

image source: https://muppet.fandom.com/wiki/Super_Grover

Historically, it's also separate from confounding factors

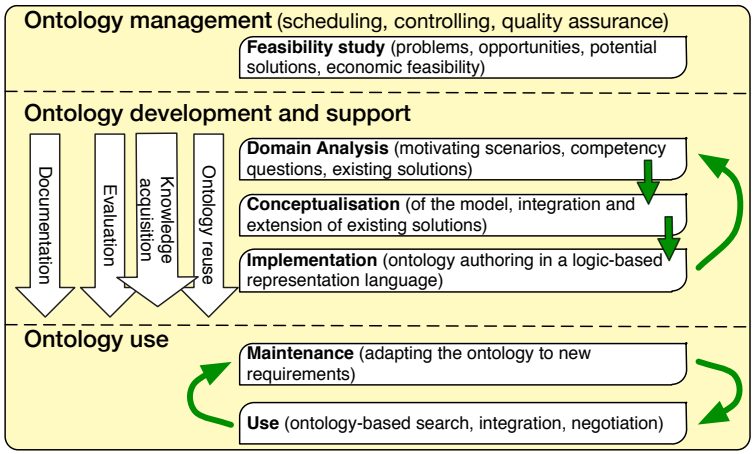
- Thus not about

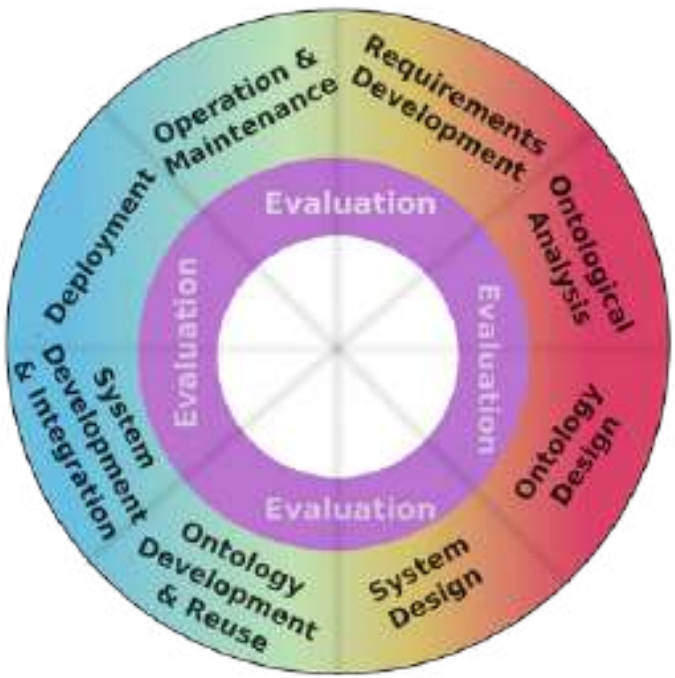
  "Is it a concept? Is a universal?"

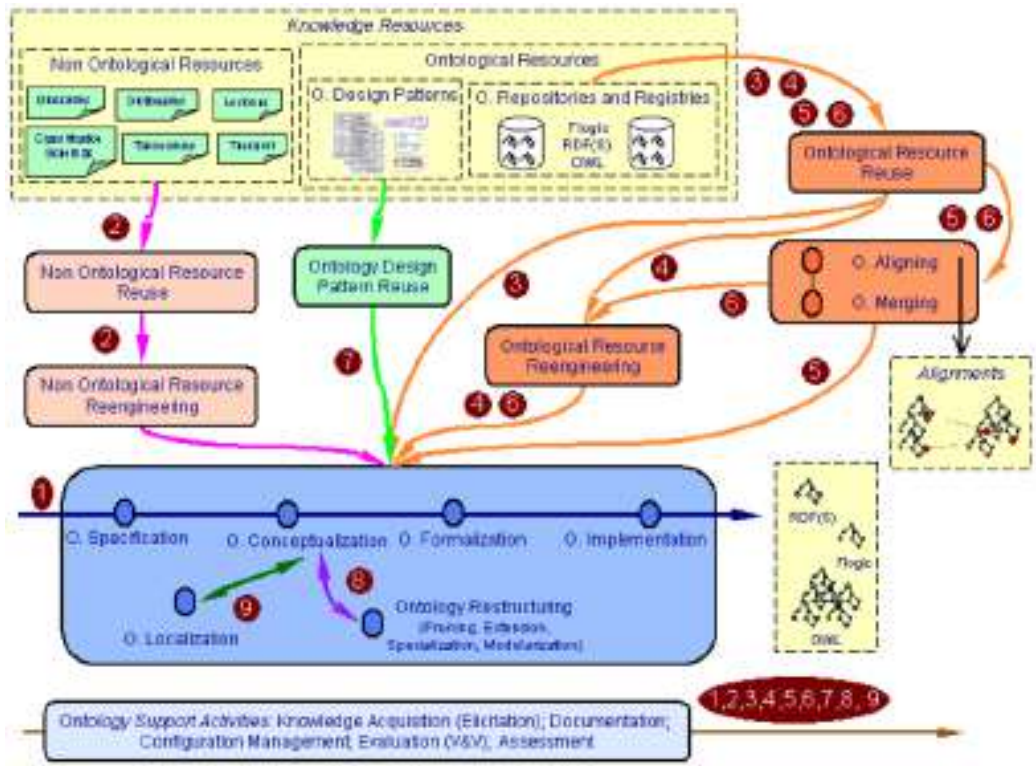
  "Is it subsumption? Is it parthood?"

- The knowledge/meaningful unit/piece of reality is assumed known and agreed upon
- And no interference of cognitive bias [Keet(2021)]
- Differently 'encoded' (i.e., formalised) due to purposes, language features, ...
- Noted then, but little done with it; more more recently

image source: `https://muppet.fandom.com/wiki/Super_Grover`







Ontology Summit 2013's lifecycle model; OE book; NeON.

Ontology development at the 'micro-level' level

- We need to get axioms into the ontology; also called *ontology authoring*

Ontology development at the 'micro-level' level

- We need to get axioms into the ontology; also called *ontology authoring*
- Uses micro-level guidelines, methods, and tools
  - Methods, such as OntoClean and OntoPartS to improve an ontology's quality
  - Tools to model, to reason, to debug, to integrate, to link to data
- Encoding biases creep in here

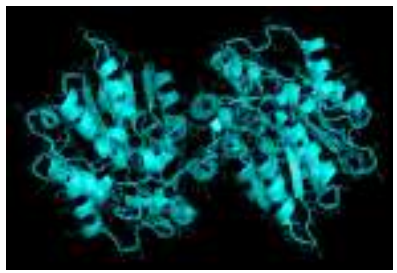
Examples

- (a) a class `Transformation` or (b) a relationship `transformsInto`?

https://pxhere.com/en/photo/823275
https://en.wikipedia.org/wiki/File:Enzyme_Polyneuridine_Aldehyde_Esterase.png
https://commons.wikimedia.org/wiki/File:Fancy_mice.jpg

Examples

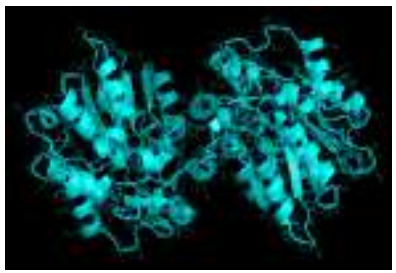
- (a) a class `Transformation` or (b) a relationship `transformsInto`?
- (a) Enzyme `is a` Protein or (b) Enzyme `has bearer` Protein?



https://pxhere.com/en/photo/823275
https://en.wikipedia.org/wiki/File:Enzyme_Polyneuridine_Aldehyde_Esterase.png
https://commons.wikimedia.org/wiki/File:Fancy_mice.jpg

Examples

- (a) a class `Transformation` or (b) a relationship `transformsInto`?
- (a) Enzyme `is a` Protein or (b) Enzyme `has bearer` Protein?
- (a) Mouse `hasColour` Colour or (b) Mouse `hasQuality` Colour that `hasQuale` Physical Region?

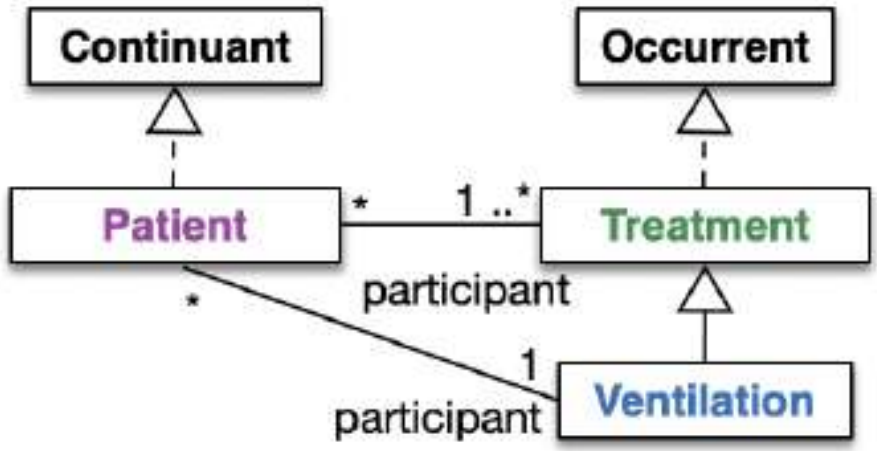




https://pxhere.com/en/photo/823275
https://en.wikipedia.org/wiki/File:Enzyme_Polyneuridine_Aldehyde_Esterase.png
https://commons.wikimedia.org/wiki/File:Fancy_mice.jpg

Comparing examples

Three different patterns with a purpose bias

- Top: biased toward a science focus
- Middle: conceptual data modelling influence or purpose (e.g., models for FAIR data sets)
- Bottom: a thesaurus-like approach useful for document annotation



Ventilation Therapy
NT Ventilator Support
BT Ventilation
UF Mechanical Ventilation
RT Patient

Key questions

- Does it matter which way it is put in the ontology?
- If so: how/where/why, and is one always better than the other(s)?

Key questions

- Does it matter which way it is put in the ontology?
- If so: how/where/why, and is one always better than the other(s)?
- Which patterns are there for representing the 'same thing'?
- Are they really equivalent alternatives of the same meaningful unit?
- Do certain patterns co-occur, to make up a style of representing something?

Key questions

- Does it matter which way it is put in the ontology?
- If so: how/where/why, and is one always better than the other(s)?
- Which patterns are there for representing the 'same thing'?
- Are they really equivalent alternatives of the same meaningful unit?
- Do certain patterns co-occur, to make up a style of representing something?
- How does all this affect use and reuse of ontologies? (w.r.t., e.g., ontology imports, FO use, data integration, literature annotation, ...)

Outline

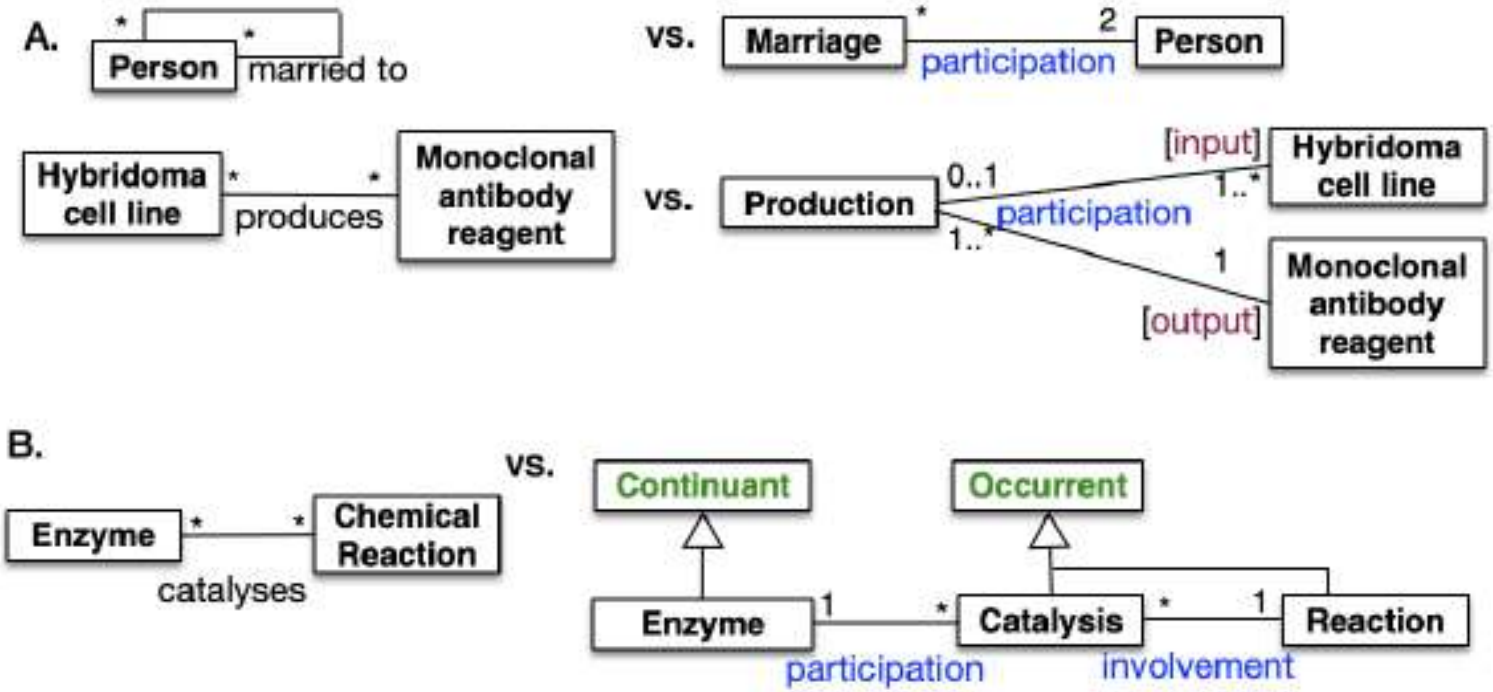
Outline

Which patterns where?

- Identify common modelling patterns (mainly from [Fillottrani and Keet(2017), Scharffe and Fensel(2008), Keet(2021)])
- Devise formalisation of ontology pattern (OP) and of OP alignment
- Formalise those common patterns and alignments
- Algorithm design for *automated finding* of the patterns and *checking correctness* of a possible pattern-based alignment or substitution

Fillottrani, P.R., Keet, C.M. Patterns for Heterogeneous TBox Mappings to Bridge Different Modelling Decisions. Proc. of ESWC'17. Springer LNCS 10249, 371-386.

