# MAPPING WITH DESCRIPTION LOGIC

An Approach to Data Integration

Geraldine Wade MD
Clinical Informatics Consulting
October 28, 2016

# Outline

- Definition
- Challenges
- Use Case
- Purpose
- Method
- Comparison with single concept mapping
- Data quality
- Examples
- Summary

# Definition

Mapping with Description Logic is…..

*….the use of source code sets or legacy terms that map to post-coordinated expressions using SNOMED CT description logic components*

*….to enable the integration of source terms into a single corpus that can be more easily managed and maintained.*

# Challenges

- Existing data (legacy, interface, complex terms, special data sets)

- Evolving requirements
  a. local (immediate inclusion of terms/outside of release cycle)
  b. national (data sharing, quality reporting etc.)

- Implementation issues
  a. speed of term inclusion
  b. need for harmonization with other code sets
  c. CDS, patient safety

- Terminology management (limited resources, data quality)

*SNOMED is becoming more important (either on front end or back end)

# Use Case

- Increasing importance of SNOMED CT adoption
- Need to migrate existing terms (legacy, interface, local) to align with SNOMED CT
- Need to be able to expand your local extension (immediate need for term addition)
- Maintaining “fit for purpose”
- Simplify implementation/ terminology management
- Focus on data quality

GOAL: To transform the source data elements into the SNOMED CT corpus without data loss

# Purpose (why use DL method)

- More complete data representations of source terms through the use of SNOMED CT DL components
- Using SNOMED CT DL components allow for alignment with existing SNOMED CT content
- Source concepts can be directly integrated into the SNOMED CT corpus
- Data integration into a single corpus
- Can use classifier for terminology management/ equivalence testing
- Improve data quality

# Method

- Map source terms using description logic and post-coordinated expressions
- Apply rules for consistency for groups of similar concepts (consider source domains for context)
- Select proximal parent, defining and qualifying relationships (i.e. post-coordinated map)
- Add to SNOMED CT corpus in Protégé
- Evaluate using data quality metrics…..correctness, completeness, consistency, timeliness

# DL Method

Advantages:

- Uses DL [formal knowledge representation]
- Conforms to OWL EL2 profile [for SNOMED CT/LOINC]
- New concepts are represented and modeled in a compatible data structure (alignment with SNOMED CT)
- Allows for the immediate inclusion of local terms
- Provides constraints, data quality checks/classifier (terminology management)
- Improves data quality from source legacy data

…more complete … more correct ….more consistent
more concepts are FULLY DEFINED

# Comparison with single concept mapping

- Single concept mapping ….map to best match  to a single SNOMED CT concept
- Human subjectivity (BT, NT, =)(closest match), different mapping rules …unless exact match, there is no semantic equivalency…the degree of “correctness” varies
- How do you apply CDS to BT or NT maps? What will be missed.
- Context of source term often not available
- Maps usually maintained separately from SNOMED CT
- Risk of data loss, overall diminished data quality

# Data quality

Correctness, Consistency, Completeness and Timeliness

*"An ontology that provides a framework of constraints and relationships between concepts"*

- applicable to structured EHR data
- reusable across domains and use cases

Correctness: measures whether data is likely correct between matching datasets
Consistency: accuracy of values and representation of output data
Completeness: are there missing components?
Timeliness: time of the observation, version

*Johnson S. A Data Quality Ontology for the Secondary Use of EHR Data. , AMIA 2015 Annual Symposium*

# Examples

**Map to single target**

| A | B | C | D | F |
|---|---|---|---|---|
| Local_Code | Local_Description | Target_Code | Target_Description | Usage_Coun |
| 234436 | BLADDER NEOPLASM MALIGNANT MULLERIAN MIXED TUMOR | 93689003 | Primary malignant neoplasm of bladder (disorde | |
| 234453 | BLADDER NEOPLASM MALIGNANT MESODERMAL MIXED TUMOR | 399326009 | Malignant tumor of urinary bladder (disorder) | |
| 234690 | BLADDER NEOPLASM MALIGNANT MESENCHYMOMA | 93689003 | Primary malignant neoplasm of bladder (disorde | |
| 234645 | BLADDER NEOPLASM MALIGNANT PARAGANGLIOMA | 93689003 | Primary malignant neoplasm of bladder (disorde | |

