

Terminology Systems and Protege

Ameen Abu-Hanna

Ronald Cornet

Medical Informatics

University of Amsterdam

Monica Crubezy

Samson Tu

Stanford Medical Informatics

Stanford University

Outline

- Understanding Terminology Systems
- Protege counter-parts

Terminology Systems

Many T systems

- ICD
- NHS Clinical Terms
- SNOMED
- ...

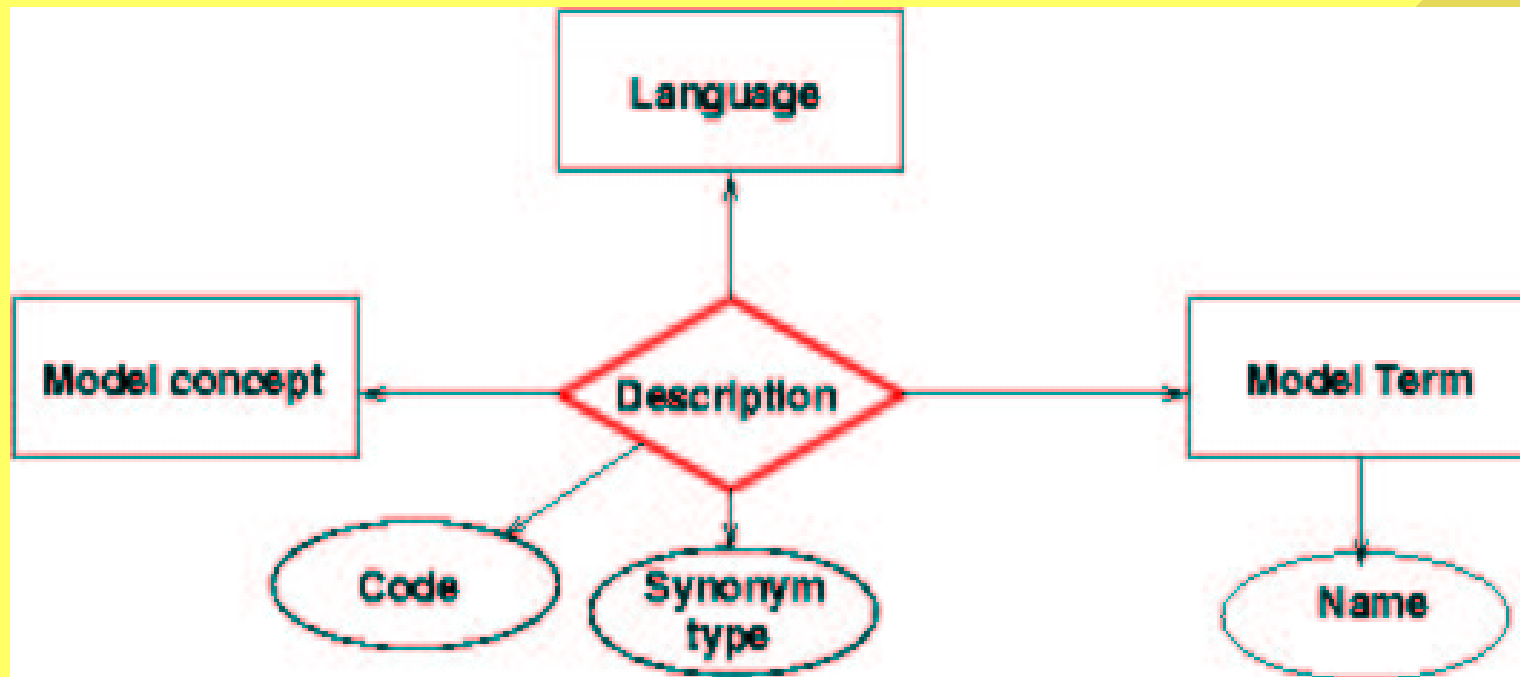
Desiderata

- Concept oriented
- Explicit Rels
- Composition Rules
- Domain Completeness
- Multiple classification
- Use of definitions
- Synonyms
- Multilingual
- Non redundant
- Unique codes
- Non ambiguous
- Non vague
- Context-free codes

Desiderata

- Concept oriented
- Explicit Rels
- Composition Rules
- Domain Completeness
- Multiple classification
- Use of definitions
- Synonyms
- Multilingual
- Non redundant
- Unique codes
- Non ambiguous
- Non vague
- Context-free codes

Meta Model



FOL constraint on meta-model

For every language there is exactly one preferred term per concept

for-all c in Concept
for-all l in Language
exists-! t in Term #meaning exactly one

(and (Description(c, l, t)
 Synonym-type(<c,l,t>)="preferred"))

FOL constraints on meta-model

Different concepts cannot have the same preferred term

```
for-all c1 in Concept
for-all l in Language
for-all t in Term
```

```
(==> (and Description(c1, l, t)
      Synonym-type(<c1, l, t>="preferred"))
```

```
(not-exists c2 in Concept
 (and Description(c2, l, t)
      Synonym-type(<c2, l, t>="preferred"
      c2 != c1)))
```

How to specify a new TS?

