

**Observational Health Data  
Sciences and Informatics (OHDSI)**

**Natural Language Processing (NLP)  
Working Group**

**11/14/2018**

# AGENDA

- ❖ Introduction of new members
- ❖ Updates on ongoing projects
  - ❖ Rules for defining term\_exists – led by Stephane Meystre
  - ❖ Mapping of CUIs to standard terminology – Juan Banda
  - ❖ Mapping of Note Types to LOINC/standard vocabulary – Karthik Natarajan, Ruth Reeves, Jon Duke and Hua Xu– Report type list discussion
  - ❖ Landscape Analysis of section identifier systems and proposal of a standard terminology for use – Hua Xu, Karthik Natarajan
  - ❖ Examples and rules for term\_temporal – led by George Hripsack (Sunny)
  - ❖ Standardization of term\_modifiers and values – Hua Xu

# NOTE\_NLP & NOTE

Field	Required	Type
note_nlp_id	Yes	Big Integer
note_id	Yes	integer
section_concept_id	No	integer
snippet	No	varchar(250)
offset	No	varchar(50)
lexical_variant	Yes	varchar(250)
note_nlp_concept_id	No	integer
note_nlp_source_concept_id	no	integer
nlp_system	No	varchar(250)
nlp_date	Yes	date
nlp_date_time	No	datetime
term_exists	No	varchar(1)
term_temporal	No	varchar(50)
term_modifiers	No	varchar(2000)

Field	Required	Type
note_id	Yes	integer
person_id	Yes	integer
note_date	Yes	date
note_datetime	No	datetime
note_type_concept_id	Yes	integer
note_class_concept_id	Yes	integer
note_title	No	varchar(250)
note_text	Yes	RBDMS dependent text
encoding_concept_id	Yes	integer
language_concept_id	Yes	integer
provider_id	No	integer
note_source_value	No	varchar(50)
visit_occurrence_id	No	integer

# Normalization of the term\_modifier

- 1. define the modifier list;
- 2. define normalized value of each modifier;

# term\_modifiers

## General:

- Negated
- Subject
- Certainty
- Conditional
- Temporal
- discoveryTechnique??
- Confidence??

## Disease specific:

- Body location
- Severity
- Course

## Medication specific:

- Dose
- Form
- Frequency
- Duration
- Route
- Necessary
- Unit

## Test specific:

- Value
- Unit
- Reference Range

## Procedure specific

- Body location
- Device
- Duration
- Method

# General

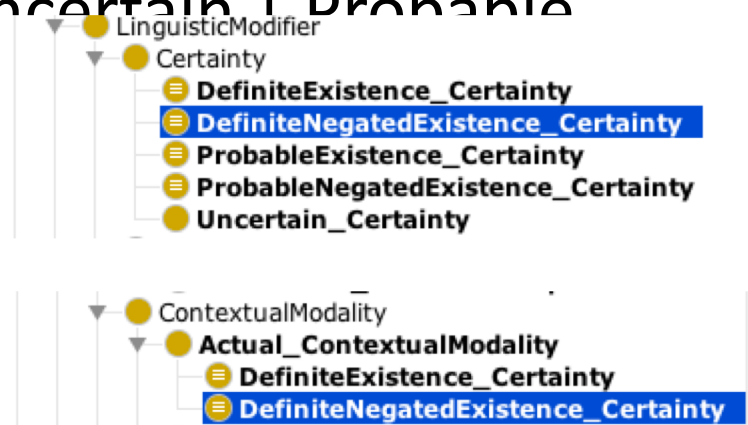
- **Negation**, Boolean; True means negated;
- **Certainty**, String Enum;

Example in text

— Definite | Probable | Uncertain | Probable

that may be a **UNC** UNC\_OF **problem** definite N

Reference



Question: merge negation into certainty?

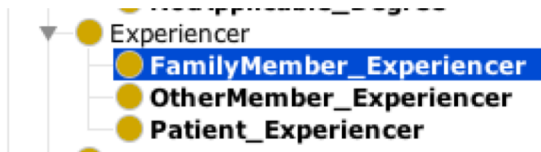
# General

- **Subject**, String Enum;

– Question: granularity?

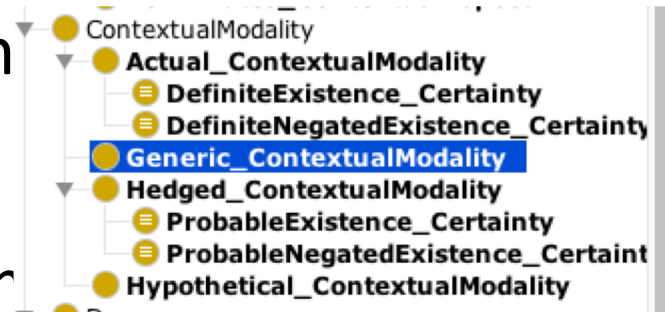
- *Family | Other | Patient*      **Or**      *father | mother | roommate | friend | patient...*

Reference



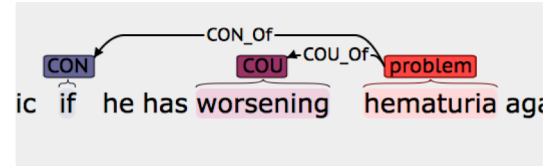
# General

- **Generic**, boolean;
  - True means generic mention
- **Conditional**, boolean;
  - True when has conditional m



