A Method for the Automated Mapping of Laboratory Results to LOINC

Lee Min Lau MD PhD, Kate Johnson MT (ASCP) MHR,
Kent Monson MS MT (ASCP) CLS, Siew Hong Lam MD, Stanley M. Huff MD
3M Health Care, Murray, Utah and Intermountain Health Care, Salt Lake City, Utah

Abstract

LOINC is emerging as the standard for laboratory result names, and there is great interest in mapping legacy terms from laboratory systems to it. However, the mapping task is non-trivial, requiring significant resource commitment and a good understanding of the LOINC identifying attributes for the laboratory result names. Because the number of results in a laboratory system may range from around 500 to 2000 or more, manual, one-by-one matching, even with the aid of the RELMA matching tool provided by LOINC, is time consuming and laborious. Moreover, human variation may introduce mapping inconsistencies or errors. Through our experience mapping the results from a variety of laboratory systems to LOINC, an automated mapping method has been developed and is described in this paper. This method allows for data from the laboratory information system to be provided in a manner familiar to the submitting technician, and makes use of parsing and logic rules, combined with synonyms, attribute relationships and mapping frequency data, to perform automated matching to LOINC.

Introduction

The 3M Longitudinal Data Repository (LDR) is based upon the VOSER approach, with the Healthcare Data Dictionary (HDD) as its foundation [1]. Whenever possible, controlled medical vocabularies and coding classifications, such as LOINC, SNOMED, UMLS and ICD9CM, are used as either "starter sets" or reference sources in populating the HDD. The starter set for laboratory results is LOINC™ [2]. LOINC stands for Logical Observation Identifier Names and Codes. In LOINC, laboratory results are named in a consistent and comprehensive manner, according to six attributes:

1. Component or analyte, e.g. sodium, glucose;
2. Property, e.g. substance concentration, mass rate;
3. Time, e.g. random (point in time) or 24-hour;
4. System or specimen or sample, e.g. serum, urine;
5. Scale or precision, e.g. quantitative, ordinal;
6. Method, e.g. electrophoresis, immune blot.

Each unique combination of the six parts would constitute a unique laboratory result and be given a unique LOINC identity code.

When a health care institution implements the 3M LDR, the organization's codes are mapped to the concepts in the HDD. Then, when the local codes and the associated patient data are sent in HL7 transactions from the institution's legacy systems to the CDR, the HDD concept identifier is stored along with the legacy system information and patient data. The HDD mapping thus allows different health care institutions to continue using their legacy system and codes, yet achieve standardization with established terminologies. In the present climate of mergers and acquisitions, the HDD mapping also allows for standardization across disparate legacy systems from different facilities now under the same enterprise umbrella.

For laboratory results, the laboratory system's internal code values are mapped to the LOINC codes in the HDD. To date, the local codes of seventeen laboratory systems from eleven health care institutions have been mapped to LOINC. The laboratory systems mapped have come from ADAC, Cerner, ALG, Citation and Sunquest, among other vendors. The number of local result codes from each laboratory system ranged from under a thousand to nearly three thousand. Through the experience, some lessons have been learned, and an automated mapping process is designed to maximize the efficiency and accuracy of the mapping.

As can be expected, the mapping of local laboratory results to LOINC codes is not a trivial task. It requires significant resource commitment from both 3M and its customers, and the speediness of the mapping completion has an impact on the implementation schedule. More importantly, the accuracy and consistency of the mapping is critical to ensure correctly normalized and integrated data. For instance, it is expected that the commonly used laboratory results would be found in LOINC, moreover, that multiple institutions should be mapping to the same results. This is important for the exchange and pooling of data [3].

Unfortunately, laboratory system data is seldom, if ever, provided in a form that translates easily to the LOINC definitions. For instance, LOINC names the

1067—5027/00/$5.00

property of a laboratory result according to the IUPAC definitions described in the Silver Book [4]. Examples include Mass Concentration (MCNC), Substance Concentration (SCNC), Number Rate (NRAT), Volume Ratio (VRTO), etc. Laboratory systems have not been seen to provide their result data with this type of property description. Type of scale (precision), such as ordinal, nominal, quantitative, etc., is another part of the LOINC name that is not intuitive to laboratory personnel. Last but not least, LOINC may be known to the laboratory technician, but usually not in sufficient details to understand its methodology or even its goals.

