A Vision of the Journey Ahead: Using Public Health Notifiable Condition Mapping to Illustrate the Need to Maintain Value Sets

Shaun Grannis MD, MS, Daniel Vreeman PT, DPT, MS
The Regenstrief Institute and Indiana University School of Medicine; Indianapolis, IN

Abstract

Maintaining value sets is a necessary and distinct activity apart from maintaining recognized controlled vocabularies. As an illustration of one such value set, we evaluate the CDC and Regenstrief versions of the notifiable condition mapping tables (NCMT) and illustrate they are not synchronized. We used practical informatics approaches including heuristic queries and similarity measures to accurately identify more than 800 new candidate reportable LOINC codes. To successfully maintain value sets we must establish a clear strategy for coordinating the value sets and process for disseminating among stakeholders. These stakeholders will likely be distinct from, but interface with, the existing standards development organizations (SDOs).

Introduction

Interest in and resources for standardizing electronic healthcare transactions are at unprecedented levels as manifest in the many initiatives being advanced by the Office of the National Coordinator for Health Information Technology (ONC). Examples of these initiatives include offering incentives to providers and hospitals for demonstrating meaningful use of health information technology;[1] funding regional HIT extension centers to offer technical assistance and guidance to support and accelerate meaningful use of Electronic Health Records (EHRs);[2] and funding improvement and expansion of health information exchange (HIE) services.[3] These initiatives share a common goal: to improve the quality and efficiency of healthcare by promoting interoperability and exchange of data.

Achieving this goal is not easy: it requires that electronic healthcare information be recorded and exchanged in computer-interpretable formats using agreed upon standardized semantic and syntactic content.[4] The goal is further complicated by the reality that healthcare incorporates a wide variety of workflows with each workflow using varying combinations of data elements, business rules, and data exchange methods.

A number of efforts have attempted to standardize, harmonize and recommend specifications for health information interoperability and exchange. Most recently, the health information technology standards panel (HITSP) has created a host of interoperability specifications that describe standards-based approaches for exchanging electronic healthcare data.[5] Common among both the HITSP interoperability specifications and the harmonization efforts that preceded HITSP (such as the consolidated health informatics initiative -- CHI[6]), is the recognized use of familiar, widely available controlled medical terminology (semantic) standards for capturing healthcare data, which include LOINC, ICD, and SNOMED. The necessity of standards such as these in healthcare interoperability specifications is unsurprising.

What may be surprising to some is the central theme of this paper: operationalizing many healthcare workflows with interoperability specifications requires use of more than just the commonly available and familiar terminology standards. A *value set* is a collection of concepts drawn from one or more terminology systems and grouped together for a specific purpose. It may be a simple list of concepts drawn from a single code system, or it might be constituted by expressions drawn from multiple code systems. The vocabulary standards realm is a large, complex environment and comprehending the myriad terminologies and the growing numbers of value sets defined for specific workflows and messaging standards is a resource-intensive undertaking.

If we are to sustain standardized electronic health exchange, then the value sets associated with the widely available standards must also be continuously maintained and harmonized in conjunction with their affiliated standards. However, responsibility for creating value sets and metadata may lie outside the standards development organizations (SDO's) because such information is typically use-case-specific.

While these notions are recognized among standards developers and informaticians, and efforts supporting maintenance of value sets and metadata are underway,[7-9] it may not be as widely recognized among less technically inclined HIT stakeholders, including policy makers. To illustrate the need for routine coordinated maintenance of these affiliated value sets, we will describe the concrete example of automated electronic laboratory reporting of public health notifiable conditions

Reporting of public health notifiable conditions is a requisite for successfully managing the public health disease burden in a community. However, clinical care processes under-report public health notifiable conditions for a variety of reasons: reporters are overburdened and/or under-resourced; reporters lack knowledge or willingness; clinical data is scattered across disparate settings in different formats, which makes completing a report difficult.[10] Using a standards-based messaging and vocabulary infrastructure (HL7 and LOINC), the Regenstrief Institute has implemented and maintained an HIE-based, automated electronic laboratory reporting (ELR) and case-notification system for over 10 years.[11] The system, called the notifiable condition detector (NCD), receives more than 350,000 real-time HL7 version 2 clinical transactions daily, including laboratory studies, diagnoses, and transcription from more than 50 organizations, national labs and local ancillary service organizations. The NCD demonstrated a 4-fold greater detection rate than traditional physician-based reporting methods.[12]

