*open*EHR Terminology binding

Thomas Beale

openEHR

Problem overview

What is ‘terminology binding’?

A *formally expressible* connection between information model representation and terminology representation of clinical statements recorded in the EHR

What is the ‘binding problem’?

We need to know how to control the use of terminology within structured data so that it achieves what we want:

- Provides basis for querying
- Economically feasible

First, we need to know how to structure data so it:

- Doesn’t violate ontological truths;
- Is mappable to ontological concepts;
- Supports data entry, storage, querying, reuse

Which ‘structured’ data?

Two kinds:

- Legacy proprietary: structures are all different
- Shared, standardised: agreed structures and information model, within a community of users (can be more than one such community).

The second kind we can standardise on.

Shared clinical data generally include structure and many data types.

Data are structured

Clinical statements are naturally structured, e.g.

- *lab results*: list / tree structure; normal ranges;
  - Microbiology is usually a large tree structure
- *vital signs*: timing and multiple data points;
  - BP: (2 data points + patient state) x time-series
- *physical examination*: structured by anatomy
  - E.g. Endoscopy of colon
- *assessments*: structured according to e.g. temporal model of disease course;
- *orders*: timing info, structured medication info;
- *actions*: timing, medication structured info

Other sources of structure

Data capture: at the user interface, the elements of a clinical statement are naturally distinct, e.g. procedure, site, protocol, time...

Document structures: reports, referrals etc are also structured, including audit info, sections.

For querying: data items that are queried for separately are usually separated, e.g. procedure type and body site.

Data have many types

Clinical statement data includes instances of:

- Text
- Coded terms
- Quantity, including units, proportions, counts, etc
- URIs
- Booleans
- Date, time, date/time, duration
- Parseable text, e.g. Units, medication timing
- Other more complex types

What *should* be SNOMED-coded?

- Answers which are:
  - textually expressible
  - whose value range is
    - Best modelled by as ontological description (i.e. discrete categorisation),
  - likely to be independently queried later on.
  - E.g. types of disease; blood types; but not general patient story (not expressible as just concepts)

I.e. a subset of textual data, which are a subset of all data

What *could* be SNOMED-coded?

- Questions which:
  - Need to be queried on using an *agreed reference coding standard*.

Example: ‘serum sodium’ (in context of blood film result of patient) does not need any coding to be 100% reliably queryable in *open*EHR environment. However, for the data to be re-usable by ANYONE later on, SNOMED-coding makes sense.

Understanding the binding problem

One thing complicates the task...SITUATION

Examples:

- list of body positions is not the same as list of body positions pertinent to measuring BP;
- valid Rh blood types differs depending on whether for blood collection or transfusion;
- almost all scales, e.g. Apgar, GCS, Borg, Barthel etc define their own value sets for common phenomena, which differ from contextless value sets of the same / similar phenomena in naming and number of divisions.

Value sets in scales

Figure 1. Modified Borg scale.

SCALE	SEVERITY
0	No Breathlessness* At All
0.5	Very Very Slight (Just Noticeable)
1	Very Slight
2	Slight Breathlessness
3	Moderate
4	Some What Severe
5	Severe Breathlessness
6	
7	Very Severe Breathlessness
8	
9	
10	

FEEDING
0 = unable
5 = needs help cutting, spreading butter, etc., or requ
10 = independent

BATHING
0 = dependent
5 = independent (or in shower)

GROOMING
0 = needs to help with personal care
5 = independent face/hair/teeth/shaving (implements

DRESSING
0 = dependent
5 = needs help but can do about half unaided
10 = independent (including buttons, zips, laces, etc.)

BOWELS
0 = incontinent (or needs to be given enemas)
5 = occasional accident
10 = continent

BLADDER
0 = incontinent, or catheterized and unable to manage alone
5 = occasional accident
10 = continent

TOILET USE
0 = dependent

Score of 0	Score of 1	Score of 2	Component of Acronym
Skin colour			
blue all over	white at extremities body pink	pink all over	Appearance
Heart rate			
absent	slow	fast	Pulse
Reflex response			
no response to stimulation	grimacing when stimulated	crying and coughing	Grimace
Muscle tone			
limp	some bending or stretching of limbs	active movement	Activity
Breathing			
absent	weak or irregular	good and your baby is crying	Respiration

openEHR

Binding and *open*EHR

Where is binding relevant in *open*EHR?

*open*EHR Archetypes - essentially, maximum data sets, i.e. all data points for a given domain 'recording' concept (not its ontological 'description').

- Examples:
  - Vitals signs: BP, Heart-rate etc
  - Labs – very structured, well understood
  - Physical exam – e.g. Pain, symptom....numerous!
  - Scales, e.g. GCS, Apgar, Barthel – ordinal data
- Terminology need: globally invariant mappings; broad value sets e.g. 'infectious agent'

Where is binding needed?

*open*EHR Templates - essentially, use-case specific content specifications; consist of data points from archetypes

- Examples:
  - Discharge summary
  - Lab report
  - Encounter note
- Terminology need: define local / region-specific or specialty-specific value sets and constraints, e.g. ‘lung infection’
- NOT JUST TO SNOMED CT!

Kinds of binding - today

- Compositional expressions already used
- Direct binding to concept points
- Archetype local value sets → direct binding – value set specific to archetype
- Ref set binding for data points that correspond to reusable value sets
- Templates can have direct binding to SCT terms, with static value set defined in archetype or ref set reference

Kinds of binding - future

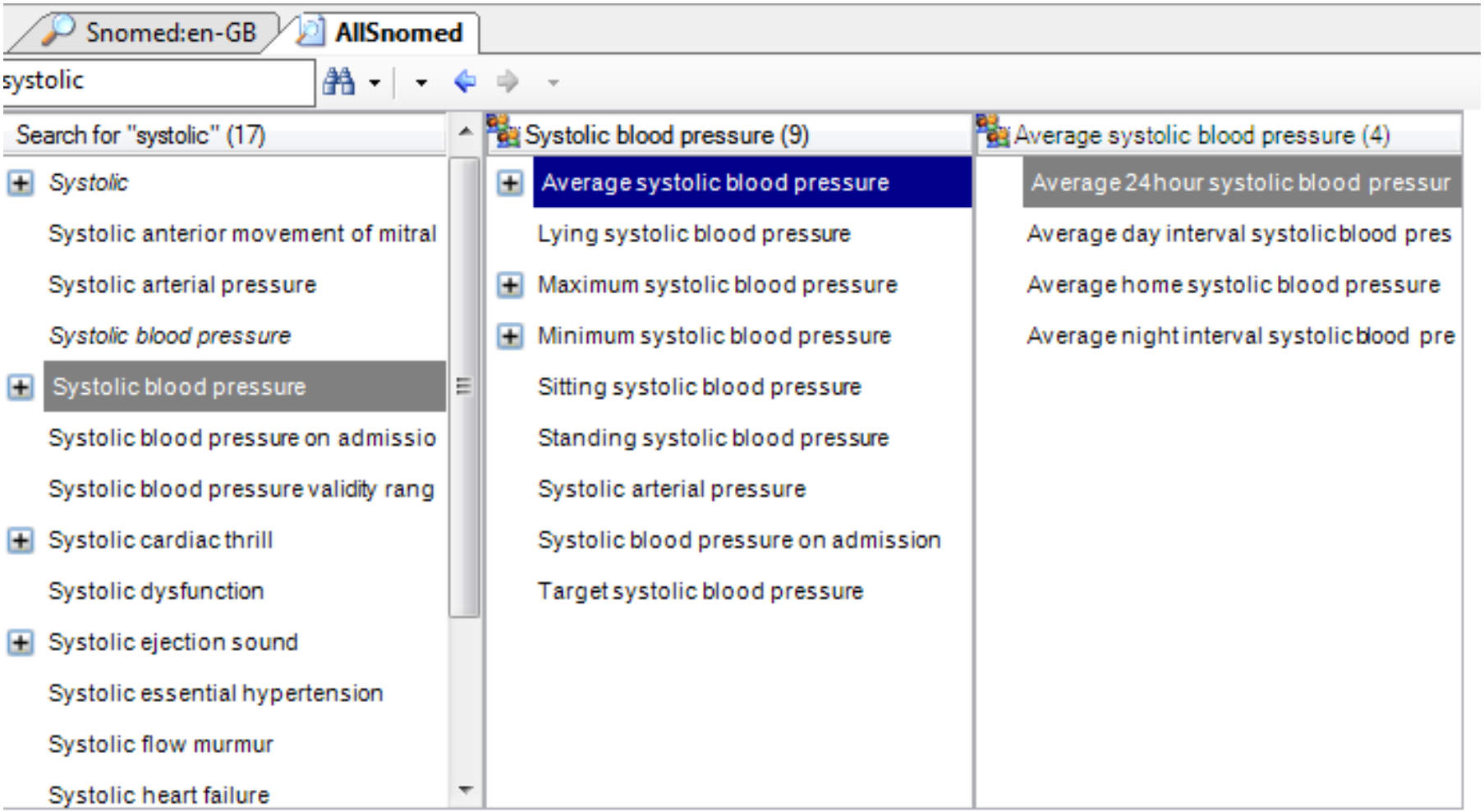
- Context-dependent bindings
- SCT Compositional constraints
- SCT Composition pattern mapping?

