Automatic generation of consistency constraints for an OWL representation of the FMA

Olivier Dameron, Julie Chabalier



EA 3888 - "Conceptual modeling of biomedical knowledge" Université de Rennes 1, France http://www.ea3888.univ-rennes1.fr

Context

- Integrate human anatomical knowledge into applications
 - diagnosis of bullet injuries [Virtual Soldier Project]
 - grading tumors (lung, glioma)
 - localization of biological entities (iron metabolism)
 - ontologies
- Integrate ontologies of anatomy with other ontologies
 - physiology
 - pathology
 - cellular components
- Perform automatic reasoning

Requirements

- Ontology of anatomy
 - Anatomy = fundamental discipline for medicine
- Representation formalism
 - usable by applications
 - supporting imports
 - supporting reasoning

Available resources

- FMA = reference ontology for anatomical knowledge
 - 75,000 classes + 2.5 millions relations
 - represented in frames
- Previous work on the conversion of the FMA into OWL
 - Golbreich, Zhang and Bodenreider [Web Semantics 2006]
 - Dameron, Rubin and Musen [AMIA 2005]
 - Noy and Rubin [SMI technical report 2007]

Objective

- Guarantee consistency
 - no internal contradiction
 - this is the least that we can expect
 - requires OWL-specific constraints (disjunction...)
 - adequately represents reality
- Generate OWL-specific constraints to improve reasoning
 - complement to the work by Noy and Rubin
 - focus on what is implicit in the FMA-frames ontology

Methods

- Leverage the FMA building principles
 - in the hierarchies and relations
 - in the naming conventions
- 1 building principle -> 1 pattern
 - Check consistency
 - e.g.: direct subclass of "Region of chest wall"
 - Generate additional OWL-specific constraints
 - e.g.: Lung / Left_lung / Right_lung
- Python script for applying pattern

FMA-OWL core architecture

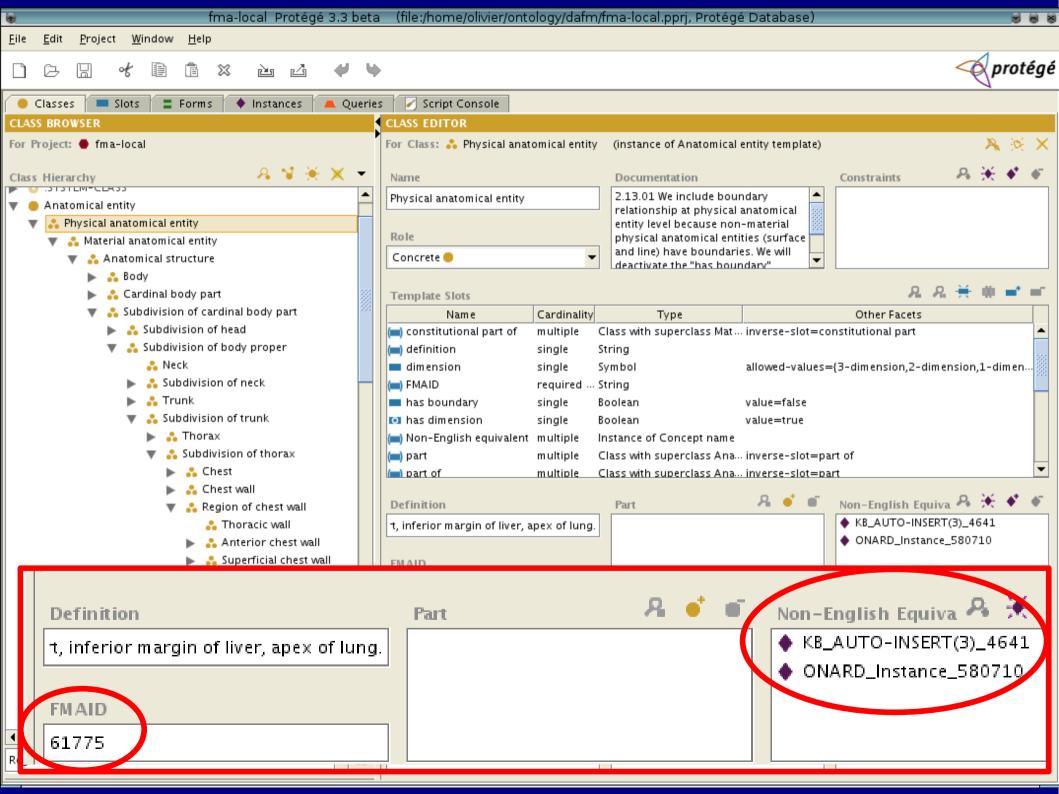
Core architecture

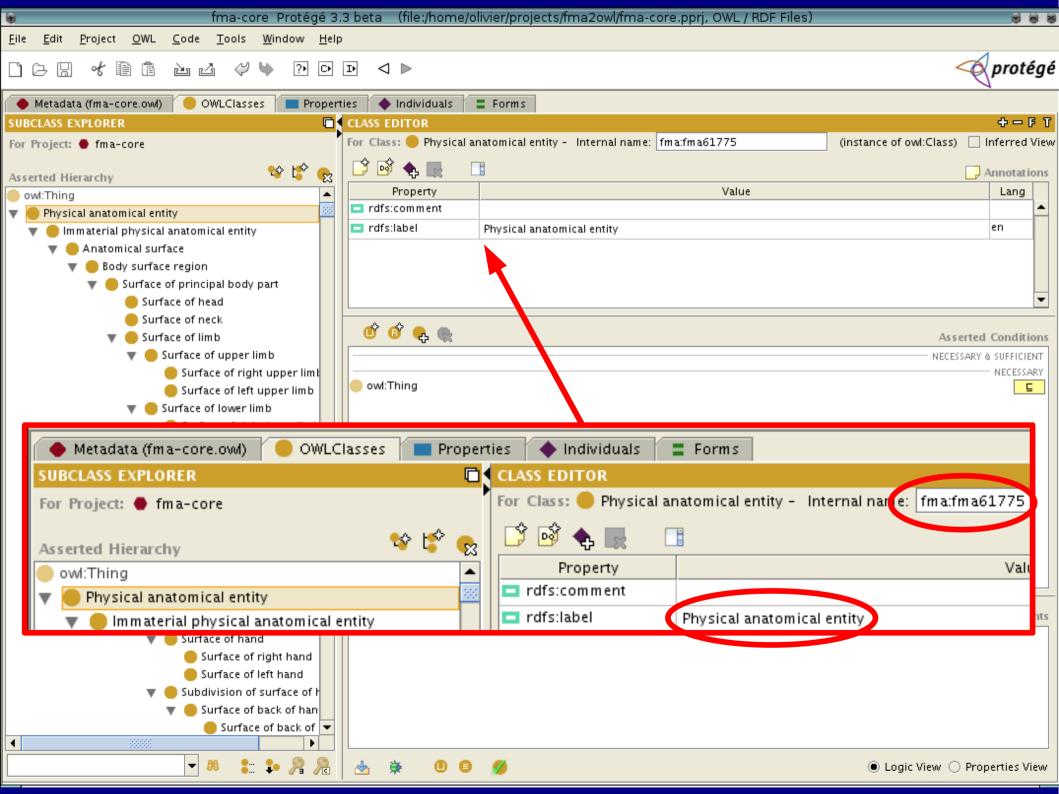
- fma-core.owl = straightforward conversion of the FMA in owl
- 1 owl file / constraint type
 - fma-disjunction.owl
 - imports fma-core.owl
 - import fma-disjunction-lateralization.owl
 - imports fma-disjunction-enumeration.owl ...
 - fma-multipleInheritance.owl
 - imports ...