To produce these results, the NCD leverages mappings between standardized test codes and conditions for which that test may be reportable. These mappings are exemplified in the Centers for Disease Control and Prevention's (CDC) PHIN notifiable condition mapping table (NCMT),[13] which was initially created by a multi-stakeholder partnership whose members included representatives from the CDC, the Council of State and Territorial Epidemiologists (CSTE) and the Regenstrief Institute. The CDC offers their NCMT in downloadable format[14] and Regenstrief also shares their NCMT upon request. The NCMT associates each potentially reportable LOINC code with the nationally notifiable disease[15] for which that code is potentially reportable; non-reportable LOINC codes are not present in the NCMT. Some LOINC codes are specific for a single disease, whereas others are less specific. For example, LOINC code 14470-9 is an enzyme immunoassay test code for Chlamydia trachomatis and is solely reportable for "Chlamydia trachomatis genital infection" while LOINC code 11475-1 is a general test code for "microorganism identified" and is potentially reportable for many diseases, depending on the test results.

The NCMT improves the case detection processing efficiency in two ways. First, rather than scanning all transactions using computationally expensive algorithms, results that are not potentially reportable can be bypassed by assuming that a clinical transaction is potentially reportable only when the transaction contains a LOINC code from the NCMT. Second, by leveraging the disease list associated with a given LOINC code as recorded in the NCMT, the accuracy of the case detection methods can be improved by focusing and tailoring the logic to the disease(s) expected to be found for a given test.[16]

Case detection systems that wish to leverage the NCMT must map local test codes to LOINC codes. Regenstrief maintains an up-to-date translation table that maps local test codes to LOINC codes for the Indiana Network for Patient Care (INPC), an operational statewide health information exchange.[17] After proprietary test codes in each INPC transaction are automatically translated into LOINC codes, the NCD uses the LOINC code to determine whether a result is potentially reportable by cross-referencing the NCMT. Only those transactions whose LOINC codes exist in the NCMT are evaluated; all others are bypassed. While this approach improves processing efficiency, if a potentially reportable LOINC code is missing from the NCMT its absence may result in false-negatives (missed cases).

To avoid missed cases, the NCMT must be updated on a regular basis because NCMT mappings change over time for various reasons. First, the list of nationally notifiable infectious diseases is revised periodically. For example, a disease may be added to the list as a new pathogen emerges, or a disease may be deleted as its incidence declines. Second, reporting criteria for federal or state entities may be updated at differing frequencies. Third, new LOINC test codes are created on a routine basis and among these additions may be tests that identify notifiable diseases whose codes must be added to the NCMT along with the test's associated disease(s).

Regenstrief currently manages their NCMT LOINC-to-disease mappings in partnership with local and state public health partners within Indiana. New potentially reportable LOINC codes are most commonly identified when our public health partners convey details of a case that the NCD failed to identify via automated methods. Using this information we trace the root cause, which often resides in the NCMT: either a condition was absent for an existing LOINC code, or the LOINC code was absent for a given condition. As an example of the former, a LOINC code for urine culture was present but lacked a mapping to MRSA. For the latter, a new LOINC code for Hepatitis A was recently created but was not added to the NCMT. Regenstrief has also acquired additional reportable LOINC codes from HIE-

public heath collaboratives in Washington and New York states.

Approaches for addressing the changing milieu have been described to varying degrees,[18,19] though none characterize the feasibility of implementing the approach, nor do they contemplate the necessary coordination among stakeholders. In the remainder of this paper we highlight the informatics and strategic approaches to identifying and communicating newly created potentially reportable LOINC codes. Specifically, we first describe and evaluate approaches to identifying new LOINC codes for tests that may contain reportable results by using various axes of LOINC and string comparators; we then discuss processes and potential stakeholders needed to move forward.