Intensive Care

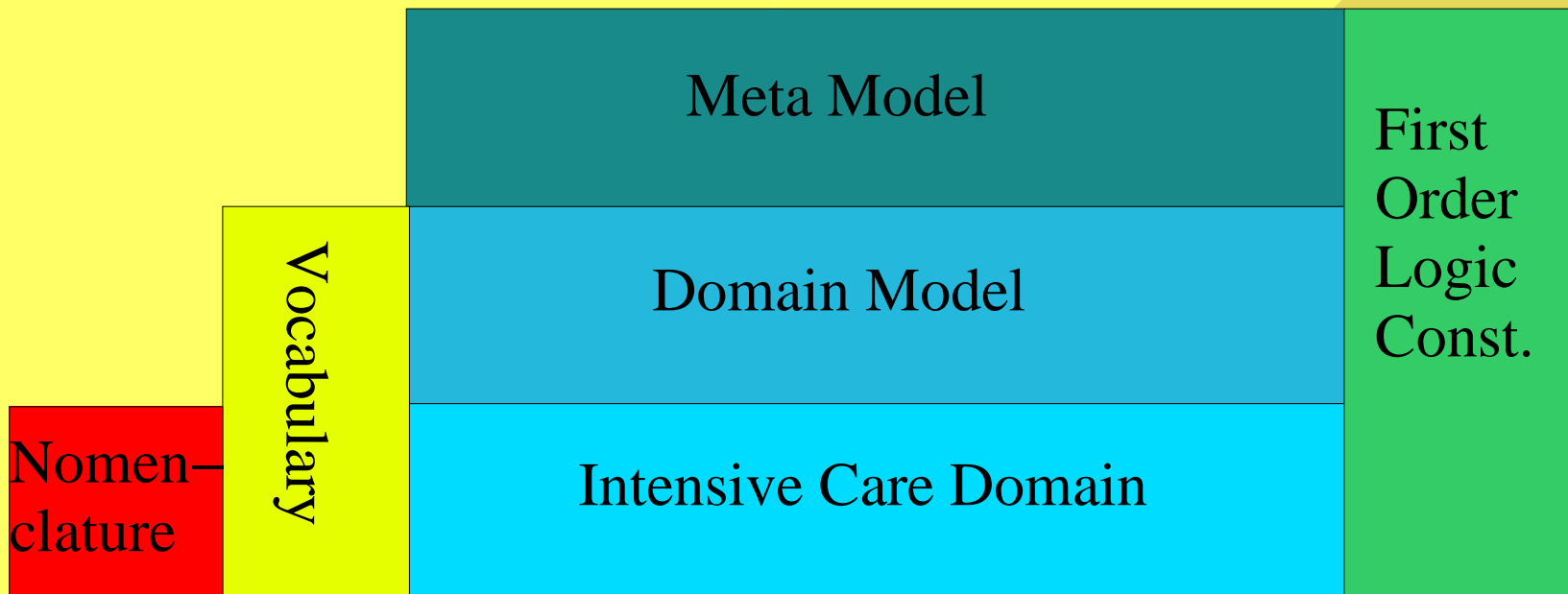
Intensive Care Knowledge Base

± 2500 concepts: Diseases

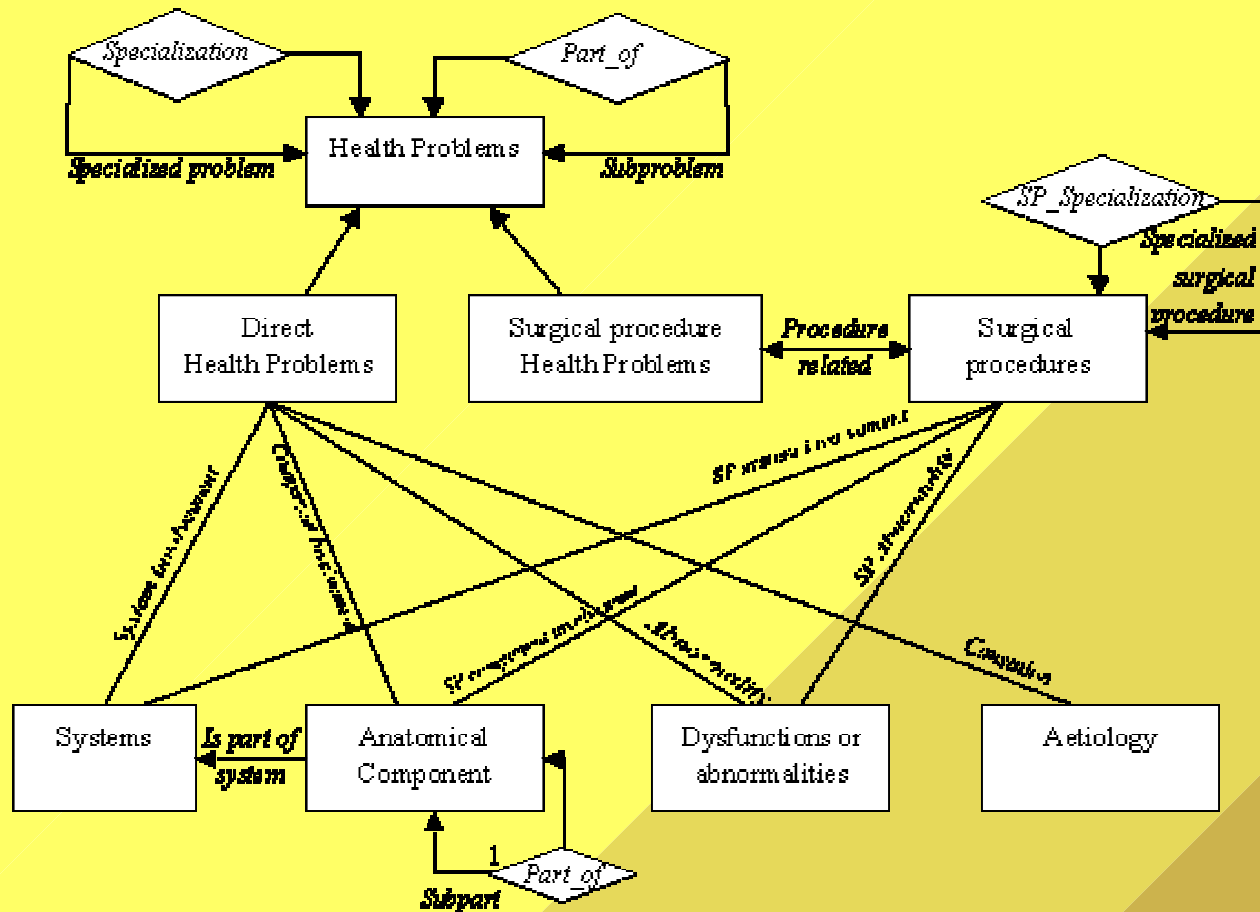
Anatomy, Systems, Abnormalities, Etiologies