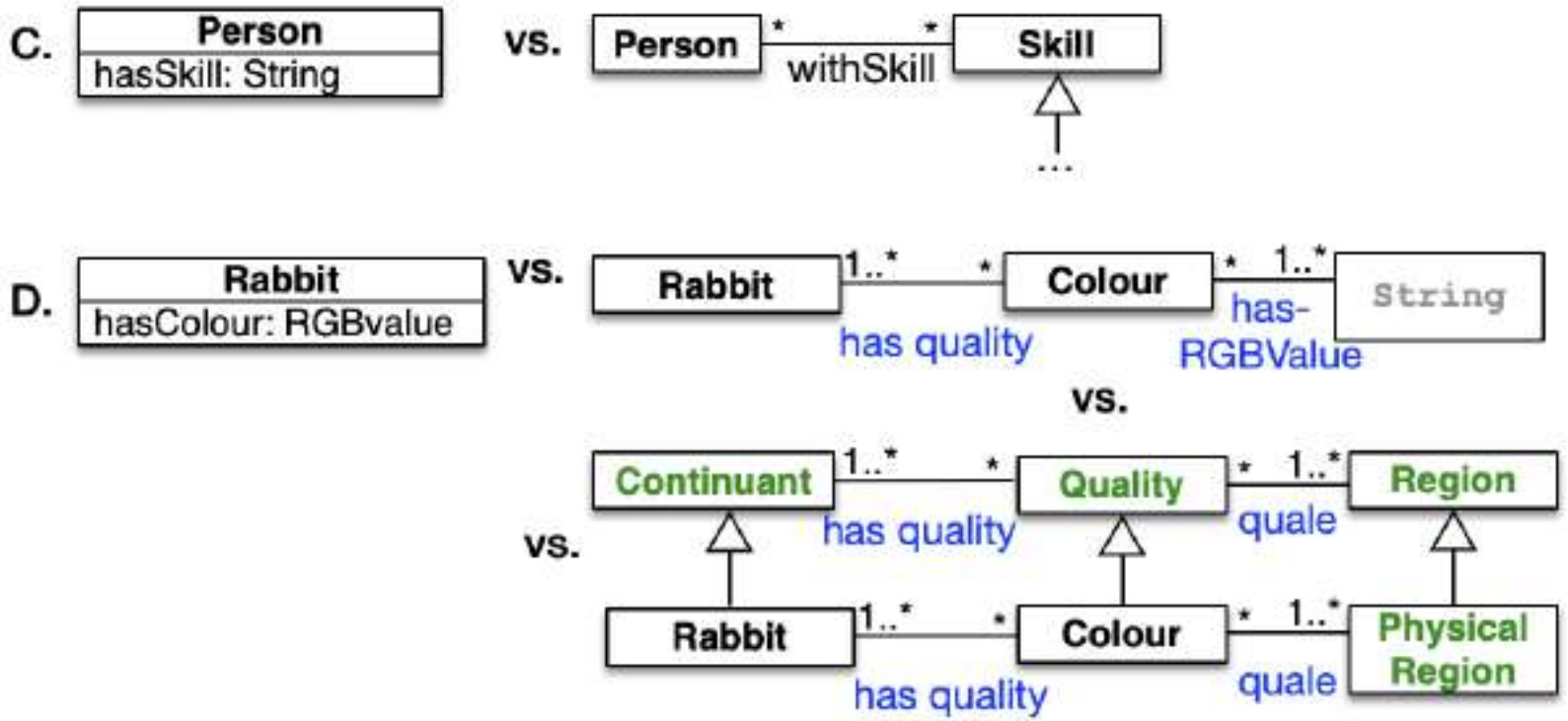
To reify or not to reify?



Considerations:

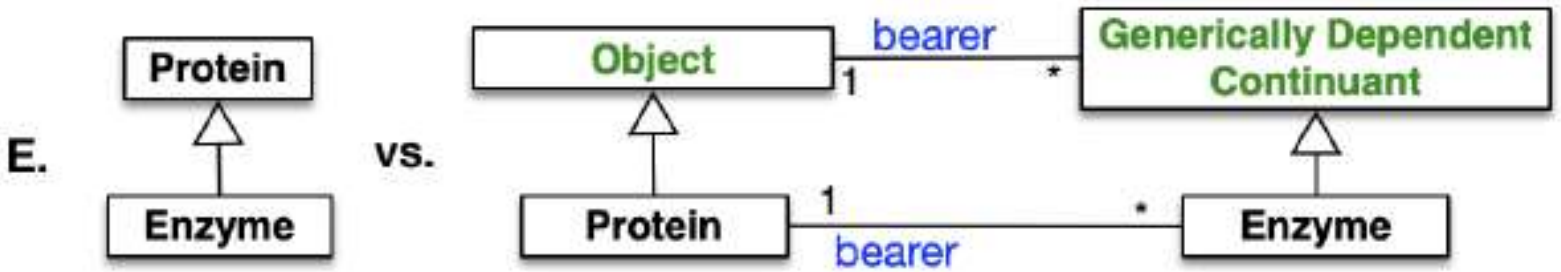
- Are the more precise cardinality constraints needed?
- The foundational ontology to align it to

Attributions as attributes or as classes?



- Pros and cons of data properties
- What kind of thing is the one hidden in the data property?
- The foundational ontology to align it to

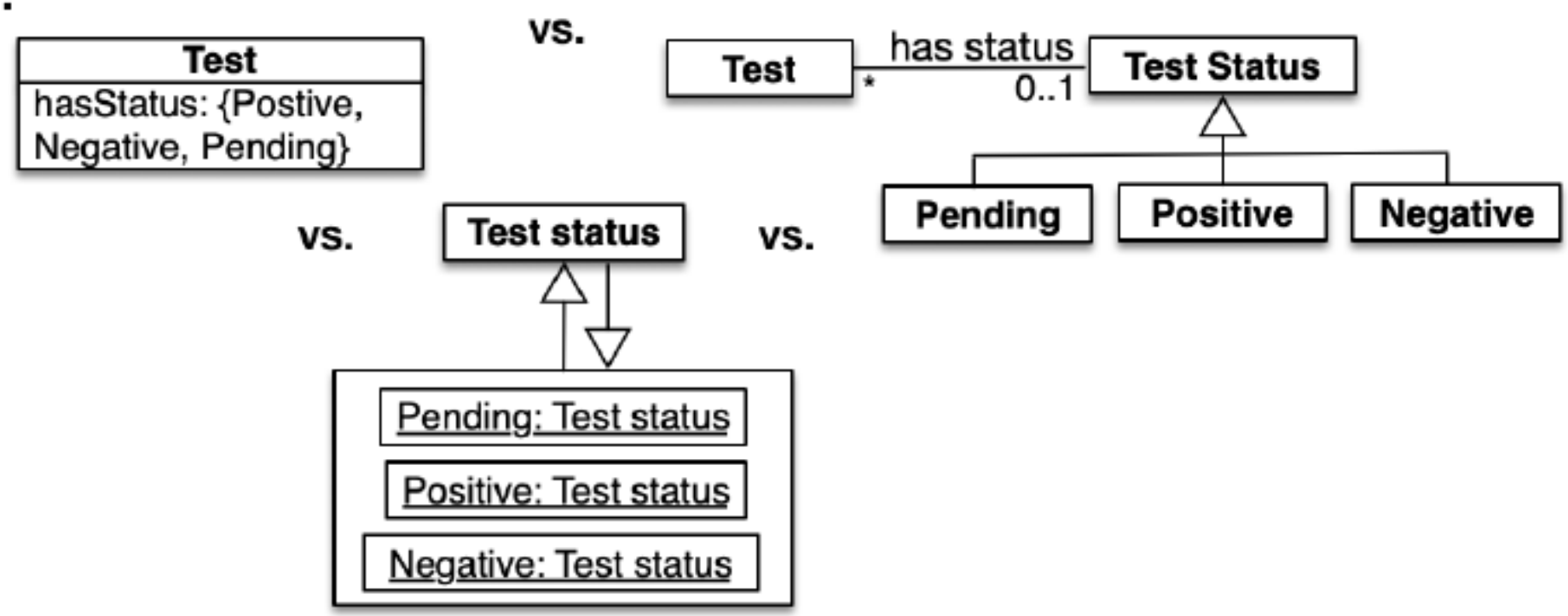
Roles or subclasses?



- Foundational ontology approach uses the bearer ('inheres in'); conceptual data models the subclassing
- Property inherence wrt subclasses (e.g., Protein has as property molecular weight and secondary structure, and enzyme too)
- The foundational ontology to align it to

Values, one-off with individuals, or subclasses?

F.



- Data property versus 'something else' (+ pros and cons of data properties, again)
- Individuals in an ontology? One-off/nominals?
- Extensibility and stability of the representation
- Computational costs

Pattern management, first step: formalise them

- Language of pattern instantiation (OWL, some DL, ...)
- Language for patterns with vocabulary $\mathcal{V}$, meta-level (second-order) elements (or stereotypes)
- Ontology pattern, with name, elements from $\mathcal{V}$, pattern axiom components, pattern's full formalisations; e.g.:
  - pattern name: *basic all-some*
  - pattern elements: $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{R}_1$
  - pattern axiom components: $\sqsubseteq$, $\exists$
  - pattern's full formalisation $\mathcal{C}_1 \sqsubseteq \exists \mathcal{R}.\mathcal{C}_2$

  Example instantiation: $\texttt{Butterfly} \sqsubseteq \exists \texttt{derivedFrom.Chrysalis}$

Fillottrani, P.R., Keet, C.M. Patterns for Heterogeneous TBox Mappings to Bridge Different Modelling Decisions. Proc. of ESWC'17. Springer LNCS 10249, 371-386.

Class vs. Object Property (case A)

- *alignment pattern name*: *class-OP*
- *pattern elements*: $\mathcal{C}_1, \mathcal{C}_2, \mathcal{R}_1$ from O, $\mathcal{C}'_3, \mathcal{C}'_4, \mathcal{C}'_5, \mathcal{R}'_2, \mathcal{R}'_3$ from O'
- *alignment patterns' contexts*:
  - *pattern P in O* (the one on the left): $\exists \mathcal{R}_1.\mathcal{C}_2 \sqsubseteq \mathcal{C}_1$ and $\exists \mathcal{R}_1^{-}.\mathcal{C}_1 \sqsubseteq \mathcal{C}_2$;
  - *pattern P' in O'* (the one on the right): $\exists \mathcal{R}'_2.\mathcal{C}'_4 \sqsubseteq \mathcal{C}'_3$, $\exists \mathcal{R}'^{-}_2.\mathcal{C}'_3 \sqsubseteq \mathcal{C}'_4$, $\exists \mathcal{R}'_3.\mathcal{C}'_5 \sqsubseteq \mathcal{C}'_3$, $\exists \mathcal{R}'^{-}_3.\mathcal{C}'_3 \sqsubseteq \mathcal{C}'_5$, $\mathcal{C}'_3 \sqsubseteq (\exists \mathcal{R}'_2)$, and $\mathcal{C}'_3 \sqsubseteq (\exists \mathcal{R}'_3)$.

Outline

Complications

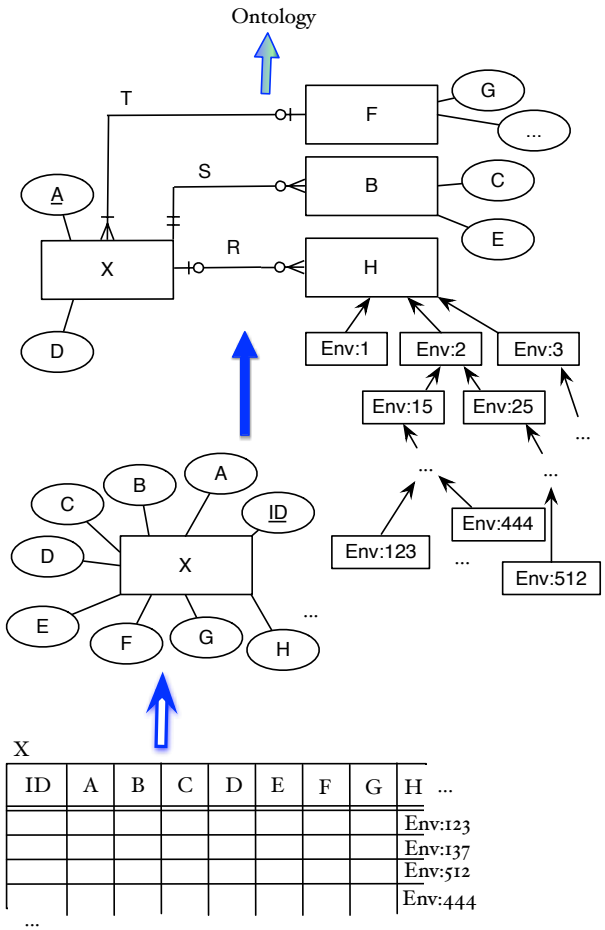
- Different representation decisions in different ontologies impedes:
  - ontology interoperability and alignment
  - ontology reuse
  - ontology development: in automation and, e.g., CQ translations
  - deployment in ontology-driven information systems

Ontology development bottom-up vs few core relations from FOs

- NLP-based approaches:
  - Based on SVO phrases in English – proliferation of candidate OPs
  - Mistaken belief of 'verb will become name of object property'

Ontology development bottom-up vs few core relations from FOs

- NLP-based approaches:
  - Based on SVO phrases in English – proliferation of candidate OPs
  - Mistaken belief of 'verb will become name of object property'
- From conceptual data model of a database (or its schema) to an ontology
  - Favours flat structure, many properties
  - Little automated reasoning for 'debugging'



CQs example

- *Which [CE1] [OPE] [CE2]?* as a so-called "archetype" [Ren et al.(2014)]

  For/With axiom type $CE_1 \sqsubseteq \exists OPE.CE_2$

  fill in the slots; e.g., "Which [animal] [eats] [fruit]?"

  axiom to add or to verify: $\texttt{Animal} \sqsubseteq \exists \texttt{eats.Fruit}$

Wisniewski, D., Potoniec, J., Lawrynowicz, A., Keet, C.M. Analysis of Ontology Competency Questions and their Formalisations in SPARQL-OWL. Journal of Web Semantics, 2019, 59:100534.

CQs example

- *Which [CE1] [OPE] [CE2]?* as a so-called "archetype" [Ren et al.(2014)]

  For/With axiom type $CE_1 \sqsubseteq \exists OPE.CE_2$

  fill in the slots; e.g., "Which [animal] [eats] [fruit]?"

  axiom to add or to verify: $\texttt{Animal} \sqsubseteq \exists \texttt{eats.Fruit}$
- Can be, and has been, modelled differently
- Practically: m:n relation between CQ sentence pattern and SPARQL-OWL query & fomalisation

$\Rightarrow$ Need for decoupling of question from axiom pattern [Wisniewski et al.(2019)]

- Rather: 'entity chunks' and 'predicate chunks' in a sentence; decide later on whether EC becomes a C, CE or a OPE
- More management of CQs with axioms and queries

Wisniewski, D., Potoniec, J., Lawrynowicz, A., Keet, C.M. Analysis of Ontology Competency Questions and their Formalisations in SPARQL-OWL. Journal of Web Semantics, 2019, 59:100534.

CQs example (abridged notation)

What data are measured for gait assessment? // simple all-some in Dem@care
What software can perform [task x]? // is_executed_in [subclass of] IAO_0000064 in SWO
What information is clinically relevant for dedicated physical activities (i.e. exercises)?
// all-some + subclass of ExerciseSummary Dem@care
What software can read a .cel file? // with has_specified_data_input and
has_format_specification in SWO

"What EC1 PC1 EC2": [
"SELECT * WHERE { ?placeholder_PPx1 <subClassOf> :b1 ; <subClassOf> :b0 . :b0
<onProperty> :b2 ; <someValuesFrom> ?placeholder_PPx2 . ?placeholder_PPx2
<subClassOf> :b3 }",
"SELECT * WHERE { ?placeholder_PPx1 <subClassOf> :b0 . :b0 <onProperty> :b1 ;
<someValuesFrom> :b2 . ?placeholder_PPx1 <subClassOf> :b3 }"
"SELECT * WHERE { ?x0 <subClassOf> :b5 ; <subClassOf> :b0 . :b0 <onProperty> :b6 ;
<someValuesFrom> :b1 . :b1 <intersectionOf> :b2 . :b2 <first> :b7 ; <rest> :b3 .
:b3 <first> :b4 . :b4 <onProperty> :b8 ; <someValuesFrom> :b9 . :b3 <rest> <nil> }",],

Potoniec, J., Wisniewski, D., Lawrynowicz, A., Keet, C.M. Dataset of Ontology Competency Questions to SPARQL-OWL Queries Translations. Data in Brief, 2020, 29: 105098. https://github.com/CQ2SPARQLOWL/Dataset

CQs example axiom patterns

- CQ *What EC1 PC1 EC2?* with several formalisation options used:

  $C_1 \sqsubseteq \exists R.C_2$

  $C_1 \sqsubseteq \exists R.C_2$ and $C_2 \sqsubseteq C_3$

  $C_1 \sqsubseteq \exists R.C_2$ and $C_2 \sqsubseteq \exists S.C_3$

Keet, C.M., Mahlaza, Z., Antia, M.-J. CLaRO: a Controlled Language for Authoring Competency Questions. MTSR'19. Springer CCIS vol. 1057, 3-15.
Antia, M.-J., Keet, C.M. Assessing and Enhancing Bottom-up CNL Design for Competency Questions for Ontologies. CNL'21. ACL, 1.11.