Methods

We downloaded version 2.29 of the LOINC standard from http://loinc.org on February 11, 2010 with 57,475 unique codes. The Regenstrief NCMT was extracted from the operational NCD on February 11, 2010. The CDC NCMT was downloaded on February 11, 2009 from the CDC website. LOINC codes and NCMT's were loaded into version 8.1 of PostgreSQL database (http://postgresql.org) for analysis. We labeled each reportable LOINC code as present only in the Regenstrief NMCT, present only in the CDC NCMT, or present in both. We also included the distinct (and often different) lists of conditions mapped to each LOINC code for Regenstrief and the CDC NCMT.

To identify LOINC codes that were candidates for being reportable but were currently not labeled as such, we linked currently unlabeled LOINC codes to LOINC codes labeled as reportable by using various combinations of the LOINC axes *Component* (what is measured, evaluated, or observed), *System* (context or specimen type within which the observation was made), and *Method* (procedure used to make the measurement or observation).

Subsets of these linked, unlabeled codes were then manually reviewed to identify which were potentially reportable. We also cross-referenced the newly identified potentially reportable LOINC codes with the INPC local code to LOINC mappings for determining which of these codes were actively used in the INPC. These newly identified potentially reportable LOINC codes with active mappings in the operational HIE represent likely missed cases.

To indentify undiscovered potentially reportable LOINC codes using an alternative method we used a hybrid string comparator implemented in version 5.8.8 of Perl (http://www.perl.org) that incorporates three well known similarity functions: the Jaro-Winkler comparator, the Levenshtein edit distance, and the Longest Common subsequence.[20] We hypothesize that tests having a LOINC component name similar to the nationally notifiable disease name may increase the likelihood the test is reportable for that condition. We generated string similarity scores by comparing all 151 of the 2010 nationally notifiable disease names (available at http://www.cdc.gov/ncphi/disss/nndss/phs/infdis.htm) with all 57,475 LOINC codes, for a total of 8,678,725 comparisons (151 x 57,145).

Results

The Regenstrief and CDC NCMT's used for this evaluation contained 4,871 and 5,645 LOINC-to-condition mappings, respectively. Figure 1 shows the Venn diagram of unique LOINC codes for the Regenstrief and CDC NCMT's. Although the CDC NCMT contained more records, the Regenstrief NCMT contained more unique LOINC codes (3,691 compared with 3,498). The Regenstrief NCMT had an average of 1.3 conditions per LOINC code while the CDC NCMT had an average of 1.6 conditions per LOINC code.
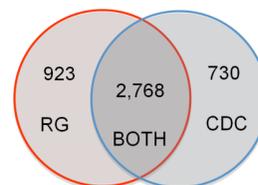

Figure 1: Venn diagram showing the relationship between unique reportable LOINC codes from the current CDC and Regenstrief (RG) NCMT files. This illustrates how the once-aligned files have diverged over time. Regenstrief has more unique reportable LOINC codes while the CDC has more LOINC-to-condition mappings.

Using the LOINC axes Component, System, and Method we identified 685 candidate LOINC codes from the Regenstrief NCMT and 507 candidates from the CDC NCMT. These were all manually reviewed for accuracy and 3 false-positive mappings were noted in the Regenstrief NCMT file stemming from an erroneous mapping of a Helicobacter Pylori LOINC code to Campylobacter disease. All 507 CDC candidate LOINC codes were deemed reportable by manual review. Using this simple but accurate approach, a combined 779 new reportable LOINC codes were identified using both the CDC and Regenstrief NCMT

files. Figure 2 shows a Venn diagram illustrating the distribution of candidate LOINCs for each NCMT.
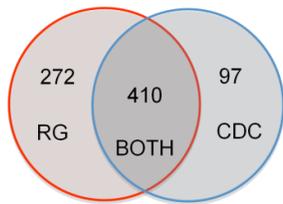

Figure 2: Venn diagram showing the relationship between the 779 newly identified reportable LOINC codes. Using LOINC axes of Component, System, and Method the Regenstrief (RG) NMCT identified 272 novel reportable LOINC codes, the CDC NCMT identified 97 novel codes, and both identified 410 shared codes.

We cross-referenced the 682 newly identified Regenstrief potentially reportable LOINC codes with the INPC local code-to LOINC mappings and found that 136 of the 682 are actively used in the INPC. Similarly we found that 29 of the 507 newly identified CDC codes are actively used in the INPC. Adding these potentially reportable LOINC codes to the NCD system will likely improve condition detection accuracy by decreasing missed cases.

