Exploring the use of ontologies and automated reasoning to manage selection of reportable condition lab tests from LOINC

Karen Eilbeck*, Jason Jacobs, Sunanda McGarvey, Cynthia Vinion and Catherine Staes

[1] Department of Biomedical Informatics, 26 South 2000 East, HSEB
Salt Lake City, UT 84112-5775, USA.
[2] CDC Century Center, Bldg 2500, Rm 2308, Mail Stop E78, Atlanta, GA. 30329, USA

ABSTRACT

Epidemiologists publish criteria for laboratory tests that must be reported to public health agencies in order to initiate public health control measures. There are efforts to publish value sets of standard laboratory test names using Logical Observation and Identifier Names and Codes (LOINC®) codes to enable automated systems to use the codes to identify reportable events. Unfortunately, the set of lab tests (and thus codes) vary by state, are difficult to manually curate, and may be missing desired or include undesired tests. Previously, we developed an ontology that classified the terminology used to describe LOINC®-coded tests for Chlamydia. To test the extensibility of this model, we extended the ontology to handle tests for tuberculosis. The requirements for tuberculosis laboratory test reporting in Utah and New York City gathered for the CDC's Reportable Conditions Knowledge Management System (RCKMS) project were reviewed. This provided the basis for manual queries to LOINC® to garner all possible tests for tuberculosis, and examination of these tests revealed new terms to add to the ontology. For each test, we created a new ontology term with a logical definition, and used the HermiT reasoner to automatically classify the tests into an ontology of tests. We used the new ontology to query for epidemiological selection, and compared to manually selected result sets. The LOINC® database provides structure that is useful to develop an application ontology to support epidemiologists with the task of managing sets of codes that meet reporting criteria. The automated classification strategy we propose is reproducible and extendable to address new diseases and problems found as the ontology is improved.

1 INTRODUCTION

Public health authorities such as local and state public health departments and the Centers for Disease Control and Prevention (CDC) monitor the population for over 80 reportable conditions such as elevated blood lead level, anthrax, measles and tuberculosis. Timely reporting of these conditions allows for proper treatment of individuals, implementation of control measures to prevent the occurrence of new events, investigation of outbreaks, and assessment of preventive measures. With increasing use of electronic health records (EHR) and pressure for their Meaningful Use (MU) (1), there have been efforts to automate systems to improve reporting from health care organizations to public health entities. Before a report is made however, the *reporting logic,* that is the rules for deciding when a condition is reportable, must be communicated from public health epidemiologists to health care providers and systems. Different jurisdictions represented by different state and major city health departments have differing criteria for what warrants a public health report. The criteria may include: a) Laboratory criteria that define a set of reportable laboratory names and results (e.g., positive tuberculin skin test); b) Clinical criteria, including signs and symptoms (e.g., prolonged and productive cough) or clinical findings (e.g., abnormal chest x-ray); or c) Epidemiologic criteria (e.g., recent travel outside the US). For each criterion in each jurisdiction, there is a *value set* of artifacts that satisfy the requirements. For laboratory criteria, the value sets are most often lists of laboratory test names and reportable results or a list of codes corresponding to entries in the Logical Observation and Identifier Names and Codes® (LOINC) database (available at http://loinc.org).

The laboratory domain of the LOINC® database includes names and identifiers and a structured definition for each test. Each test has a universal code, a name, and a unique value for six axes: component, property, timing, system, scale and method. It has over 46,000 entries and is updated regularly. Traditionally, each jurisdiction has posted their own list of laboratory reporting criteria on a website. In 2012, the CDC published value sets of LOINC® codes, in a Reportable Conditions Mapping Table (RCMT), for all jurisdictions (2). The value sets were curated manually via queries to the LOINC® database. Difficulty arises when either the selection logic or the content of the database change, and when there are differences in the logic between jurisdictions (which frequently occurs). Also the criteria defined by epidemiologists do not match perfectly to the

* To whom correspondence should be addressed:
keilbeck@genetics.utah.edu

organizational structure of LOINC®. This leads the epidemiologist to perform multiple queries in order to garner all the codes fitting a set of criteria. Updating these values sets is a time intensive task. The problem concerns: How to automate the definition of sets of reportable laboratory tests for each jurisdiction, and thus, how to translate epidemiological requirements to a set of LOINC® codes.

While hierarchical and ontological approaches to organizing the contents of LOINC® have been applied (3, 4), the use of ontologies to bridge the semantic gap between epidemiological requirements and the database's content has not been resolved. Previously, we implemented an ontology for chlamydia laboratory tests and explored the consequence of different selection logic on the resulting set of tests (5). This allowed us to visually represent the effect of more and less stringent requirements, and propose the use of ontologies to manage this portion of the reporting workflow.