CQs example axiom patterns

- CQ *What EC1 PC1 EC2?* with several formalisation options used:

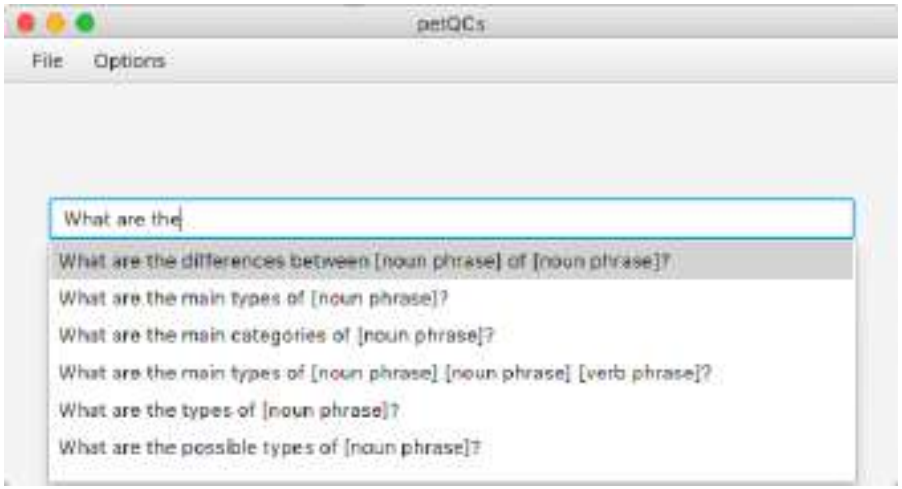
  $C_1 \sqsubseteq \exists R.C_2$

  $C_1 \sqsubseteq \exists R.C_2$ and $C_2 \sqsubseteq C_3$

  $C_1 \sqsubseteq \exists R.C_2$ and $C_2 \sqsubseteq \exists S.C_3$

- CQ template language to assist: CLaRO [Keet et al.(2019)]

  `https://github.com/mkeet/CLaRO`



Keet, C.M., Mahlaza, Z., Antia, M.-J. CLaRO: a Controlled Language for Authoring Competency Questions. MTSR'19. Springer CCIS vol. 1057, 3-15.
Antia, M.-J., Keet, C.M. Assessing and Enhancing Bottom-up CNL Design for Competency Questions for Ontologies. CNL'21. ACL, 1.11.

What to do with them?

- Swap one for the other (ontology reuse, ontology-driven CDM, etc.)
- Heterogeneous alignments in case of ontology reuse
- Don't add both to your ontology
- Whichever you choose, use it consistently throughout the ontology

Outline

Questions

- If those decisions are made systematically, a ‘style’ emerges
- Once we know the styles, converters can be developed

Fillottrani, P.R., Keet, C.M.. Dimensions Affecting Representation Styles in Ontologies. KGSWC’19. Springer CCIS vol. 1029, 186-200.

Questions

- If those decisions are made systematically, a 'style' emerges
- Once we know the styles, converters can be developed

$\Rightarrow$ What are the styles?

$\rightarrow$ What are the dimensions contributing to a style?

$\rightarrow$ What are the parameters' values/variations to choose from?

Fillottrani, P.R., Keet, C.M.. Dimensions Affecting Representation Styles in Ontologies. KGSWC'19. Springer CCIS vol. 1029, 186-200.

Indicative definition

Definition (Representation style, refined version)

A *representation style* is set of features used for representing a particular piece of conceptualisation or (understanding of) reality, for which there may be different (meaningfully equivalent, but not necessarily logically equivalent) ways to represent it in a logic that supports the style. The representation style is a justification-based positive design decision in at least one scenario.

Structure of the inventarisation

- Two 'levels': 1) predominantly theoretical (on the type of artefact one aims to create irrespective of the practical considerations), 2) engineering decisions (2a) tooling, 2b) language).
- For each dimension:
  - label/name/key phrase
  - description
  - examples
  - ≥ 2 'traits' with name + description
- Result: 10 dimensions with 28 traits

Theoretical (1/2)

- *Degree of adherence to ontological principles in representing the knowledge*. extremes: the "foundational ontology way" and "applied way" (arguably, a logic-based conceptual data model).

  `Theoretical:` predominantly or entirely with ontological principles, such as qualities, reification of processes, inherence of roles, no data properties.

  `Applied:` predominantly or entirely with decisions for applications, such as attributes/data properties and processes as relations.

  `Mixed:` the ontology contains both such decisions.

Theoretical (2/2)

- *Granularity of relations*. limit oneself to a few core relations, such as parthood, participation, causality, and membership vs. declaring relations for every subtle distinction, such as a structural parthood as subtype of part-of and a celebrates that refines participates in.

  `Parsimony:` when there are no refinements of the basic relations.

  `Abundance:` when there are refinements on the basic relations or when there are domain-specific relations, or both.

Engineering (1/3)

- *Modular vs monolithic*; with or without module management

  `Monolithic:` there is one file, with no imports or mergers.

  `Modul., ext.:` at least one ontology is imported or merged such that the import has maintained its IRI; hence, it is associated with the process of ontology reuse.

  `Modul., int.:` at least one ontology is imported, such that it is associated with the process of decomposition of a domain.

Engineering (2/3)

- *General Concept Inclusions (GCIs) vs only named entities* on the left-hand side of the inclusion;

  `Explicit GCIs:` they have been declared by the modeller, such as $\texttt{Property} \sqcap \exists \texttt{propertyOf}.\texttt{Presential} \sqsubseteq \texttt{Presential}$ in gfo-basic.

  `Hidden GCIs:` they have not been declared by the modeller explicitly, but they are there indirectly through other axioms: there is a pair $\texttt{A} \equiv \texttt{C}$ and $\texttt{A} \sqsubseteq \texttt{D}$ and `C` and `D` are complex class expressions[1].

  `No GCIs:` they have not been declared explicitly or implicitly.

Engineering (3/3)

- *Values/instances/classes interplay*. Representation of certain entities that may be deemed different kind of elements, depending on one's modelling viewpoint, practicalities, and which constructors are available in the language

  `Nominals:` as described in option 1; i.e., `Week` $\equiv$ {`Sunday`, ..., `Saturday`} where Sunday etc. are individuals.

  `Enumerated:` as described in option 2; i.e., the values of a data property `onWeekday` can be one of the values `Sunday`, ..., `Saturday`.

  `Class-inst.:` as described in option 3; i.e., they are all classes appropriately related, and a 'Sunday 6 January 2019' is an instance of `Sunday` etc.

  `Mixed:` any two or three appear in the ontology.

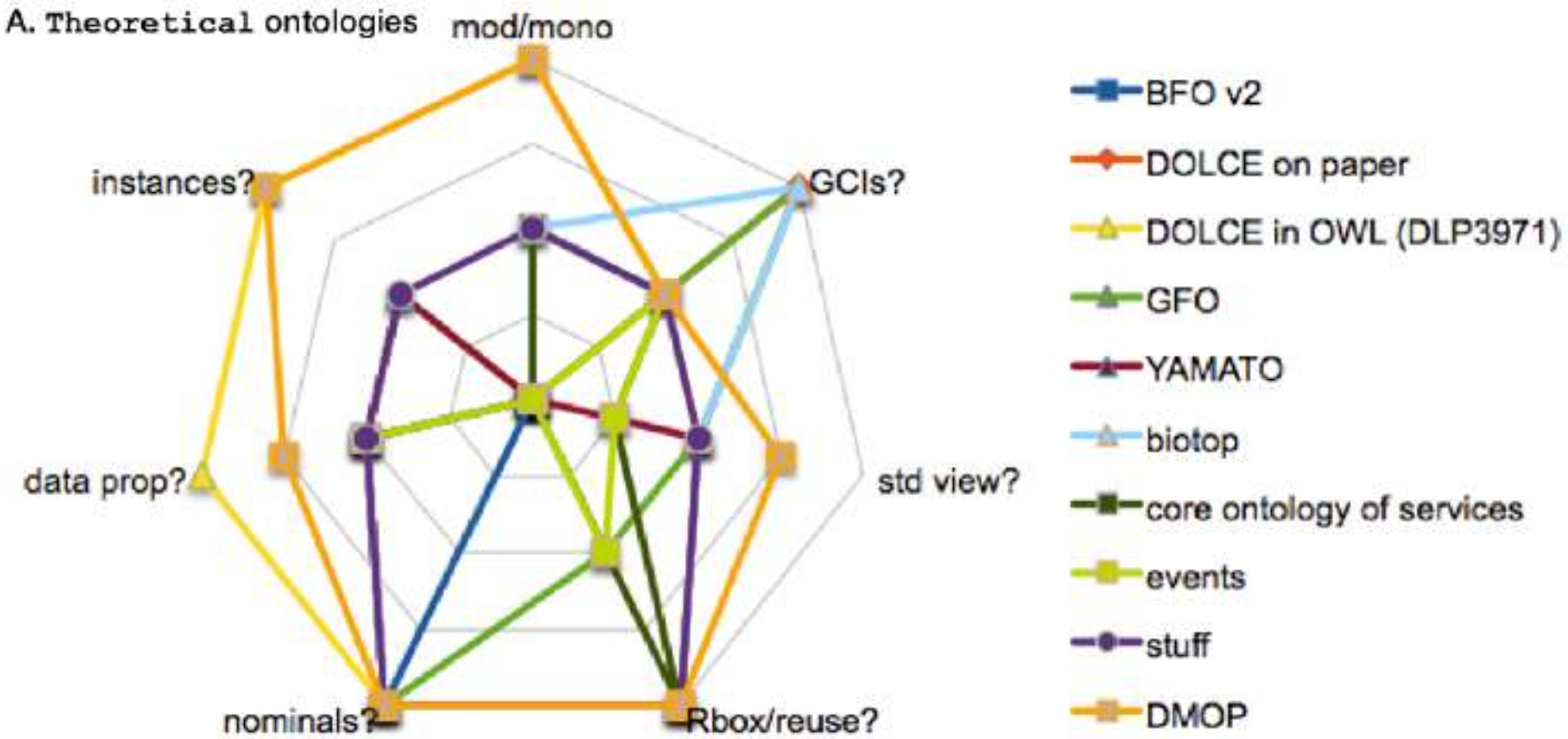
  `-:` not applicable.

Question & Procedure

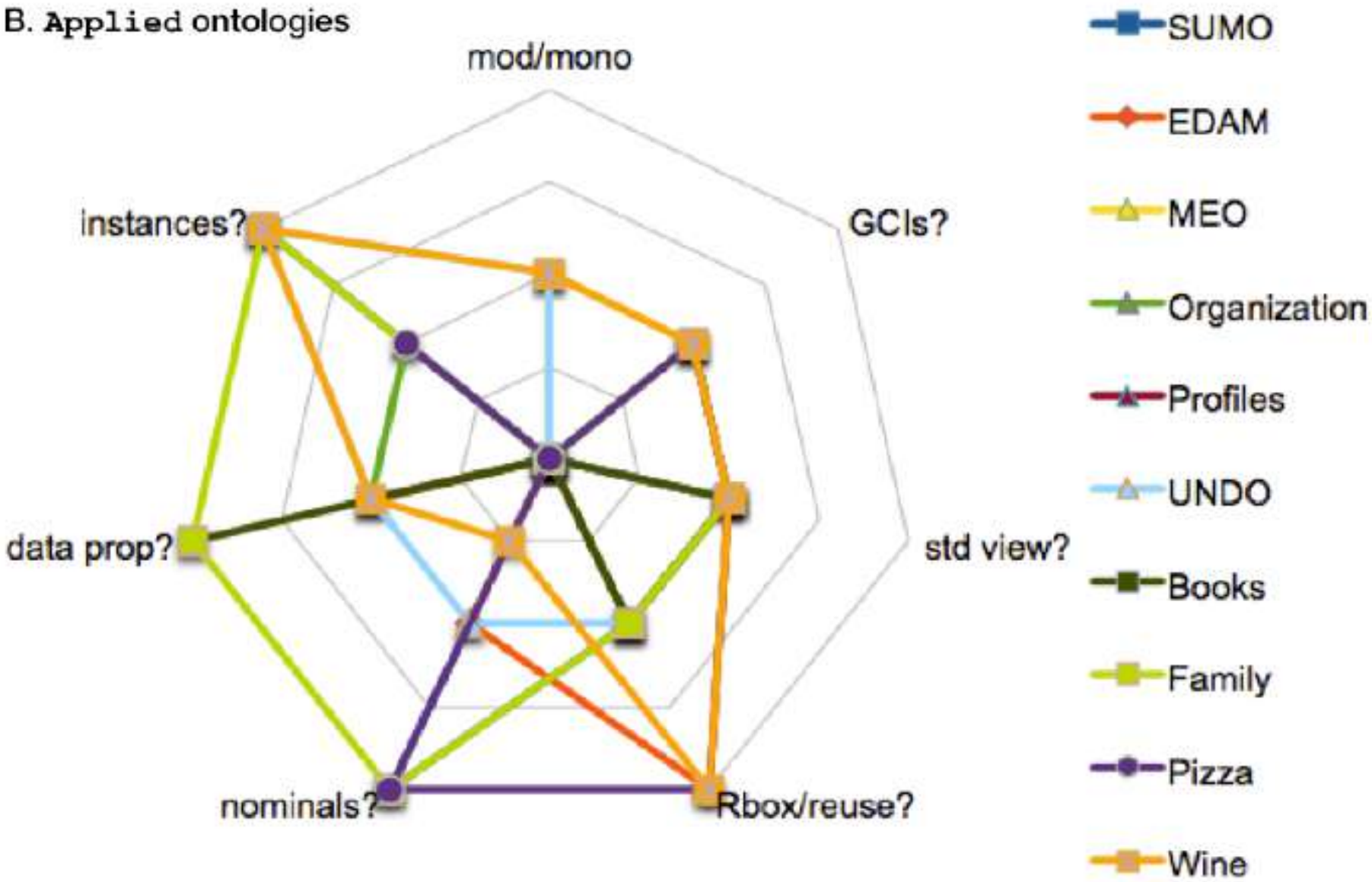
⇒ Are they useful features to check an ontology for?

1. Select different types of ontologies: foundational, core, domain, and tutorial.
2. Classify them manually; carried out by two people independently.
3. Check for inter-annotator agreement and whether the dimensions suffice; if there is disagreement, either:
   - 3a. Harmonise and move to Step 4, below;
   - 3b. Resolve conflict in classification, refine either the affected dimension's value or description thereof, and return to Step 2.
4. Analyse the data on expected consequences and on whether any recurring combinations of dimensions emerged.

