

# Requirements for pre- and post-coordinated ontology engineering

Daniel Schober, Stefan Schulz  
UKLFR



# Overview

- Introduction to problem
- Pre-coordination requirements
- Post-coordination requirements
  - Disease cODP
    - Pneumonia example
- Conclusions

# An example (not DL)

## LIONC

- Express “Weight at birth”
  - by **single pre-coordinated ID**
    - LOINC 8339-4
  - by **post-coordination via multiple IDs**
    - Weight | LOINC 3141-9, Weight circumstance | LOINC 8337-8, Birth | SNOMED F-88005

## Problem

- LOINC semantics does not allow to detect such equivalences
- In SNOMED-CT (DL) this is easier, but it is not clear to what extend & which areas need to be pre-coordinated

# Introduction

- OWL 2 RL profile of Description Logics (DL)
  - Abundantly used, i.e. in OBO Foundry
  - Constructors
    - class inclusion ( $\sqsubseteq$ ), class equivalence ( $\equiv$ ),
    - class disjointness using negation ( $\neg$ ),
    - existential quantification ( $\exists$ ), value restriction ( $\forall$ ), conjunction ( $\sqcap$ ), disjunction ( $\sqcup$ )
- Ontologies are neither fully pre-coordinative nor post-coordinative
- They are a mix, but the degrees towards one or the other is often not based on the final application

## Our hypothesis

- Different use cases need different design principles for shaping ontologies

# Design requirements for pre-coordination

- Ontologist creates class definitions in design phase
- Provides complex compositional terms
  - via fully defined classes
- First pre-coordinate for abundantly used classes
  - Avoid combinatorial explosion & performance issues
- Uses equivalence statements
  - Implementing Aristotelian definitions
- **Low level of domain knowledge required by user**

# Design requirements for post-coordination

- User creates class definitions at annotation-time
- Ontology design must prevent nonsensical & ambiguous & redundant coordinations
- **User must be guided by constraints and patterns**
  - requires comprehensive axiomatization
  - i.e. using value restrictions and negations
- **Much domain knowledge required by user**

# Describing the Pneumonia Domain

Pneumonia is an inflammation of lung tissue

Major classification axes (sortals):

- Pathological/anatomical localisation
  - the localization of the disease in the lung and its tissues
- Disease course
  - acute or chronic
- Etiological characteristics
  - Causes: infections, physical, chemical...
  - Pre-existing conditions, which are complicated by pneumonia
- Environmental characteristics
  - where it was acquired (community or hospital)

# Pre-coordination requirements

- **Focus of pre-coordination**
  - Stating equivalences between one and multiple atomic IDs linked by logics constructors
  - Used for reasoning at OE time, e.g. consistency checks

i.e. formalizing examples from above ...

*Pneumonia*             $\equiv$  *Inflammation*  $\sqcap$   $\exists$  **has-participant.** *LungTissue*  
*AcutePneumonia*    $\equiv$  *Pneumonia*  $\sqcap$   $\exists$  **bearer-of.** *AcuteQuality*  
*BacterialPneumonia*  $\equiv$  *Pneumonia*  $\sqcap$   $\exists$  **has-agent.** *BacteriaPopulation*  
*ViralPneumonia*     $\equiv$  *Pneumonia*  $\sqcap$   $\exists$  **has-agent.** *VirusPopulation*

All being straight aristotelian definitions



# Pre-coordination Requirements

## Needed Constructors

- equivalence operator ( $\equiv$ ), existential restriction ( $\exists$ ) and conjunction ( $\sqcap$ )  
→ limited OWL EL flavour is enough

## But, ... we can't express...

- Class delineation
  - Allows Pneumonia in Kidney
    - as *LungTissue* and *KidneyTissue* have not been made disjoint
- Object property cardinalities
  - Can't specify whether *AcutePneumonia* can have, besides *acute*, further qualities, e.g. *chronic*
- Allowed values (range restrictions)
  - Can't specify whether the **has-agent** in *BacterialPneumonia* can also be filled by, e.g. *VirusPopulation*
- Toplevel categories
  - whether Pneumonia is a *Process*, *PathologicalStructure*, or both

*This will restrict use cases to simple IR tasks encoded by post-coordinated expressions in pre-coordinated queries (or vice-versa)*

# Pre-coordination requirements

## Taxonomic hierarchies

- Provide SubclassOf definitions for inferring e.g.