Taxonomy

Original FMA taxonomy

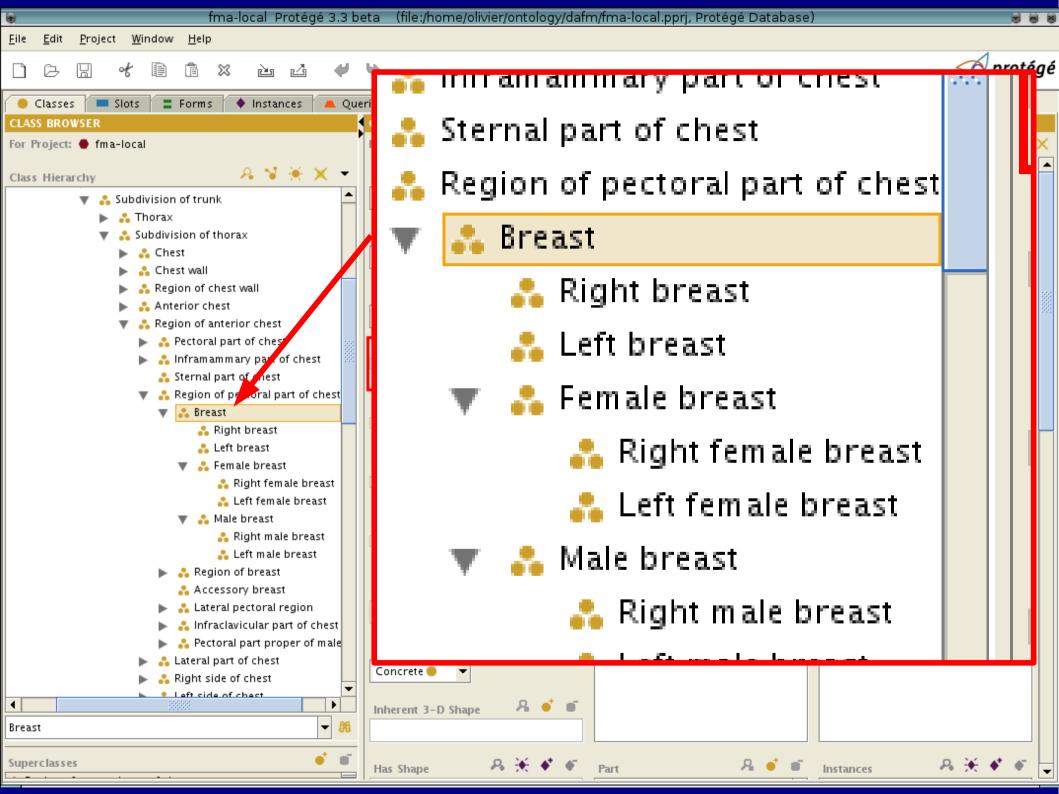
- Single inheritance
- We reused the method by Noy and Rubin
 - frame classes become OWL classes
 - FMAID are used as rdf:ID (prefixed by fma)
 - class name and synonyms become rdfs:label