Type 1 binding – direct

Direct binding

- WHEN: we want to associate a terminology concept with a data item that we want to be able to query
- Ex: systolic BP
- Generally an archetype path → code binding
- Each path acts like a post-coordination
  - E.g. 24 hour average systolic pressure

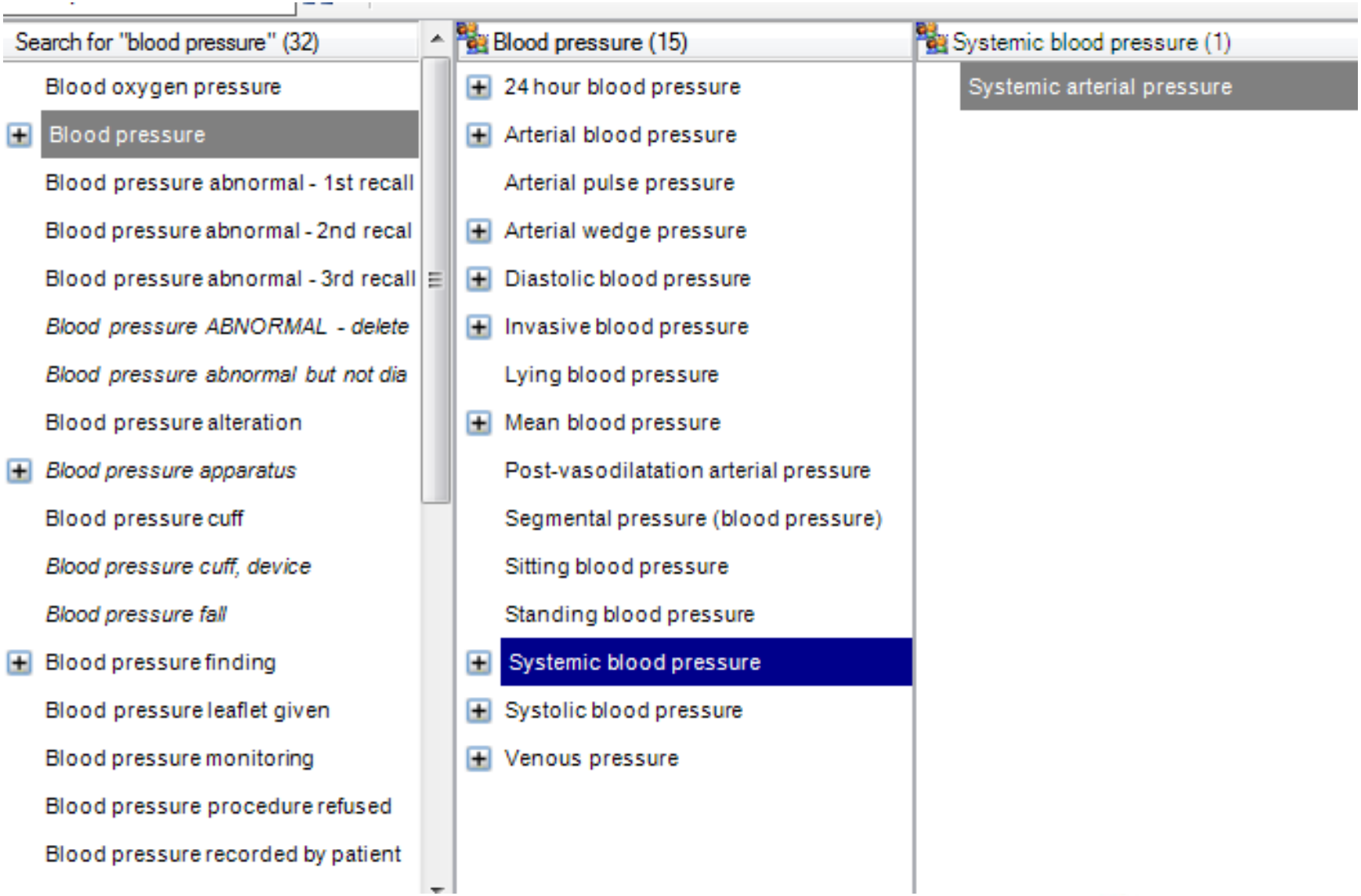
Which SCT concept do we pick?



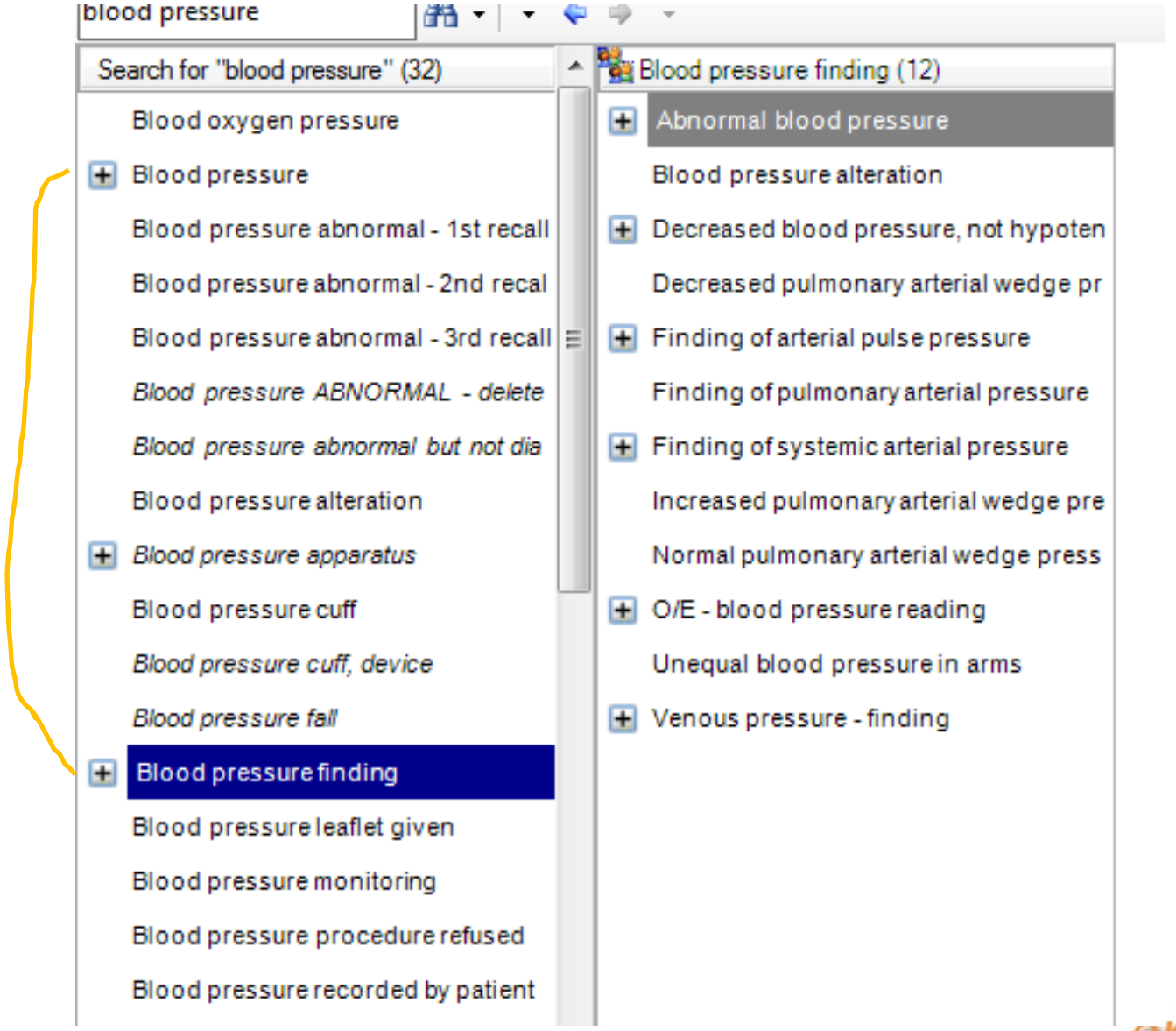
If we bind |systolic blood pressure| (usually means instantaneous), SNOMED-driven queries would pick up 24h av, max, min etc

