

Issues in Mapping LOINC Laboratory Tests to SNOMED CT

Olivier Bodenreider, MD, PhD

U.S. National Library of Medicine, NIH, Bethesda, MD

olivier@nlm.nih.gov

Abstract

Comprehensive clinical terminologies such as SNOMED CT tend to overlap with specialized terminologies such as LOINC (e.g., for the domain of laboratory procedures). Terminological systems such as the UMLS are often used to bridge between terminologies. However, the integration of LOINC in the UMLS and with other terminologies remains suboptimal. We mapped concepts for laboratory tests from LOINC to pre-coordinated SNOMED CT concepts, based on shared relations to other concepts. As LOINC is finer-grained than SNOMED CT, several LOINC codes tend to map to the same SNOMED CT concept. However, a large proportion of LOINC codes could not be mapped to SNOMED CT through this approach, because of underspecified definitions in SNOMED CT and a lack of fine-grained, pre-coordinated concepts in SNOMED CT.

Introduction

Biomedical terminologies and ontologies have proliferated in the past decade, not only for biology, but also for clinical medicine [1]. Terminologies such as SNOMED CT provide a large coverage of the domain of clinical medicine and often overlap with other large general terminologies (e.g., MeSH) and with specialized terminologies (e.g., LOINC).

In clinical information systems, terminologies such as SNOMED CT, used in patient records, need to be interoperable with terminologies used in subsystems, such as laboratory systems (e.g., LOINC). Terminology integration systems, such as the Unified Medical Language System (UMLS) play an important role in creating *post hoc* mappings between these terminologies and contribute to the interoperability of systems relying on these terminologies. A key element to identifying equivalent concepts across terminologies in the UMLS is the lexical resemblance among concept names. As a consequence, concepts whose names are not amenable to natural language processing, such as the names of laboratory tests in LOINC, generally cannot be mapped to equivalent concepts in other terminologies. However, both SNOMED CT and LOINC provide formal definitions

for their concepts in the form of a rich set of relations to other concepts. Comparing such sets of relations also provides the basis for comparing these concepts, provided there are enough shared relations between the two terminologies.

The objective of this paper is to analyze the issues in mapping concepts for laboratory tests from LOINC to existing, pre-coordinated SNOMED CT concepts, based on their descriptions (i.e., their relations to other concepts) and to evaluate the proportion of such mappings that can be derived automatically. Although SNOMED CT supports post-coordination, this study is purposely limited to the mapping between pre-coordinated concepts in LOINC and SNOMED CT.

The development of these terminologies is often supported by public funding, and harmonization between these terminologies has recently become a requirement from some funding agencies. Therefore, this study can also be considered a contribution to harmonizing SNOMED CT, the most comprehensive clinical terminology, with LOINC, the leading terminology for laboratory tests. While a few studies have explored the integration of LOINC and SNOMED [2, 3], the two terminologies have not been harmonized yet.

Background

The general problem area of this study is ontology matching, i.e., the identification of equivalent (or related) concepts across ontologies. Among the approaches developed for aligning ontologies, the two major families of techniques exploit the lexical resemblance among concept names (lexical alignment) and the structural resemblance among sets of relations in which the concepts are involved (structural resemblance). A review of these methods is beyond the scope of this paper and the interested reader is referred to [4] for further information.

In the case of LOINC, as mentioned earlier, the names of laboratory tests are not amenable to natural language processing techniques, including edit distance, stemming and normalization, because LOINC strings are created by concatenating with colons the names of the concepts to which a laboratory test is

related (e.g., *Sodium:SCnc:Pt:Ser/Plas:Qn*). Therefore, the technique of choice for aligning LOINC laboratory tests to other terminologies relies not on lexical, but on structural resemblance between the concepts across terminologies.

Materials

LOINC. The Logical Observation Identifiers, Names, and Codes (LOINC) is a vocabulary for laboratory tests and clinical observations [5, 6]. The two main types of entities in LOINC are laboratory tests and clinical observations, on the one hand, and the entities necessary for their description (sometimes referred to as “parts”), on the other. In fact, LOINC “part” concepts (e.g. *sodium*) serve as building blocks for the description of tests and observation, in association with a set of semantic relations. For example, *Sodium:SCnc:Pt:Ser/Plas:Qn*, the laboratory test in which the molar concentration of sodium is measured in the plasma (or serum) is identified by 2951-2. The list of relations of this concept to other concepts (“parts”) is shown in Table 1. For example, the “part” concept *Sodium* is linked to this test by the relationship *component*.