Malignant mesodermal mixed tumor of the bladder →
Malignant tumor of urinary bladder (disorder)

Single target constraint → data loss

## Map to post-coordinated expression (DL approach)

| A | B | C | D | F |
|---|---|---|---|---|
| Local_Code | Local_Description | Target_Code | Target_Description | Usage_Coun |
| 234436 | BLADDER NEOPLASM MALIGNANT MULLERIAN MIXED TUMOR | 93689003 | Primary malignant neoplasm of bladder (disorde | |
| 234453 | BLADDER NEOPLASM MALIGNANT MESODERMAL MIXED TUMOR | 399326009 | Malignant tumor of urinary bladder (disorder) | |
| 234690 | BLADDER NEOPLASM MALIGNANT MESENCHYMOMA | 93689003 | Primary malignant neoplasm of bladder (disorde | |
| 234645 | BLADDER NEOPLASM MALIGNANT PARAGANGLIOMA | 93689003 | Primary malignant neoplasm of bladder (disorde | |

Post coordinated expression → no data loss

399326009 | malignant tumor of urinary bladder | : 116676008 | associated morphology | = 112684005 | mesodermal mixed tumor

A service of the U.S. National Library of Medicine | National Institutes of Health

# UMLS Terminology Services

## Metathesaurus Browser

Basic View | Report View | Raw View

**Concept: [C2212529] malignant mesodermal mixed tumor of bladder**

**Semantic Types**

Neoplastic Process [T191]

**Atoms (2)** string [AUI / RSAB / TTY / Code]

malignant mesodermal mixed tumor of bladder [A13849442/MEDCIN/PT/234453]

Attributes (1) Name | Value | RSAB

Relations (4) REL | RELA | RSAB [SType1 - SType2] STypeId | String | CUI

CHD | isa | MEDCIN [SCUI - SCUI] 39016 | malignant neoplasm of bladder | C0005684
RN | isa | MEDCIN [SCUI - SCUI] 399326009 | Malignant tumor of urinary bladder | C0005684
RO | has_finding_site | MEDCIN [SCUI - SCUI] 89837001 | Urinary bladder structure | C0005682
RO | has_associated_morphology | MEDCIN [SCUI - SCUI] 112684005 | Mesodermal mixed tumor | C1334603

malignant mesodermal mixed tumor of bladder (diagnosis) [A13877088/MEDCIN/FN/234453]