Theory-oriented ontologies



Applied ontologies



Observations

- SUMO with characteristics of an applied ontology
- No reuse of relations typically with 'Hierarchy, bare' or 'flat';
- (more) data properties in Applied ontologies cf Theoretical
- Applied ontologies with more use of nominals, instances, and data properties
- Several ontologies declare domain and range axioms for OPs, but these declarations are mostly not reflected in other CEs
- Number of theoretical ontologies with none or only hidden GCIs is higher than expected, suggesting a low level of usage for advanced ontology expressions
- Explicit or hidden GCI are mostly absent from applied ontologies (except MEO)

Outline

So far (in this talk)

- Example on ventilators
  - different purposes
- CQ:axiom or CQ:query is n:m rather than 1:1
  - Complicates validation efforts in ontology authoring
  - Makes finding ontologies for use and reuse harder
- Ontology reuse across ontologies (alignment, integration, heterogeneous mappings)
- Encoding bias for purpose: *choose* a style to fit the purpose

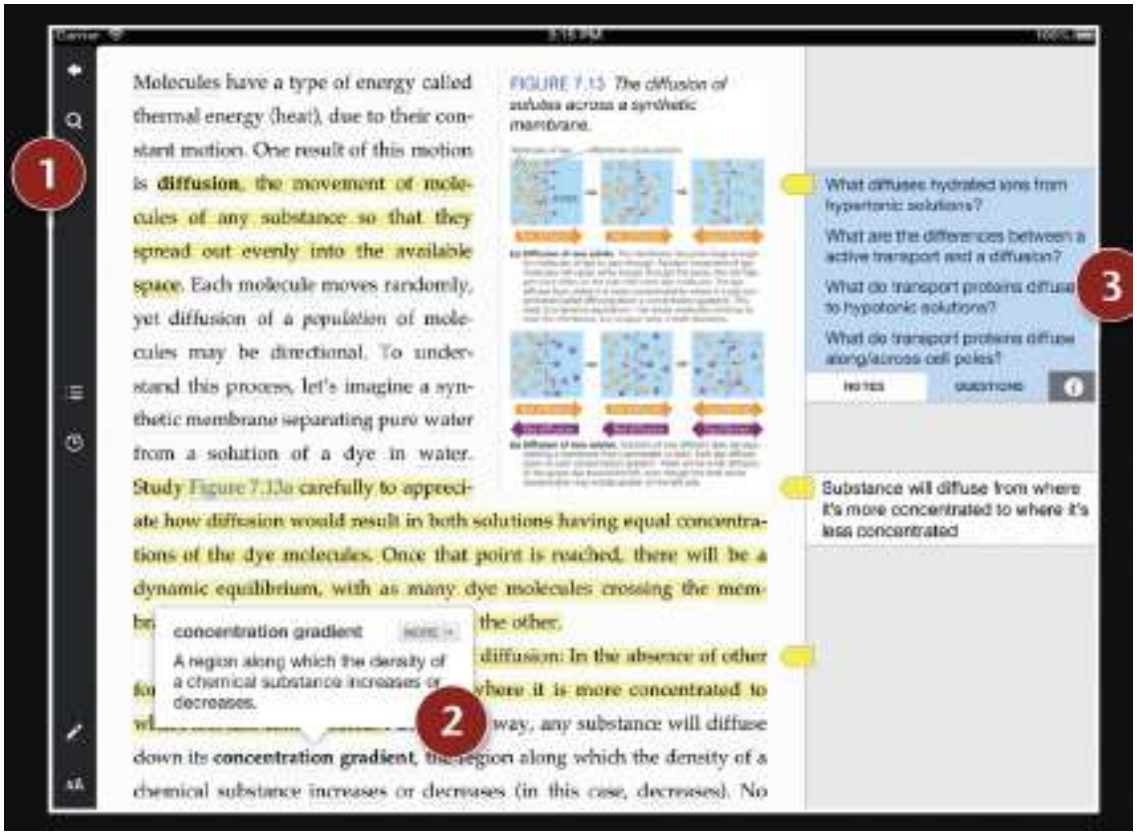
Intelligent "Inquire biology" textbook

- Idea: annotate textbook with ontology, generate questions automatically, mark automatically

Intelligent "Inquire biology" textbook

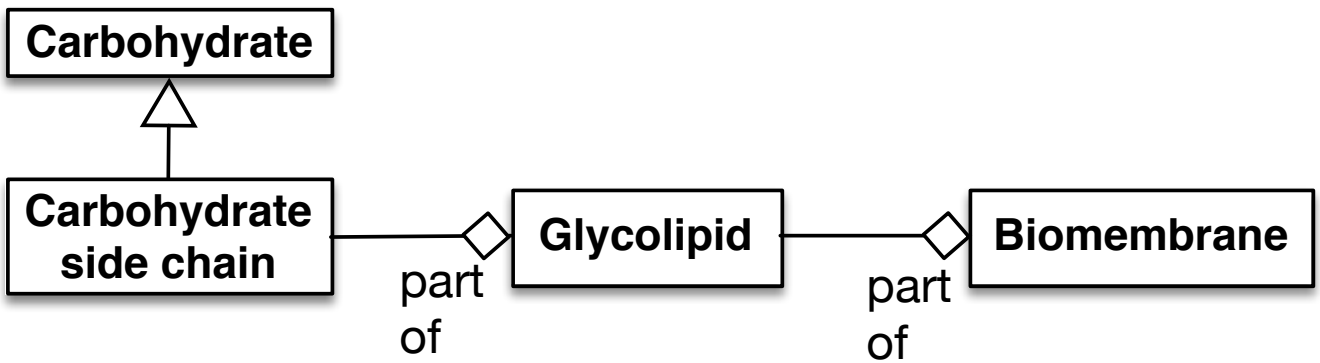
- Idea: annotate textbook with ontology, generate questions automatically, mark automatically

  Figure: Annotated p132 of Biology (9th edition) by Campbell and Reece [Chaudhri et al.(2013)]:



Intelligent "Inquire biology" textbook

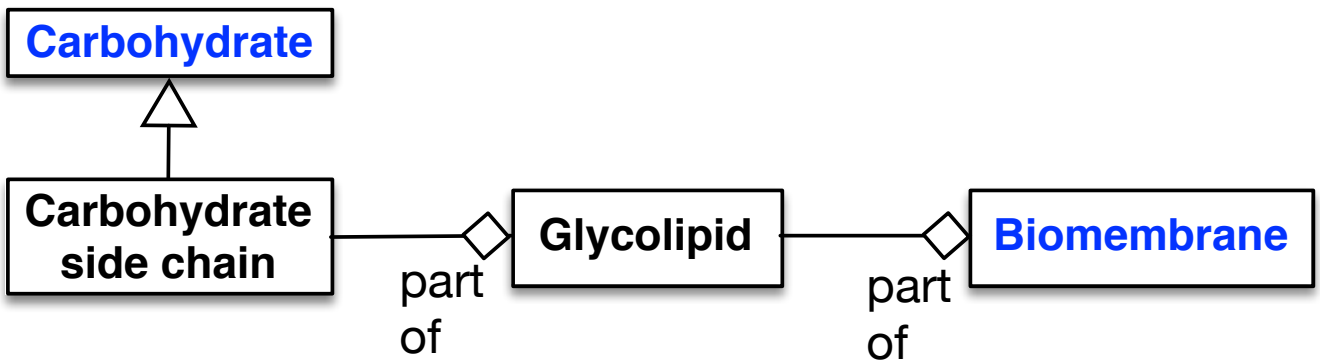
Q: What is the relation between a carbohydrate and a biomembrane?



(example from and based on [Chaudhri et al.(2013)])

Intelligent "Inquire biology" textbook

Q: What is the relation between a carbohydrate and a biomembrane?

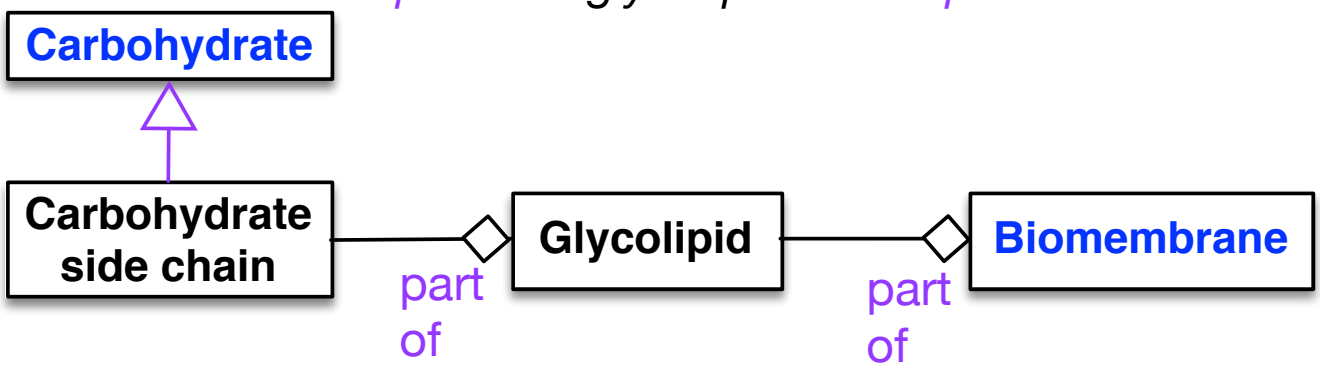


(example from and based on [Chaudhri et al.(2013)])

Intelligent "Inquire biology" textbook

Q: What is the relation between a carbohydrate and a biomembrane?

A: A carbohydrate side chain is a carbohydrate that is part of a glycolipid that is part of a biomembrane



(example from and based on [Chaudhri et al.(2013)])

Try to do that for *any* ontology

- Question templates for different types of educational questions

  Is a <T_NOUN> <OP_IS_PARTICIPLE_BY> <Quantifier_some> <T_NOUN>?

  What does a <T_NOUN> <OP_VERB>?

Raboanary, T., Wang, S., Keet, C.M. Generating Answerable Questions from Ontologies for Educational Exercises. MTSR'21. Springer CCIS vol. 1537, 28-40.
https://github.com/mkeet/AQuestGO

Try to do that for *any* ontology

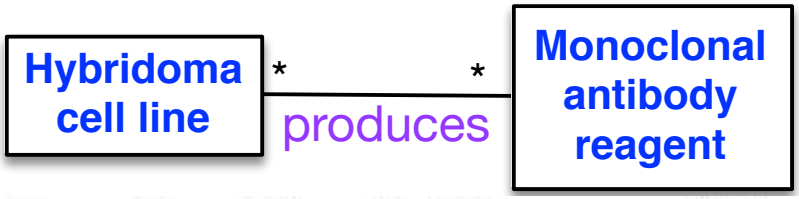
- Question templates for different types of educational questions

  Is a <T_NOUN> <OP_IS_PARTICIPLE_BY> <Quantifier_some> <T_NOUN>?

  What does a <T_NOUN> <OP_VERB>?

- Notion of "axiom prerequisites"
- Assumes a certain way of representing something and of verbalising something; the easiest:
  - The 'applied' option, (questions about endurants, dispositions etc are out of scope for a subject domain)
  - The 'abundance' of relations
  - Modularity invariant (provided same naming scheme)
  - Values/instances/classes interplay: variant-dependent code adds complexity to the algorithms

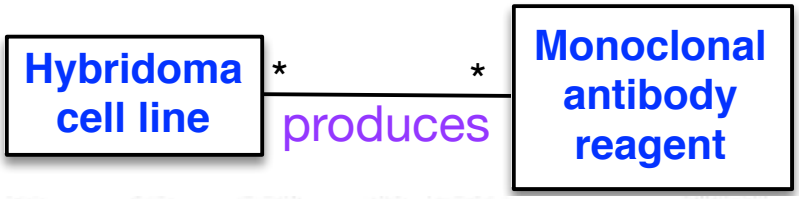
Raboanary, T., Wang, S., Keet, C.M. Generating Answerable Questions from Ontologies for Educational Exercises. MTSR'21. Springer CCIS vol. 1537, 28-40.
`https://github.com/mkeet/AQuestGO`



$$\texttt{MonoclonalAntibodyReagent} \sqsubseteq \exists \texttt{produced\_by}.\texttt{HybridomaCellLine}$$

$$\texttt{produces} \equiv \texttt{produced\_by}^{-}$$

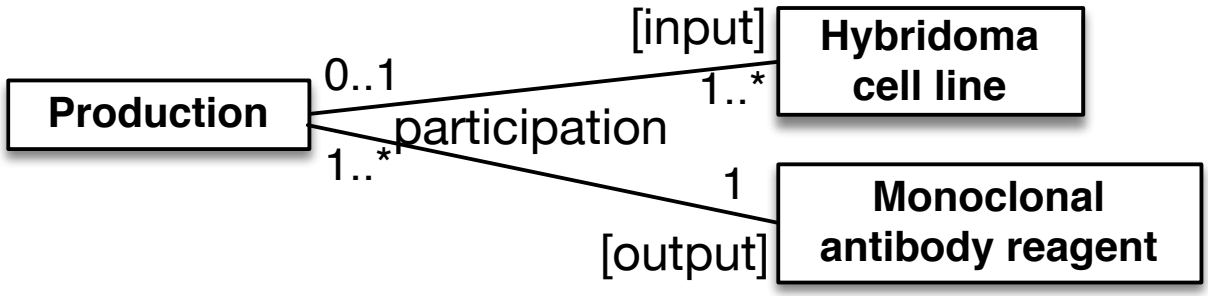
Q: *Is Monoclonal antibody reagent produced by some Hybridoma cell line?*
A: Yes
Q: *What does a Hybridoma cell line produce?*
A: Monoclonal antibody reagent



$\mathtt{MonoclonalAntibodyReagent} \sqsubseteq \exists \mathtt{produced\_by.HybridomaCellLine}$

$\mathtt{produces} \equiv \mathtt{produced\_by}^{-}$

Q: *Is Monoclonal antibody reagent produced by some Hybridoma cell line?*
A: Yes
Q: *What does a Hybridoma cell line produce?*
A: Monoclonal antibody reagent

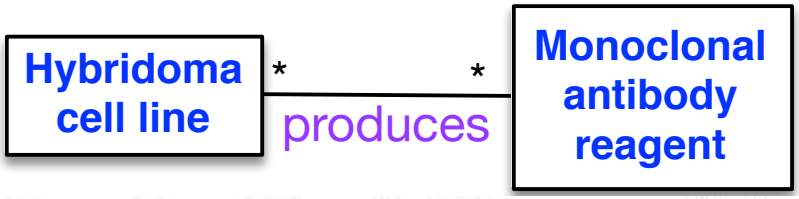


$\mathtt{MonoclonalAntibodyReagent} \sqsubseteq \exists \mathtt{participates\_in.Production}$

$\mathtt{HybridomaCellLine} \sqsubseteq \forall \mathtt{participates\_in.Production}$

.... something with the [input] and [output] roles ....