Table 1. List of the relation of the laboratory test 2951-2 to “part” concepts in LOINC

Relation-ship	Part ID	Part name
Component	LP15099	Sodium
Property	LP6860	SCnc – Substance Concentration (per volume)
Time	LP6960	Pt – Point in time (Random)
System	LP7576	Ser/Plas – Serum or Plasma
Scale	LP7753	Qn – Quantitative
Method	--	--

More formally, each laboratory test is described in reference to the analyte measured (component), the property under investigation, the time aspect, the origin of the sample (system) and the type of scale used. Additionally, the method used is reported when appropriate. The LOINC terminology does not use any particular formalism, such as description logics. However, the formal definitions provided by LOINC all conform to the 6-axis template presented in the example above and make use of named semantic relations, which makes them amenable to automatic processing. In addition to simple tests, LOINC also defines complex concepts, including panels (i.e., collections of tests) and concepts involving a challenge (e.g., glucose measurement, 90 minutes after oral

administration of 75g of glucose). The total number of tests and observations is 50,809, of which 37,767 correspond to laboratory tests and 34,767 to simple laboratory tests. The total number of “part” concepts is 44,314, of which 13,794 are used as value for the 6 main axes. All LOINC concepts are integrated in the UMLS. While the “part” concepts are generally well integrated with equivalent concepts from other terminologies, concepts for laboratory tests and clinical observations are not, due to the peculiarity of their names. The version of LOINC used in this study is version 2.22.

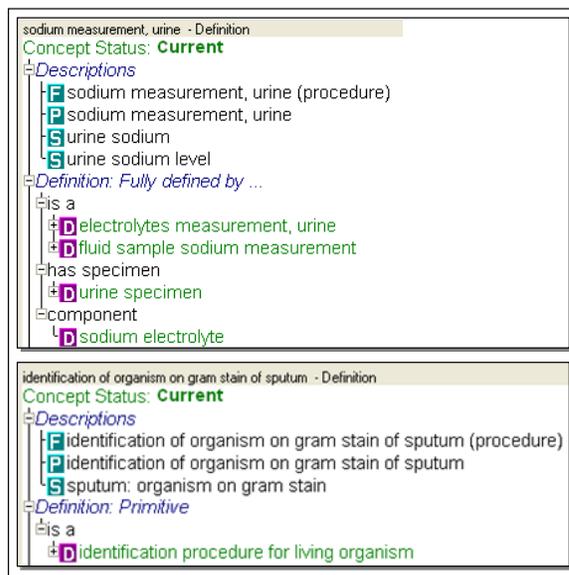


Figure 1. Example of laboratory procedures in SNOMED CT (fully defined - top; primitive and minimally modeled - bottom)

SNOMED CT is a comprehensive concept system for healthcare that provides broad coverage of clinical medicine, including anatomy, diseases, and procedures (laboratory procedures and others) [7]. SNOMED CT uses description logics for its representation. In practice, as in LOINC, SNOMED CT concepts can be used as building blocks for describing other SNOMED CT concepts. For some SNOMED CT concepts, the set of relations to other concepts provided is necessary and sufficient to fully define the concept. Other concepts, called primitives, are incomplete definitions, sometimes limited to one subclass (isa) relation. Unlike LOINC, SNOMED CT does not use a fixed template for the description of laboratory tests, but uses whatever sets of relations are appropriate. Examples of fully defined and primitive SNOMED CT laboratory concepts are shown in Figure 1, along with their relations to other concepts. Of the 310,311 active concepts in SNOMED CT, 9,511 correspond to laboratory procedures. The total

number of distinct concepts involved in the description of laboratory procedure concepts is 5,608. All SNOMED CT concepts are integrated in the UMLS. The version of SNOMED CT used in this study is dated July 31, 2007.