This lack of understanding may lead to undesirable mapping results. For instance, some institutions submitted the result with specimen of blood while others used serum or plasma, thus mapping to different LOINC codes. Although in some cases the testing is truly done on blood versus on serum, often it is because the institution failed to differentiate between the specimen collected and the specimen tested. Although both LOINC codes are valid, they are independent – LOINC does not provide relationships among or grouping the codes. Therefore, unless one recognizes this problem while mapping and made the decision to "correct" the institution's specimen submission and map to only one of the codes, the mapped results will not be comparable. In a sense, LOINC is like a messaging standard; being a naming convention means it allows combinations of values in the six fields that may not be truly realistic.

LOINC provides a tool, RELMA™ (Regenstrief LOINC Mapping Assistant) to assist in the mapping of local codes to LOINC. Although RELMA is of considerable help in looking up possible matches, it does so for each local code one at a time, and a human review is required to choose from among the candidate matches. This one-at-a-time mapping requires significant time and effort, and can also be affected by inconsistency or inaccuracy as a natural result of human variation.

Method

For these reasons, discussed above, we found it desirable to evolve a LOINC laboratory result matching methodology to automate as much of the process as possible. Because each LOINC code is based on a unique combination of six attributes, each code can be thought of as having a unique set of six relationships, one to each attribute. This approach is modeled after the computable (machine readable), formal definitions of SNOMED RT [5, 6]. In RT, each axis of SNOMED will have its own unique set of defining roles. For instance, in addition to the Is-A relationship, a diagnosis is defined by its Associated Topography, Associated Morphology and Associated Etiology relationships; whereas a procedure is defined by its Administered Substance, Uses Equipment, Has Object, Associated Topology, Has Measured Component, Administers Energy and Has Scale Type relationships. An example of a LOINC code and its six attributes is given in Table 1, and its relationship set in Table 2.

Therefore, a first step would be to create the sets of relationships for the existing LOINC codes. The HDD's relationship table is used to house these relationships. Similarly, a laboratory result from a legacy system can also be given its set of relationships according to the six LOINC attributes. Then, a straightforward automated comparison of the "to be matched" laboratory result's relationship set to all the LOINC relationship sets, attribute to attribute, would identify the exact LOINC code match.

The information in Tables 1 and 2 are presented as text for ease of reading. In practice, unique identifiers are contained in the tables instead. In addition to the LOINC code to attribute relationships, each independent value of an attribute is defined as a concept and placed in the right domain, in the HDD relationship table, with Is-A relationships. An example is shown in Table 3, and, for ease of reading, the text representation is included in parenthesis.

Table 1. A LOINC Example

LOINC Code	*LOINC Name*	*Component/ Analyte*	*Property*	*Time*	*System/ Specimen*	*Scale*	*Method*
2159-2	CREATININE:MCNC:PT:AMN:QN:	Creatinine	Mass Concentration	Point In Time	Amniotic Fluid	Quantitative	

Table 2. Relationship Set for LOINC Code 2159-2

Concept A	*Relationship*	*Concept B*
LOINC 2159-2	Has Component	Creatinine
LOINC 2159-2	Has Property	Mass Concentration
LOINC 2159-2	Has Time	Point In Time
LOINC 2159-2	Has System	Amniotic Fluid
LOINC 2159-2	Has Scale	Quantitative
LOINC 2159-2	Has Method	Null Method

Table 3. Relationships for the LOINC Attributes

Concept A	*Relationship*	*Concept B*
3 (Creatinine)	9 (Is A)	10 (Component)
11 (Metanephrine)	9 (Is A)	10 (Component)
12 (Creatine Kinase)	9 (Is A)	10 (Component)
13 (CK MB)	9 (Is A)	10 (Component)
14 (Hepatitis A IgM)	9 (Is A)	10 (Component)
4 (Mass Concentration)	9 (Is A)	15 (Property)
16 (Mass Rate)	9 (Is A)	15 (Property)
17 (Catalytic Concentration)	9 (Is A)	15 (Property)
18 (Arbitrary Concentration)	9 (Is A)	15 (Property)
5 (Point In Time)	9 (Is A)	19 (Time)
20 (24 Hour)	9 (Is A)	19 (Time)
6 (Amniotic Fluid)	9 (Is A)	21 (System)
22 (Urine)	9 (Is A)	21 (System)
23 (Serum)	9 (Is A)	21 (System)
7 (Quantitative)	9 (Is A)	24 (Scale)
25 (Ordinal)	9 (Is A)	24 (Scale)
8 (Null Method)	9 (Is A)	26 (Method)
27 (Electrophoresis)	9 (Is A)	26 (Method)

Next, the laboratory results to be matched to LOINC are requested from the system, in a form more familiar to the laboratory technician than the LOINC format. Explanations and examples are provided to help. For instance, instead of asking for the property of the laboratory results, the data type (number, text, titer, etc.), data value examples (for when the data type is text), and the unit with which the result is reported are requested for. A data submission sample is shown in Table 4.