- **Hepatitis:** *isa* **Disease**,
location **Liver**,
abnormality **Inflammation**
- **Infective Hepatitis:** *isa* **Hepatitis**,
cause **Micro-organism**

Knowledge Components in DICE

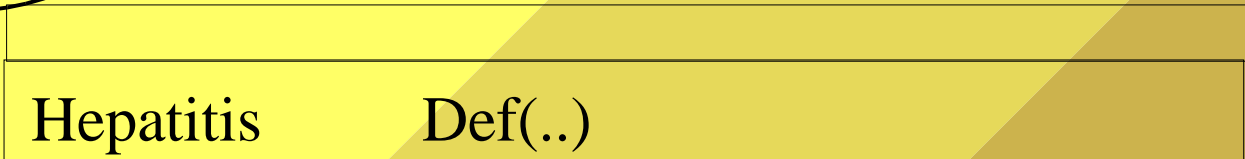
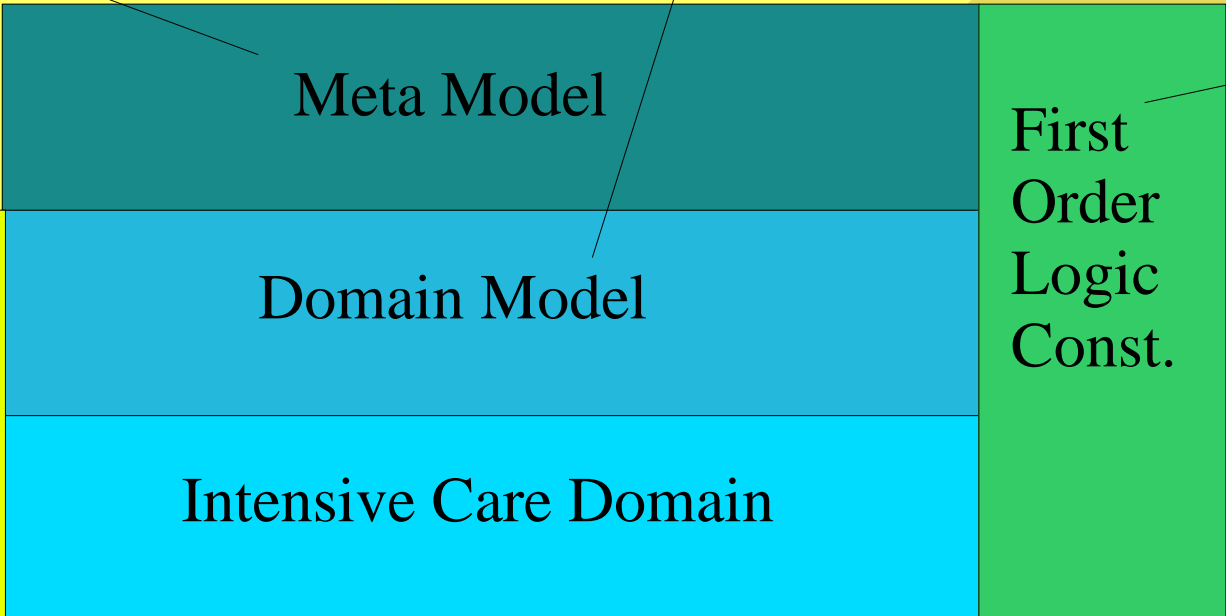
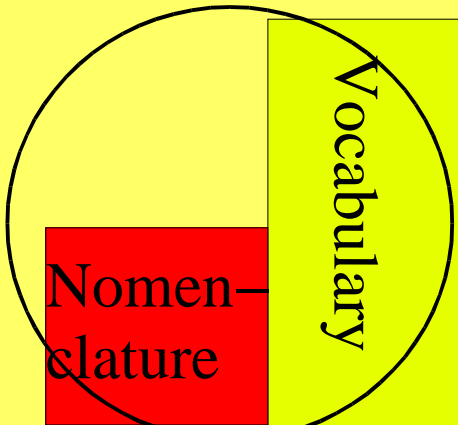
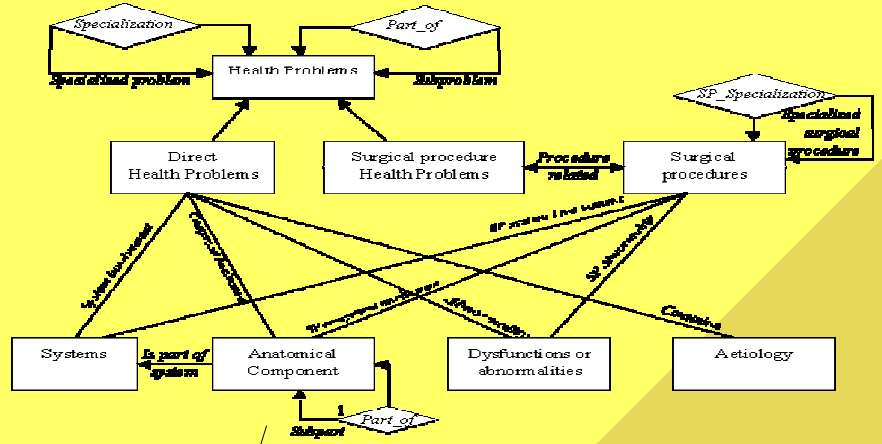
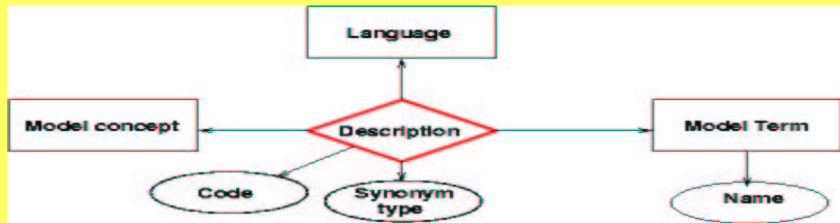


Domain Model in DICE



Nomenclature/Defs in DICE

Disease	Anatomical Component	Abnormality	Aetiology
Hepatitis	Def: Liver	Def: Infection	OR: (HepV, E-BV, CytomegaloV)



Protege?