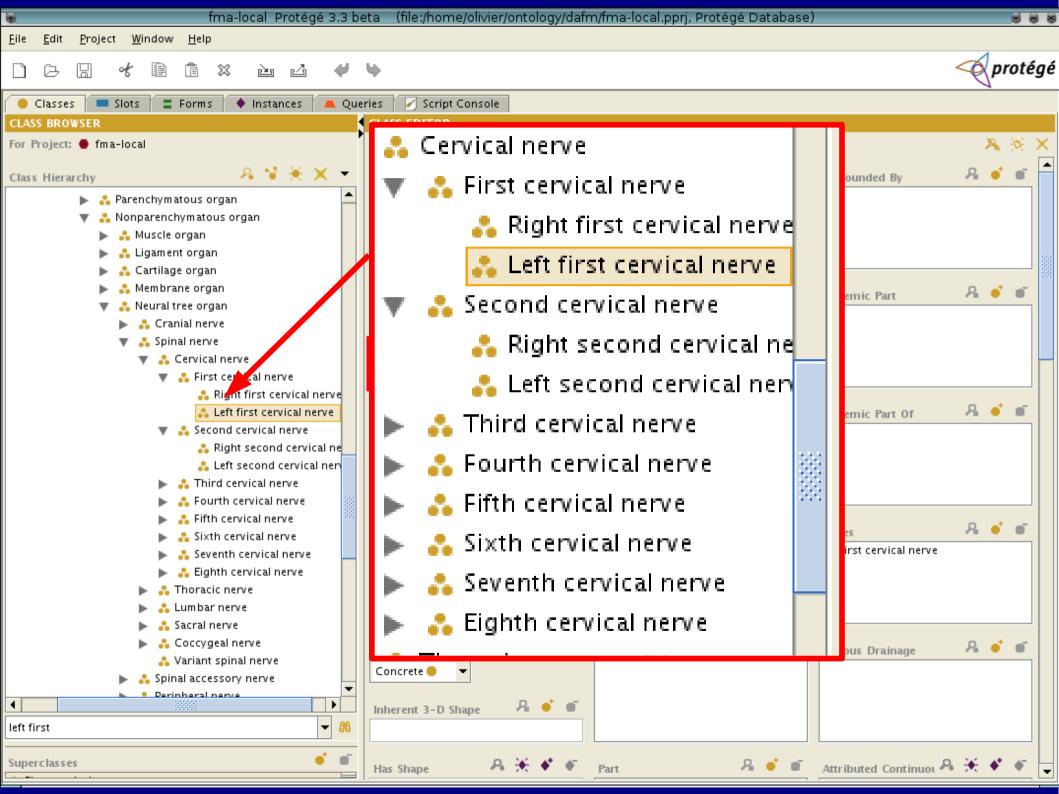




Multiple inheritance

- Left/right and male/female
 - 3 classes (Breast, Areola, Nipple)
- Left/right and enumeration
 - 65 classes (e.g. Left first cervical nerve)
- Upper/lower and enumeration
 - 5 classes (e.g. Upper first molar socket)