openEHR

can ‘systolic’ be post-coordinated?



|Bp| v |bp finding|



openEHR

/data/events[any event]/data/items[Systolic]
/data/events[any event]/data/items[Systolic]/value
/data/events[any event]/data/items[Diastolic]
/data/events[any event]/data/items[Diastolic]/value
/data/events[any event]/data/items[Comment]
/data/events[any event]/data/items[Comment]/value
/data/events[any event]/data/items[Mean Arterial Pressure]
/data/events[any event]/data/items[Mean Arterial Pressure]/value
/data/events[any event]/data/items[Pulse Pressure]
/data/events[any event]/data/items[Pulse Pressure]/value
/data/events[any event]/state
/data/events[any event]/state/items[Position]
/data/events[any event]/state/items[Position]/value
/data/events[any event]/state/items[Position]/value/defining_code
/data/events[any event]/state/items[Tilt]
/data/events[any event]/state/items[Tilt]/value
/data/events[any event]/state/items[Exertion]
/data/events[any event]/state/items[Sleep status]
/data/events[any event]/state/items[Sleep status]/value
/data/events[any event]/state/items[Sleep status]/value/defining_code
/data/events[any event]/state/items[Confounding factors]
/data/events[any event]/state/items[Confounding factors]/value
/data/events[24 hour average]
/data/events[24 hour average]/data
/data/events[24 hour average]/data/items[Systolic]
/data/events[24 hour average]/data/items[Systolic]/value
/data/events[24 hour average]/data/items[Diastolic]
/data/events[24 hour average]/data/items[Diastolic]/value

271649006 |systolic blood pressure|
OR
72313002|systolic arterial pressure|
OR
399304008|Systolic blood pressure on admission|
OR....

314449000| Average 24hour systolic blood pressure|

Archetype paths

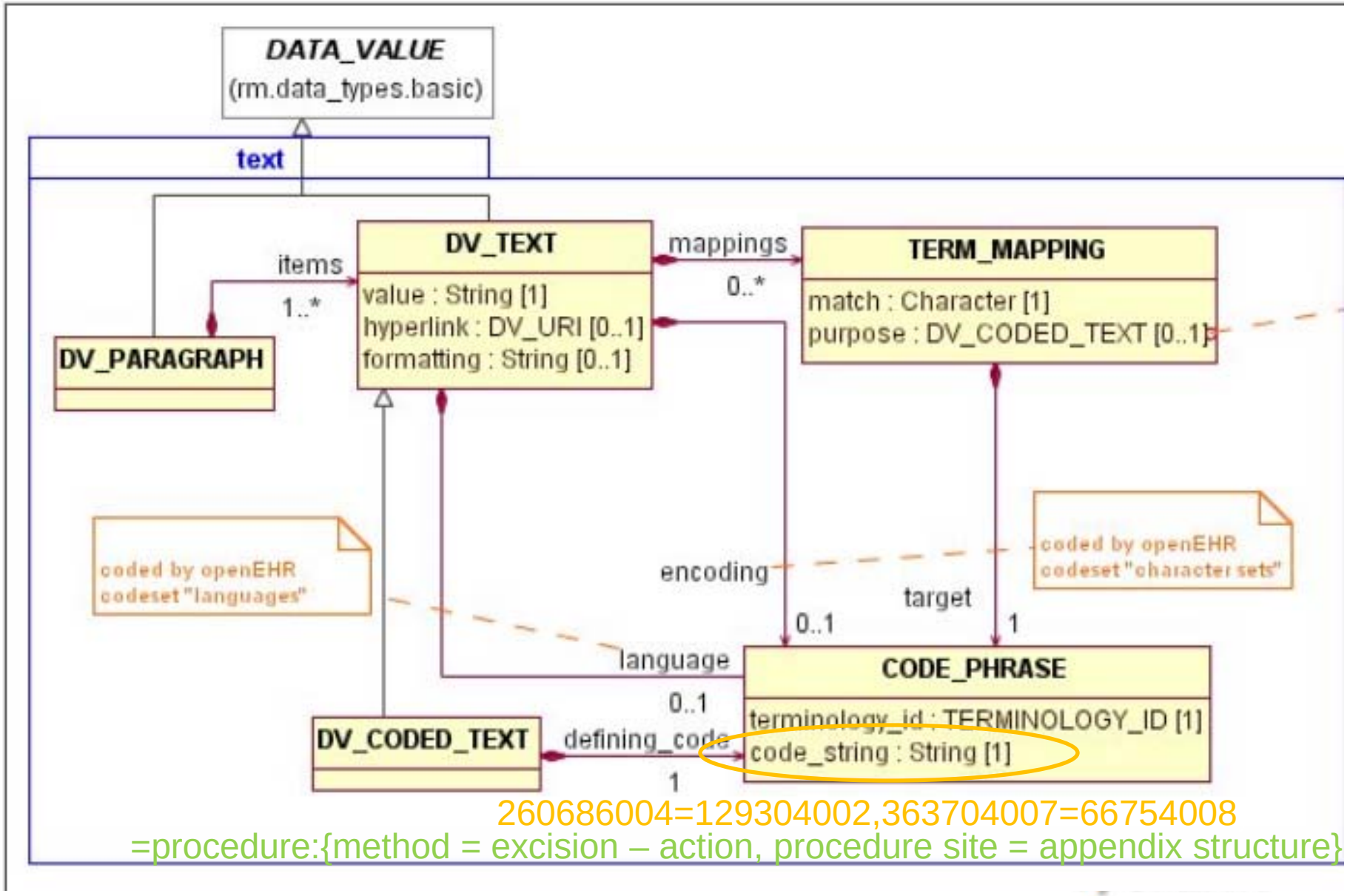
openEHR

Considerations

- Many parts of SNOMED, excessive precoordination makes it difficult to know what to choose
- Basic problem: whatever binding modeller chooses, query author might choose a different concept, and the results may not be correct.

Type 2 binding – Compositional expressions

*open*EHR supports expressions



Considerations

What if information system uses pre-coordinated term? A different post-coordination? Will querying work?