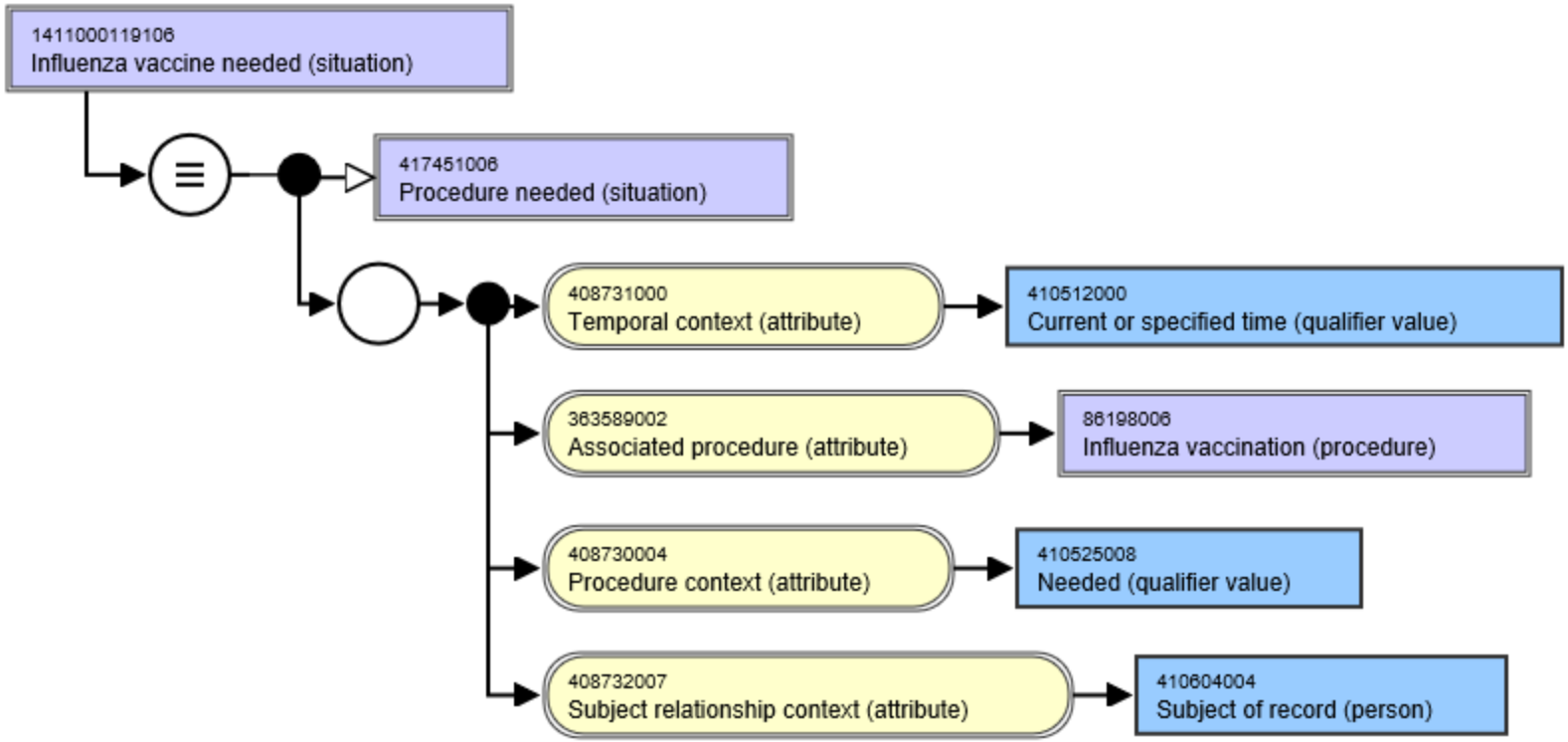
## Single vs post-coordinated map

Source: 132976| Need for rabies vaccination (MEDCIN term)

→ requires rabies vaccination course (finding)

Closest single map → inconsistency/domain change

# Look for similar concept



A service of the U.S. National Library of Medicine | National Institutes of Health

# UMLS Terminology Services

## Metathesaurus Browser

Basic View | Report View | Raw View

Finding [T033]

**Atoms (13)** string [AUI / RSAB / TTY / Code]

need for rabies vaccination [A13748031/MEDCIN/PT/132976]

Attributes (1) Name | Value | RSAB

Relations (3) REL | RELA | RSAB [SType1 - SType2] STypeId | String | CUI

CHD | isa | MEDCIN [SCUI - SCUI] 197086 | need for vaccination | C2193787
RN | isa | MEDCIN [SCUI - SCUI] 417451006 | Procedure needed | C1562512
RO | has_associated_procedure | MEDCIN [SCUI - SCUI] 34631000 | Rabies vaccination | C0042205

need for rabies vaccination (treatment) [A18093007/MEDCIN/FN/132976]

need for vaccination rabies [A13996816/MEDCIN/SY/132976]

Requires rabies vaccination course (finding) [A15087253/MTH/PN/NOCODE]

OS Requires rabies vacc course [A1266818/RCD/OA/65W6.]