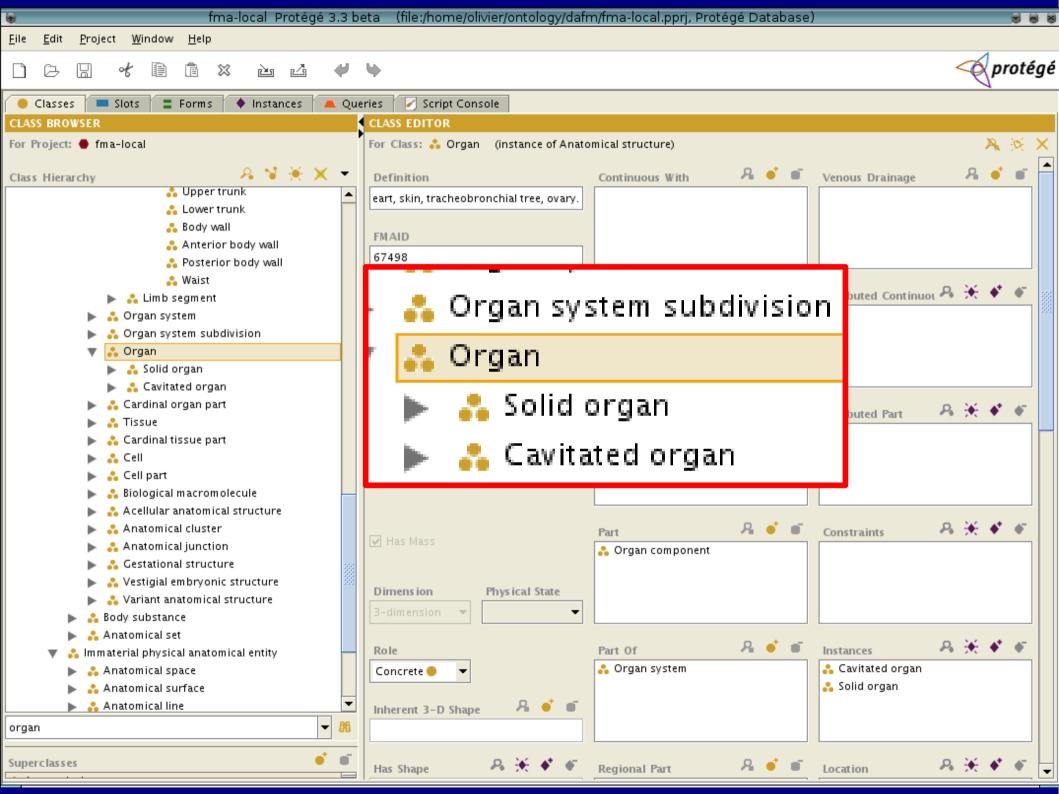


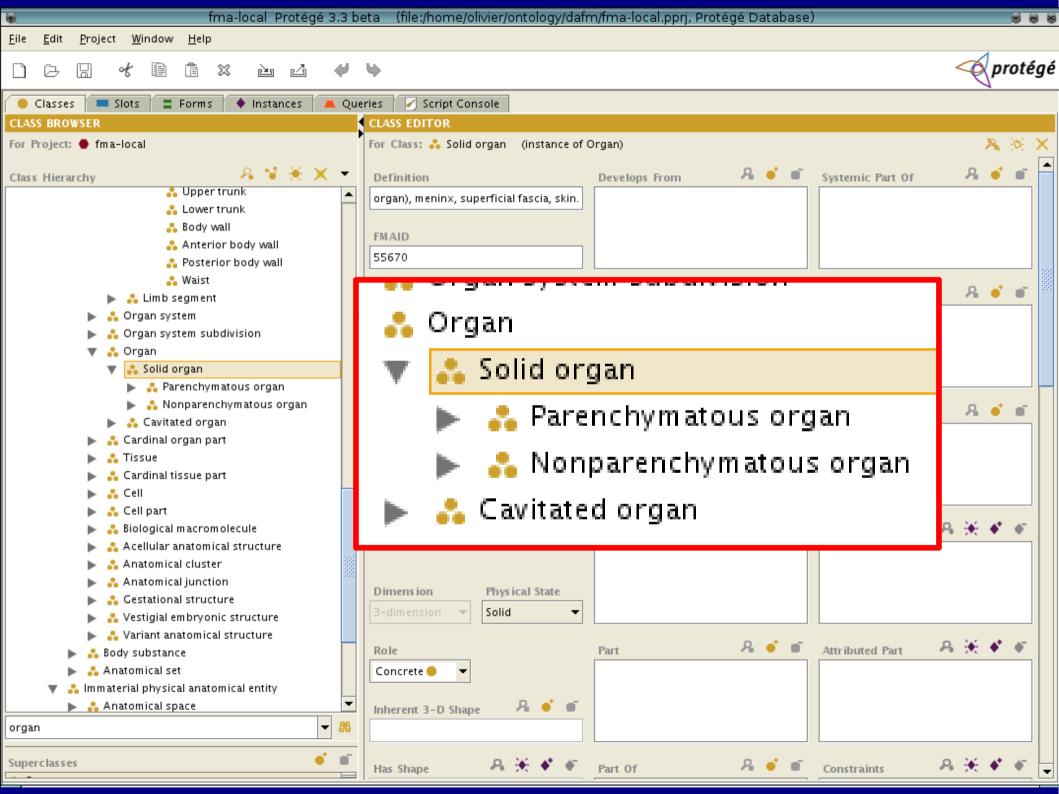
OWL-specific constructs

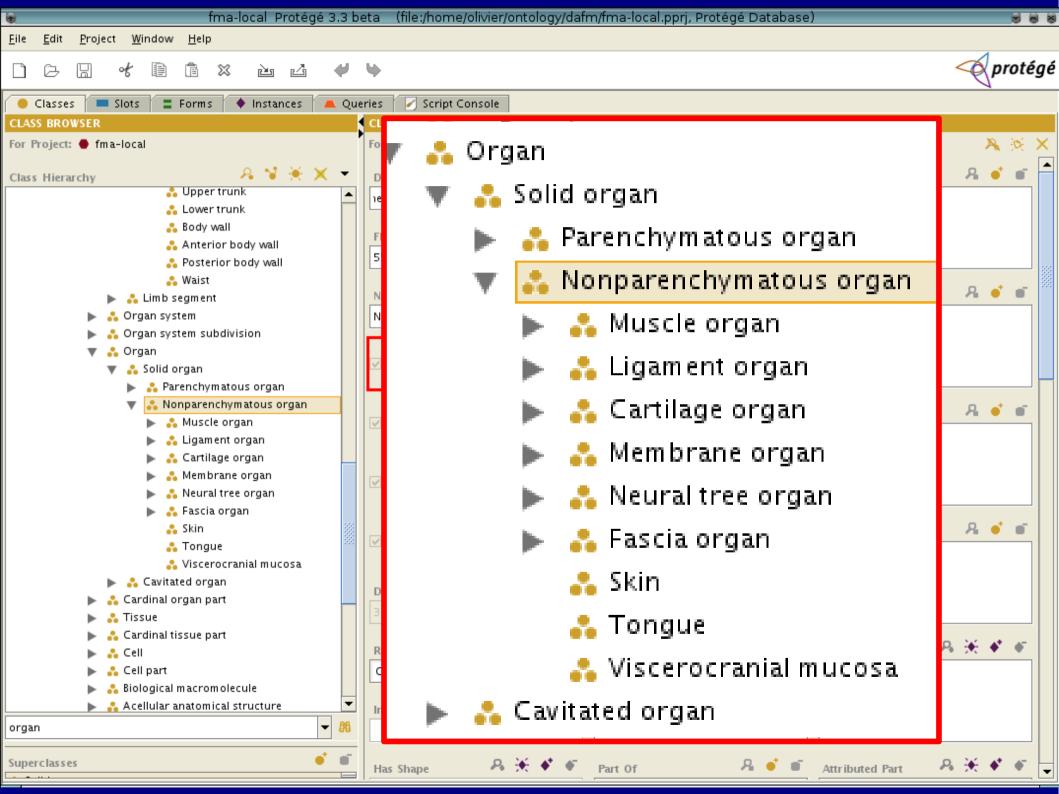
Disjunction

Two classes are disjoint iff they do not have any instance in common

By default in the FMA, all siblings are disjoint



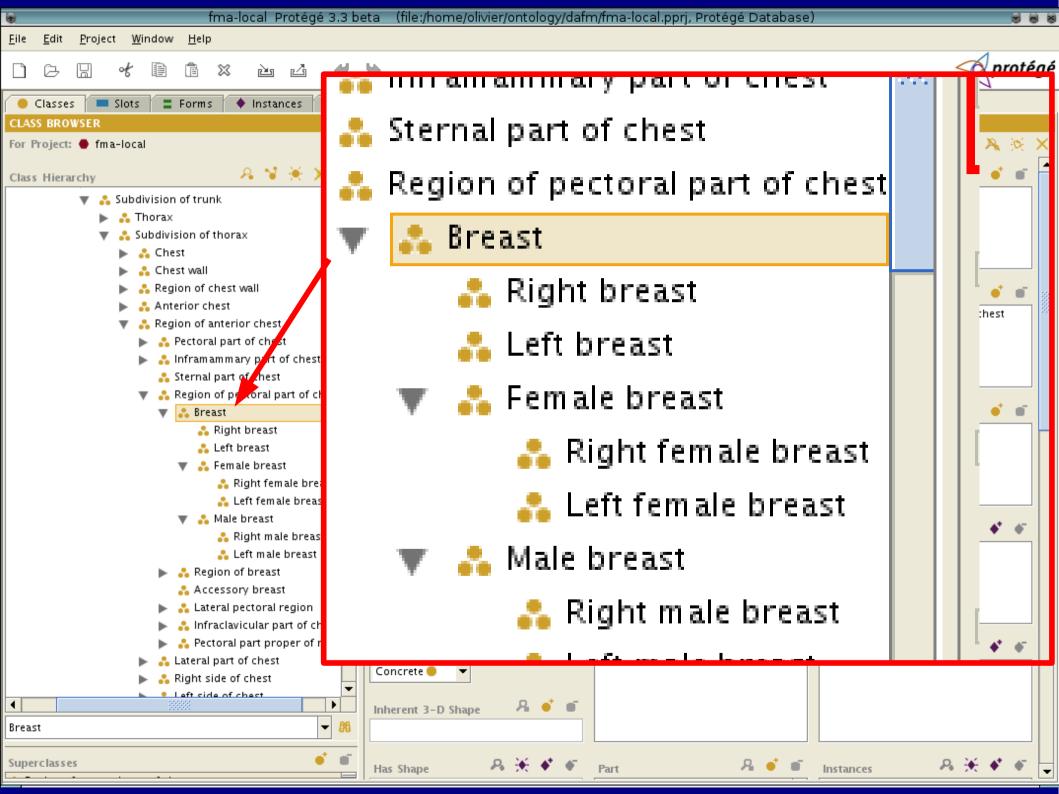




Disjunction

Two classes are disjoint iff they do not have any instance in common

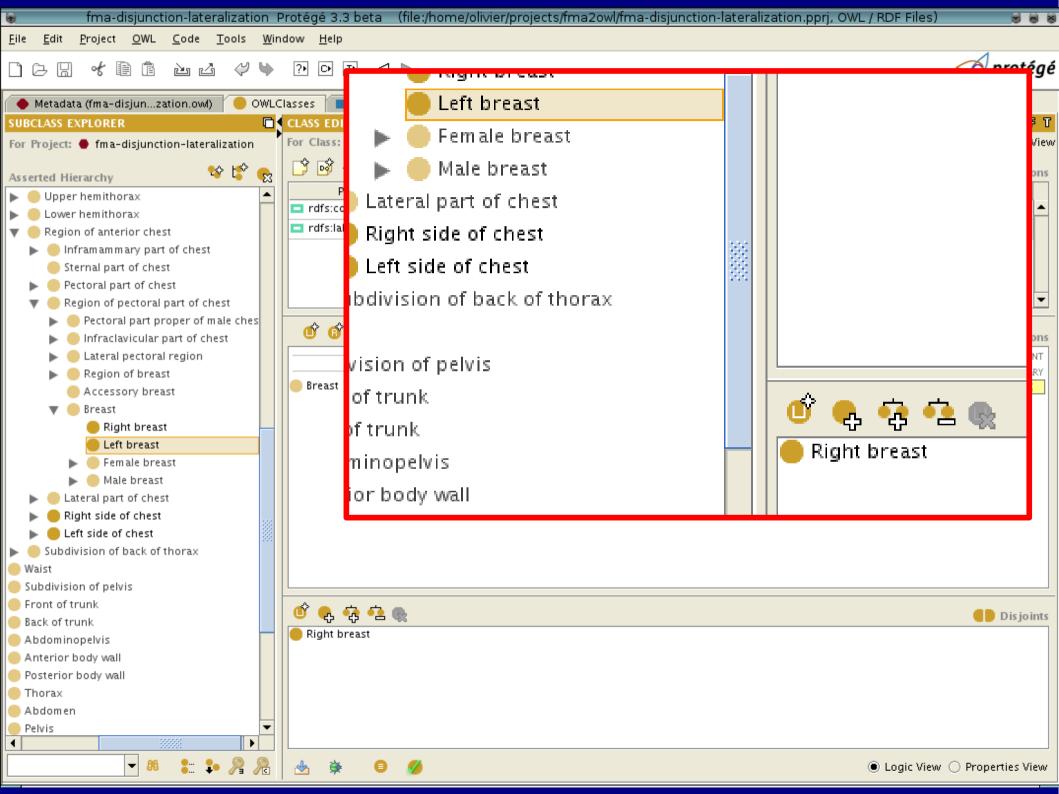
- By default in the FMA, all siblings are disjoint
- Sometimes they are not :-(