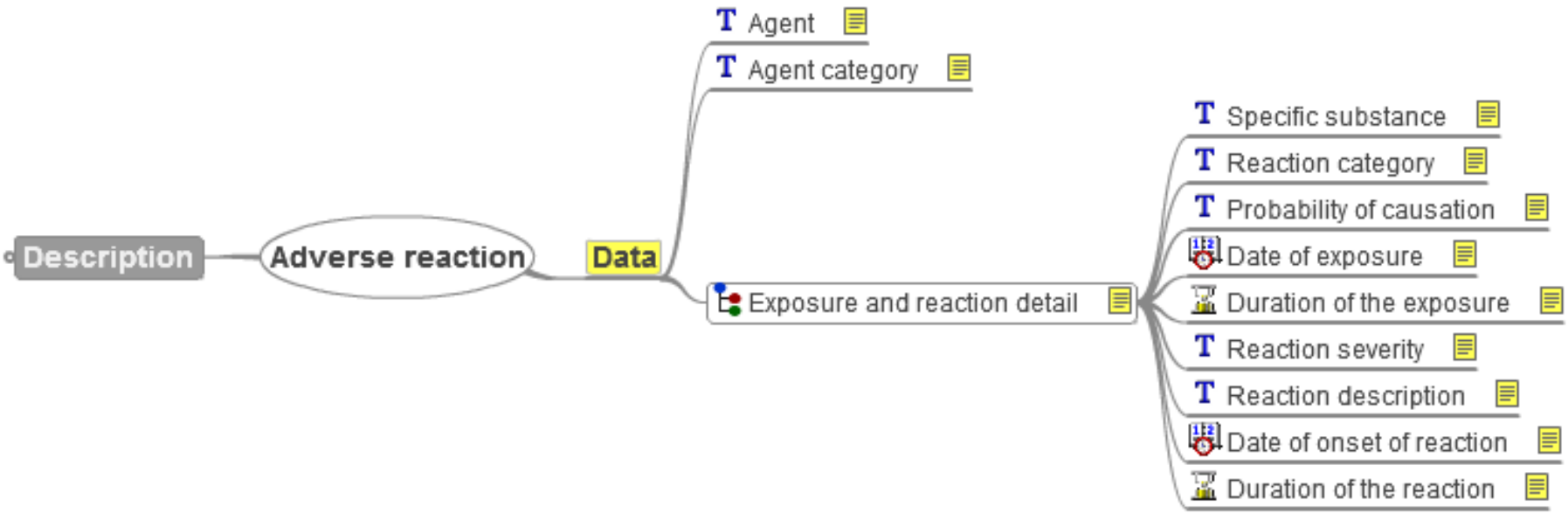
Relies heavily on normal form & equivalence working correctly.... and being economic to implement!

Type 3 binding – archetype internal value set

Internal value set

- WHEN: *situationally* dependent values,
  - e.g. Position of patient *for blood pressure measurement*
  - E.g. Set of breathing values *for Apgar*
- WHEN: poor/no matches available in SCT
- OR: good matches available, but no refset/subset available or desired, e.g. local use only
- Currently VERY COMMON in archetypes, including for scales

Adverse reaction (mindmap view)



openEHR

Attribute view

Archetype: Adverse reaction (openEHR-EHR-EVALUATION.adverse.v1)

Header | **Data**

unordered)

Specific substance Text Occurrences: 0..1 (optional)	The specific substance that caused the reaction if different from the agent (e.g. brands or members of a class).	Free or coded text
Reaction category Coded Text Occurrences: 0..1 (optional)	The type of reaction experience by the person as determined by the clinician.	• Intolerance [Leads to unpleasant symptoms which are sufficient to avoid use in the future.] • Sensitivity [Leads to morbidity which is potentially threatening to the wellbeing of the person.] • Allergy [Leads to an IgE mediated reaction.] • No reaction [Person has been exposed with no reaction.]
Probability of causation Coded Text Occurrences: 0..1 (optional)	Degree of certainty that the agent was the cause of the reaction.	• Certain/Highly likely [A reaction to the agent is deemed to be or have been present by the clinician.] • Probable [The reaction is deemed to be the probable cause of the reaction.] • Possible [The agent is deemed to be a possible cause of the reaction.]
Date of exposure Date/Time Occurrences: 0..1 (optional)	The date (+/- time) when the person became exposed to the agent.	
Duration of the exposure Duration Occurrences: 0..1 (optional)	The duration of the exposure to the agent.	
Reaction severity	The category of the reaction.	• Mild [A reaction which causes little

Previous Next

ADL view

```
EVALUATION[at0000] matches {  -- Adverse reaction
    data matches {
        ITEM_TREE[at0002] matches {    -- structure
            items cardinality matches {1..*; unordered} matches {
                ELEMENT[at0003] matches {      -- Agent
                    value matches {
                        DV_TEXT matches {*}
                    }
                }
                ELEMENT[at0010] occurrences matches {0..1} matches {      -- Agent category
                    value matches {
                        DV_CODED_TEXT matches {
                            defining_code matches {
                                [local::
                                at0011,     -- Food
                                at0012,     -- Animal
                                at0013,     -- Medication
                                at0014,     -- Other chemical or substance
                                at0031,     -- Non-active ingredient of medication
                                at0033,     -- Imaging dye or media
                                at0034]     -- Environmental
                            }
                        }
                    }
                }
                CLUSTER[at0019] occurrences matches {0..*} matches {      -- Exposure and reaction detail
                    items cardinality matches {1..*; unordered} matches {
                        ELEMENT[at0032] occurrences matches {0..1} matches {      -- Specific substance
```

```
}
ELEMENT[at0015] occurrences matches {0..1} matches {      -- Reaction category
        value matches {
                DV_CODED_TEXT matches {
                        defining_code matches {
                                [local::
                                at0016,     -- Intolerance
                                at0017,     -- Sensitivity
                                at0018,     -- Allergy
                                at0030]     -- No reaction
                        }
                }
        }
}
ELEMENT[at0004] occurrences matches {0..1} matches {      -- Probability of causation
        value matches {
                DV_CODED_TEXT matches {
                        defining_code matches {
                                [local::
                                at0005,     -- Certain/Highly likely
                                at0006,     -- Probable
                                at0007]     -- Possible
                        }
                }
        }
}
ELEMENT[at0020] occurrences matches {0..1} matches {      -- Date of exposure
        value matches {
                DV_DATE_TIME matches {*}
        }
```

```
        }
    }
    ELEMENT[at0021] occurrences matches {0..1} matches {      -- Duration of the exposure
        value matches {
            DV_DURATION matches {*}
        }
    }
    ELEMENT[at0023] occurrences matches {0..1} matches {      -- Reaction severity
        value matches {
            DV_CODED_TEXT matches {
                defining_code matches {
                    [local::
                    at0024,     -- Mild
                    at0025,     -- Disabling
                    at0026]     -- Life threatening
                }
            }
        }
    }
    ELEMENT[at0022] matches {       -- Reaction description
        value matches {
            DV_TEXT matches {*}
        }
    }
    ELEMENT[at0027] occurrences matches {0..1} matches {      -- Date of onset of reaction
        value matches {
            DV_DATE_TIME matches {*}
        }
    }
    ELEMENT[at0028] occurrences matches {0..1} matches {      -- Duration of the reaction
```

Agent category binding...

Archetype term	SCT candidates
at0011|food|	406465008|food allergen|, 255620007|Foods| Note 149 top-level concepts containing 'food'
at0012|animal|	39866004|animal| Note 241 top-level concepts containing 'animal'; no 'animal allergen'
at0013|medication|	119 top-level terms containing 'medication' (heavily pre-coordinated), but no |medication|...!
at0014|Other chemical or substance|	33565001|chemical agent| ??? 167 top-level concepts containing 'chemical'
at0031|Non-active ingredient of medication|	Nothing with 'ingredient'
at0032|imaging dye or media|	Nothing suitable
at0033|environmental|	Some approximate matches...

Considerations

- Should we bind to SNOMED at all?
  - Codes could be useful, since we might want to find adverse reactions caused by ‘environment’ or ‘food’
- How to bind, or model?
  - Currently, the archetype defines the value set
  - Could bind each internal code to an SCT code
    - Difficulties finding candidate concepts
  - Could we use a ref set instead?
    - This archetype has 4 internal coded value sets...what happens with 2000 archetypes?