Using the string similarity method we identified 147 LOINC codes that produced nearness scores equal to or greater than 0.80 and were not found by the previous approach. Manual review of this set found 122 reportable LOINC codes not identified using the deterministic LOINC axes approach. Examples of these codes are listed in Table 1.

Table 1: Examples of reportable LOINC codes found using string nearness between LOINC component and notifiable condition name. The first 3 records represent true-positives (reportable tests), while Cyclosporine represents a false-positive.

LOINC	LOINC Component	Notifiable Condition	Similarity
45082-5	Chlamydia trachomatis rRNA	Chlamydia trachomatis infection	0.838
49374-2	Hepatitis C virus RNA	Hepatitis C virus chronic	0.812
41857-4	Vibrio parahaemolyticus DNA	Vibrio parahaemolyticus	0.896
15103-5	Cyclosporine	Cyclosporiasis	0.834

Discussion

These results suggest that value sets such as the NCMT will diverge if not maintained on a routine basis in conjunction with its affiliated standards. Further, we demonstrate that it is feasible to create an accurate informatics approach to identifying potentially reportable LOINC codes. However, the greatest challenge to maintaining the NCMT may not lay with developing the informatics component consisting of search algorithms, SQL queries and nearness comparators. Instead, the greatest challenge may be coordinating the management of this information across

multiple stakeholders. The focused nature of value sets and metadata such as the NCMT suggests that in many cases responsibility for maintaining these data lies outside of the standards development organization, resting instead with the stakeholders and subject matter experts who derive value from them. In other cases the SDO may be the appropriate steward of the updates because the SDO already has a forum for convening the value-set stakeholders. The common Laboratory Order Value Set from LOINC (http://loinc.org/usage) is a recent example whose development was closely tied to the development of the standard itself.

While we used NCMT as a specific case, the need to maintain value sets applies to other areas, with another notable example being quality reporting.[21] Quality measures rely on controlled vocabularies and because all controlled vocabularies evolve over time, the quality initiatives must harmonize their rules with this evolution.

To successfully maintain value sets, several prerequisites are needed. First, working relationships among stakeholders must exist. For the NCMT, relationships already exist among many of the stakeholders. CSTE currently collaborates with CDC, Regenstrief and other organizations to facilitate information sharing within the informatics community. Second, a clearly designated entity with appropriate resources must be charged with maintaining the value set by enabling a suitable maintenance schedule. Third, the value set maintainer must keep in close contact with affiliated SDO's. Fourth, a clear method for distributing the value set is needed.

Given these prerequisites, a potential process for coordinating the NCMT may proceed as follows: First, a new LOINC version is released twice per year by Regenstrief or CSTE releases a new reportable condition list. Second, new candidate reportable LOINC codes are identified using practical, manageable search strategies that may include deterministic searches, string similarity functions, and other processes that leverage the LOINC hierarchy (Regenstrief and/or CSTE and/or CDC). Third, SME's (most likely from CSTE and CDC) review the candidate codes and identify those that are truly reportable, and determine under what circumstances and for what conditions the test is reportable. Fourth, the NCMT is updated and disseminated (Regenstrief and/or CDC). A centralized distribution mechanism should be preferred to a highly distributed mechanism that may create greater coordination challenges.

Conclusions

First, we must recognize that maintaining value sets is a necessary and distinct activity apart from maintaining recognized controlled vocabularies. As an example of this, we illustrated that the CDC and Regenstrief versions of the notifiable condition mapping tables (NCMT's) are not synchronized. Second, we described practical informatics approaches to aid in keeping it up to date by identifying candidate reportable LOINC codes. Finally, to successfully maintain value sets we must establish a clear strategy for coordinating the value sets and process among stakeholders. These stakeholders will likely be distinct from, but interface with, the existing standards development organizations (SDO's).