Disjunction patterns

- X / NonX or X / Non-X (84)
 - Non-nucleated cell
- Left X / Right X (3736)
 - Left lung
- X left Y / X right Y (13989)
 - Skin of right breast
- Male X / Female X (25)
 - Male breast
- X male Y / X female Y (75)
 - Right side of male chest

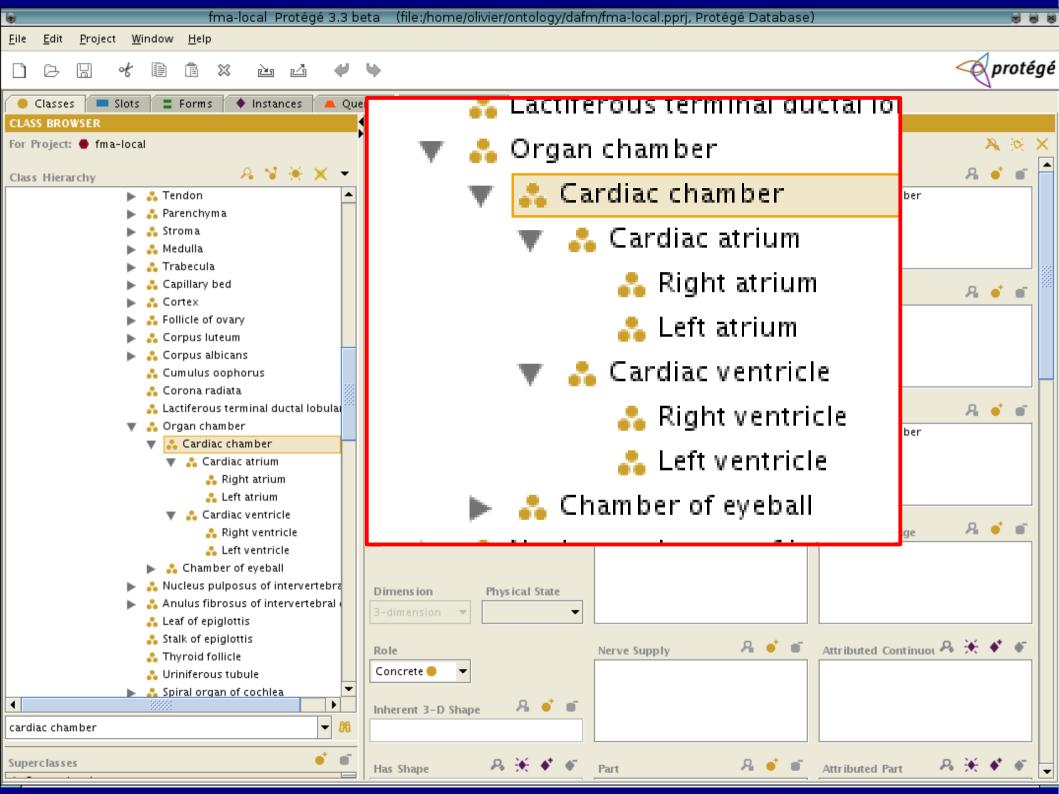
- enumerations (140)
 - First cervical nerve
- upper / (middle) / lower (99)
 - Upper lobe of lung



Coverage

The subclasses of a class provide an exhaustive decomposition of this class (partition)