Type 4 binding – ref set

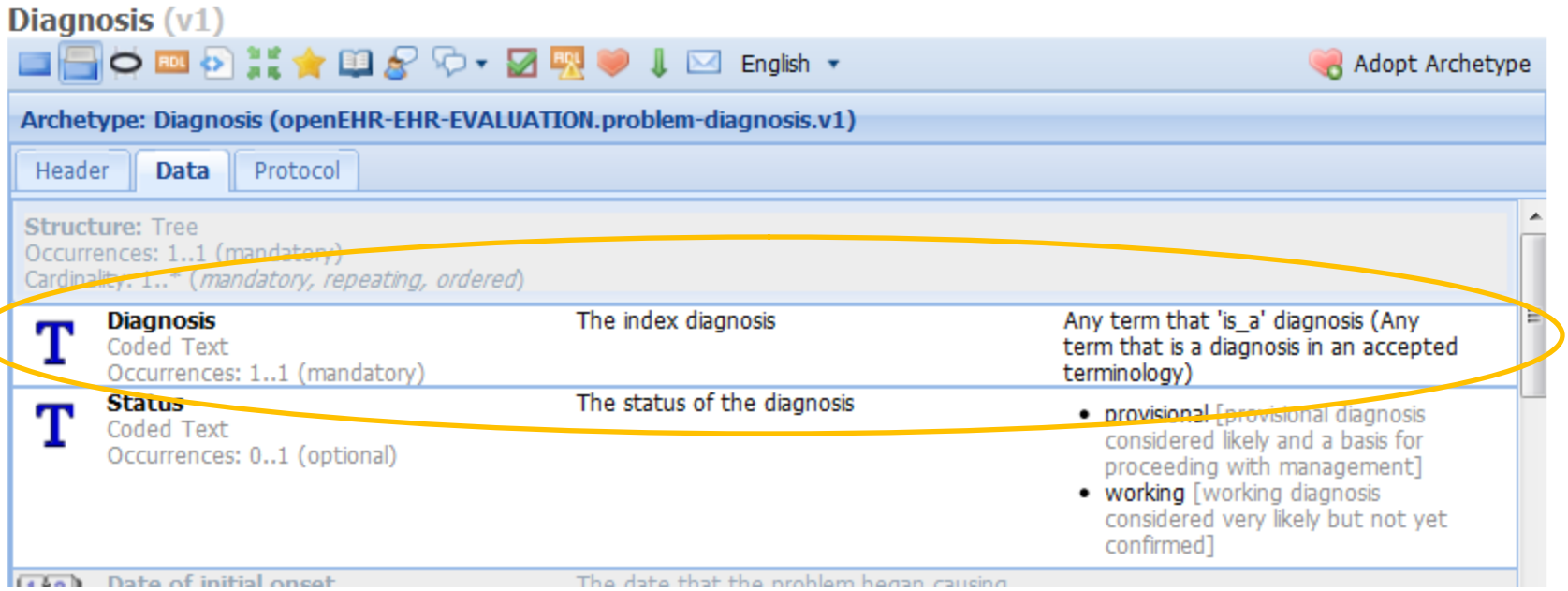
Ref set binding

This is for data points that correspond to context-independent domain concepts, e.g.

- Pain character
- Infectious agent

The archetype or template can include an ac-code that binds to an external resource, such as a ref-set id/URI.

Problem/diagnosis



openEHR

Diagnosis (v1)

English

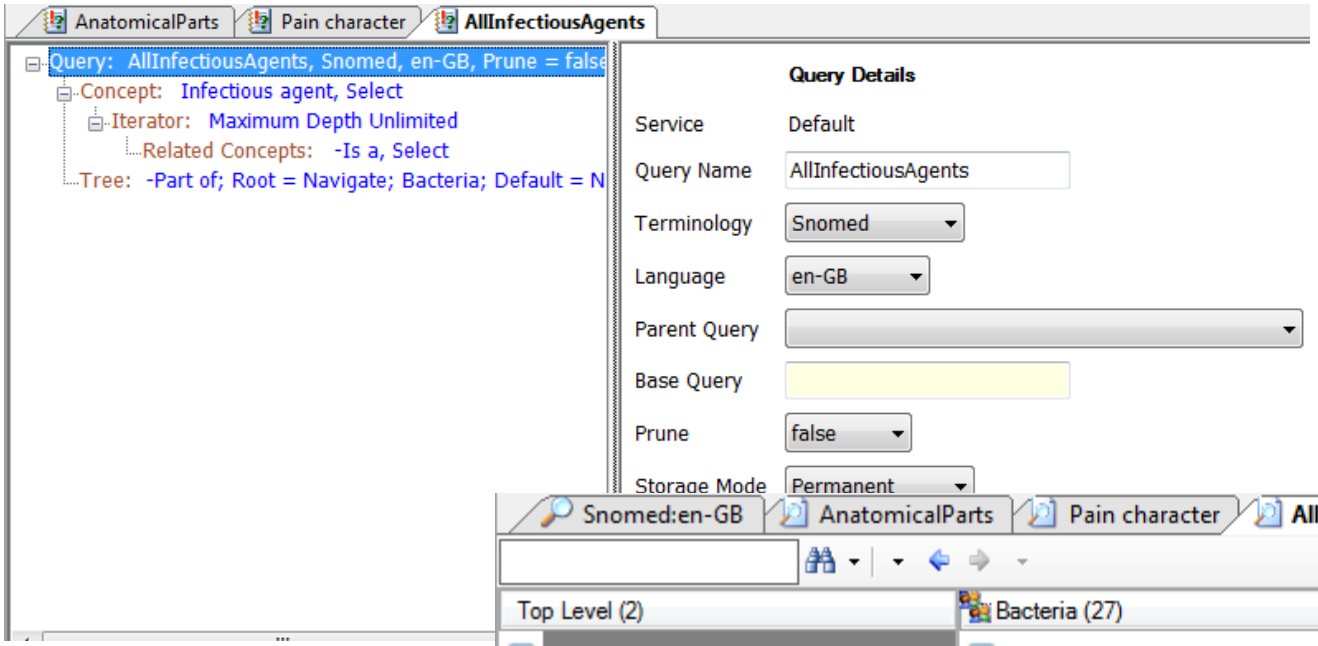
```
data matches {
    ITEM_TREE[at0001] matches {    -- structure
        items cardinality matches {1..*; ordered} matches {
            ELEMENT[at0002.1] matches {    -- Diagnosis
                value matches {
                    DV_CODED_TEXT matches {
                        defining_code matches {[ac0.1]}        -- Any term that 'is_a' diagnosis
                    }
```

```
["en"] = <
    items = <
        ["ac0.1"] = <
            text = <"Any term that 'is_a' diagnosis">
            description = <"Any term that is a diagnosis in an accepted terminology">
        >
        ["ac0000"] = <
            text = <"Any term that describes a body site">
            description = <"An anatomical structure with qualifiers">
        >
    >
```

```
constraint_bindings = <
    ["SNOMEDCT"] = <
        items = <
            ["ac0.1"] = <http://www.terminology.org?refsetid=20293847593>
        >
    >
>
```