The patient was referred to the Lupus clinic.

problem ← attrOf generic





# General

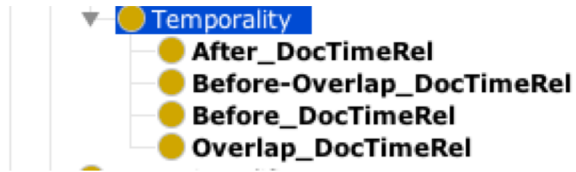
- Course

– unmarked | changed | increased | decreased |  
improved | worsened | resolved



# General

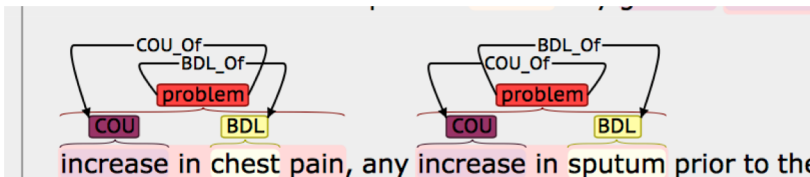
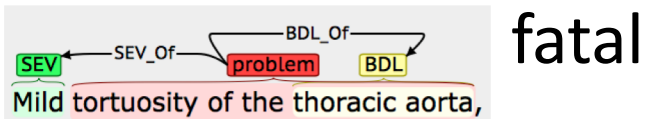
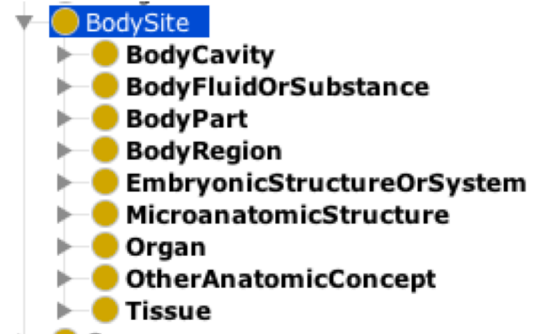
- **Temporal**, String;
  - Date, Time, Duration, Frequency;



- **Discovery technique**, String?
- **Confidence score**, double?

# Disease specific

- Body location
  - Map to UMLS CUI | standard
- Severity
  - unmarked | slight | moderate | severe | life



7 | ADMISSION PHYSICAL EXAMINATION:

# Medication specific

- Medication specific:

- Dose, double

- Frequency, String

- Duration, String

- Form, String enum

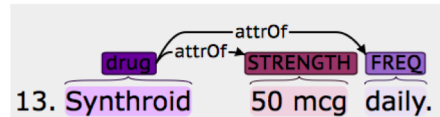
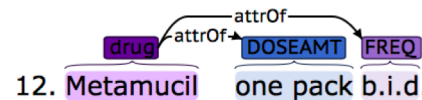
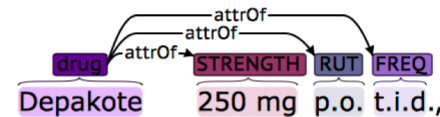
- Route, String enum

- Unit, Standard concept id?

- Necessary

- Start date?

- End date?



# Test specific

- Value, String;
  - Value could be int, double, percentage, range, or String

5 e , his CD4 count was less than 200 last winter .

The diagram shows a yellow box labeled 'test' with an arrow labeled 'valueOf' pointing to a blue box labeled 'labvalue'. Below this, the text 'e , his CD4 count was less than 200 last winter .' is shown with 'less than 200' underlined in blue, corresponding to the 'labvalue' box.

5 Urine HCG was negative, c  
replacement, increase bub

The diagram shows a yellow box labeled 'test' with an arrow labeled 'valueOf' pointing to a blue box labeled 'labvalue'. Below this, the text 'Urine HCG was negative, c' is shown with 'negative' underlined in blue, corresponding to the 'labvalue' box.

The patient was transfused from a Hct 29 to 34 on the day of discharge .

The diagram shows a yellow box labeled 'test' with an arrow labeled 'valueOf' pointing to a blue box labeled 'labvalue'. Below this, the text 'The patient was transfused from a Hct 29 to 34 on the day of discharge .' is shown with '29 to 34' underlined in blue, corresponding to the 'labvalue' box.

5 LDL 93 mg/dl 0-129

The diagram shows a yellow box labeled 'test' with an arrow labeled 'valueOf' pointing to a red box labeled 'labvalue'. The red box has an arrow labeled 'referenceOf' pointing to a blue box labeled 'reference'. Below this, the text '5 LDL 93 mg/dl 0-129' is shown with '93 mg/dl' underlined in red and '0-129' underlined in blue.

7 CHD Risk 3.02 L 3.90-5.80

The diagram shows a yellow box labeled 'test' with an arrow labeled 'valueOf' pointing to a red box labeled 'labvalue'. The red box has an arrow labeled 'referenceOf' pointing to a blue box labeled 'reference'. Below this, the text '7 CHD Risk 3.02 L 3.90-5.80' is shown with '3.02' underlined in red and 'L 3.90-5.80' underlined in blue.

- Unit, String enum;
- Reference Range, String;

# Procedure specific

- Body location

- Map to UMLS CUI | standard

- Device

- Map to UMLS CUI | standard vocabulary;

- Duration, String

- Method, String