The UMLS is a terminology integration system for biomedicine. The most recent version of the UMLS (2007AC) used in this study integrates 143 terminologies, including LOINC and SNOMED CT. The UMLS identifies equivalent terms across terminologies and groups them into one UMLS concept. As mentioned earlier, natural language processing of the terms plays an important role in the identification of equivalent terms by the UMLS, and terms that are not amenable to natural language processing are less likely to be linked to potentially synonymous terms from other terminologies. In fact, only 6 LOINC concepts for laboratory tests (1) and clinical observations (5) are mapped to some concept in SNOMED CT. In contrast, the “part” concepts from LOINC tend to be well integrated with equivalent concepts from other terminologies. Of the 13,794 LOINC “part” concepts used as value for the 6 main axes, 4,501 (33%) are mapped to SNOMED CT through the UMLS. While it provides equivalence relations among terms across vocabularies, the UMLS does not provide equivalence across relationships.

Methods and results

We first examined the relationships linking laboratory test concepts to other concepts in LOINC and SNOMED CT and established correspondences between relationships across the two terminologies. We then used these common relationships and the UMLS mapping between concepts to align concepts sharing similar relations.

Analyzing relations for laboratory tests LOINC and SNOMED CT

The list of the 6 relationships linking laboratory test concepts to “part” concepts in the LOINC template is shown in the first column of Table 1. In SNOMED CT, laboratory test concepts can be linked to other SNOMED CT concepts through sixteen relationships, including **component**, **has specimen** and **procedure site – direct**. Based on the documentation available for the two terminologies and after manual inspection of a sample of concepts, we established the following correspondence between relationships in LOINC and SNOMED.

The relationship **component** in LOINC corresponds to **component** in SNOMED CT, linking a laboratory procedure to the analyte. The relationship **system** in

LOINC links the laboratory test to the substance in which the analyte is measured. In SNOMED CT, this relationship can be represented by a combination of relationships, linking the laboratory test first to a specimen (**has specimen**), and then linking the specimen to a substance (**specimen substance**). Alternatively, the relationship **system** in LOINC is sometimes represented by **procedure site – direct** in SNOMED CT (e.g., for skin tests). The relationship **scale** in LOINC corresponds to the relationship **scale type** in SNOMED CT. Finally, the relationship **time** in LOINC corresponds to the relationship **time aspect** in SNOMED CT. No correspondence was found in SNOMED CT for the relationship **property** in LOINC. Quite counterintuitively, the relationship **method** in LOINC does not correspond to the relationship **method** in SNOMED CT.

Mapping based on shared relations

The four LOINC relationships having a correspondence in SNOMED CT (**component**, **system**, **scale** and **time**) are potentially useful for mapping laboratory tests from LOINC to SNOMED CT. Of these, **time** is used in the definition of only one SNOMED CT laboratory concept and does not practically contribute to the mapping.

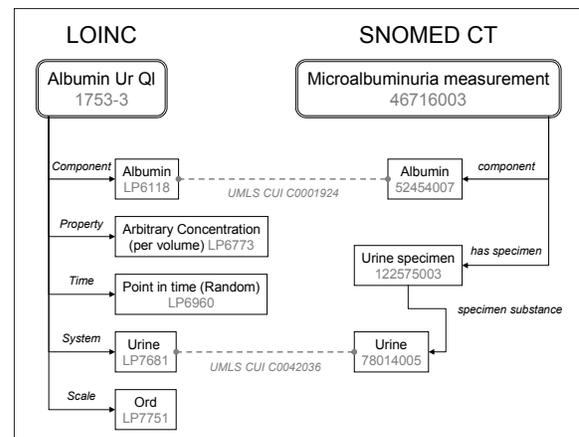


Figure 2. Mapping from LOINC to SNOMED CT based on shared relations

The mapping based on shared relations between LOINC and SNOMED CT associates a LOINC laboratory test concept L having the relations **component** to c_L , **system** to sy_L and **scale** to sc_L to a SNOMED CT concept S having the relations **component** to c_S , **has specimen** to sp_S (and **substance** between sp_S and sy_S), and **scale type** to sc_S . Additionally, there must be a correspondence between the following concepts: c_L and c_S , sy_L and sy_S , and sc_L and sc_S . In practice, this correspondence is assessed by the fact that the two concepts in a pair share the same UMLS

concept unique identifier (CUI). As mentioned earlier, a relation between S and sy_S through **procedure site – direct** also supports the mapping of L to S in lieu of the indirect relation S **has specimen** sp_S **specimen substance** sy_S . The mapping based on shared relations is illustrated in Figure 2.