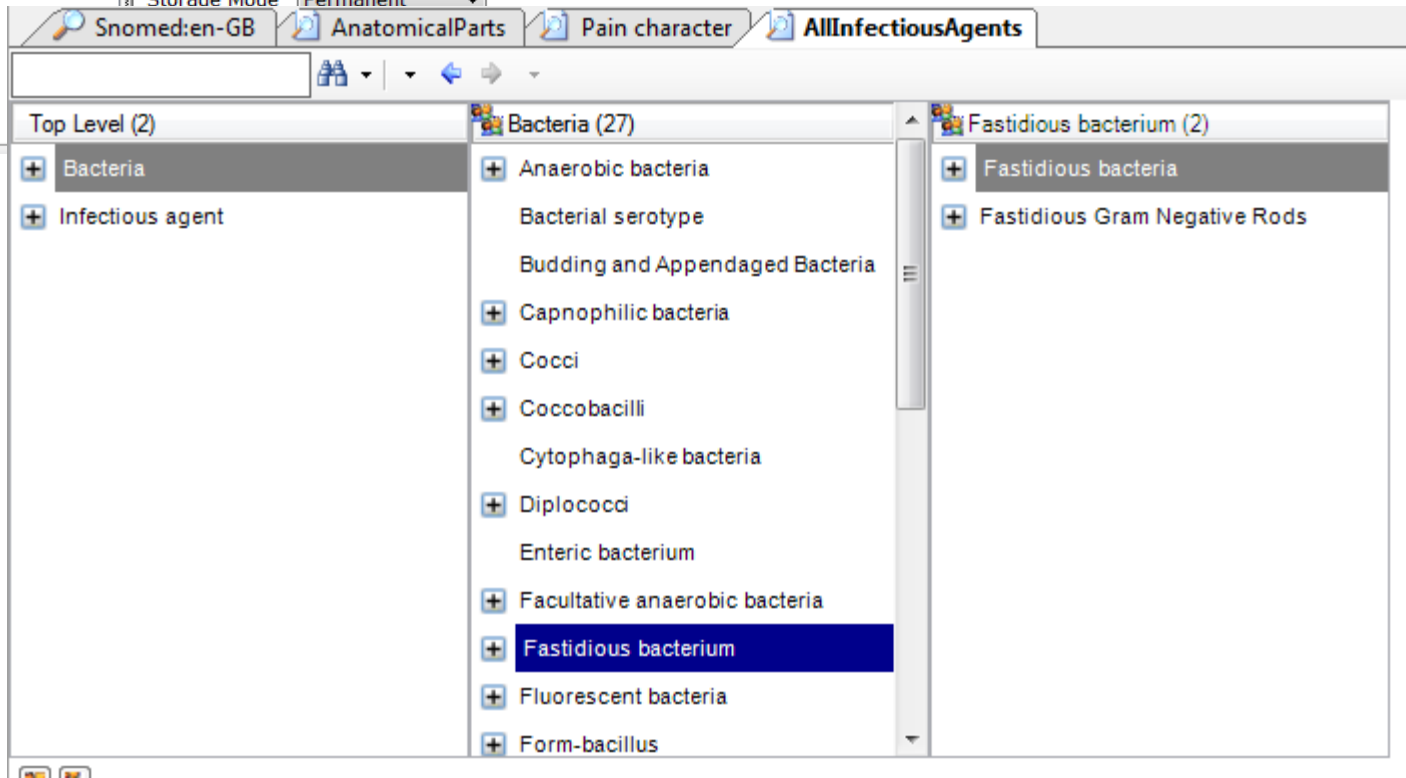

openEHR

All Infectious Agents ref set



←Definition

Result →



Future approaches

1 Compositional constraints

- WHEN: we already agree on using single post-coordinated code phrase
- E.g. Want to force information capture of site to include laterality, where it is defined.
- Can express a SNOMED constraint for this that forces laterality to exist.
- This capability does not yet exist in *open*EHR, but is very easy to add into the C_CODE_PHRASE constraint class
- Requires a solid definition of SNOMED constraint grammar

2 Context-dependent bindings

- WHEN: terminology codes incorporate contextual value, e.g. patient sex, pathology challenge, time, etc
  - E.g. Bindings to LOINC codes can depend on ‘protocol’, i.e. LOINC ‘challenge’
  - Some SNOMED concepts are specific to patient gender or other attributes

E.g. Lying systolic bp

The following not legal ADL today, but it could be....

/data/events[at0006|any event|]/data/items[Systolic]
WHEN /data/events[at0006|any event|]
/state /items[at0008|Position|] = at1003|Lying|
➔ 407556006|Lying systolic blood pressure|

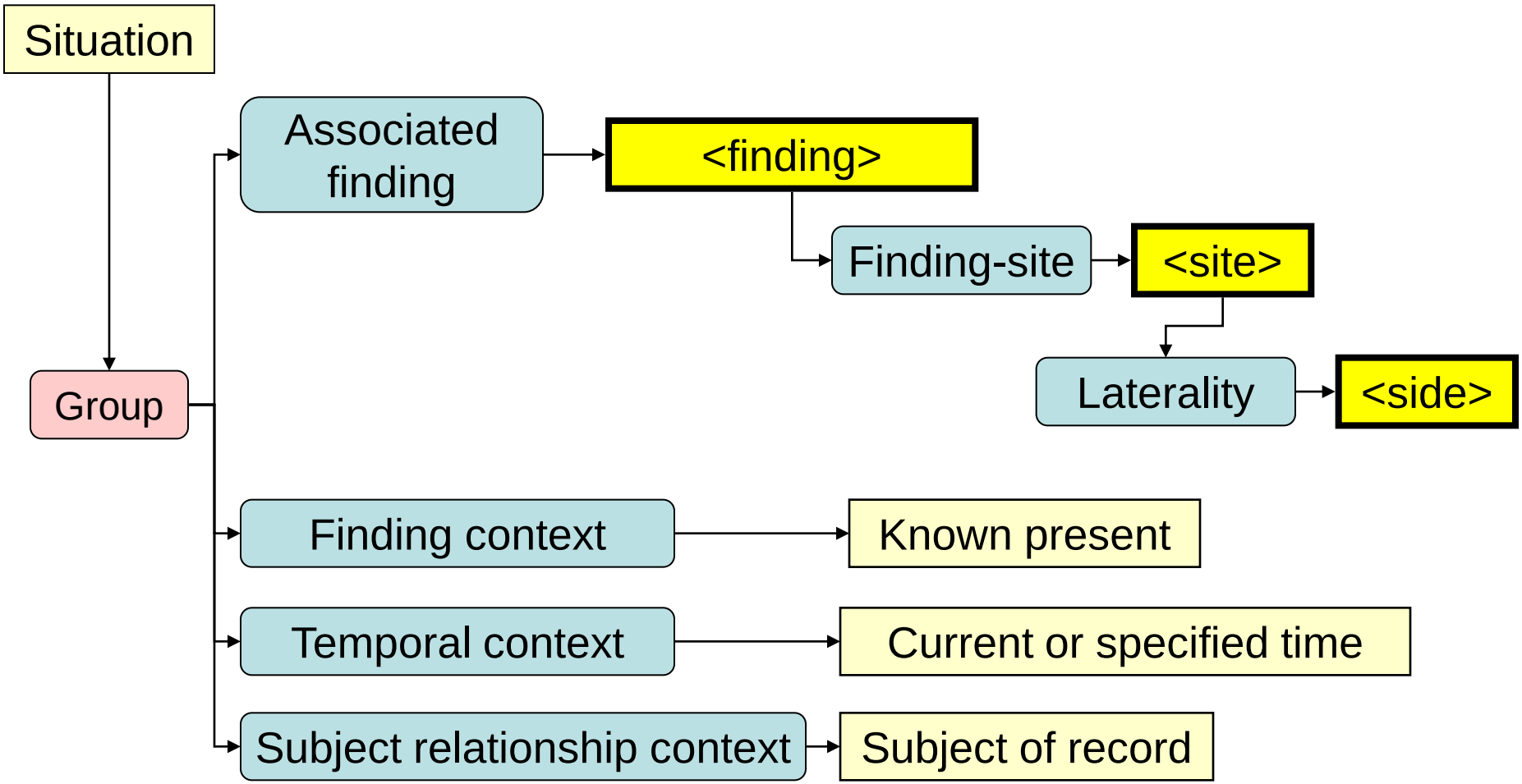
2 Context-dependent bindings

- Would need small syntax addition in ADL to connect a condition (FOPL expression on archetype data) to a concept in terminology.
- Considerations for SNOMED:
  - Excessive precoordination makes concept selection difficult; query author might select another concept