OS Requires rabies vaccination course [A1266819/RCD/OP/65W6.]

requiere una dosis de vacuna antirrábica [A6152861/SCTSPA/PT/170543008]

requiere una dosis de vacuna antirrábica (hallazgo) [A6152860/SCTSPA/FN/170543008]

Requires rabies vaccination course [A3269801/SNOMEDCT_US/PT/170543008]

OS Requires rabies vaccination course [A4874126/SNOMEDCT_US/OF/147797002]

## Single vs post-coordinated map

417451006 | procedure needed | : { 363589002 | associated procedure | = 34631000 | rabies vaccination | }

Can put as rows in table or post-coord expression

Post coordinated expression → consistency/no domain change/aligns with current SNOMED CT content

# Data Integration

How to we align these expressions within SNOMED CT?

399326009 | malignant tumor of urinary bladder | : 116676008 | associated morphology | = 112684005 | mesodermal mixed tumor

417451006 | procedure needed | : { 363589002 | associated procedure | = 34631000 | rabies vaccination | }

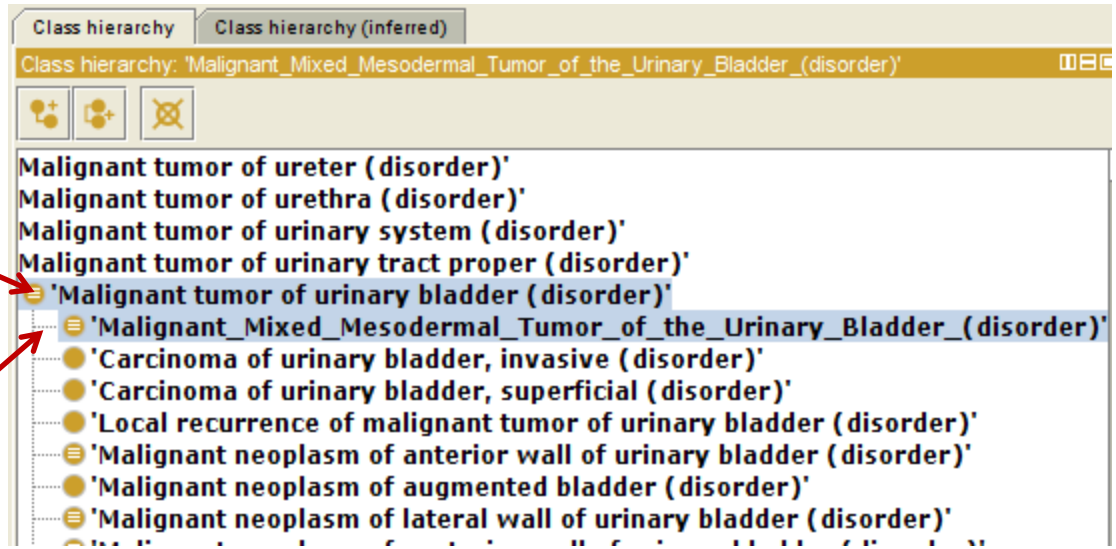
*Interface terminology can be integrated with the SNOMED CT corpus