The mapping based on three shared relations (**component**, **system** and **scale**) is likely to yield too few results, because there are very few laboratory concepts in SNOMED CT for which all three relations are represented. In order to increase recall, and at the risk of degrading precision, we also compute mappings ignoring either **system** or **scale**, or both. (It does not make sense to ignore relations involving **component**, which is central to laboratory tests.)

In practice, we compute the number of LOINC to SNOMED CT mappings based on the presence of 3 common relations (involving c_L , sy_L , and sc_L), two common relations (involving c_L and sy_L , and c_L and sc_L), and for common relations involving only c_L . No cases of mappings were found in which scale (sc_L) was involved. The number of mappings based on c_L and sy_L , and on c_L alone are listed in Tables 2 and 3, respectively. For example, there are 61 cases in which a pair (c_L , sy_L) corresponds to 2-10 LOINC concepts and 1 SNOMED CT concept.

Table 2. Mapping based on shared component and system relations

		SNOMED CT			
		0	1	2-10	>10
LOINC	0	--	0	0	0
	1	6,038	3	1	0
	2-10	12,097	61	13	0
	>10	4,295	112	61	5

Table 3. Mapping based on shared component relations alone

		SNOMED CT			
		0	1	2-10	>10
LOINC	0	--	0	0	0
	1	3,626	379	57	9
	2-10	3,678	1,074	270	13
	>10	188	152	183	37

In a majority of cases, LOINC concepts cannot be mapped to SNOMED CT on the basis of shared relations, even when considering component relations alone. When a mapping is found, it often links several LOINC concepts to one (or few) SNOMED CT concepts. In a few cases, however, one LOINC concept is mapped to several SNOMED CT concepts.

Discussion

The mapping based on shared relations is not satisfactory, because recall is insufficient when all three relations are used and, conversely, precision is very low when only one or two relations are used. In order to enhance the performance of the mapping process, we performed a failure analysis. When examining the formal definitions of the 9,511 laboratory procedures in SNOMED CT, it is apparent that few of them are directly and fully compatible with the definition of equivalent concepts in LOINC. As shown in Figure 3, 1,697 laboratory procedures do not exhibit any of the three main relations studied (component, system and scale). A majority of the remaining concepts only exhibits the component relation. This alone explains the poor performance we observed.

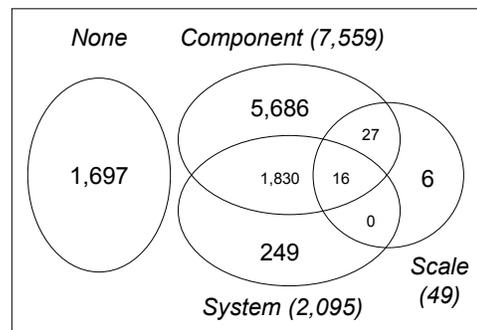


Figure 3. Number of laboratory procedures in SNOMED CT using relations equivalent to component, system and scale in LOINC

Component. The definition of laboratory tests is often more precise in LOINC than in SNOMED CT. In fact, only a fraction of the values for **component** in LOINC have a correspondence in SNOMED CT through the UMLS. Such laboratory tests from LOINC can thus not be mapped directly to SNOMED. However, by exploiting the hierarchy of the components in LOINC, it might be possible to find a mapping in SNOMED CT to a concept more generic than the original LOINC concept. In this case, it becomes possible to map the original LOINC laboratory test concept to a more generic laboratory test concept in SNOMED CT.