3 Compositional pattern approach

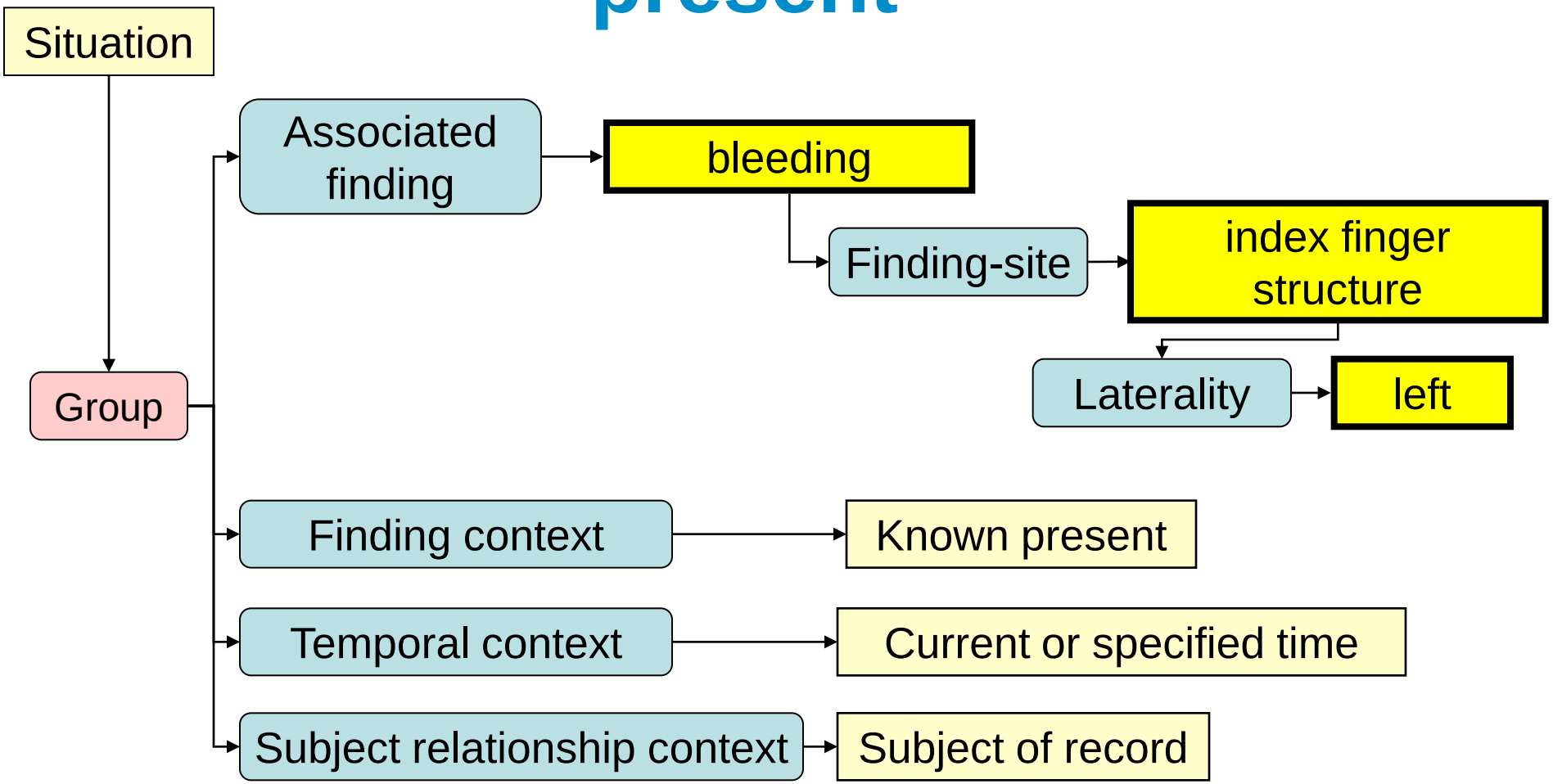
- WHEN: there are multiple attributes in IM (some may be post-coordinated), that we want to code together rather than separately
- The emergence of patterns for Compositions of complex clinical statements may be useful in solving the binding problem
- Beginning looks promising
- Questions:
  - how will this work evolve?
  - HOW MUCH COMPOSITION IS ENOUGH?

Clinical finding present + site + side



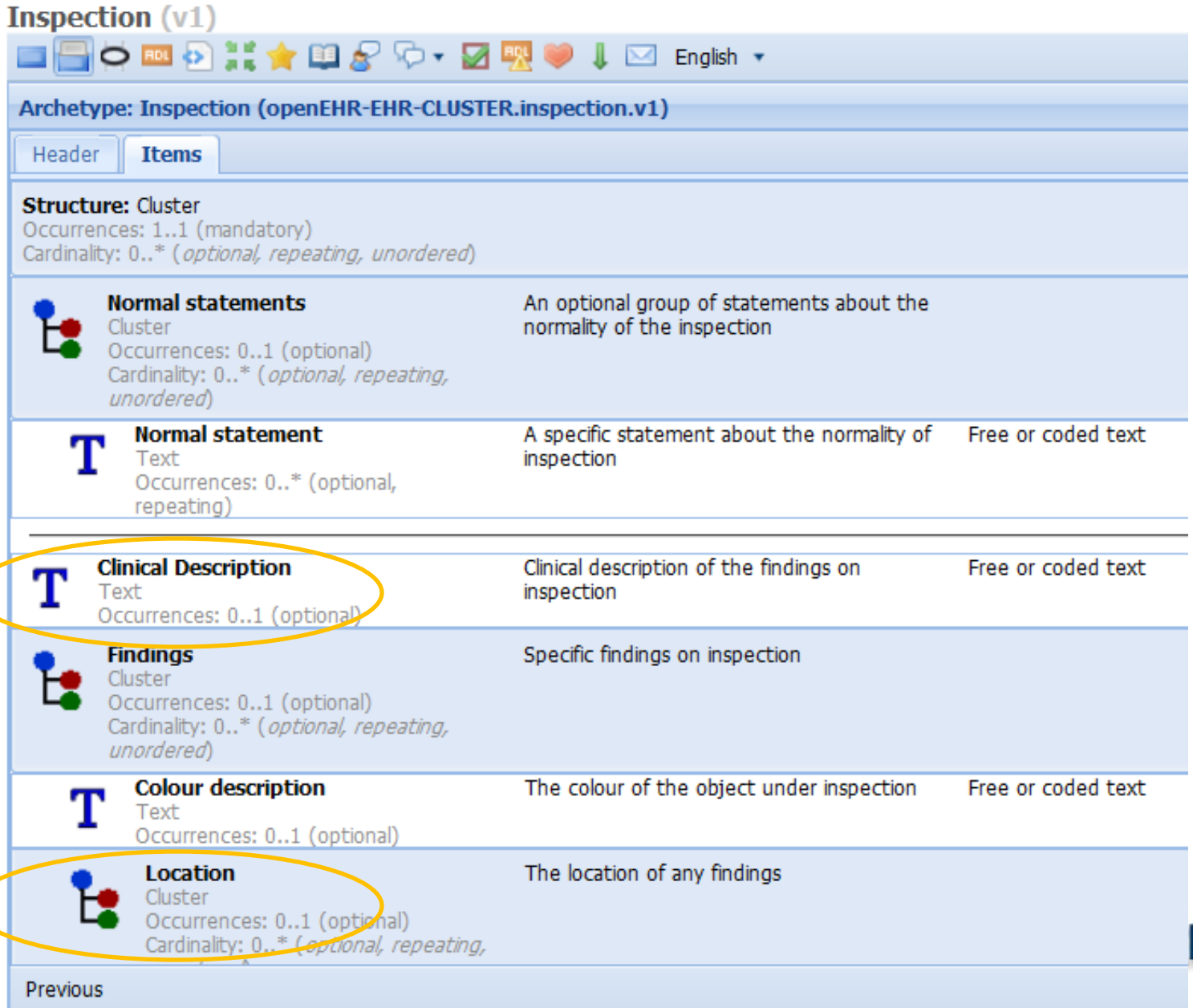
Clinical-finding-present-with-site-and-side (<finding>,<site>,<side>)

Bleeding of left index finger present



Clinical-finding-present-with-site-and-side (bleeding, index finger structure, left)

Inspection archetype



Inspection (v1)

English

Archetype: Inspection (openEHR-EHR-CLUSTER.inspection.v1)