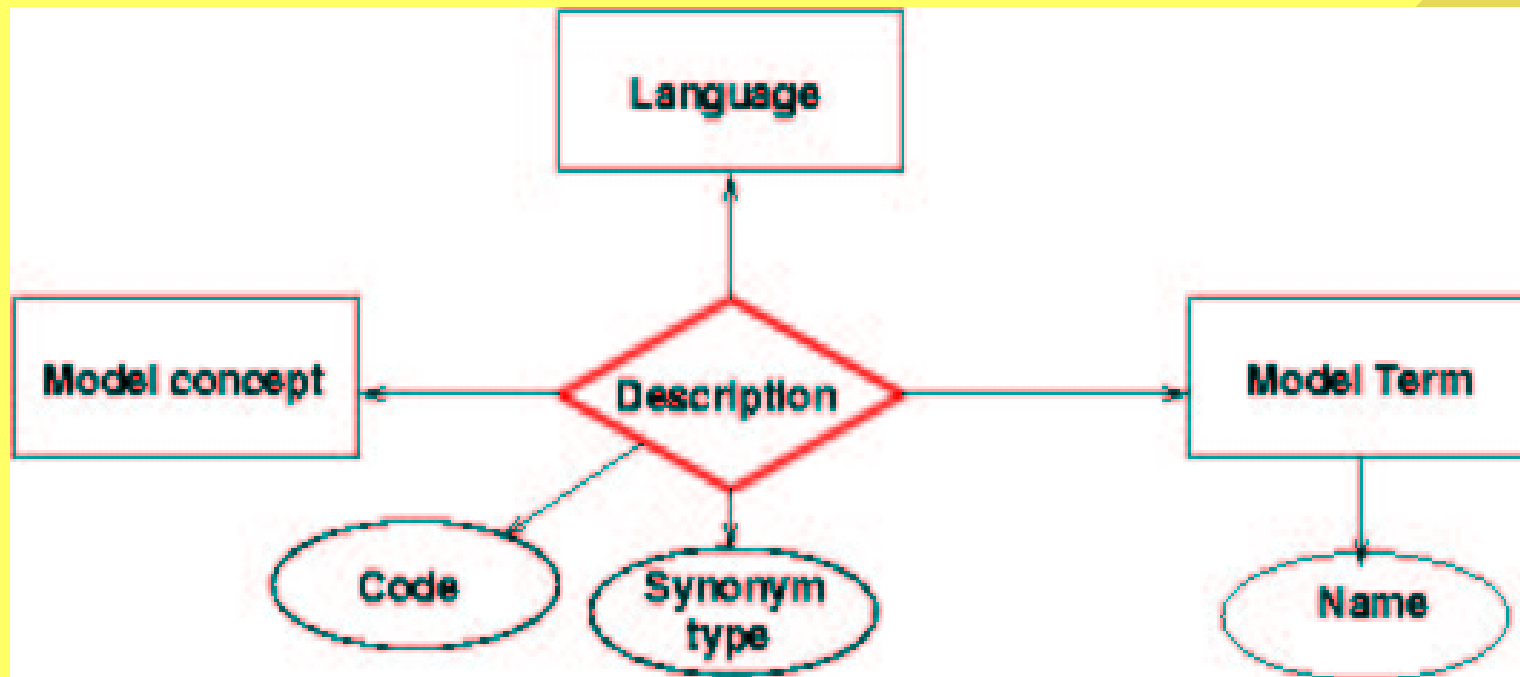
Knowledge model in Protege

- **Concept oriented**
- **Explicit Rels**
- Composition Rules
- **Domain Completeness**
- **Multiple classification**
- **Use of definitions**
- Synonyms
- Multilingual
- Non redundant
- Unique codes
- Non ambiguous
- Non vague
- Context-free codes

Knowledge model in Protege

- **Concept oriented**
- **Explicit Rels**
- **Composition Rules**
- **Domain Completeness**
- **Multiple classification**
- **Use of definitions**
- Synonyms
- Multilingual
- Non redundant
- Unique codes
- Non ambiguous
- Non vague
- Context-free codes