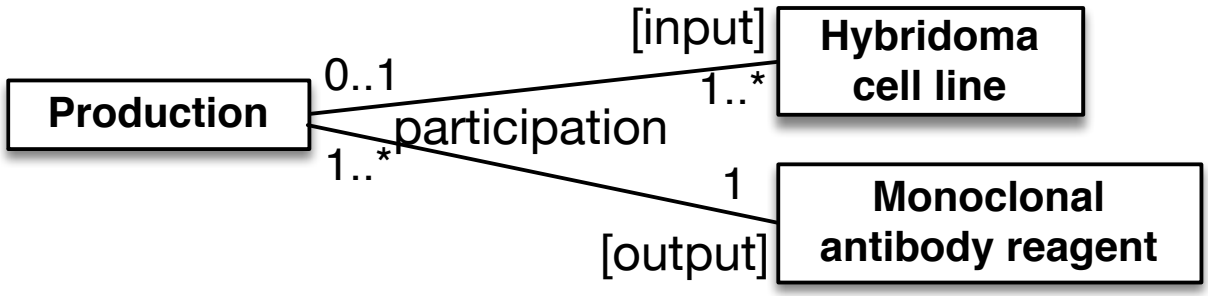
Algorithm TBD, for every ontology pattern



$\mathtt{MonoclonalAntibodyReagent} \sqsubseteq \exists \mathtt{produced\_by}.\mathtt{HybridomaCellLine}$

$\mathtt{produces} \equiv \mathtt{produced\_by}^{-}$

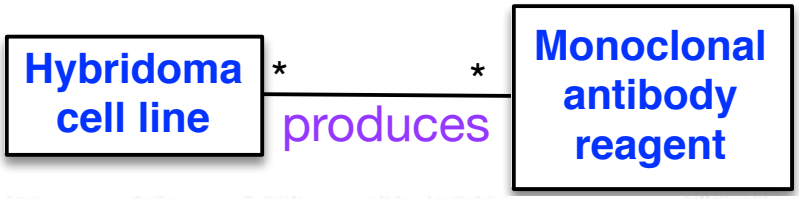
Q: *Is Monoclonal antibody reagent produced by some Hybridoma cell line?*
A: Yes
Q: *What does a Hybridoma cell line produce?*
A: Monoclonal antibody reagent



$\mathtt{MonoclonalAntibodyReagent} \sqsubseteq \exists \mathtt{participates\_in}.\mathtt{Production}$

$\mathtt{HybridomaCellLine} \sqsubseteq \forall \mathtt{participates\_in}.\mathtt{Production}$

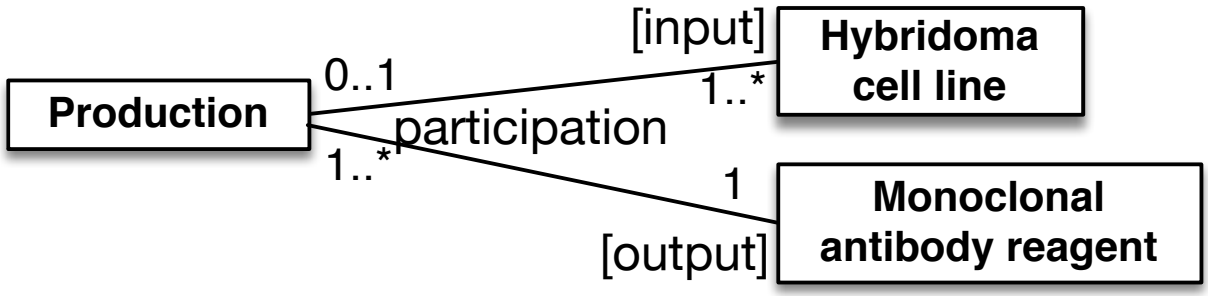
Q: *What does a Hybridoma cell line participate in?*
A: Production



$\mathtt{MonoclonalAntibodyReagent} \sqsubseteq \exists \mathtt{produced\_by.HybridomaCellLine}$

$\mathtt{produces} \equiv \mathtt{produced\_by}^{-}$

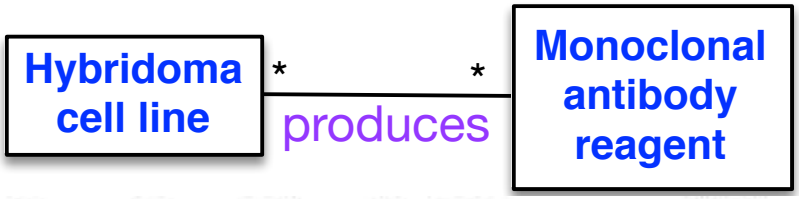
Q: *Is Monoclonal antibody reagent produced by some Hybridoma cell line?*
A: Yes
Q: *What does a Hybridoma cell line produce?*
A: Monoclonal antibody reagent



$\mathtt{MonoclonalAntibodyReagent} \sqsubseteq \exists \mathtt{participates\_in.Production}$

$\mathtt{HybridomaCellLine} \sqsubseteq \forall \mathtt{participates\_in.Production}$

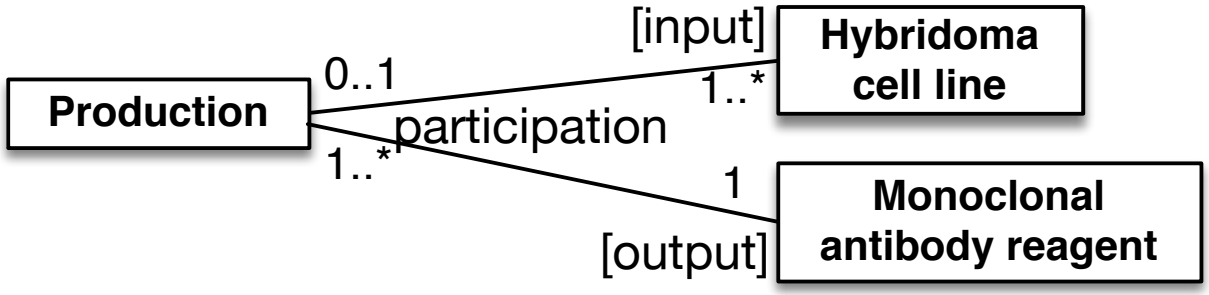
Q: *What does a Hybridoma cell line participate in?*
A: ~~Production~~



$$\mathtt{MonoclonalAntibodyReagent} \sqsubseteq \exists \mathtt{produced\_by.HybridomaCellLine}$$

$$\mathtt{produces} \equiv \mathtt{produced\_by}^{-}$$

Q: *Is Monoclonal antibody reagent produced by some Hybridoma cell line?*
A: Yes
Q: *What does a Hybridoma cell line produce?*
A: Monoclonal antibody reagent



$$\mathtt{MonoclonalAntibodyReagent} \sqsubseteq \exists \mathtt{participates\_in.Production}$$

$$\mathtt{HybridomaCellLine} \sqsubseteq \forall \mathtt{participates\_in.Production}$$

Q: *What does a Hybridoma cell line participate in?*
A: Production of monoclonal antibody reagents

Correct but awkward or confusing sentences

(BioTop) Does a material object project onto an immaterial three dimensional physical entity?

(BioTop) A taxon quality projects onto a taxon value region. True or false?

note: taxon value region is an "abstract region in which the values of biological taxa are located (cf. Schulz et.al ISMB 2008)."

(SO) Which physical endurant has a state that is only a stuff state?

Correct but awkward or confusing sentences

(BioTop) Does a material object project onto an immaterial three dimensional physical entity?

(BioTop) A taxon quality projects onto a taxon value region. True or false?

note: taxon value region is an "abstract region in which the values of biological taxa are located (cf. Schulz et.al ISMB 2008)."

(SO) Which physical endurant has a state that is only a stuff state?

(SO) Does a mixed stuff have a part stuff that is a stuff?

$\texttt{MixedStuff} \equiv \texttt{Stuff} \sqcap \exists \texttt{hasPartStuff}.\texttt{Stuff}$

(BioTop) Which condition has a life that is some life?

$\texttt{condition} \equiv \texttt{disposition} \sqcup \texttt{function} \sqcup \texttt{material object} \sqcup \texttt{process}$

$\texttt{material object} \sqsubseteq \exists \texttt{has life}.\texttt{life}$

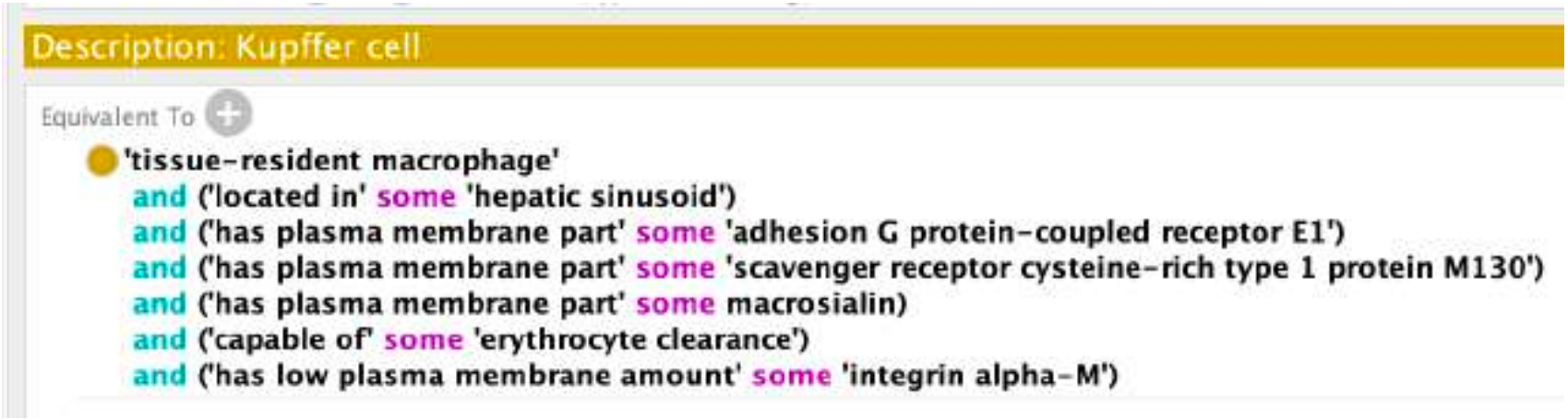
Better results with certain domain ontologies

- An applied style
- With abundance in granularity of relations
- Not just/mainly a ‘bare’ hierarchy, not just domain & range axioms

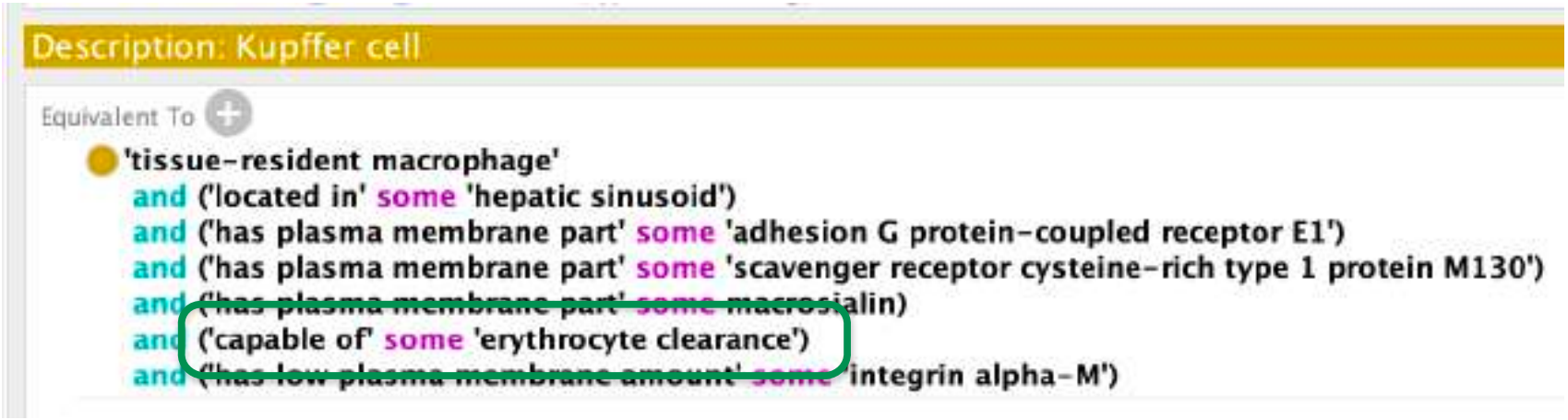
Better results with certain domain ontologies

- An applied style
- With abundance in granularity of relations
- Not just/mainly a ‘bare’ hierarchy, not just domain & range axioms
- ‘hasX’ and other OP naming considerations
- Naming in ontology often within-context, but out-of-context in the self-standing questions

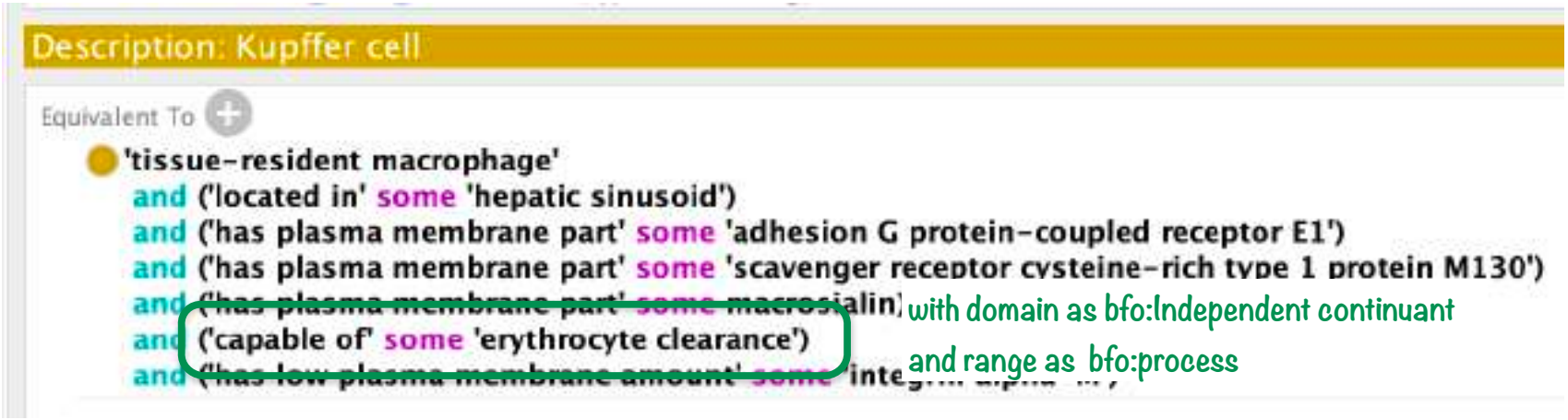
Cell types in CL [Osumi-Sutherland et al.(2021)]



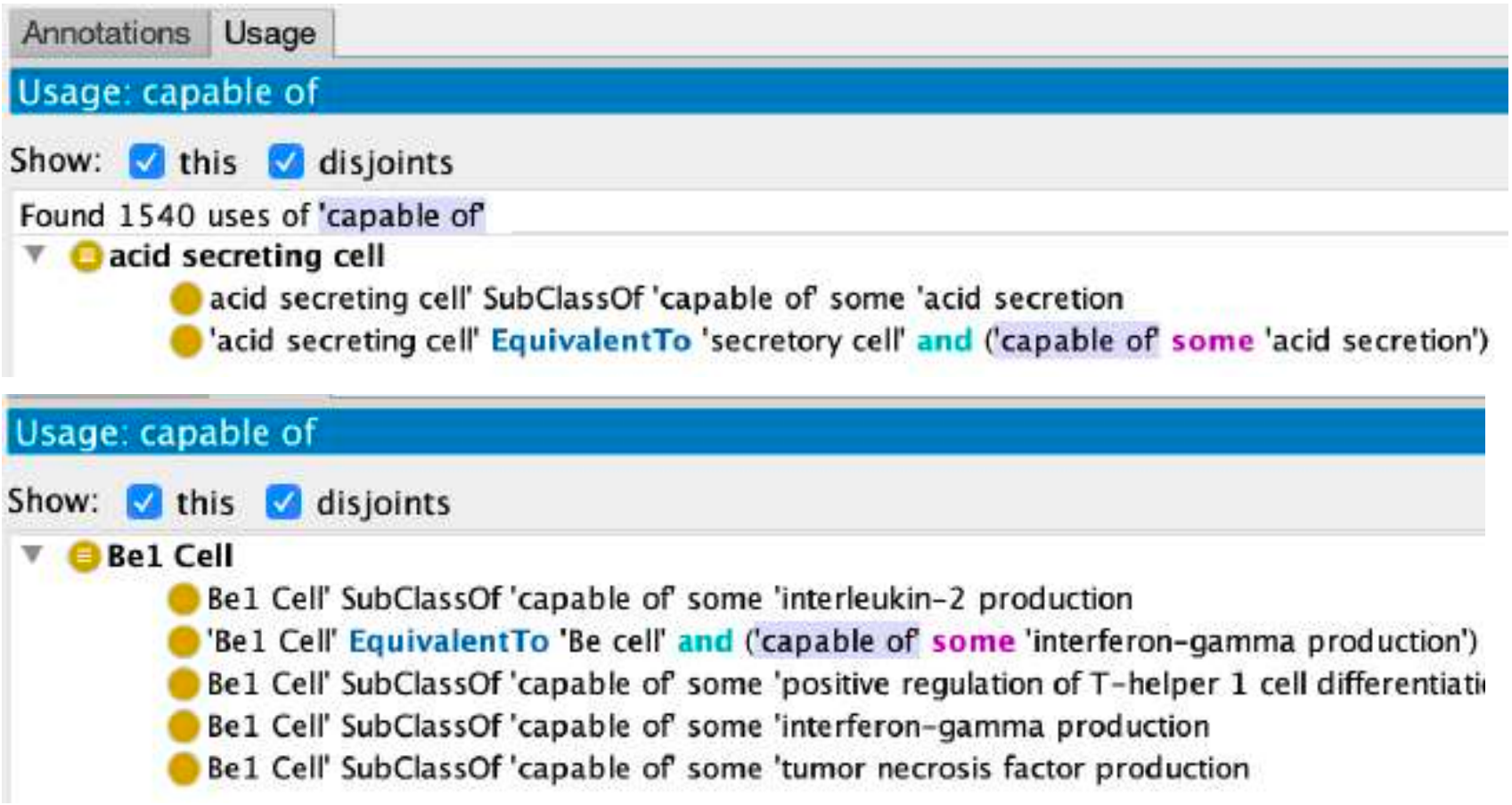
Cell types in CL [Osumi-Sutherland et al.(2021)]



Cell types in CL [Osumi-Sutherland et al.(2021)]



Cell types in CL [Osumi-Sutherland et al.(2021)]



Cell type – alternative encodings

- ⋆ *[an IC]* $\sqsubseteq \exists$capable of.*[a process]*

  versus

1. Kupffer cell $\sqsubseteq \exists$clears.erythrocyte
2. acid secreting cell $\sqsubseteq \exists$secretes.acid
3. Be1 cell $\sqsubseteq \exists$produces.interferon-gamma
4. type I NK T cell $\sqsubseteq \exists$binds.glycosphingolipid
5. vomeronasal organ $\sqsubseteq \exists$detects.pheromone
6. ....