In order to generate a set of relationships to the LOINC attributes, for each of the laboratory results, the attribute information must first be derived from the submitted data. One approach is to combine a comprehensive synonym set for the attribute concepts with parsing and logic rules. For instance, Table 5 shows some synonyms for the component attribute values of Metanephrine and Creatine Kinase; whereas Table 6 shows some synonyms for the system attribute values of Urine and Amniotic Fluid. Therefore, the data in the result name and specimen columns are compared to these tables, respectively, to arrive at those concepts that correctly identify the component and system attributes. The synonyms are obtained from the laboratory systems that have been mapped, supplemented by information culled from LOINC and other reference sources, such as textbooks, laboratory manuals and user guides. In order to improve on the matching effectiveness, a tool is provided for manual matching of any submitted term not currently in the synonym tables, and the matched term is then added to the synonym tables for future mapping. Note that for result name and specimen, any mention of time (e.g. "24H") is ignored in the synonym matching to identify the component and system attributes, respectively (see Tables 5 and 6). This is because timing is its own separate attribute and thus does not matter in the matching for component and system, even though terms like "24H" are commonly included in the result names.

Table 4. Laboratory Results to be Mapped to LOINC

Result Code	*Result Name*	*Specimen*	*Data Type*	*Data Value Examples*	*Unit*	*Timing*	*Method*
1000	CREATININE	AMNIOTIC FL	NUM		MG/DL		
2000	24H METANEPH	URINE	NUM		MG/24H		
3000	CK	SERUM	NUM		U/L		
4000	CK.MB	SERUM	%				ELECTROPHORESIS
5000	HAVAB IGM	SERUM	TEXT	POSITIVE/NEGATIVE			

Table 5. Synonyms for the Component Attribute

Concept ID	*Concept Name*	*Synonym*
11	Metanephrine	METANEPH
11	Metanephrine	24H METANEPH
12	Creatinine Kinase	CK
12	Creatinine Kinase	CPK
12	Creatinine Kinase	CK TOTAL

Table 6. Synonyms for the System Attribute

Concept ID	*Concept Name*	*Synonym*
22	Urine	U
22	Urine	UR
22	Urine	24 U
22	Urine	24 UR
6	Amniotic Fluid	AMN FL
6	Amniotic Fluid	AMNIOTIC FL
6	Amniotic Fluid	AMN

The synonyms are also used to support basic parsing and logic to derive or double-check other attributes. For instance, if the Time attribute is not explicitly stated in the timing column but included in the result name, then the timing description is parsed out. For instance, the "24H" is parsed out from "24H METANEPH" when the submission neglects to state it explicitly in the timing column (see Table 4). Otherwise, a null value in the Timing would default to a "Point In Time" (random) value for the time attribute. The value for the time attribute is also double-checked using the unit column – e.g. the denominator is "24 Hours" (or a synonym); and the specimen column – e.g. "24 Hour Urine".

The data type (number, text, titer, etc.) of the result value, data value examples (for when the data type is text), and the reporting unit columns are used to derive the property and scale attributes. For instance, when the data type is Number, the scale is Quantitative. Then, according to a table of rules maintained for this purpose, the unit with which the laboratory result is reported will point to its property. A sample is shown in Table 7. Again, any submitted unit term not currently in the synonym tables will be manually matched and added, and its property rule will also be added for future mapping. The data type column is also checked for units, e.g. "%", that are commonly placed there in error.

Table 7. Deriving the Property Attribute From the Reporting Unit

Concept ID	*Concept Name*	*Property*
28	MG/DL	4 (Mass Concentration)
29	G/L	4 (Mass Concentration)
30	MG/24H	16 (Mass Rate)
31	NG/MIN	16 (Mass Rate)

Lastly, if the method column is left blank, the method attribute defaults to "Null Method". The end result of the data manipulation is shown in Table 8. From this table, a set of six attribute relationships is generated for each laboratory result (see Table 9 for an example) and compared to those relationship sets already in the HDD. This matching is done automatically, and helped by another set of rules, best described as "mapping tips". For instance, Result Code 3000 in Table 4 is submitted with the specimen of Serum, thus the "Has System" relationship generated for it is to Serum (see Table 8). A "mapping tip" states that if no LOINC match exists for this relationship, given that all other attribute relationships matched, to check if the LOINC "Has System" relationship is to "Serum/Plasma". If so, that LOINC code is considered a match.