Meta Model

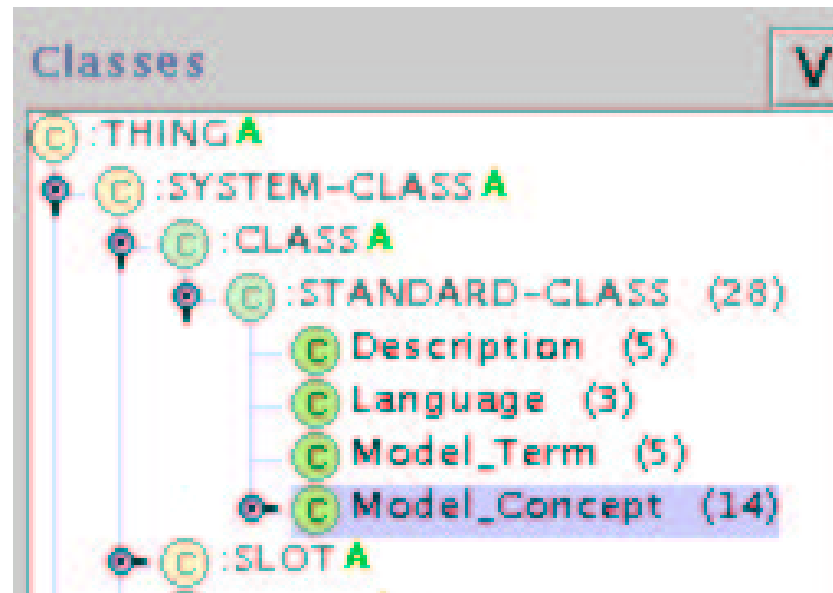


Meta Model in Protege: Using **Meta Classes**

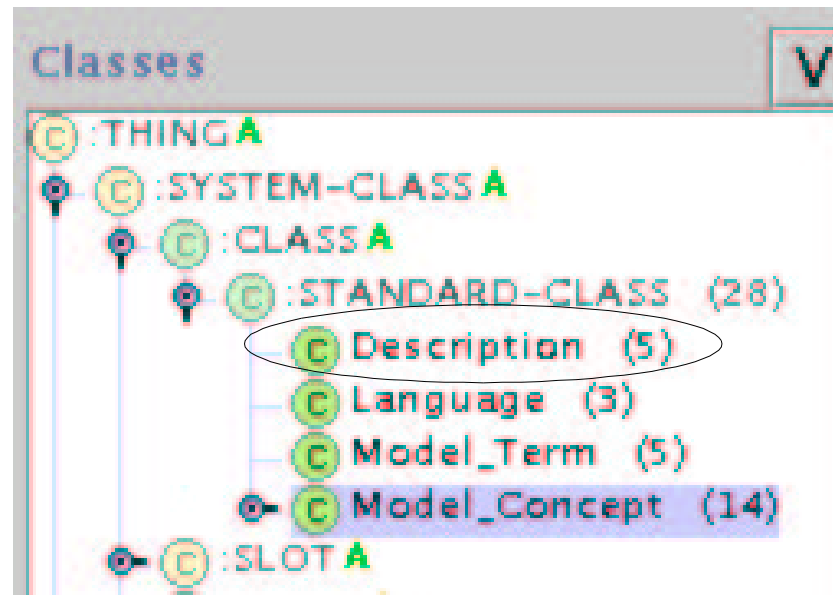
Standard class

- Description
- Language
- Model concept
- Model term

Meta Classes

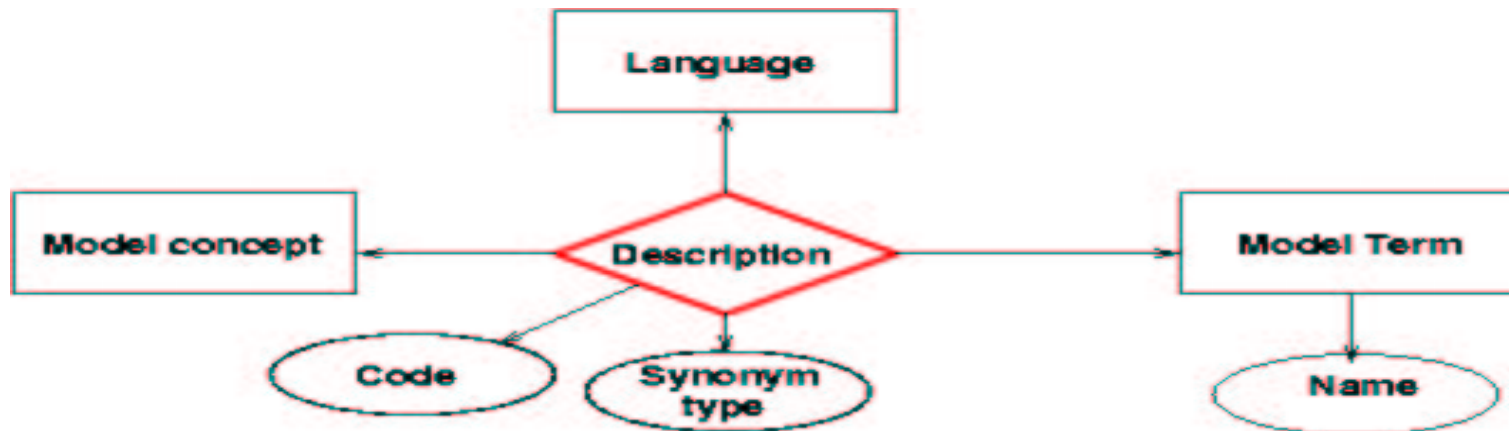


Meta Classes

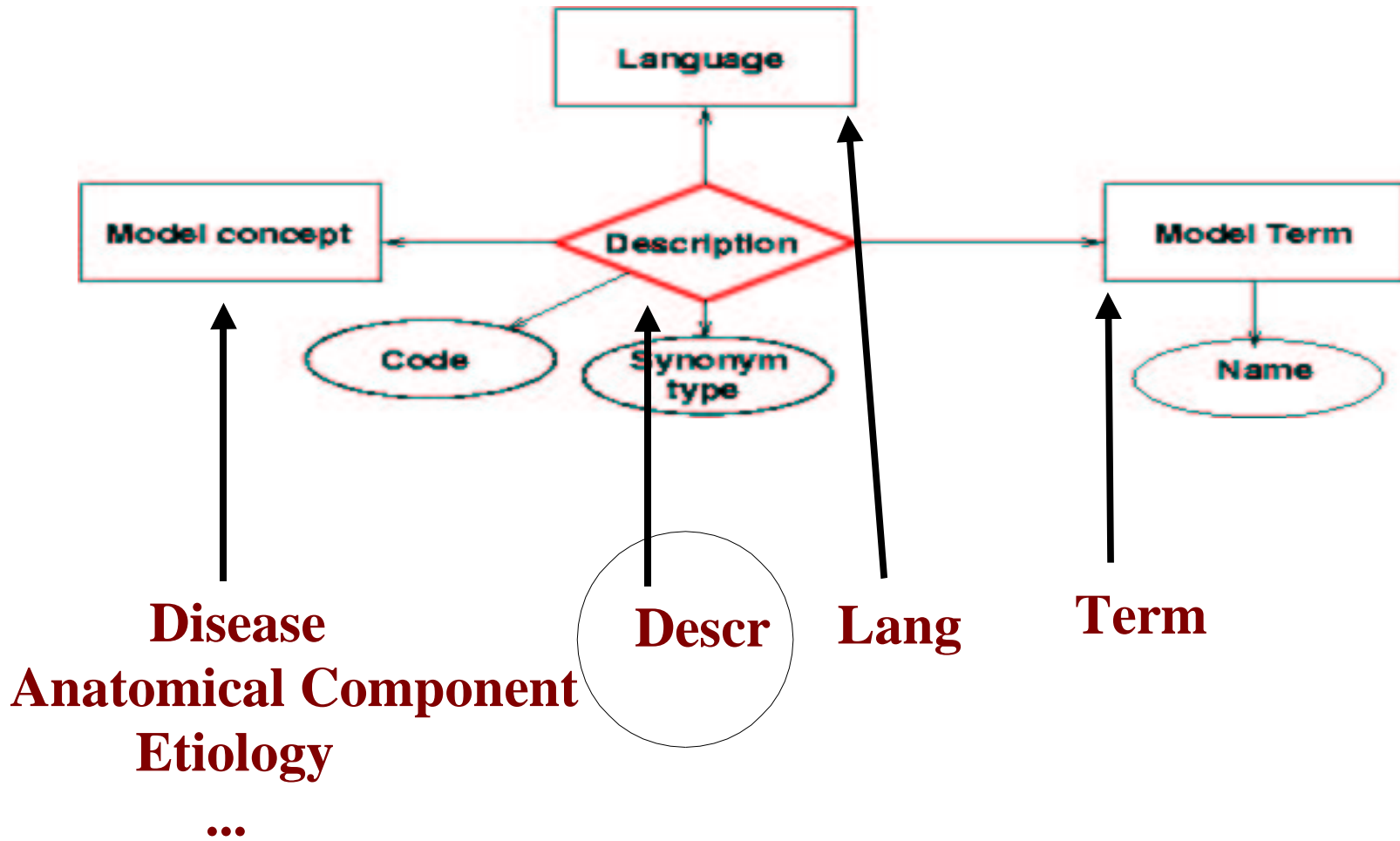


Description

Template Slots				V	V	C	X	+	-
Name	Type	Cardinality	Other Facets						
S code	String	single							
S synonym_type	Symbol	required single	allowed-values={preferred,synonym} ...						
S description.modelterm	Instance	required single	classes={Model_Term}						
S description.language	Instance	required single	classes={Language}						
S description.modelconcept	Instance	required single	classes={Model_Concept}						



Instance Classes



Instances of Meta Classes

The screenshot shows the Protégé-2000 interface for editing the 'Descr' class. The 'Instances' tab is active, showing the following details:

- Name:** Descr
- Role:** Abstract
- Template Slots:**

Name	Type	Cardinality	Other Facets
localcode	Integer	required single	
ronalcode	Integer	required single	
- Instance Fields (highlighted in red):**
 - Description.modelconcept:** Code: 000
 - Description.language:** (empty field)
 - Description.modelterm:** (empty field)
 - Synonym Type:** (dropdown menu)

PAL: Codes are unique

The screenshot shows a window titled "Codes are unique (PAL-CONSTRAINT)". The interface is divided into several sections:

- Name:** A text field containing "Codes are unique".
- Description:** A text area containing the text: "Two different instances of the Description metaclass cannot have the same value for their code slot."
- Statement:** A code editor with a toolbar (check, list, undo, redo, copy, paste, close) and the following code:

```
