

## 11.2 Terminology / Information Model Boundary Issues

When performing data analytics over clinical data, it is important to understand the interdependency between the terminology and the structural information model. For example, it is not sufficient to find a diagnosis of 56265001 |heart disease|, and make the assumption that the patient has heart disease. Instead, the surrounding information model must be considered to discover whether this is, for example, a confirmed diagnosis for the patient themselves, a suspected or preliminary diagnosis for the patient, or perhaps a family history of heart disease in the patient's paternal grandfather. Contextual or qualifying information about a code may appear in a variety of places, including:

- Within the information model – for example, a section heading titled "Family History"
- In the same coded data element – for example, precoordinated as "394886001|suspected heart disease|" or postcoordinated as "56265001 |heart disease|: 408729009 |finding context| = 415684004|suspected|"
- In a separate coded data element – for example, Diagnosis = 56265001 |heart disease|, Type = 148006 |preliminary diagnosis|

By understanding where and how this contextual or qualifying information is represented, more appropriate queries can be created.

When the same semantics may be represented in both the information model *and* the terminology, there is also a risk of ambiguity as to how these two representations should be combined. This is clearly demonstrated by models in which both the information model and the terminology can represent 'negation' or 'absence'. Does the combination of 'negation' in the information model and 'absence' in the terminology indicate:

- Double negative,
- Redundant restatement of the negative, or
- Additional emphasis of the negative?

It is important in these situations to have clear rules about how the semantics in the information model and the terminology should be combined.

The challenge often becomes even greater when heterogeneous data sources are integrated. When different information models represent the same semantics using different combinations of structure versus terminology, retrieval and reuse may miss similar information. To avoid false negatives or false positives in the query results, the integration and/or analytics processes must resolve these differences.

For example, in Figure [figure 11.2-0](#) below, the system on the left uses the 'Family history' structural heading to indicate that the selected disease is a family history, while the system on the right precoordinates this within the terminology. When integrating or querying across these data sources, these semantics need to be harmonized to ensure accurate queries can be performed.

Family history:	Problems:
Asthma <input type="checkbox"/>	Family history of asthma <input type="checkbox"/>
Cerebrovascular disease <input type="checkbox"/>	Family history of cerebrovascular disease <input type="checkbox"/>
Depression <input type="checkbox"/>	Family history of depression <input type="checkbox"/>
Diabetes mellitus <input checked="" type="checkbox"/>	Family history of diabetes mellitus <input checked="" type="checkbox"/>
Ischemic heart disease <input type="checkbox"/>	Family history of ischemic heart disease <input type="checkbox"/>
Sickle cell anemia <input type="checkbox"/>	Family history of sickle cell anemia <input type="checkbox"/>
Thalassemia <input type="checkbox"/>	Family history of thalassemia <input type="checkbox"/>

Figure 11.2-1: Two ways of recording family history of diabetes mellitus

Even when the same information model is used, different systems may populate this model with differing levels of precoordination. For example, the three clinical systems shown below in [figure 11.2-0](#) each collect data about a 'suspected lung cancer' diagnosis in a different way. For this reason, when given a common data model (as shown in [figure 11.2-0](#)), different systems may populate this in different ways. When this occurs, queries must be careful to consider all possible representations of the data, to ensure that contextual and qualifying information about each code is correctly interpreted.

The figure shows three distinct user interface forms for recording a suspected lung cancer diagnosis:

- General Practice (Green header):** Features a 'Problem/Dx' section with a dropdown for 'Cancer', a 'Body Site' dropdown for 'Lung', and a 'Status' section with radio buttons for 'Suspected' (selected), 'Confirmed', and 'Not found'. 'OK' and 'Cancel' buttons are at the bottom.
- Clinic (Orange header):** Features a 'Problem/Diagnosis' section with a 'Prob/Dx Name' dropdown for 'Suspected cancer' and a 'Body Site' dropdown for 'Lung'. 'OK' and 'Cancel' buttons are at the bottom.
- Public Hospital (Red header):** Features a 'Diagnosis' section with a 'Name' dropdown for 'Suspected lung cancer'. 'OK' and 'Cancel' buttons are at the bottom.