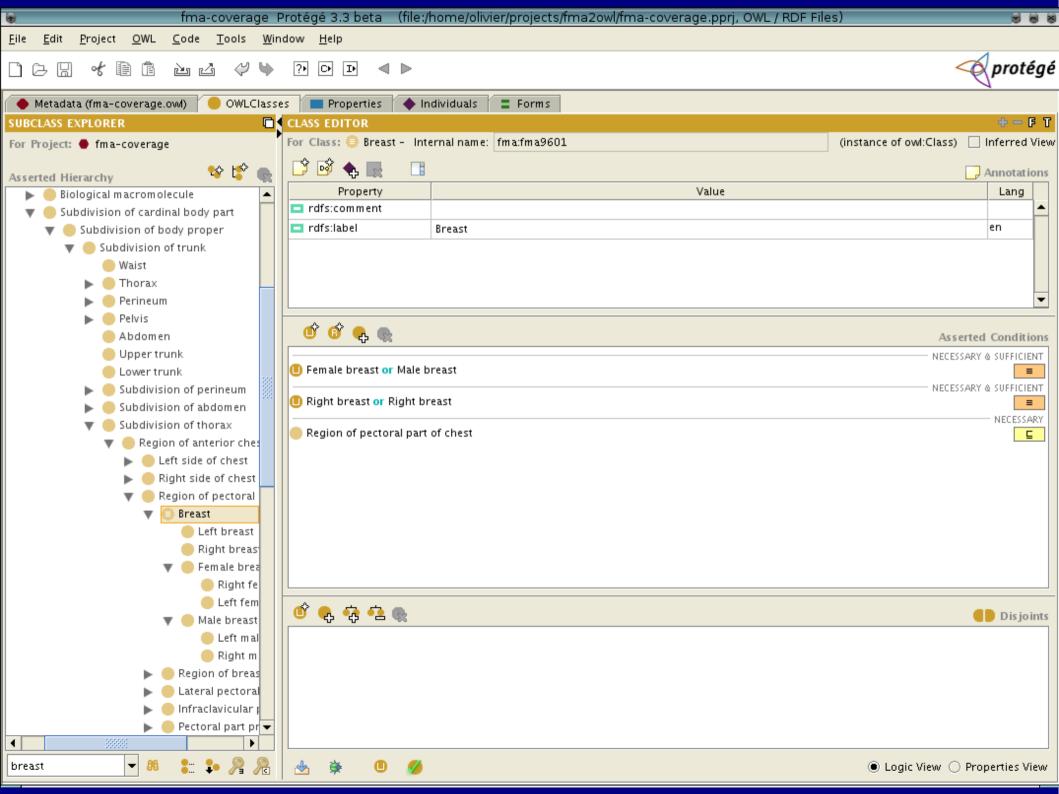
 Most of the time, using the union of all the direct subclasses is enough



Coverage

The subclasses of a class provide an exhaustive decomposition of this class (partition)

- Most of the time, using the union of all the direct subclasses is enough
- But sometimes this is sub-optimal (but still logically valid)



Coverage patterns

- Left / right
 - Breast
- Male / female
 - Breast
- Enumeration
 - Internal intercostal muscle (1 - 11)

- Upper / (middle) / lower
 - Lobe of lung
- X / Non-X
 - Cell (nucleated vs Non-nucleated)

We reused the multiple inheritance patterns

Consistency checking

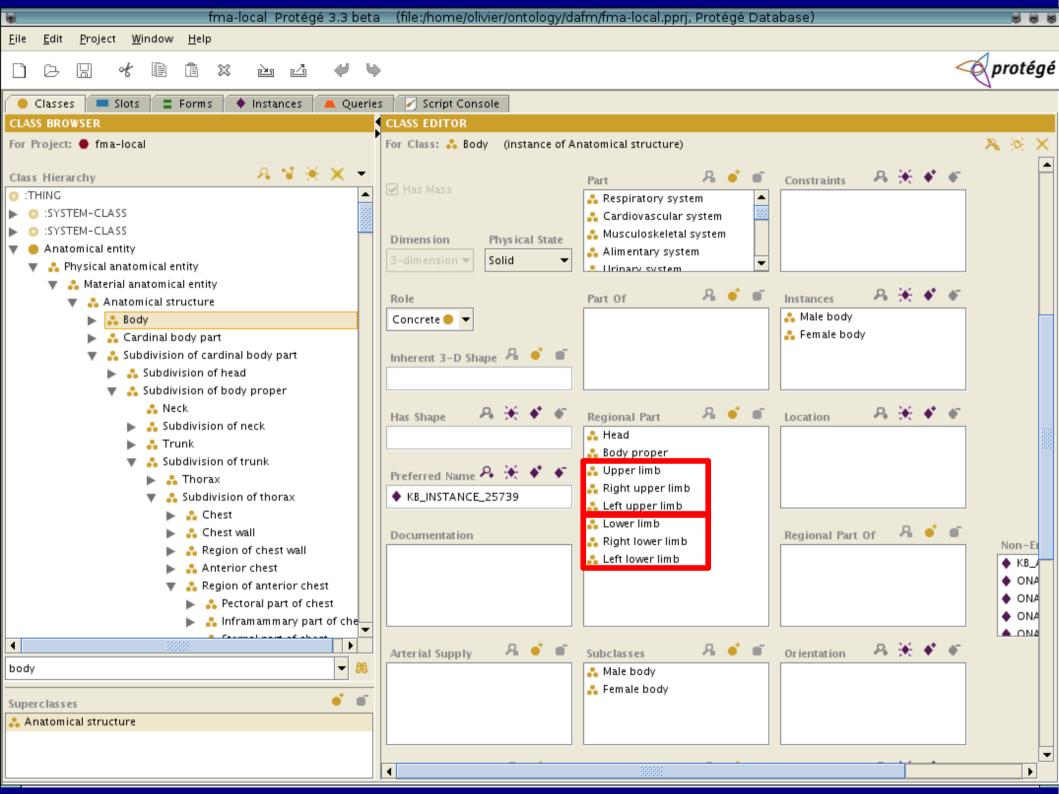
check that among the direct XXX parts of an anatomical entity, none is a superclass of the other