Effect on queries

- Cell Annotation Platform (Celltype.info) only allows search by celltype to find data sets: no effects
- Others that use CL also have mostly only simple term search

Effect on queries

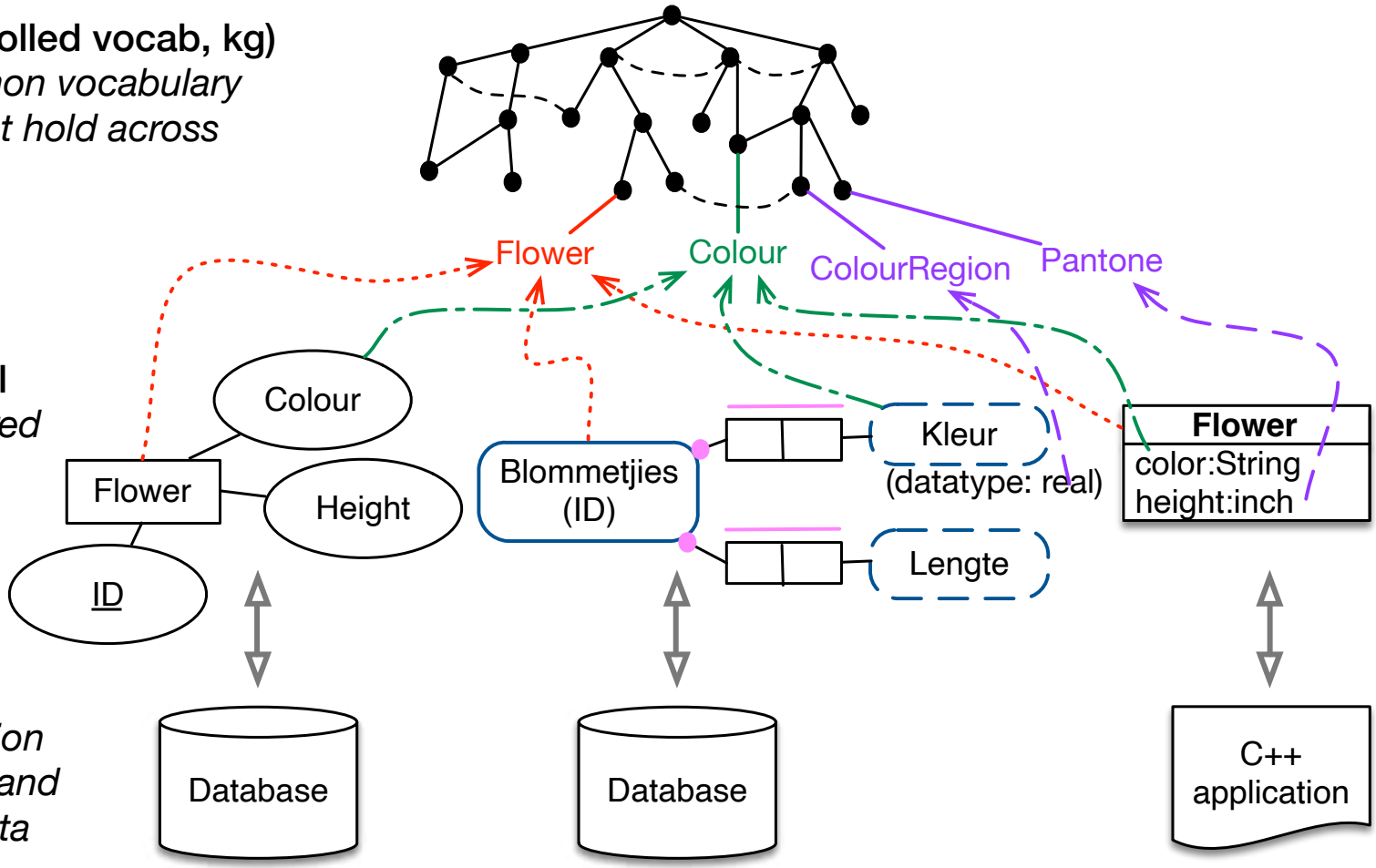
- Cell Annotation Platform (Celltype.info) only allows search by celltype to find data sets: no effects
- Others that use CL also have mostly only simple term search
- But, if you were to want to search for, e.g.:
  - the function/role of a cell; or
  - all cells that perform a certain role
  - all datasets that are from experiments on erythrocyte clearance by macrophages
  - projects on capabilities of some cell type
- the 'capable of' version will be much easier cf over very many possible names for relations: fewer names to remember, more structured, consistency in approach

Ontology-based querying – Preliminaries

Ontology (or controlled vocab, kg)
provides the common vocabulary and constraints that hold across the applications

Conceptual model
shows what is stored in that particular application

Implementation
the actual information system that stores and manipulates the data

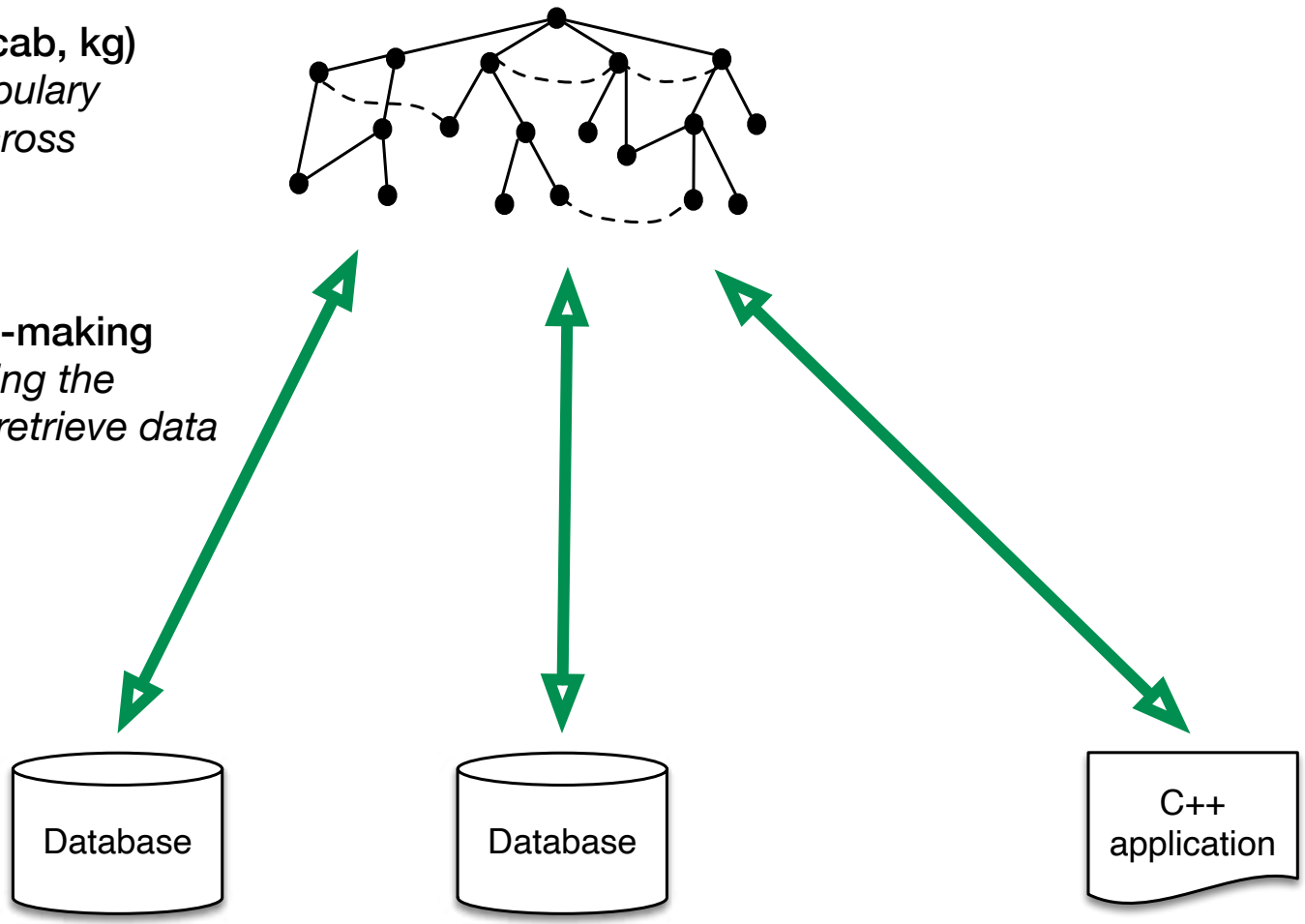


Ontology-based querying – Preliminaries

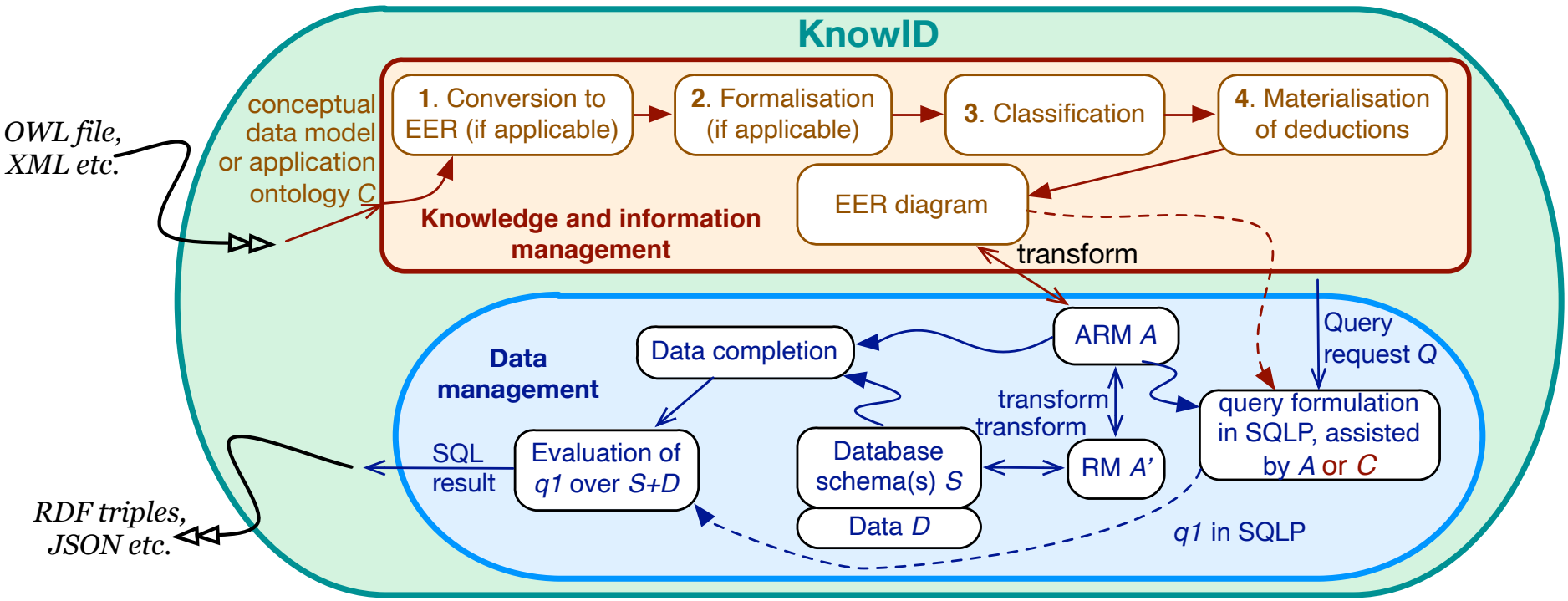
Ontology (or controlled vocab, kg)
provides the common vocabulary and constraints that hold across the applications

Queries for decision-making
formulate queries using the knowledge graph to retrieve data

Implementation
the actual information system that stores and manipulates the data



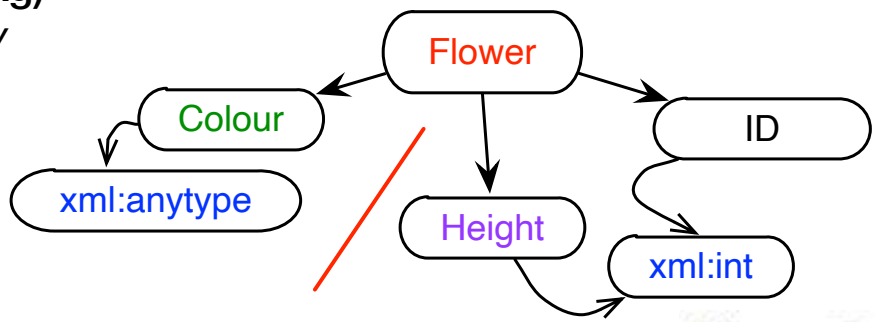
Ontology-based querying with some form of OBDA/I



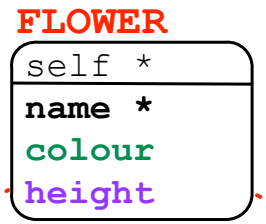
Fillottrani, P.R., Keet, C.M. KnowID: An architecture for efficient Knowledge-driven Information and Data access. Data Intelligence, 2020, 2(4): 487-512.