(forall ?descr1
  (forall ?descr2
    (=> (and (/= ?descr1 ?descr2)
              (own-slot-not-null code ?descr1)
              (own-slot-not-null code ?descr2))
          (/= (code ?descr1) (code ?descr2))))))
```
- Range:** A text area containing the following code:

```
(defrange ?descr1 :FRAME Description)
(defrange ?descr2 :FRAME Description)
```

PAL: 1 preferred term per concept

```
(forall ?language
  (forall ?description1
    (forall ?description2
      (=> (and (language.description ?language ?description1)
              (language.description ?language ?description2)
              (/= ?description1 ?description2)
              (= (description.modelconcept ?description1)
                (description.modelconcept ?description2))
              (synonym_type ?description1 (coerce-to-symbol "preferred"))))
        (and (not (synonym_type ?description2 (coerce-to-symbol "preferred"))))
          (exists ?term1
            (and (description.modelterm ?description1 ?term1)
                 (modelterm.description ?term1 ?description1)
                 (not (exists ?term2
                       (and (/= ?term2 ?term1)
                            (description.modelterm ?description1 ?term2)
                            (modelterm.description ?term2 ?description1))))))))))
```

PAL: 1 concept per preferred term

```
(forall ?desc1
  (forall ?desc2
    (=> (and
      (/= ?desc1 ?desc2)
      (synonym_type ?desc1 (coerce-to-symbol "preferred"))
      (synonym_type ?desc2 (coerce-to-symbol "preferred"))
      (= (description.language ?desc1)
         (description.language ?desc2))
      (/= (description.modelconcept ?desc1)
          (description.modelconcept ?desc2))))
    (/= (description.modelterm ?desc1)
        (description.modelterm ?desc2))))))
```