- part: 118
- constitutional part: 26
- regional part: 86
- systemic part: 34

 If we consider the regional parts of "Body", we find:

```
- "Head",
```

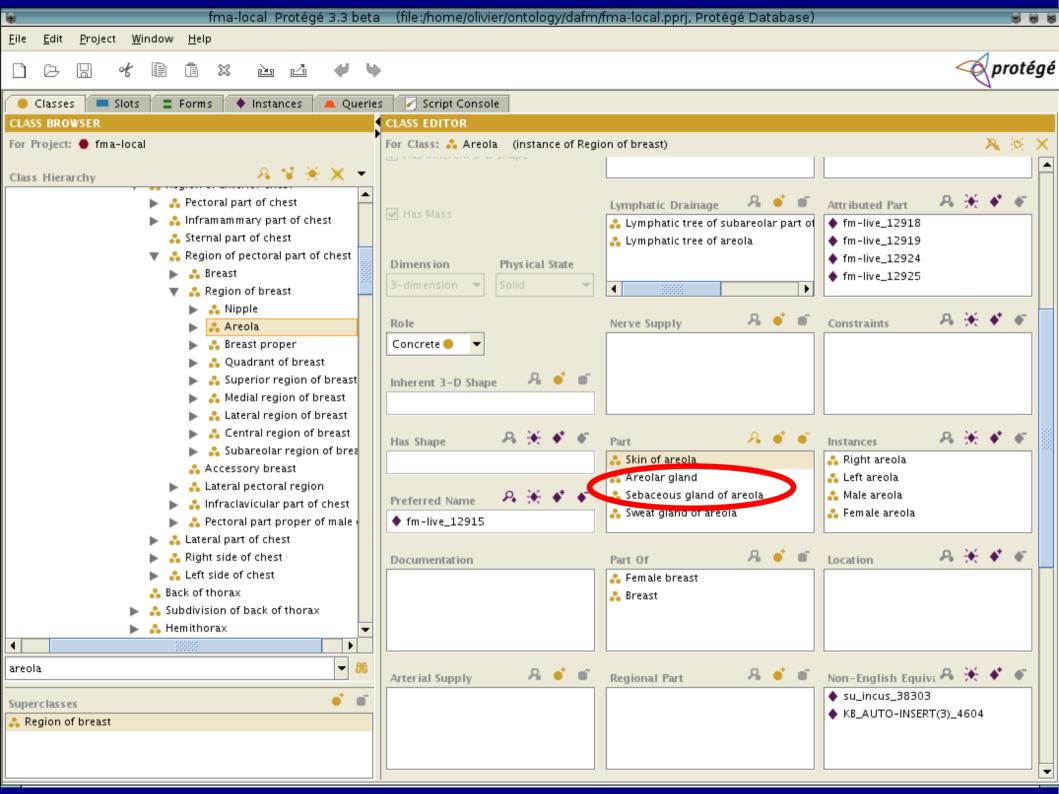
- "Body proper",
- "Upper limb", "Left upper limb", "Right upper limb",
- "Lower limb", "Left lower limb", "Right lower limb".

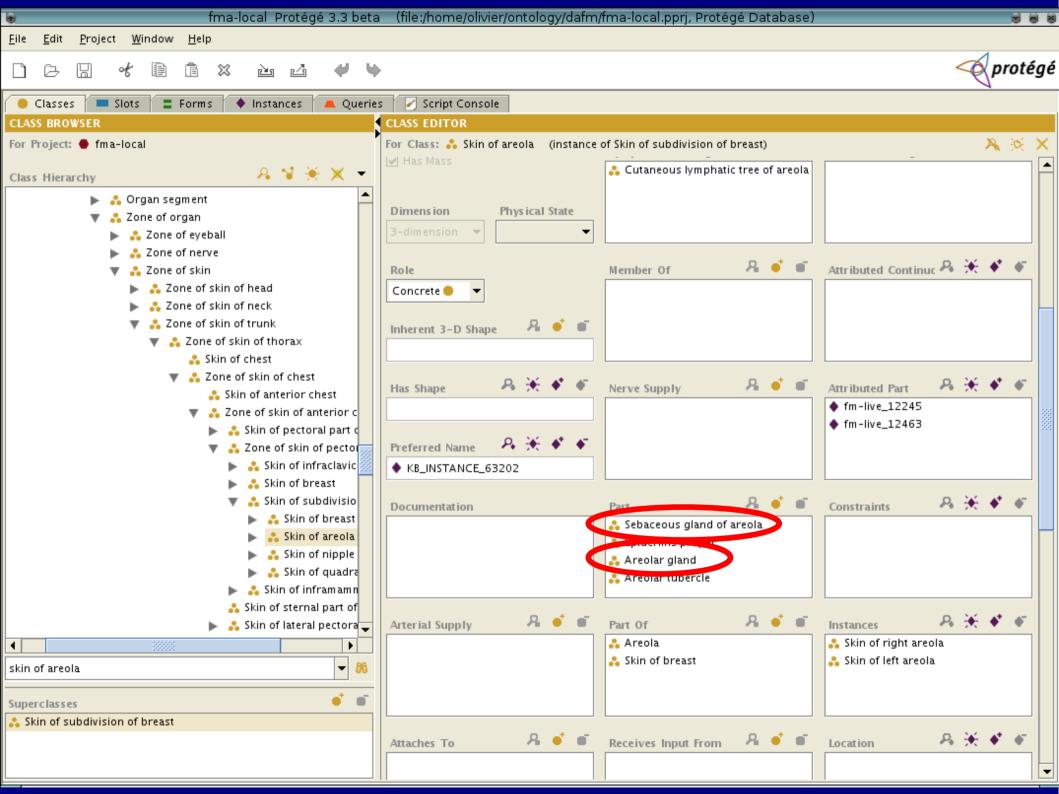


- If we consider that this represents one exhaustive decomposition of the body into regional parts, it makes sense to specify that there has to be one left lower limb and one rigt lower limb and one left upper limb... But then, "Upper limb" and "Lower limb" are not necessary (but still correct).
- Now, if we consider that these are the only possible elements of this partition, we only need "Head", "Body proper", "Upper limb" and "Lower limb"... and the left/right upper/lower limbs are again not necessary (but admittedly again still correct).

check that among the direct XXX parts of an anatomical entity, none is a part (possibly indirect) of the other

ex: Areola has among its parts "Skin of areola",
"Sebaceous gland" and "Areolar gland". It turns out that "Sebaceous gland" and "Areolar gland" are already parts of "Skin of areola"