Header | **Items**

	Shape or distribution Cluster Occurrences: 0..1 (optional) Cardinality: 0..* (*optional, repeating, unordered*)	The shape or distribution of the object inspected	
T	**Description** Text Occurrences: 0..1 (optional)	Description of the	Free or coded text
	Symmetrical Boolean Occurrences: 0..1 (optional)	Whether the distribution is symmetrical	
T	**Contour** Text Occurrences: 0..1 (optional)	The contour of the object of inspection	Free or coded text
T	**Overlying skin or surface** Text Occurrences: 0..1 (optional)	A description of the surface	Free or coded text
T	**Surrounds** Text Occurrences: 0..1 (optional)	Description of immediate surrounds to object	Free or coded text
	Edge Cluster Occurrences: 0..1 (optional) Cardinality: 0..* (*optional, repeating, unordered*)	Findings about the edge of the object under inspection	
T	**Description** Text	Description of the edge or border of the object of inspection	Free or coded text

Previous

Inspection (v1)

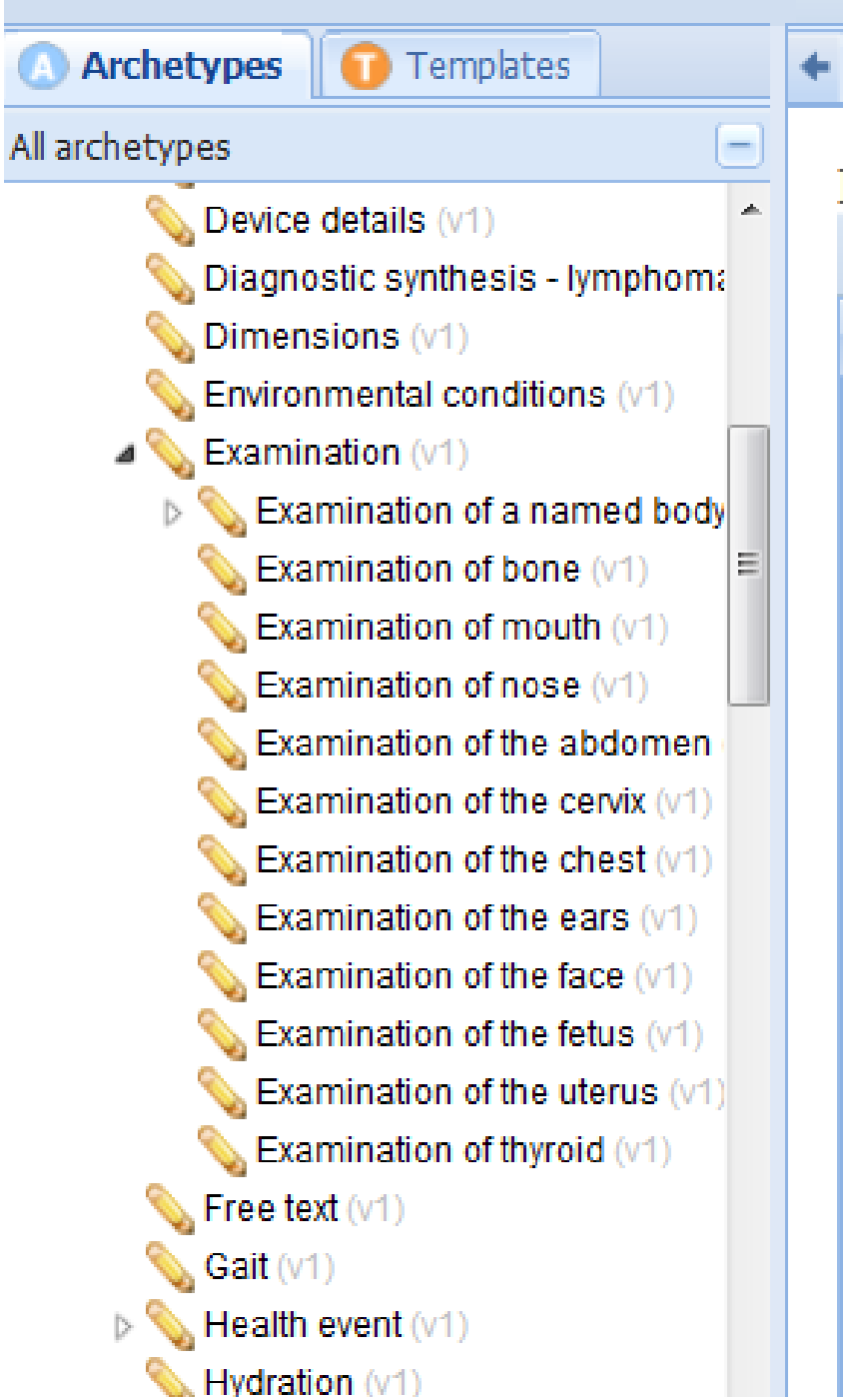
English ▾ Adopt Archety

Archetype: Inspection (openEHR-EHR-CLUSTER.inspection.v1)

Header | **Items**

Description Text Occurrences: 0..1 (optional)	Description of the edge or border of the object of inspection	Free or coded text
Definition Coded Text Occurrences: 0..1 (optional)	The nature of the edge of the object of examination	• Well defined [The edge is clearly defined] • Moderately defined [The edge is defined but less clear in places] • Poorly defined [The edge is not defined clearly in many places]
Translucent Boolean Occurrences: 0..1 (optional)	Is the object of inspection translucent?	
Detail Cluster Occurrences: 0..1 (optional) Cardinality: 0..* (*optional, repeating, unordered*)	Detailed examination of a finding on inspection	
Cluster: Exam Slot (Cluster) Occurrences: 0..1 (optional)	Exam details	**Include:** openEHR-EHR-CLUSTER.exam.v1 Included
Drawing Multimedia Occurrences: 0..1 (optional)	A drawing of the findings on inspection	image/cgm, image/gif, image/png, image/tiff, image/jpeg

Previous


Archetypes
Templates
All archetypes
Device details (v1)
Diagnostic synthesis - lymphoma
Dimensions (v1)
Environmental conditions (v1)
Examination (v1)
Examination of a named body
Examination of bone (v1)
Examination of mouth (v1)
Examination of nose (v1)
Examination of the abdomen
Examination of the cervix (v1)
Examination of the chest (v1)
Examination of the ears (v1)
Examination of the face (v1)
Examination of the fetus (v1)
Examination of the uterus (v1)
Examination of thyroid (v1)
Free text (v1)
Gait (v1)
Health event (v1)
Hydration (v1)

openEHR

Binding ....

- Archetype(s) are far more detailed (but mostly optional data points)
- Two data points match:
  - /items[Clinical description] = Finding
  - /items[Findings]/items[Location]/items[Description] = Site
- Mismatches:
  - Second is 0..* - e.g. a burn could be in multiple locations → pattern only allows 1 location
  - Archetype location assumes laterality included – needs variant pattern?

Possible general approach

- Map pattern *parameters* to content model *data points*; add following to archetype:

```
Concept_bindings = <
    [“SNOMEDCT”] = <
      patterns = <
            [“clinical finding”] = <
                  name = <[1213124|clinical finding present|]>
                  mappings = <
                        [02020202|finding|] = </items[at0004|Clinical description|]>
                        [33333222|site] =
                                    </items[Findings]/items[Location]/items[Description] >
                  >
            >
    >
>
```

Potential?

Could this work generally?

- Could avoid full Compositional code strings in data, and instead map pattern parameters to IM data points
  - ➔ reduces dependence
  - But how stable are the parameters?
  - Will there be enough patterns?

Summary

- *open*EHR archetype/template approach provides a semantic framework for capturing, representing and querying any data
- BIG ADVANTAGE: bindings are expressde in archetypes and templates, NOT THE DATA; can be added AFTER initial deployment
- Initial binding approaches are working, but are incomplete, and may be out of date, e.g. Internal value sets → Ref sets in the future?

Challenges

- When to code .... and not
- How to ensure binding assumption matches query authors (there will be many of the latter)
- How to choose SCT concepts.... pre-coordination problem
- Need for SCT post-coordination expression equivalence to work
- Solution that handles ICDx, LOINC, local terminologies

Questions?

Resources

http://www.openEHR.org/knowledge - archetypes

Other – see e.g. D Markwell's CFH report 2009

Acknowledgements:

Kent Spackman – pattern slides