Knowledge model in Protege

- Concept oriented
- Explicit Rels
- Composition Rules
- Domain Completeness
- Multiple classification
- Use of definitions
- Synonyms
- Multilingual
- Non redundant
- Unique codes
- Non ambiguous
- Non vague
- Context-free codes

Knowledge model in Protege

- Concept oriented
- Explicit Rels
- Composition Rules
- Domain Completeness
- Multiple classification
- Use of definitions
- Synonyms
- Multilingual
- Non redundant
- Unique codes
- Non ambiguous
- Non vague
- Context-free codes

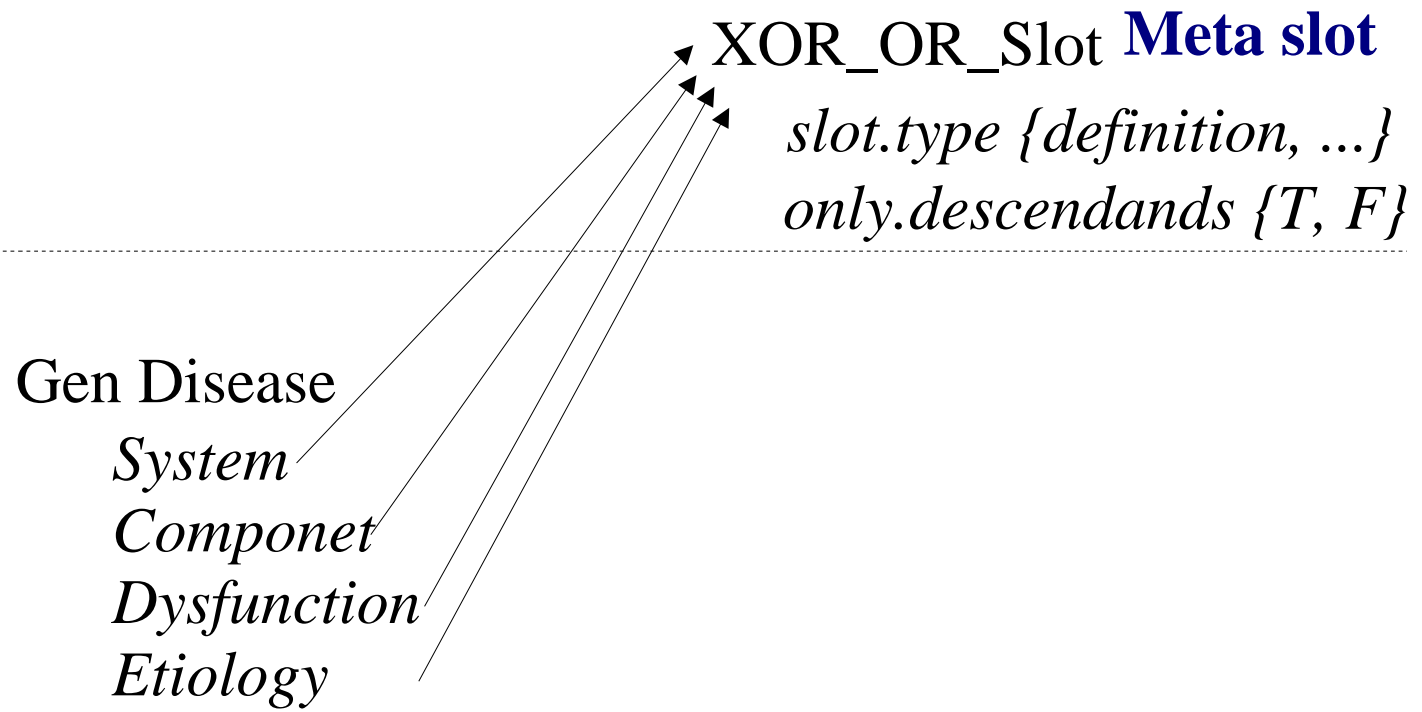
Nomenclature/Defs in DICE

Disease	Anatomical Component	Abnormality	Aetiology
Hepatitis	Def: Liver	Def: Infection	OR: (HepV, E-BV, CytomegaloV)

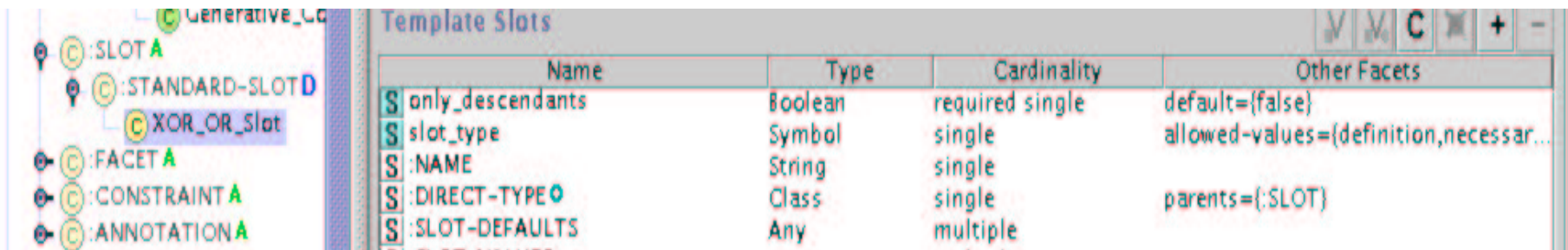
Meta Slots



Nomenclature



Meta slot: XOR_OR_Slot



Name	Type	Cardinality	Other Facets
S only_descendants	Boolean	required single	default={false}
S slot_type	Symbol	single	allowed-values={definition,necessar...
S :NAME	String	single	
S :DIRECT-TYPE	Class	single	parents=(:SLOT)
S :SLOT-DEFAULTS	Any	multiple	

Slot instances used to specify legal combinations

The screenshot displays a software interface with a class hierarchy on the left and a 'Template Slots' table on the right.