This paper describes the extension of our prototype ontology for chlamydia-related laboratory tests (5) to include tests for tuberculosis. Tuberculosis tests are more varied than those previously examined for Chlamydia. Tuberculosis is one of three conditions being explored by CDC and collaborators for a prototype Reportable Conditions Knowledge Management System (RCKMS).

2 METHODS

2.1 Explore reporting logic for tuberculosis

We identified the written requirements (selection logic) for tuberculosis reportable condition laboratory tests from the Utah Department of Health and the New York City Department of Health. The microorganisms tested for in these 2 jurisdictions are Mycobacterium tuberculosis and Mycobacterium complex, a subset which includes M. tuberculosis, M. africanum, M. bovis-BCG, M. caprae, M. canetti, M. microti, M. pinnipedii, and M. bovis. The methods of testing determined by these jurisdictions are: tuberculin skin tests, interferon gamma release assay, acid fast bacillus culture/smear, non-specific culture, drug susceptibility testing, high pressure liquid chromatography, DNA probes, PCR, and organism specific culture. The latter 3 methods had previously been cataloged in our laboratory test ontology as they were mandated for Chlamydia reporting.

2.2 Extract all Tuberculosis tests from LOINC®

Using the reporting logic as a guide, we manually queried the LOINC® database using Access. We gathered all extracted tests for tuberculosis, excluding acid fast bacillus and antibiotic resistance tests. We automatically parsed the analyte and organism from the component axis using Excel.

2.3 Add new terms to existing base ontology

To gather the new terms to add to our reportable condition ontology, we extracted five organism terms, 43 system terms, one analyte term, and 19 method terms from the set of all tests. The 188 new terms were added to the existing laboratory test ontology following a hierarchy appropriate for epidemiological searching. For example, the new term *acid fast stain* is a kind of *microscopy* and *bone* is a kind of *specimen source site*.

2.4 Create ontology terms for each LOINC® test, and classify using semantic reasoner

For each LOINC laboratory test in the set of all tests that could detect tuberculosis, we created a cross product ontology term in OBO format (6) using a Perl script to parse out the relevant fields from the LOINC® database. Each term has a LOINC ID, a name, a genus, and relationships to four differentiae (a system, an analyte, a method, and an organism). The new terms were added to the existing ontology, and classified hierarchically using the HermiT reasoner via the OBO Ontology Release Tool (OORT) (7). The output of OORT is a classified ontology in both OWL and OBO formats, and may be browsed here.(8) Figure 1 displays a laboratory test, with its differentiating properties in the ontology and its corresponding unique values in LOINC®.

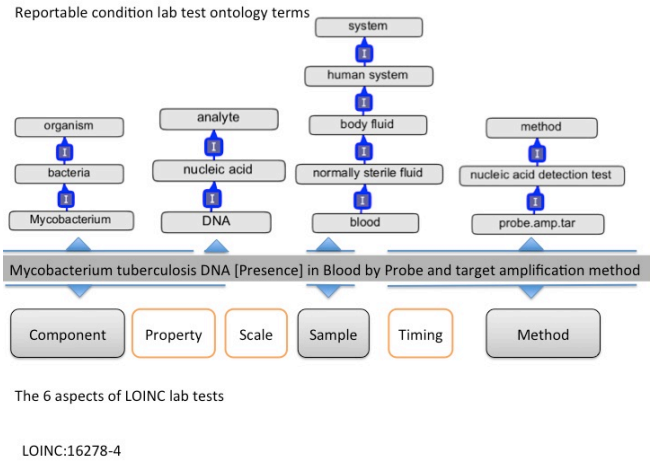

Figure 1 The LOINC laboratory test 16278-4 showing the classification of four differentiating properties using the laboratory test ontology above the grey line and the classification of the LOINC axes below the grey line.

2.5 Query the tests in the ontology and compare to manual curated lists of laboratory tests

The full ontology contains the tuberculosis laboratory tests organized into a meaningful hierarchy of general to specific tests. This ontology was queried for the value sets to select logic for tuberculosis tests, from several jurisdictions including Utah, Colorado, Washington and Delaware. Manually curated valued sets for these jurisdictions were availa-

ble from the RCKMS pilot project for comparison. The queries were performed using the search tool within the OBO-edit ontology tool (9).