Figure 11.2-2: Three ways of recording suspected lung cancer

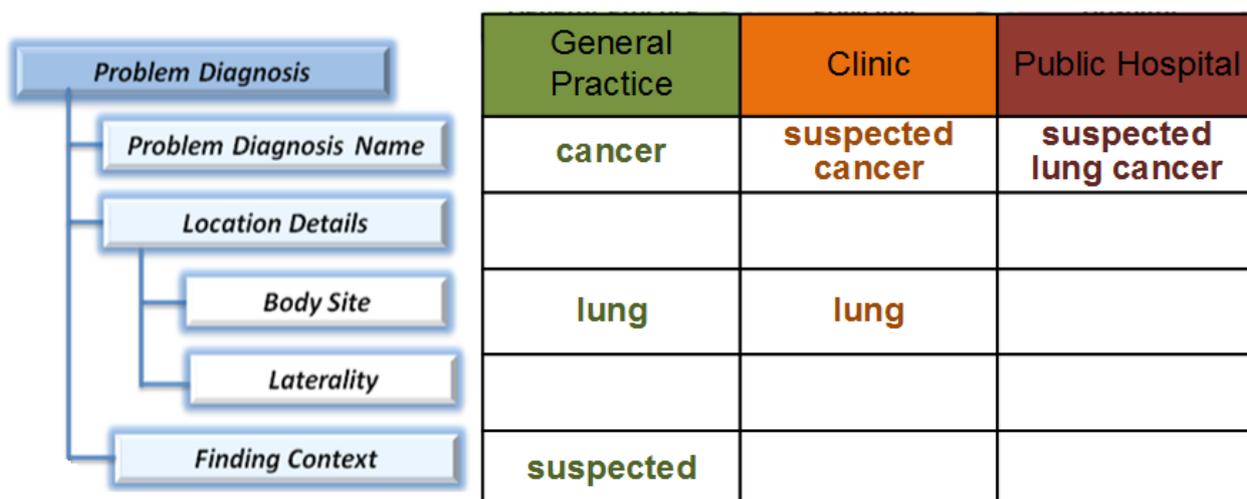


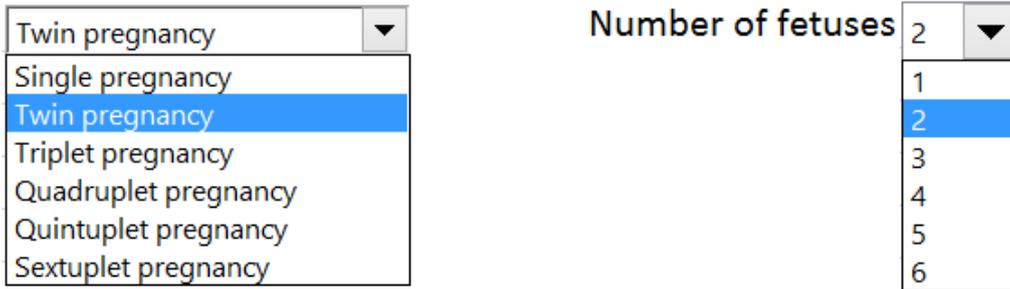
Figure 11.2-3: Three ways of populating a common Problem Diagnosis model

SNOMED CT is in the unique position to be able to resolve many of these challenges, using the techniques described in sections 6.4 [Description Logic Over Terminology](#) and 6.5 [Description Logic Over Terminology and Structure](#). For example, SNOMED CT enables the computation of equivalence and subsumption between alternative representations of data. For example, the postcoordinated expression

22253000 |pain| : 363698007 |finding site| = 56459004 |foot|

(which can be represented either in a single data element or using two separate data elements for 22253000 |pain| and 56459004 |foot|) can be automatically determined to be equivalent to the precoordinated concept 47933007 |foot pain| (stored in a single data element).

Some cases exist, however, where SNOMED CT is not currently able to automatically establish equivalence. These cases primarily relate to concepts for which the SNOMED CT concept model does not yet fully model their meaning. For example, the two approaches for representing a 'twin pregnancy' shown below ( [figure 11.2-0](#)) are currently not able to be computed as equivalent using SNOMED CT.



**Figure 11.2-4: Two non-equivalent ways of recording a twin pregnancy using SNOMED CT**

The SNOMED CT concept model continues to be extended to support equivalence and subsumption testing within an increasing number of hierarchies of SNOMED CT.

---