Class Hierarchy:

- Anatomic_Component A
 - Gen_Disease A
 - Gen_Viral_Hep
 - Gen_Hep_B
 - Aetiology A
 - Cytomegalo_Virus
 - Epstein-Barr_Virus

Template Slots Table:

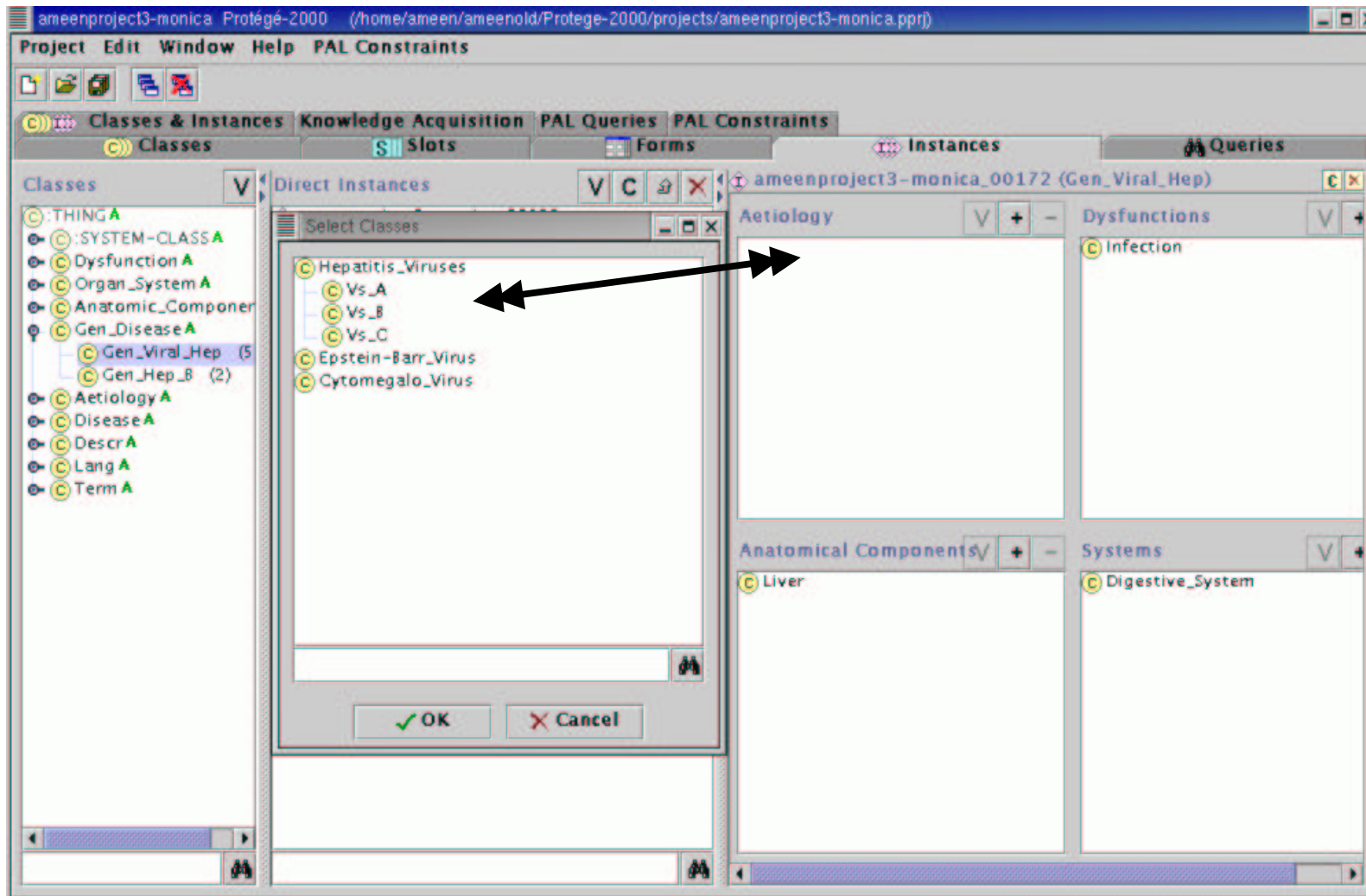
Name	Type	Cardinality	Other Facets
dysfunctions	Class	required multiple	parents={Infection}
anatomical_components	Class	required multiple	parents={Liver}
systems	Class	required multiple	parents={Digestive_System}
aetiology	Class	multiple	parents={Hepatitis_Viruses, Epstein-E...}

Specification of etiology within combination rule of Viral Hepatitis

The screenshot shows a software window titled "aetiology (XOR_OR_Slot)". The interface is divided into several sections:

- Name:** aetiology
- Value Type:** Class
- Documentation:** ** definition: don't refine
** necessary: must combine, accord to card and Desc.
** extrinsic: may combine, accord to card. and desc.
- Allowed Parents:** Hepatitis_Viruses, Epstein-Barr_Virus, Cytomegalo_Virus
- Cardinality:** required at least []
 multiple at most []
- Slot Type:** necessary (highlighted with a red circle)

Refinement at instantiation of Gen Disease



Knowledge model in Protege

- Concept oriented
- Explicit Rels
- **Composition Rules**
- Domain Completeness
- Multiple classification
- Use of definitions
- **Synonyms**
- **Multilingual**
- **Non redundant**
- **Unique codes**
- **Non ambiguous**
- Non vague
- Context-free codes

Conclusions

- **Knowledge model** of Protege forms a good basis for terminology systems.
- Meta model of Concept–Language–Term can be naturally specified by **Meta Classes** and enforced by **PAL constraints**.
- Nomenclature can be specified by **Meta Slots** but is not enforced *within* Protege.