# Add concept to SNOMED CT

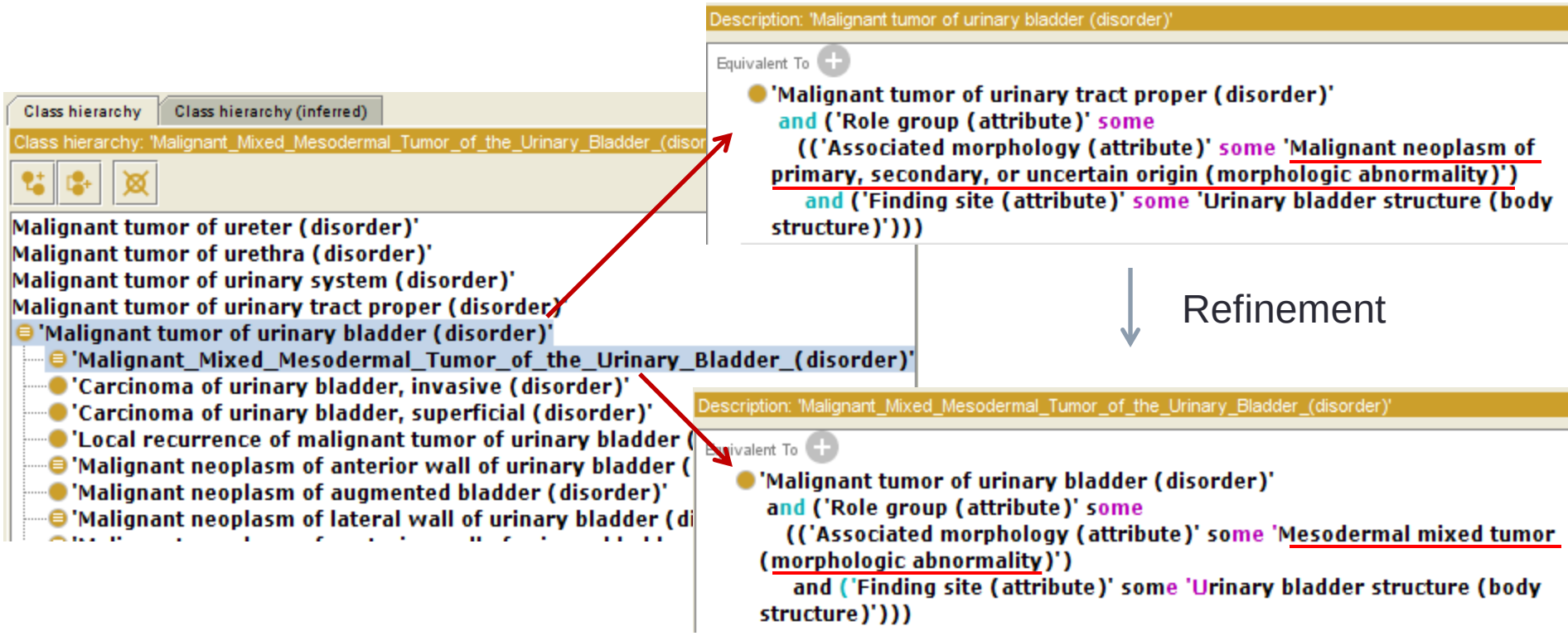
399326009 | malignant tumor of urinary bladder | : 116676008 | associated morphology | = 112684005 | mesodermal mixed tumor