3 RESULTS

The lab tests in the ontology were classified by HermiT into a general to specific subsumption hierarchy, using the differentiating properties. An example of the classification of a general test is shown in Figure 2, where two more general parent terms are inferred. In this example, ***Mycobacterium tuberculosis*** *DNA [Presence] in Bronchial specimen by Probe and target amplification method* is a kind of ***Mycobacterium sp*** *DNA [Presence] in Bronchial specimen by Probe and target amplification method*. Each LOINC® test has a transitive is_a relationship to a more general test term, and four relationships to its differentiating properties.

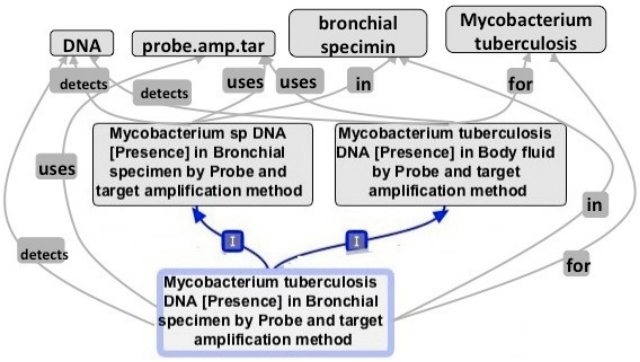

Figure 2 A representation of the classification of the lab test LOINC:14557-3 (Mycobacterium tuberculosis DNA [Presence] in Bronchial specimen by Probe and target amplification method). This test has two inferred parent terms: LOINC:21405-6 (Mycobacterium sp DNA [Presence] in Bronchial specimen by Probe and target amplification method) and LOINC:58931-7 Mycobacterium tuberculosis DNA [Presence] in Body fluid by Probe and target amplification method, which are less specific. The relations to the ontology terms for method, organism, analyte and body system are also shown.

For three reporting criteria, a comparison of the manually created value set and the value set derived from querying the ontology is shown in Table 1. For these three comparisons, every test in the manual value set was included in the ontology. For the first two criteria, querying the ontology returned a larger value set that subsumes the manually created set. The ontology-based query identified one appropriate lab test that was missed by the manual process (#6, Table 1). We found, however, that new terms for animal tests (#11-14), laboratory test orders (#7), and control specimen testing (#5) need to be added to the ontology to improve specificity. For a third criteria, the manual set includes a test (#15) that was present in the ontology but did not meet the specific conditions of the query because the manually

Table 1. Comparison of the value sets derived by querying an ontology versus manually curating a list for three criteria in tuberculosis reporting logic.

	ontology	manual	LOINC code and Name
Value set selected for 'Positive interferon gamma release assay'			
1	yes	yes	64084-7 Mycobacterium tuberculosis stimulated gamma interferon [Units/volume] corrected for background
2	yes	yes	45323-3 Mycobacterium tuberculosis tuberculin stimulated gamma interferon [Presence] in Blood
3	yes	yes	46217-6 Mycobacterium tuberculosis stimulated gamma interferon [Units/volume] in Blood
4	yes	yes	39017-9 Mycobacterium tuberculosis tuberculin stimulated gamma interferon/ Mitogen stimulated gamma interferon
5	yes	yes	46216-8 Mycobacterium tuberculosis tuberculin stimulated gamma interferon/Mitogen Mitogen stimulated gamma interferon [Units/ volume] in control Blood
6	yes	no	71773-6 Mycobacterium Tuberculosis Stimulated Gamma Interferon [Presence] In Blood
7	yes	no	71775-1 Mycobacterium Tuberculosis Stimulated Gamma Interferon Panel In Blood
Value set selected for 'Positive tuberculin skin test'			
8	yes	yes	43419-1 Tuberculosis reaction wheal [Diameter] --3 days post 1 TU intradermal
9	yes	yes	1647-7 Tuberculosis reaction wheal [Diameter] --3 days post 25 TU intradermal
10	yes	yes	1648-5 Tuberculosis reaction wheal [Diameter] --3 days post 5 TU intradermal
11	yes	no	23538-2 Tuberculosis Reaction Wheal [Diameter] --3 Days Post Dose Avian Tuberculin Intradermal
12	yes	no	23539-0 Tuberculosis Reaction Wheal [Diameter] --3 Days Post Dose Mammalian Tuberculin Intradermal
13	yes	no	32561-3 Tuberculosis Reaction Wheal [Diameter] --1 Day Post Dose Mammalian Tuberculin Intradermal
14	yes	no	23537-4 Tuberculosis Reaction Wheal [Diameter] --2 Days Post Dose Mammalian Tuberculin Intradermal
Value set selected for 'Demonstration of M. tuberculosis mycolic acids using high-pressure liquid chromatography on a culture from a clinical specimen'			
15	no	yes	67725-2 Mycobacterium sp identified in Isolate by HPLC

selected test is for mycobacterium species, not specific for M tuberculosis. The extended ontology remains appropriate for Chlamydia tests.