System. In contrast to **system** in LOINC, the specimen of laboratory procedure in SNOMED CT is often not defined with the highest precision. While most values for system in LOINC have a correspondence in SNOMED CT through the UMLS, most of them are never used as the value of **specimen substance** (through **has specimen**) in SNOMED CT. For example, *arterial blood* is used as **component** in LOINC, but SNOMED CT, while having a concept for arterial blood, tends to use higher-level concepts

(e.g., *blood*) in the relations used to define laboratory procedures. In addition to searching for exact matches in SNOMED CT for the value of **system** in LOINC, a mapping to some higher-level concept in SNOMED CT would help identify additional mappings. These mappings may not denote equivalence, but would be useful for integrating clinical data.

Another issue is that some primitive concepts have been modeled minimally in SNOMED CT (i.e., have a set of relations to other concepts sometimes limited to one subclass relation to a parent concept). In this case, the name of the laboratory test may characterize the test adequately, but the set of relations used to describe the test does not. One possible way of improving the mapping process would be to extract the information about the system from the concept name through natural language processing techniques and create the corresponding relations in SNOMED CT for mapping purposes. Such knowledge augmentation techniques have been used successfully in the alignment of anatomical ontologies [8].

Scale. In contrast to LOINC, there is no **scale type** relation defined for most laboratory test concepts in SNOMED CT. The absence of modeling of this relation often means that the test is modeled at a generic level, not referring specifically to a qualitative or a quantitative scale. To some degree, this issue resembles the mismatch between a specific system concept in LOINC and a generic system concept in SNOMED CT. The kind of mismatch in granularity observed for **system** (e.g., *arterial blood* vs. *blood*) is also present with **scale**. For example, while the concept *nominal* exists in LOINC and SNOMED CT, the finest-grained value for **scale type** in SNOMED CT is *qualitative* – an ancestor of *nominal* in the SNOMED CT hierarchy. This is the reason why none of the laboratory procedures defined with the relation **scale type** could be mapped.

Other issues. A careful analysis of the modeling choices in LOINC and SNOMED CT reveals significant differences, some of which hinder automatic mapping efforts. For example, we showed that the mapping of **system** in LOINC is often indirect in SNOMED CT, through a combination of **has specimen** and **specimen substance**. Analogously, time is represented differently in the two terminologies. For example, while both terminologies have a relationship for representing time aspects, this relation is rarely used in SNOMED CT. Moreover, SNOMED CT sometimes reifies (i.e., folds into a concept) the notion of time expressed separately in LOINC. This

phenomenon is exemplified by concepts such as *24 hour urine specimen collection* in SNOMED CT.

Conclusions

Although LOINC and SNOMED CT both cover the domain of laboratory procedures and use similar knowledge representation formalisms, the automatic mapping of laboratory procedures from LOINC to SNOMED CT based on shared relations remains incomplete and unsatisfactory. The approach we used could still be useful for assisting in the development of a manual map. To improve the performance of the mapping process, additional techniques could be used, including knowledge augmentation (i.e., extracting relations from the names of laboratory procedures) and the controlled traversal of hierarchies in SNOMED CT and LOINC.

Acknowledgements

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM). Our thanks go to Ramez Ghazzaoui who helped create and query the triple store.

References

1. Cimino JJ, Zhu X. The practical impact of ontologies on biomedical informatics. *Methods Inf Med* 2006;45 Suppl 1:124-35
2. Hsu C, Goldberg HS. Knowledge-mediated retrieval of laboratory observations. *Proc AMIA Symp* 1999;809-13
3. Spackman KA. Integrating sources for a clinical reference terminology: experience linking SNOMED to LOINC and drug vocabularies. *Medinfo* 1998;9 Pt 1:600-3
4. Euzenat J, Shvaiko P. *Ontology matching*. In. New York: Springer; 2007
5. Huff SM, Rocha RA, McDonald CJ, De Moor GJ, Fiers T, Bidgood WD, Jr., et al. Development of the Logical Observation Identifiers Names and Codes (LOINC) vocabulary. *J Am Med Inform Assoc* 1998;5(3):276-92
6. McDonald CJ, Huff SM, Suico JG, Hill G, Leavelle D, Aller R, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. *Clin Chem* 2003;49(4):624-33
7. Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. *Stud Health Technol Inform* 2006;121:279-90
8. Zhang S, Bodenreider O. Experience in aligning anatomical ontologies. *International Journal on Semantic Web and Information Systems* 2007;3(2):1-26