SNOMED CT concept

MEDCIN term

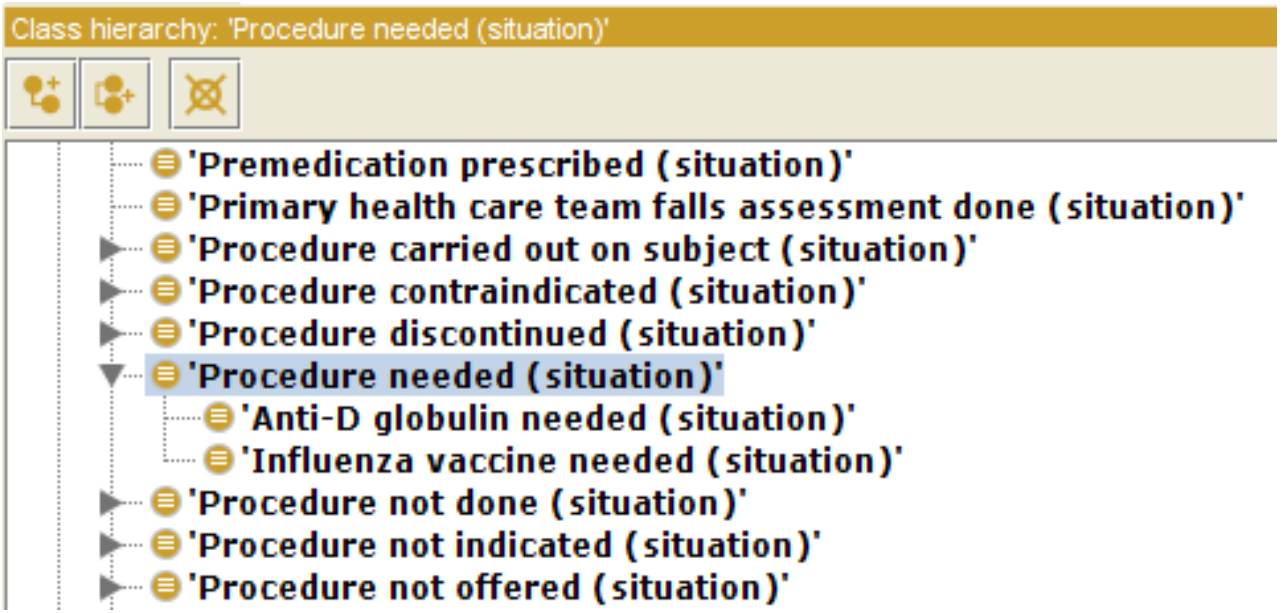


# Add relationships/classify

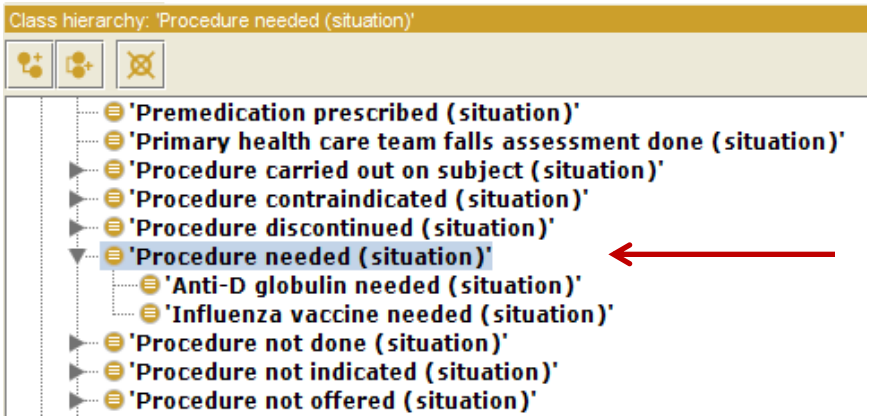


# Add concept to SNOMED CT

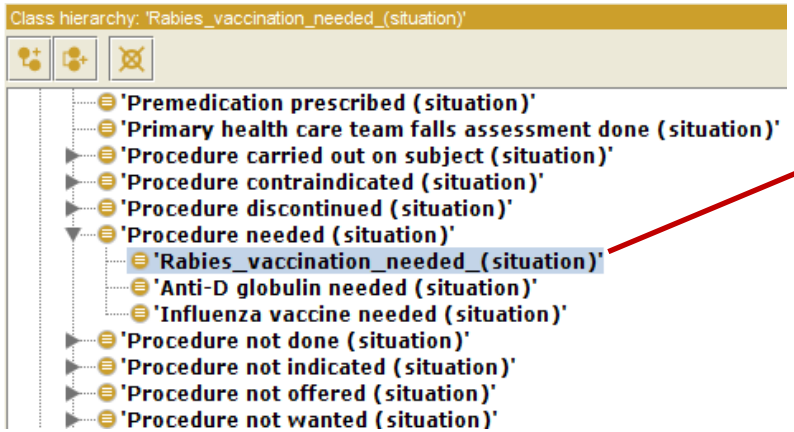
417451006 | procedure needed | : { 363589002 | associated procedure | = 34631000 | rabies vaccination | }