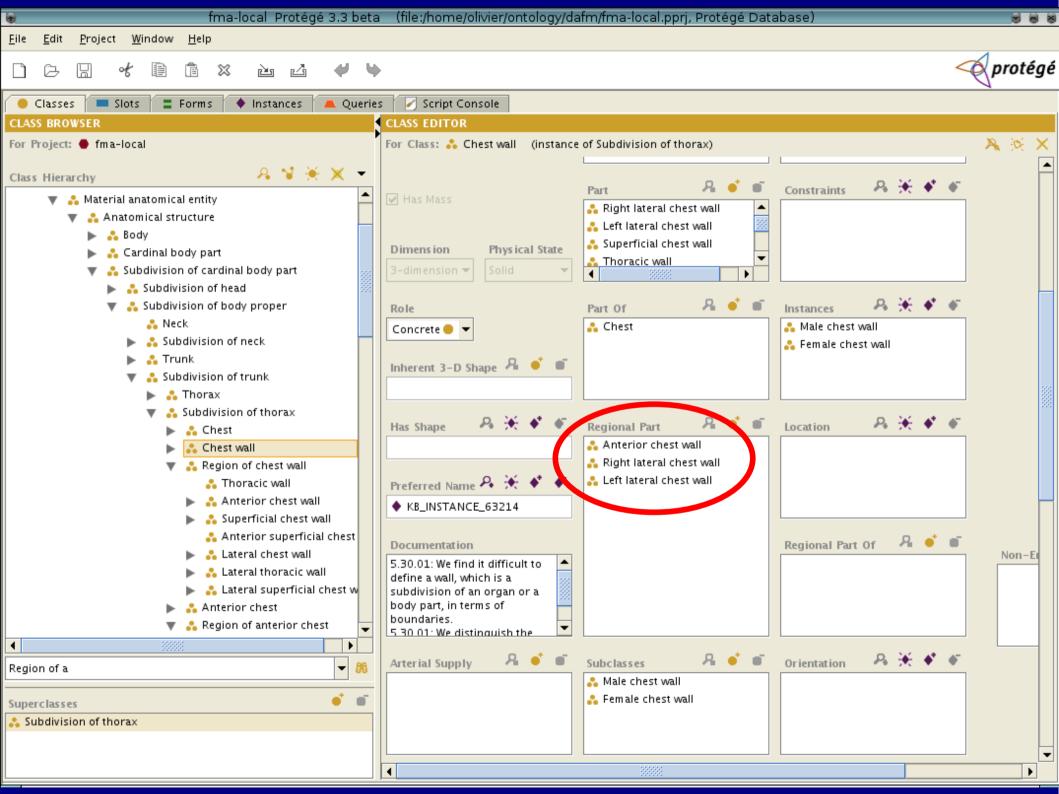


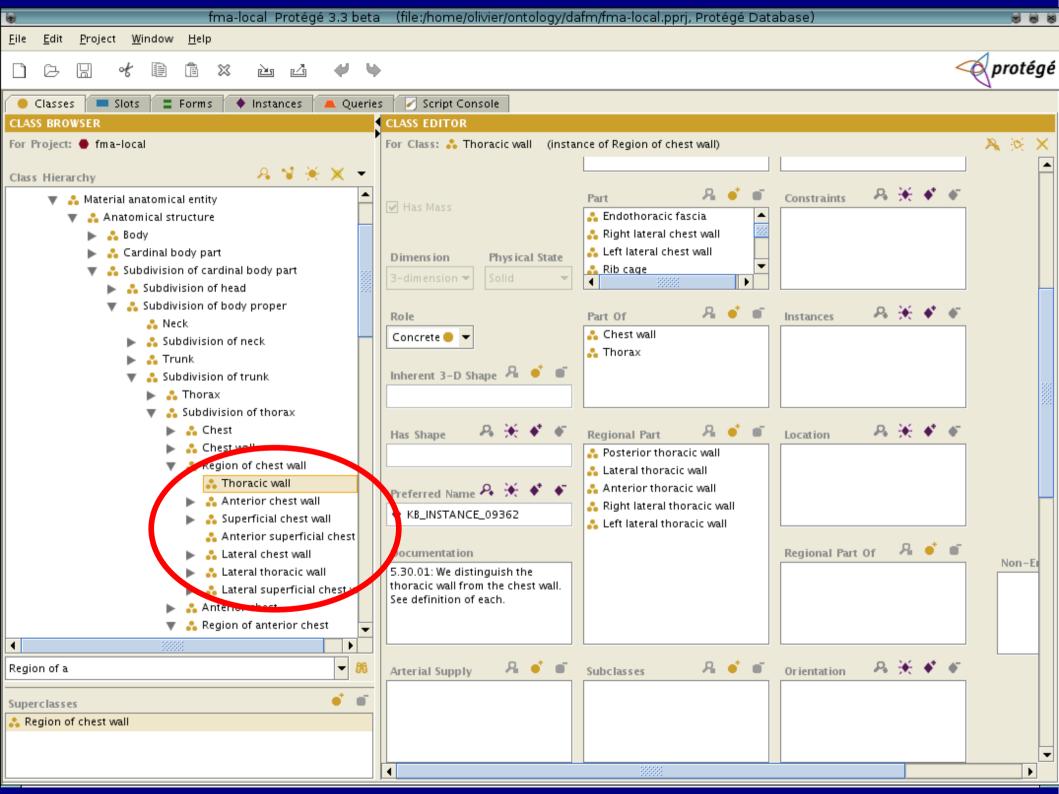
Consistency (cont.)

- For the subclasses of "Subdivision of thorax":
 - part: 24 inconsistencies (8 classes)
- For the subclasses of "Subdivision of trunk":
 - part: 42 inconsistencies (13 classes)
 - constitutional part: 4 inconsistencies (1 class)
 - regional part: 8 inconsistencies (1 class)

check that the parts of X and the subclasses of "Part of X" (if it exists) are consistent

- ex: "Region of chest wall" and the regional parts of "Chest wall"
- Procedure:
 - all the parts of X are subclasses of "Part of X"
 - all the subclasses of "Part of X" are actually parts of X





check that the parts of X and the subclasses of "Part of X" (if it exists) are consistent

- part: 24 inconsistencies
- regional parts: 11008 inconsistancies

Synthesis

Discussion

- Patterns "manually" generated
 - coverage for Vertebra
- Auxiliary classes (Part_of_X, Region_of_Y, ...):
 - exist in FMA-frames so should also exist in FMA-OWL
 - should only be defined
- FMA complex class/metaclass architecture
- Closure

Conclusion

- Consistency checking
 - is needed (the bigger your onto, the more you need it)
 - should be automated
 - detect current mistakes
 - avoid new mistakes during updates
 - ... toward semi-automatic correction?
- Generating the OWL-specific part of the FMA
 - automatic + ad hoc