*BacterialPneumonia*  $\sqsubseteq$  *BacterialInflammation*

## Mereotopological axioms

- for inferring from
    - Pneumonia  $\equiv$  Inflammation  $\sqcap$   $\exists$ has-participant. LungTissue **and**
    - LungTissue  $\sqsubseteq$   $\exists$ part-of. Lung
- Pneumonia  $\sqsubseteq$   $\exists$ has-locus. Lung

## Disjointness statements, negations and value restrictions

- *Process*  $\sqsubseteq$   $\neg$ *Structure*
- for inferring
  - *PathologicalProcess*  $\sqsubseteq$   $\neg$ *PathologicalStructure*

# Post-coordination requirements

Support and guide user to compose own post-coordinated compositions

All post-coordinated expressions by user must be

- Valid → allow only meaningful compositions & prevent nonsensical coordinations like *ViralPneumonia caused\_by BacteriaPopulation*
- Expressive → enable user to create unambiguous, clearly delineated compositions
- Reliable → support for compositions that are consistent between different modelers

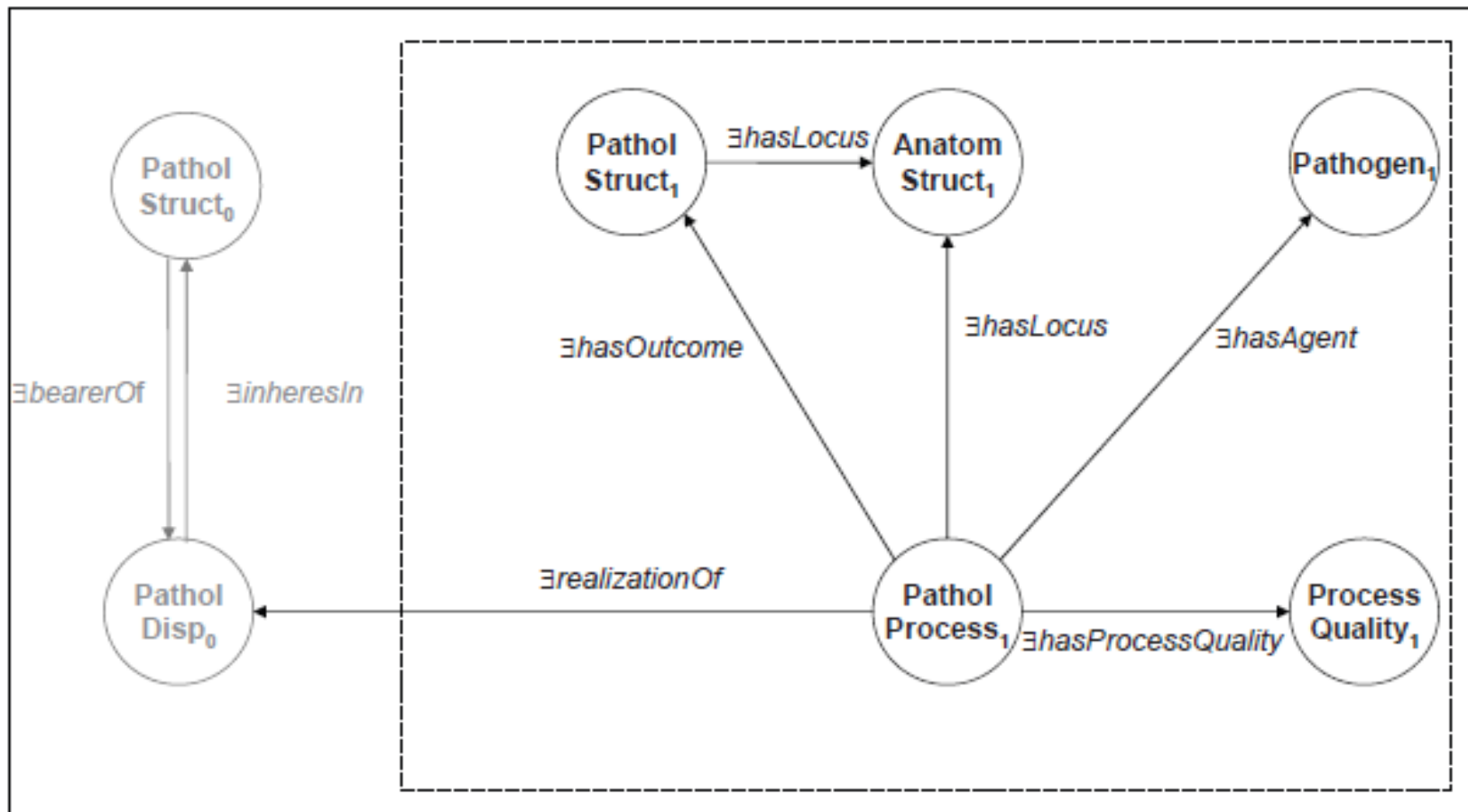
**Post-coordination needs to support users**