References

1. The Office of the Coordinator for Health Information Technology. Meaningful Use. Available at: http://healthit.hhs.gov/portal/server.pt?open=512&mode=2&objID=1325. Accessed Feb 25, 2010.
2. The Office of the Coordinator for Health Information Technology. Health Information Technology Extension Program. Available at: http://healthit.hhs.gov/portal/server.pt?open=512&mode=2&objID=1335. Accessed Feb 25, 2010.
3. The Office of the Coordinator for Health Information Technology. The State-level Health Information Exchange Consensus Project. Available at: http://healthit.hhs.gov/portal/server.pt?open=512&mode=2&objID=1242. Accessed Feb 25, 2010.
4. Walker J, Pan E, Johnston D, Adler-Milstein J, Bates DW, Middleton B. The value of health care information exchange and interoperability. Health Aff (Millwood). 2005 Jan-Jun;Suppl Web Exclusives:W5-10-W5-18.
5. Kuperman GJ, Blair JS, Franck RA, Devaraj S, Low AF. NHIN Trial Implementations Core Services Content Working Group. Developing data content specifications for the nationwide health information network trial implementations. J Am Med Inform Assoc. 2010 Jan-Feb;17(1):6-12.
6. Hufnagel SP. National electronic health record interoperability chronology. Mil Med. 2009 May; 174(5 Suppl):35-42.
7. Agency for Healthcare Research and Quality. United States Health Information Knowledge Base. Available at: http://ushik.ahrq.gov. Accessed February 25, 2010.
8. Zhu M, Mirhaji P. Semantic representation of CDC-PHIN vocabulary using Simple Knowledge Organization System. AMIA Annu Symp Proc. 2008 Nov 6:1196.
9. ISO/IEC 11179-1:2004(E): Information technology - Metadata Registries (MDR), 2004-09-15, International Organization for Standardization, Geneva, Switzerland. Available at: http://metadata-standards.org/11179. Accessed: Feb 26, 2010.
10. Doyle TJ, Glynn MK, Groseclose SL. Completeness of Notifiable Infectious Disease Reporting in the United States: An Analytical Literature Review. American Journal of Epidemiology (2002) Vol. 155, No.9:866-874.
11. Overhage JM, Suico J, McDonald CJ. Electronic laboratory reporting: barriers, solutions and findings. J Public Health Manag Pract. 2001 Nov;7(6):60-6.
12. Overhage JM, Grannis S, McDonald CJ. A Comparison of the Completeness and Timeliness of Automated Electronic Laboratory Reporting and Spontaneous Reporting of Notifiable Conditions. Am J Public Health. 2008, 98:344-350.
13. Centers for Disease Control and Prevention, Public Health Information Network (PHIN). Introduction to the PHIN Notifiable Condition Mapping Tables. Available at: http://www.cdc.gov/phin/library/documents/pdf/Introduction_to_the_PHIN_Notifiable_Condition_Mapping_Tables.pdf. Accessed Feb 11, 2010.
14. Centers for Disease Control and Prevention, Public Health Information Network (PHIN). LOINC to Condition Table. Available at: http://www.cdc.gov/PHIN/library/historical-archive.html. Accessed Feb 11, 2010.
15. McNabb SJ, Jajosky RA, et al. Summary of notifiable diseases--United States, 2006. MMWR Morb Mortal Wkly Rep. 2008 Mar 21;55(53):1-92.
16. Friedlin J, Grannis S, Overhage JM. Using natural language processing to improve accuracy of automated notifiable disease reporting. AMIA Annu Symp Proc. 2008 Nov 6:207-11.
17. Biondich PG, Grannis S. The Indianapolis Network for Patient Care (INPC): An Integrated Clinical Information System Informed by Over Thirty Years of Experience. Supplement to the J Public Health Manag Pract. Suppl 2004: S81-S86.
18. Wei L, Tokars JI, Lipskiy N, Ganesan S. An Efficient Approach To Map LOINC Concepts To Notifiable Conditions. Advances in Disease Surveillance 2007;4:172.
19. Steindel S, Loonsk JW, Sim A, Doyle TJ, Chapman RS, and Groseclose, SL. Introduction of a hierarchy to LOINC to facilitate public health reporting. AMIA Annu Symp Proc. 2002 Nov:737-11.
20. Grannis SJ, Overhage JM, McDonald C. Real world performance of approximate string comparators for use in patient matching. Stud Health Technol Inform. 2004;107(Pt 1):43-7.
21. Myles JL, Shamanski F, Witte D. The physicians quality reporting initiative: measure development, implementation and current procedural terminology coding. Adv Anat Pathol. 2010 Jan;17(1):49-52.