DISCUSSION AND CONCLUSIONS

The application ontology we developed represents an epidemiologist's view of the tuberculosis tests in LOINC®, and is not intended to be a general-purpose reference ontology of samples, analytes, or methods. Extending our initial lab test ontology to encompass the kinds of tuberculosis tests required the addition of new terms but minimal restructuring. For example, we added a *non-specific culture* term to the method portion of the ontology, where previously we had only listed *organism-specific culture* for Chlamydia reporting. This approach is therefore extensible.

Logically describing each test using differentiating principles places the tests into a hierarchy of general to specific terms, thus providing an ideal querying resource. Performing these sometimes complex hierarchical queries is not available using LOINC® alone. Although the laboratory tests in LOINC® are categorized into 6 different axes, we have found it necessary to split the Component axis into Organism and Analyte to better model epidemiological data. This ontology does not include three of the LOINC® axes: property, scale and timing, as these were not required in the selection logic we explored. This has the consequence of causing equivalent classes. Two terms may have the same set of differentiating properties in the ontology, but different axes in LOINC® such as the property measured (e.g., volume versus presence). These equivalent classes have not had an adverse effect on the queries performed to find the value sets. In the future we will look to incorporate the Scale axis into the ontology to better capture nominal tests.

In our comparison of manual and automated value sets we did not find full agreement between the two methods. The number of tuberculosis tests in each value set varied among RCMT (n=251), RCKMS (n=315), and this ontology (n=188). The automated method identified tests missed by the RCKMS manual method, and identified situations where the manually-selected tests are less specific than the reporting criteria. In addition, inspection of tests incorrectly pulled by the automated method helped us identify new terms that should be added to the ontology to improve the classification of laboratory orders and tests performed on animals and control specimens. The underlying logical structure of LOINC® provides a good candidate for ontology driven management. Manually selecting and managing the hundreds of codes required to represent reporting logic is unwieldy and imprecise. Using the ontology to query the tests provides consistency to the results and reproducibility. Revisions to the ontology are archived and additions and revisions to LOINC® can be quickly updated in the ontology and classified to the correct position. However, if the ontology is badly structured, or the fields of the tests parsed incorrectly, error can be introduced. This work demonstrates that an ontology can be used to structure and query standard lab test names consistent with an epidemiologists view of laboratory criteria. A future goal is to provide a visual tool to help epidemiologists and persons implementing electronic laboratory reporting (ELR) to see the consequences of different selection logic on the LOINC® codes included in detection logic.

ACKNOWLEDGEMENTS

The authors were partially supported by funding from the CDC (Staes) (COE in Public Health Informatics Grant No. 5P01HK000069. P.I. Matthew Samore, MD, University of Utah), CDC (Eilbeck) IPA, and National Library of Medicine (Jacobs) (NLM Training Grant No.T15LM007124).

REFERENCES

1. CMS. Meaningful Use. Balitmore [cited 2013 February 11]; Available from: http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Meaningful_Use.html.
2. CDC. Reportable Condition Mapping Table (RCMT). 2012 [cited 2013 Feb 11]; Available from: http://www.cdc.gov/ehrmeaningfuluse/rcmt.html.
3. Srinivasan A, Kunapareddy N, Mirhaji P, Casscells SW. Semantic Web Representation of LOINC: an Ontological Perspective. AMIA Annu Symp Proc. 2006:1107.
4. Steindel S, Loonsk JW. Introduction of a hierarchy to LOINC to facilitate public health reporting. AMIA Annu Symp Proc. 2002:737-41.
5. Eilbeck K, Jacobs J, Staes C. Optimize Querying of LOINC® with an Ontology: Give Me the Chlamydia Tests the Epidemiologists Want Me to Use! In: University of Hawai'i, editor. 46th Hawaii International Conf on System Sciences; Maui: Computer Society Press; 2013. p. 9.
6. Day-Richter J. OBO format version 1.2. 2006 [cited 2012 September 14]; Available from: http://www.geneontology.org/GO.format.obo-1_2.shtml.
7. OBO Ontology Release Tool. 2012 [cited 2012 September 14]; Available from: http://code.google.com/p/owltools/wiki/OortIntro.
8. LOINC ontology for chlamydia and tuberculosis. 2013 [cited 2013 April 29]; Available from: http://www.obobrowser.org/browser/obob.cgi?release=public_ICBO-OORT-TB.obo.
9. Day-Richter J, Harris MA, Haendel M, Lewis S. OBO-Edit: an ontology editor for biologists. Bioinformatics. 2007 Aug 15;23(16):2198-200.