**Post-coordination needs to embed coordination axioms**

**Ontology Design Patterns (ODPs) and upper level ontologies deliver such axioms**

# Disease cODP supporting post-coordination

- Distinction between pathological disposition, process and structure, ...



# Pattern implemented in Pneumonia ontology

- Patterns supporting post-coordination
- Rooted in BioTop upper-level ontology

Provides restrictions, e.g.

All Pneumonias can **only** be in Loci where LungTissue exists:

$$Pneumonia \sqsubseteq \forall \text{has-locus}. (\exists \text{locus-of}. LungTissue)$$

Given further restrictions, e.g. that ...

$$LungTissue \sqsubseteq \exists \text{has-locus}. Lung$$
$$Lung \sqsubseteq \exists \text{has-locus}. Thorax$$
$$\exists \text{has-locus}. Thorax \sqsubseteq \neg \exists \text{has-locus}. (Abdomen \sqcup Extremity)$$

... an ontology based annotation GUI can provide user with soleley correct Localisations possible for a Pneumonia

# Additional post-coordination pattern examples

A secondary disease is a pathological process which is the realization of a **pre-existing disposition** which inheres in a pathological structure, which exists as congenital disorder or outcome of a former pathological process

lung infarction or lung edema as a cause of pneumonia (second disease)



Pneumonia  $\sqsubseteq$

$\forall$ realization-of. (PathologicalDisposition  $\sqcap$   $\forall$ inheres-in.(LungInfarction  $\sqcup$  LungEdema))

Refining a pneumonia process by **signs and symptoms** like fever, chills, or cough

Pneumonia  $\sqsubseteq$

$\exists$ has-output. (PathologicalStructure  $\sqcap$   $\exists$ bearer-of. (PathologicalDisposition  $\sqcap$   $\forall$ has-realization. Cough  $\sqcup$  Chills  $\sqcup$  Fever  $\sqcup$  ...))

# Conclusions

- Pre-coordination mainly applies the Equivalent constructor ( $\equiv$ ) in combination with existential ( $\exists$ ) restrictions to create full aristotelian definitions
- Post-coordination also needs class inclusion constructors ( $\sqsubseteq$ ) in combination with value restrictions ( $\forall$ , closure axiom) to provide patterns that can constrain the user
- Post-coordination requires more user guidance
  - This support is provided by axiomatic templates
  - Such axioms can be found in
    1. Upper level Ontologies, like BioTop
    2. content ODPs, as stored in Pattern Databases
- These cODPs require additional semantics, extending OWL-EL to RL
  - Disjoints, negations, value restrictions
- In OWL-DL disjoints need to be used to cope with the *closed world assumption* (Not so important in Frames)
- We need additional GUI support to enforce RL implemented constraints
- Traditional large scale RL ontology reasoning is slow and might not be feasible for post-coordination at annotation time
  - Fast local, incremental reasoning methods need to be investigated

# References and Acknowledgements

## References

- Pneumonia.owl: <http://purl.org/biotop/src/pneumonia.zip>
- BioTop.owl: <http://www.purl.org/biotop/biotop.owl>
- Franz Baader et al., The Description Logic Handbook. Theory, Implementation, and Applications. 2nd edition. Cambridge, U.K.: Cambridge University Press, 2007

## Acknowledgements

- UKLFR
  - Stefan Schulz, Djamila Raufie, Martin Boeker
- The DebugIT EU project



## SNOMED CT

- Express “Insertion of a left hip prosthesis”
  - by single pre-coordinated ID
    - not available
  - by providing IDs for post-coordination via multiple IDs
    - 363704007 procedure site | = (24136001 | hip joint structure | 272741003 | laterality | =7771000 | left | ) {363699004 | direct device | =304120007 | total hip replacement prosthesis | 260686004 | method | =257867005 | insertion-action | }