Table 8. End Result of the Manipulation of the Table 4 Data Submitted for Matching

Result Code	*Result Name*	*Component/ Analyte*	*Property*	*Time*	*System/ Specimen*	*Scale*	*Method*
1000	CREATININE	3 (Creatinine)	4 (Mass Concentration)	5 (Point In Time)	6 (Amniotic Fluid)	7 (Quantitative)	8 (Null Method)
2000	24H METANEPH	11 (Metanephrine)	16 (Mass Rate)	20 (24 Hour)	22 (Urine)	7 (Quantitative)	8 (Null Method)
3000	CK	12 (Creatine Kinase)	17 (Catalytic Concentration)	5 (Point In Time)	23 (Serum)	7 (Quantitative)	8 (Null Method)
4000	CK.MB	13 (CK MB)	17 (Catalytic Concentration)	5 (Point In Time)	23 (Serum)	7 (Quantitative)	27 (Electrophoresis)
5000	HAVAB IGM	14 (Hepatitis A IgM)	18 (Arbitrary Concentration)	5 (Point In Time)	23 (Serum)	25 (Ordinal)	8 (Null Method)

Table 9. Attribute Relationship Set Generated for Result Code 1000 From Table 8

Concept A	*Relationship*	*Concept B*
Result Code 1000	Has Component	3 (Creatinine)
Result Code 1000	Has Property	4 (Mass Concentration)
Result Code 1000	Has Time	5 (Point In Time)
Result Code 1000	Has System	6 (Amniotic Fluid)
Result Code 1000	Has Scale	7 (Quantitative)
Result Code 1000	Has Method	8 (Null Method)

The "matching tips" rule set also uses a table that contains information on how frequently a LOINC code has been mapped to, calculated from the over 20,000 results submitted by the seventeen laboratory information systems we mapped. This mapping frequency information is used to suggest the most likely match when the automatching failed to find an exact match. It also helps to prevent unintentional divergence or inconsistency of mapping, for instance, matching to the LOINC code with the system attribute value of Blood in some cases and to the LOINC code with the system attribute value of Serum in other cases, all other attributes being the same, because some submissions show the collection specimen of Blood instead of the testing specimen of Serum. The frequency data is used to drive a "mapping tip" that will suggest matching to the LOINC code with the system of Serum when the "Has System" attribute relationship to Blood is seen in the automatching, for the appropriate results.

If an exact match to LOINC is not found for a laboratory result, then the synonym tables are used to obtain the LOINC term for each attribute of the result, and a proposed name in the LOINC format is built for the result (see the LOINC name in Table 1), to be submitted to LOINC for inclusion in its next version. The laboratory result and its set of attribute relationships are added to the HDD in the meantime, so that they can be used in future mapping. When the LOINC code is assigned it will be added to this result in the HDD. Meanwhile, new results can be matched to it, thus avoiding duplicate entries to the HDD or duplicate submissions to LOINC. This automated matching process can also be used to add new LOINC codes into the HDD, by pulling the new codes and attribute information from the new LOINC version into the "Submitted Data" table, then generating relationships and other required information from there to be loaded into the HDD. Thus, the "master set" of LOINC codes used in the matching is kept up to date.

Conclusion

Based on our experience mapping the result names from a variety of laboratory systems to LOINC, an automated mapping method has been developed. A manual matching component is included to add the "fallouts" from the automated matching into the "knowledge base" of synonym, relationship and rule tables, so the process can continually improve its automatching success rate. With this method, we hope to improve the efficiency and accuracy of the LOINC matching process for laboratory results.

References

1. R. Rocha et al. Designing a Controlled Medical Vocabulary Server: The VOSER Project. Computers and Biomedical Research, 27(6): 472-507, 1994

2. LOINC Users Guide. Obtained from http://www.mcis.duke.edu/standards/termcode/loinc.htm

3. L. M. Lau and S. H. Lam. Comparing the Mapping of Laboratory Results to LOINC Among Different Healthcare Institutions Using a Common Data Dictionary. Proceedings of AMIA Annual Fall Symposium, 1105, 1999.

4. The Silver Book – Compendium of Terminology and Nomenclature of Properties in Clinical Laboratory Science. International Union of Pure and Applied Chemistry. 1995.

5. SNOMED RT Documentation. College of American Pathologists. 1999.

6. Spackman K.A., Campbell K.E. (1998) Compositional concept representation using SNOMED: Towards further convergence of clinical terminologies. Proceedings of AMIA Annual Fall Symposium, 875-9, 1998.