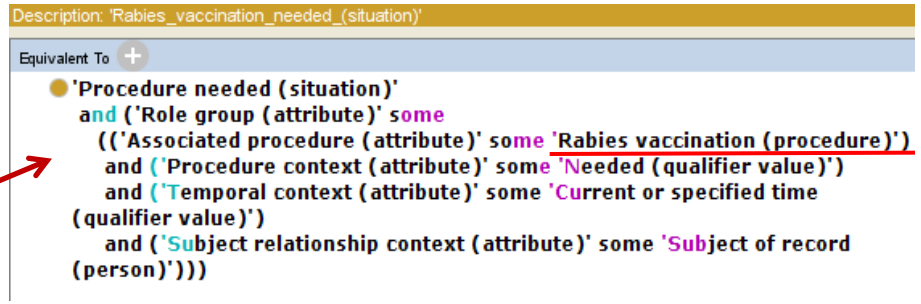


# Add procedure as Situation/classify




Parent concept



# More examples

[MEDCIN term] 141060 | routine gynecological exam visit
308335008 | patient encounter procedure | :
363702006 | has focus | = 83607001 | gynecologic examination|
, 260870009 | priority | = 50811001 | routine |

**need mapping rules to apply for consistency across all terms..like all visits…etc.*

[MEDCIN term] 275089 | MALIGNANT PROSTATE NEOPLASM STAGE III
399068003 | malignant tumor of prostate | :
47429007 | associated with | = 369837008 | t3: Prostate tumor extends through the prostatic capsule |

No data loss from the original source term.

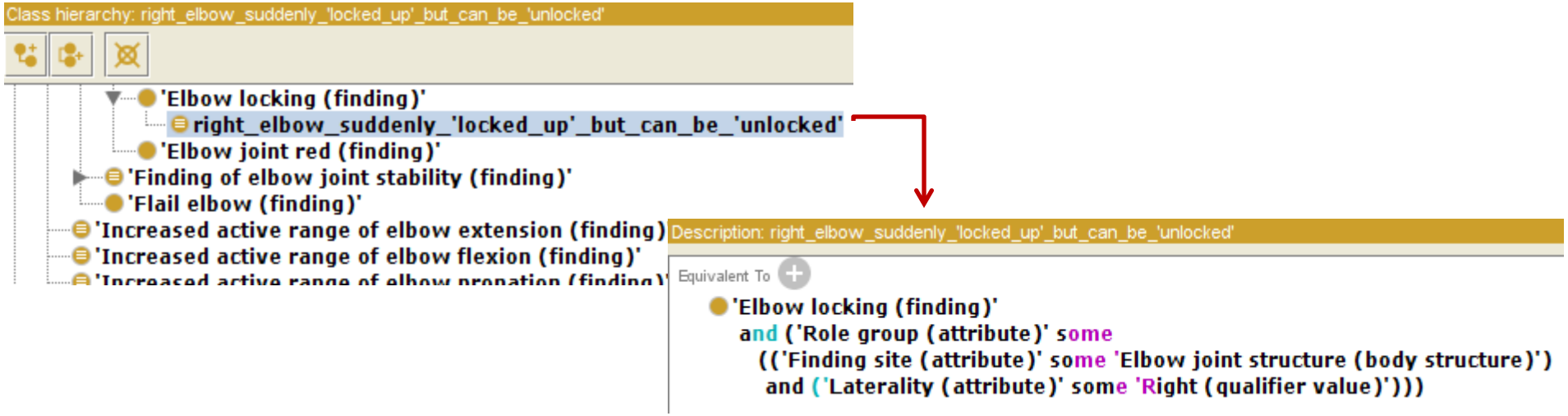
# Integrating complex terms

right elbow suddenly 'locked up' but can be 'unlocked'
[MEDCIN ID 112452]

difficult to model interface terms

# Interface terms may be complex

MEDCIN term: “right elbow suddenly ‘locked up’ but can be unlocked”




Can add as a primitive concept

# Summary

- Mapping with DL enables direct data integration into the SNOMED CT using existing DL components
- Improves data quality in terms of completeness, correctness and consistency….more FULLY DEFINED terms
- Maintenance of interface concepts in one location
- Allows for the immediate entry of local terms
- Has “built in” quality checks through use of classifier
- Terminology management
- Patient safety is the most important outcome – focus on implementation….CDS components…limited number of concepts and nodes