Connecting the knowledge to the data: OBDA or KnowID

Ontology (or controlled vocab, kg)
provides the common vocabulary and constraints that hold across the applications

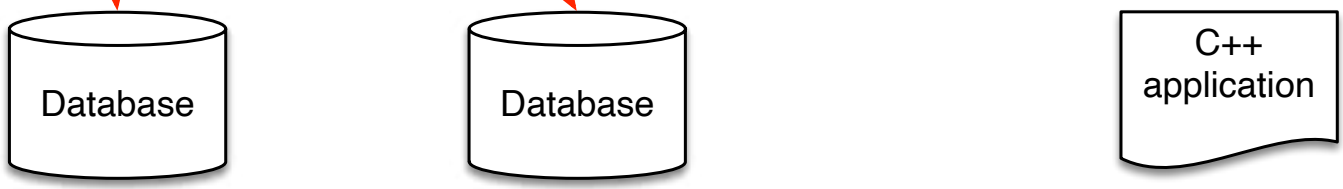


Transformation via abstract relational model
with additional virtual *identifiers*



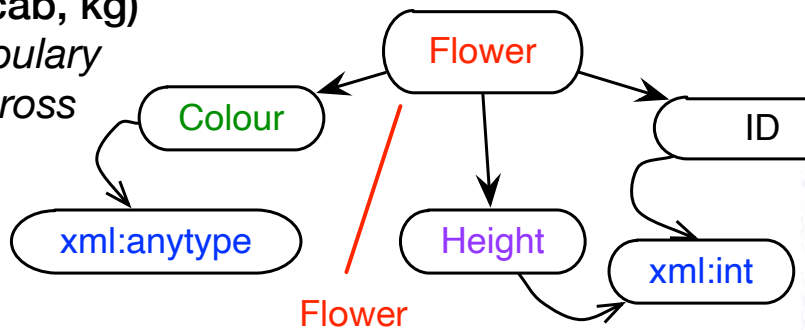
$$\begin{aligned}
&\texttt{Flower} \sqsubseteq \exists \texttt{height.Int}\\
&\texttt{Flower} \sqsubseteq \leq 1\ \texttt{height.Int}\\
&\texttt{Flower} \sqsubseteq \exists \texttt{colour.AnyType}\\
&\texttt{Flower} \sqsubseteq = 1\ \texttt{id.Int}\\
&\texttt{id.int} \sqsubseteq \texttt{Flower}\\
&\ldots
\end{aligned}$$

Implementation
the actual information system that stores and manipulates the data



Connecting the knowledge to the data: OBDA or KnowID

Ontology (or controlled vocab, kg)
provides the common vocabulary and constraints that hold across the applications

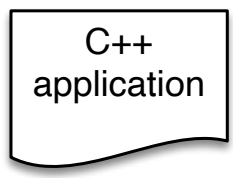


$$\begin{aligned}
&\mathtt{Flower} \sqsubseteq \exists \mathtt{height.Int}\\
&\mathtt{Flower} \sqsubseteq \leq 1\,\mathtt{height.Int}\\
&\mathtt{Flower} \sqsubseteq \exists \mathtt{colour.AnyType}\\
&\mathtt{Flower} \sqsubseteq = 1\,\mathtt{id.Int}\\
&\mathtt{id.int} \sqsubseteq \mathtt{Flower}\\
&\ldots
\end{aligned}$$

Mapping layer
links each entity to a query over the data source(s)

```
Flower
->
SELECT flowers.id
   FROM flowers
UNION
SELECT blom.name
   FROM blom
...
```

Implementation
the actual information system that stores and manipulates the data

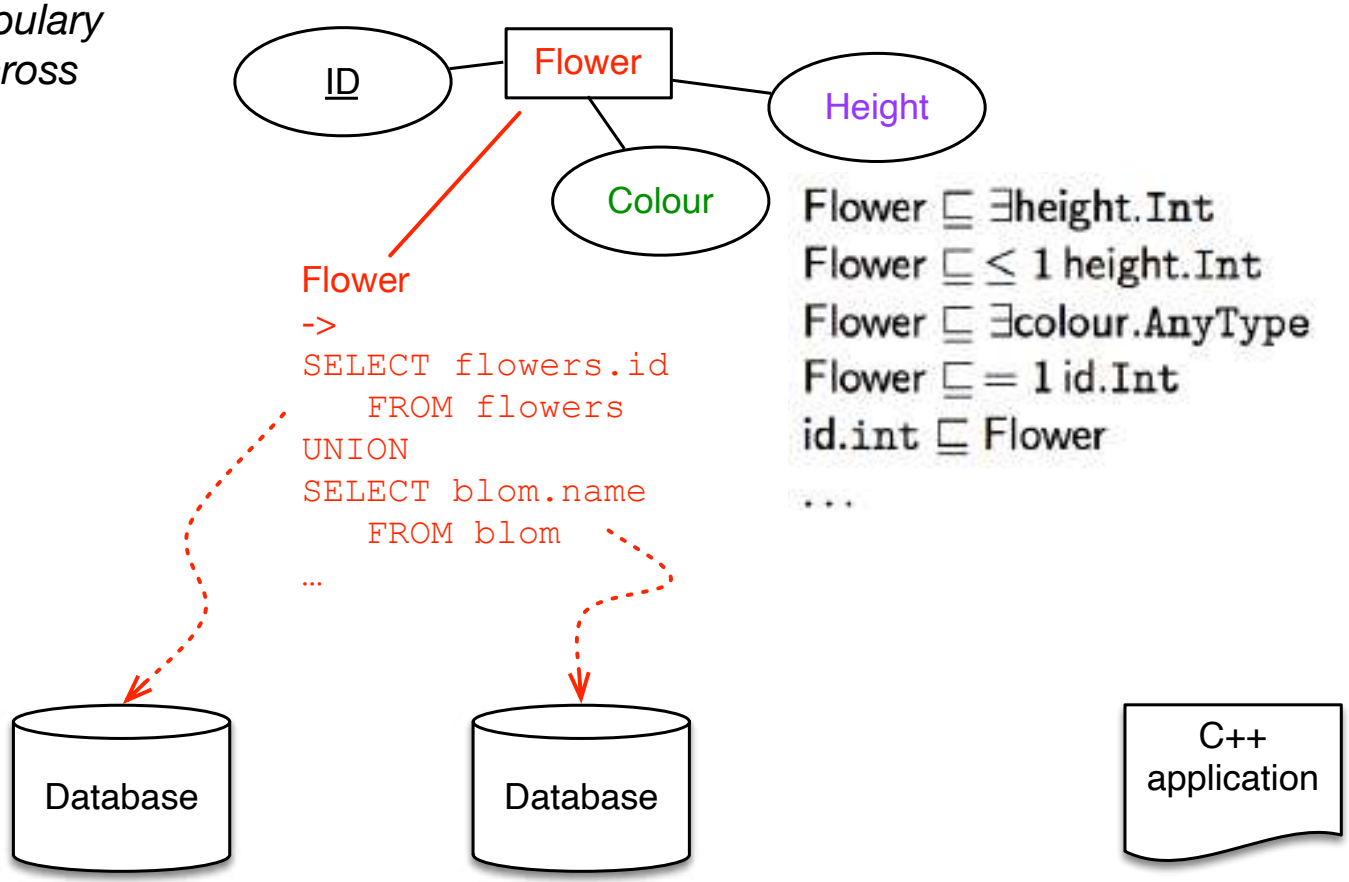


Connecting the knowledge to the data: OBDA or KnowID

Ontology (or controlled vocab, kg)
provides the common vocabulary and constraints that hold across the applications

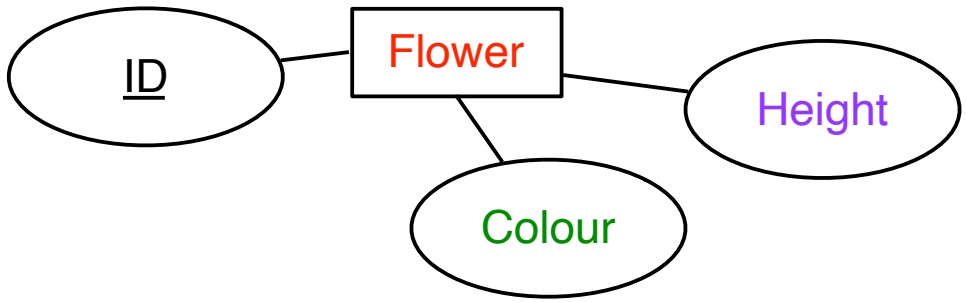
Mapping layer
links each entity to a query over the data source(s)

Implementation
the actual information system that stores and manipulates the data



Queries with OBDA models vs FO-inspired ontologies

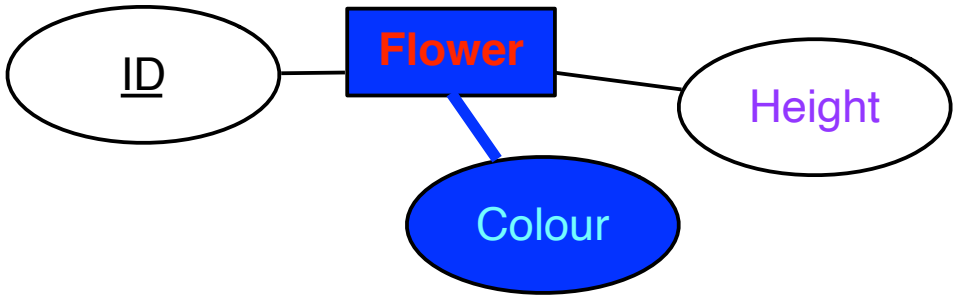
Ontology (or controlled vocab, kg)
provides the common vocabulary and constraints that hold across the applications



Calvanese D, Keet CM, Nutt W, Rodriguez-Muro M, Stefanoni G. Web-based Graphical Querying of Databases through an Ontology: the WONDER System. ACM SAC 2010.
Keet, C.M., Alberts, R., Gerber, A., Chimamiwa, G. Enhancing web portals with Ontology-Based Data Access: the case study of South Africa's Accessibility Portal for people with disabilities. OWLED'08.

Queries with OBDA models vs FO-inspired ontologies

Ontology (or controlled vocab, kg)
provides the common vocabulary and constraints that hold across the applications



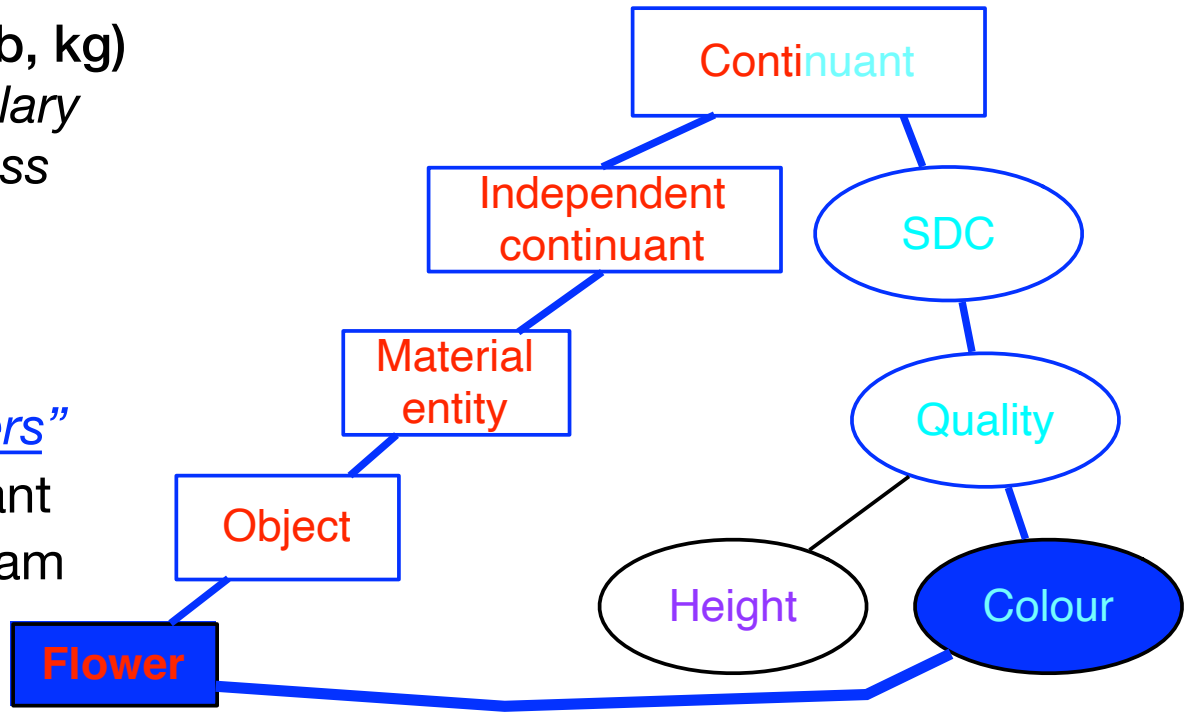
End-user query
"give me all red flowers"
just click relevant elements in the diagram

Calvanese D, Keet CM, Nutt W, Rodriguez-Muro M, Stefanoni G. Web-based Graphical Querying of Databases through an Ontology: the WONDER System. ACM SAC 2010.
Keet, C.M., Alberts, R., Gerber, A., Chimamiwa, G. Enhancing web portals with Ontology-Based Data Access: the case study of South Africa's Accessibility Portal for people with disabilities. OWLED'08.

Queries with OBDA models vs FO-inspired ontologies

Ontology (or controlled vocab, kg)
provides the common vocabulary and constraints that hold across the applications

End-user query
"give me all red flowers"
"*just*" process relevant elements in the diagram



Calvanese D, Keet CM, Nutt W, Rodriguez-Muro M, Stefanoni G. Web-based Graphical Querying of Databases through an Ontology: the WONDER System. ACM SAC 2010.
Keet, C.M., Alberts, R., Gerber, A., Chimamiwa, G. Enhancing web portals with Ontology-Based Data Access: the case study of South Africa's Accessibility Portal for people with disabilities. OWLED'08.

Better results with certain domain ontologies

- Applied style – *de facto* technically a conceptual data model
- Mainly domain & range axioms for OBDA (language limitations for performance)
- Ontology vocabulary naming schemes don't matter

Outline

Summary

- Encoding biases do exist; at least 6 groups of ontology patterns
- Combine into styles, find, and swap, based on 10 dimensions with 28 traits
- Affects ontology development: bottom-up development algorithms, alignment, CQ management, etc.
- Affects ontology use, e.g.:
  - Ontology-driven EdTech
  - Ontology-based data access
- Will not resolve itself, but needs better management

Key questions revisited

- Does it matter which way it is put in the ontology?
- If so: how/where/why, and is one always better than the other(s)?
- Which patterns are there for representing the 'same thing'?
- Are they really equivalent alternatives of the same meaningful unit?
- Do certain patterns co-occur, to make up a style of representing something?
- How does all this affect use and reuse of ontologies? (w.r.t., e.g., ontology imports, FO use, data integration, literature annotation, ...).

Key questions revisited

- Does it matter which way it is put in the ontology?

→ yes

- If so: how/where/why, and is one always better than the other(s)?

→ no

- Which patterns are there for representing the 'same thing'?

→ at least those 19

- Are they really equivalent alternatives of the same meaningful unit?

→ no

- Do certain patterns co-occur, to make up a style of representing something?

→ yes

- How does all this affect use and reuse of ontologies? (w.r.t., e.g., ontology imports, FO use, data integration, literature annotation, ...).

→ in multiple ways

Other considerations

- Encoding bias affects usability and reusability
- If ‘applied’ style, with application decisions, defeats the original purpose of ontologies
- If ‘theoretical’ style, harder to use in applications
- Mindful of purpose(s) of ontologies vs other artefacts (thesauri, conceptual data models, ...)

Other considerations

- Encoding bias affects usability and reusability
- If ‘applied’ style, with application decisions, defeats the original purpose of ontologies
- If ‘theoretical’ style, harder to use in applications
- Mindful of purpose(s) of ontologies vs other artefacts (thesauri, conceptual data models, ...)
- Adjust modelling language for modeller’s own good?
- Systematise ‘categories’ of ontologies and make developer categorise theirs?
- Better educational material?
- Patch it up with guidelines, methods, and tools to mitigate the worst?

Collaborators (on the works included in this talk), funding

- Collaborators: Diego Calvanese and Werner Nutt (FUB, Italy), Pablo Fillottrani (UNS, Argentina), Agnieszka Ławrynowicz (PUT, Poland), Jedrec Potoniec (PUT, Poland), David Toman (UW, Canada), Dawid Wisniewski (PUT, Poland)
- Current and former students: Mary-Jane Antia, Leighton Dawson, Zola Mahlaza, Bradley Malgas, Toky Raboanary, Giorgio Stefanoni, Steve Wang
- Funding from EU, HPI, NRF

References I

V.K. Chaudhri, B. Cheng, A. Overholtzer, J. Roschelle, A. Spaulding, P. Clark, M. Greaves, and D Gunning.
Inquire biology: A textbook that answers questions.
AI Magazine, 34(3):55–72, 2013.

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.

C. M. Keet, Z. Mahlaza, and M.-J. Antia.
CLaRO: a controlled language for authoring competency questions.
In E. Garoufallou et al., editors, *13th Metadata and Semantics Research Conference (MTSR'19)*, volume 1057 of *CCIS*, pages 3–15. Springer, 2019.
28-31 Oct 2019, Rome, Italy.

C.M. Keet.
An exploration into cognitive bias in ontologies.
In E.M. Sanfilippo et al., editors, *Proceedings of Cognition And OntologieS (CAOS'21), part of JOWO'21*, volume 2969 of *CEUR-WS*, page 17p, 2021.
13-16 September 2021, Bolzano, Italy.

Asiyah Yu Lin, Yuki Yamagata, William D. Duncan, Leigh C. Carmody, Tatsuya Kushida, Hiroshi Masuya, John Beverley, Biswanath Dutta, Michael DeBellis, Zoë May Pendlington, Paola Roncaglia, and Yongqun He.
A community effort for COVID-19 ontology harmonization.
In Janna Hastings and Adrien Barton, editors, *Proceedings of the International Conference on Biomedical Ontologies 2021 (ICBO'21)*, volume 3073 of *CEUR Workshop Proceedings*, pages 122–127. CEUR-WS.org, 2021.

References II

D. Osumi-Sutherland, C. Xu, M. Keays, et al.

Cell type ontologies of the human cell atlas.
Nature Cell Biology, 23:1129–1135, 2021.

Yuan Ren, Artemis Parvizi, Chris Mellish, Jeff Z. Pan, Kees van Deemter, and Robert Stevens.

Towards competency question-driven ontology authoring.
In *Extended Semantic Web Conference (ESWC'14)*, LNCS. Springer, 2014.

Francois Scharffe and Dieter Fensel.

Correspondence patterns for ontology alignment.
In A. Gangemi and J. Euzenat, editors, *International Conference on Knowledge Engineering and Knowledge Management*, volume 5268 of *LNAI*, pages 83–92. Springer, 2008.

M. Uschold and M. Gruninger.

Ontologies: principles, methods and applications.
Knowledge Engineering Review, 11(2):93–136, 1996.
URL `https://doi.org/10.1017/S0269888900007797`.

Dawid Wisniewski, Jedrzej Potoniec, Agnieszka Lawrynowicz, and C. Maria Keet.

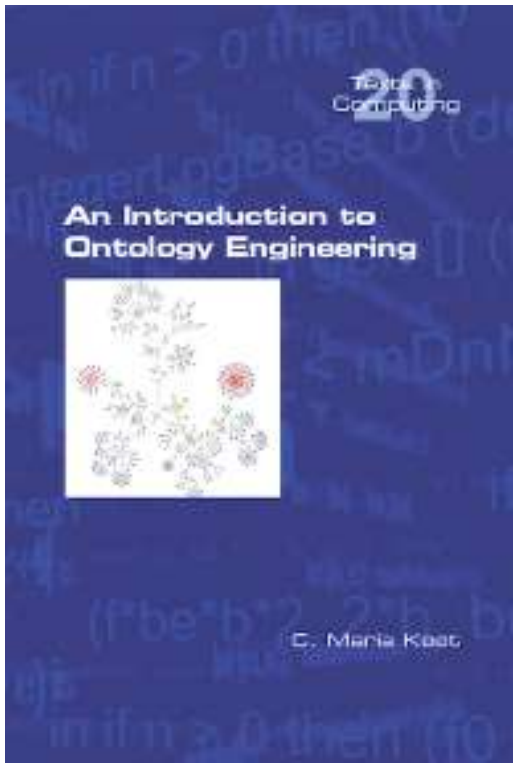
Analysis of ontology competency questions and their formalisations in sparql-owl.
Journal of Web Semantics, 59(100534):19p, 2019.

Thank you!

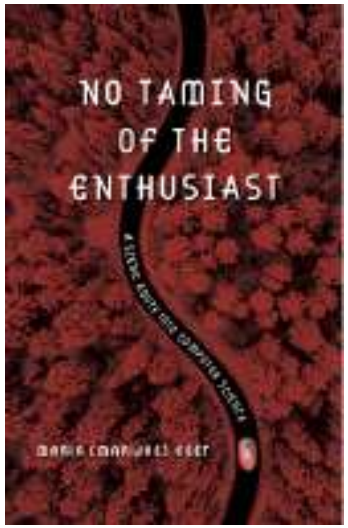
Questions?

My award-winning open textbook

`https://people.cs.uct.ac.za/~mkeet/OEbook/`



A memoir



Extra slides with more examples

Actual example for case F (1/2)

Requirement: harmonise the CIDO and CODO COVID-19 ontologies [Lin et al.(2021)]

- CODO: `laboratory testfinding` $\equiv$ {`positive, pending, negative`}
- CIDO: `positive COVID-19 diagnosis` $\sqsubseteq$ `COVID-19 diagnosis`, `presumptive positive COVID-19 diagnosis` $\sqsubseteq$ `COVID-19 diagnosis`, and `negative COVID-19 diagnosis` $\sqsubseteq$ `COVID-19 diagnosis`

Actual example for case F (1/2)

- Ontological issues
  - a finding of some fact vs. a diagnosis (i.e., conclusion drawn from the fact)
  - but when taken in context, intention is the same

Actual example for case F (1/2)

- Ontological issues
  - a finding of some fact vs. a diagnosis (i.e., conclusion drawn from the fact)
  - but when taken in context, intention is the same
- Solution options:
  1. Change CODO to use CIDO's style
  2. Change CIDO to use CODO's style
  3. Joint outside option with attribute + values instead

Class vs. Object Property (case A)

- *alignment pattern name*: *class-OP*
- *pattern elements*: $\mathcal{C}_1, \mathcal{C}_2, \mathcal{R}_1$ from O, $\mathcal{C}'_3, \mathcal{C}'_4, \mathcal{C}'_5, \mathcal{R}'_2, \mathcal{R}'_3$ from O'
- *alignment patterns' contexts*:
  - *pattern P in O* (the one on the left): $\exists \mathcal{R}_1.\mathcal{C}_2 \sqsubseteq \mathcal{C}_1$ and $\exists \mathcal{R}_1^-.\mathcal{C}_1 \sqsubseteq \mathcal{C}_2$;
  - *pattern P' in O'* (the one on the right): $\exists \mathcal{R}'_2.\mathcal{C}'_4 \sqsubseteq \mathcal{C}'_3$, $\exists \mathcal{R}'^-_2.\mathcal{C}'_3 \sqsubseteq \mathcal{C}'_4$, $\exists \mathcal{R}'_3.\mathcal{C}'_5 \sqsubseteq \mathcal{C}'_3$, $\exists \mathcal{R}'^-_3.\mathcal{C}'_3 \sqsubseteq \mathcal{C}'_5$, $\mathcal{C}'_3 \sqsubseteq (\exists \mathcal{R}'_2)$, and $\mathcal{C}'_3 \sqsubseteq (\exists \mathcal{R}'_3)$.
- *pattern alignment*:
  - *homogeneous mappings*: between $\mathcal{C}_1$ and $\mathcal{C}'_4$ and between $\mathcal{C}_2$ and $\mathcal{C}'_5$, which may be subsumption or equivalence relations.
  - *heterogeneous alignments*: $\exists \mathcal{R}_1 \sqsubseteq \mathcal{C}'_3$, $\exists \mathcal{R}_1^- \sqsubseteq \mathcal{C}'_3$, $\mathcal{C}'_3 \sqsubseteq \exists \mathcal{R}_1 \sqcap \exists \mathcal{R}_1^- \sqcap (\leq 1\, \mathcal{R}_1) \sqcap (\leq 1\, \mathcal{R}_1^-)$.

Example: ontology interoperability and alignment [Fillottrani and Keet(2017)]

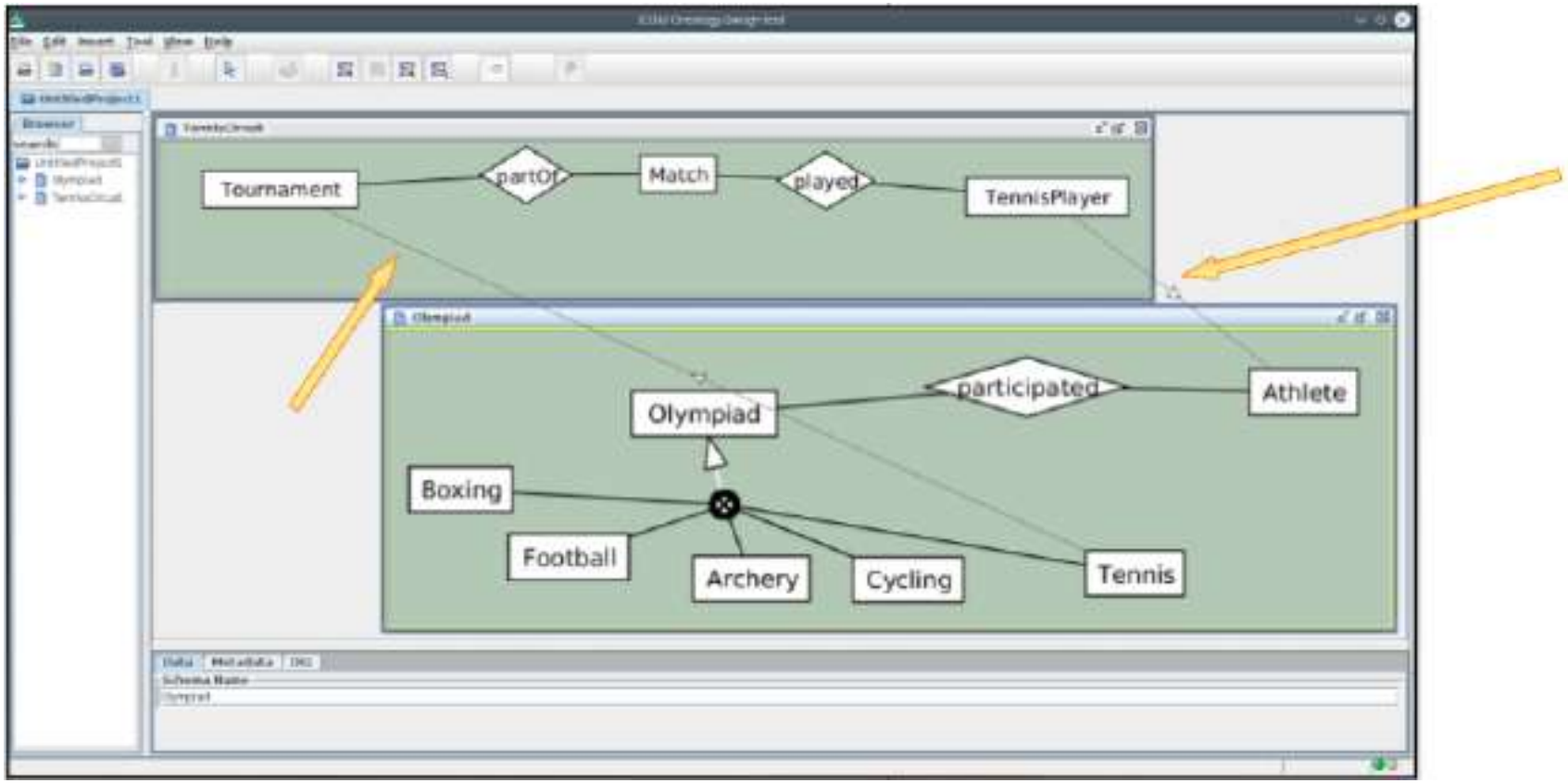


Figure: First: Mapping and searching example ('Case A': Class↔OP); Second: checking and accept/reject alignment

Example: ontology interoperability and alignment [Fillottrani and Keet(2017)]

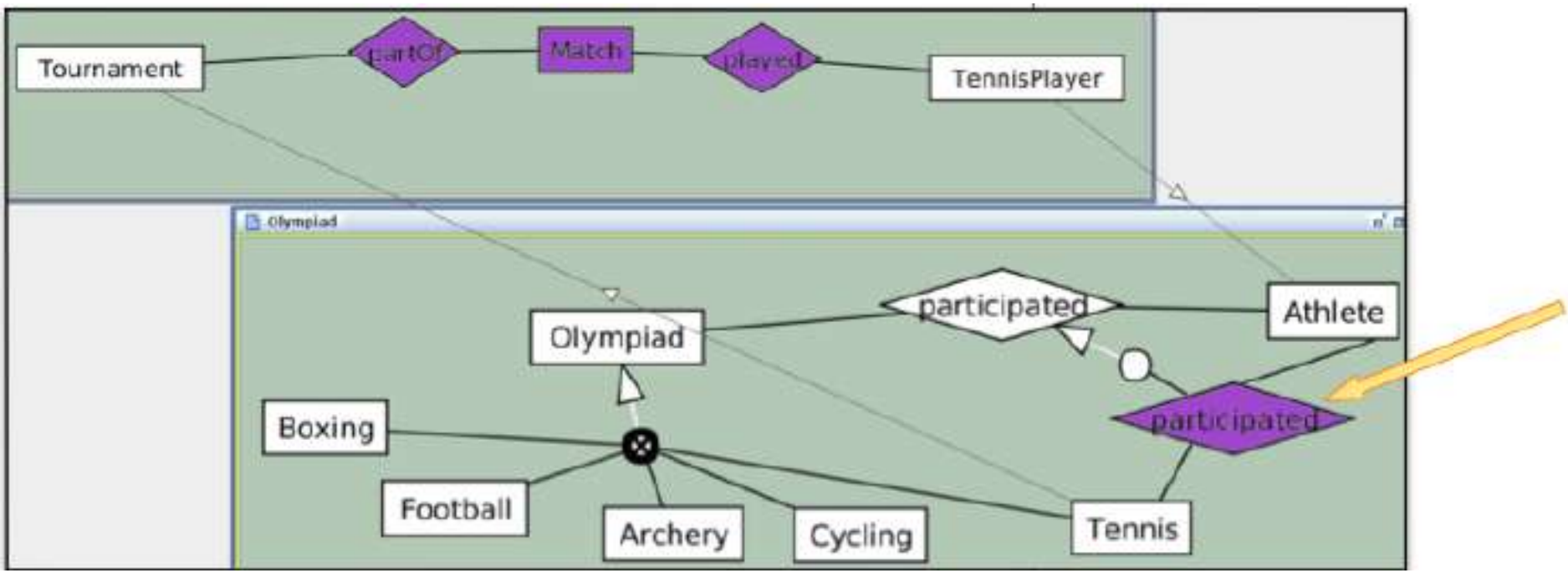


Figure: First: Mapping and searching example ('Case A': Class↔OP); Second: checking and accept/reject alignment

CQs example (abridged notation)

Is [this animal] a herbivore? // a simple ASK for “X is a Y” in AWO
Is [this software] open source development? // uses property has_licence in SWO
is [it] scriptable? // impossible with explicit knowledge in SWO
is [it] extensible? // with has_licence, has_clause, ‘derivative software allowed’, and ‘EPCC’ in SWO

"Is EC1 EC2": [
"SELECT * WHERE { :b0 <subClassOf> :b1 . :b1 <onProperty> ?p ; <someValuesFrom> :b2 }"
"SELECT * WHERE { ?placeholder_PPx1 <subClassOf> :b2 ; <subClassOf> :b0 . :b0 <onProperty> :b3 ; <someValuesFrom> :b1 . :b1 <onProperty> :b4 ; <someValuesFrom> :b5 }",
"SELECT * WHERE { ?placeholder_PPx1 <subClassOf> :b0 ; <subClassOf> :b1 }",
"SELECT * WHERE { …… }",],

"SELECT * WHERE { :b0 <subClassOf> :b1 . :b1 <onProperty> ?p ; <someValuesFrom> :b2 }": [
"Is EC1 EC2",
"PC1 I PC1 EC1 PC1 EC2",
"What types of EC1 are EC2"

Potoniec, J., Wisniewski, D., Lawrynowicz, A., Keet, C.M. Dataset of Ontology Competency Questions to SPARQL-OWL Queries Translations. Data in Brief, 2020, 29: 105098. https://github.com/CQ2SPARQLOWL/Dataset