

Mapping Department of Defense Laboratory Results to Logical Observation Identifiers Names and Codes (LOINC®)

**Lee Min Lau MD PhD, Pam D. Banning, Kent Monson MS,
Elva Knight, Pat S. Wilson, Shaun C. Shakib MPH,
3M Health Information Systems, Salt Lake City, Utah**

Abstract

The Department of Defense (DoD) has used a common application, Composite Health Care System (CHCS), throughout all DoD facilities. However, the master files used to encode patient data in CHCS are not identical across DoD facilities. The encoded data is thus not interoperable from one DoD facility to another. To enable data interoperability in the next-generation system, CHCS II, and for the DoD to exchange laboratory results with external organizations such as the Veterans Administration (VA), the disparate master file codes for laboratory results are mapped to Logical Observation Identifier Names and Codes (LOINC) wherever possible. This paper presents some findings from our experience mapping DoD laboratory results to LOINC.

Introduction

The Department of Defense (DoD) has invested significant effort in its Electronic Health Record (EHR), the Composite Health Care System (CHCS), which has been in use throughout all DoD facilities for nearly a decade. Over 500 hospitals and clinics are supported by 101 host systems. Each host system has a set of master files used to encode patient data in CHCS. When content is loaded into a CHCS master file (e.g. a list of laboratory result names), a unique Internal Entry Number (IEN) is automatically generated as the identifier (primary key). The creation of CHCS master files is not coordinated across all DoD host systems. Therefore, an item would receive different IENs in different DoD host systems (e.g. "1" may mean "Serum Sodium" at one host site but "Hematocrit" at another). Patient data in the CHCS database, encoded with the master file codes at each host system, is thus not encoded the same way across all DoD facilities.

The DoD requires data interoperability within its organization. It also needs to exchange clinical data with the Veterans Administration (VA) in order to provide services in a seamless fashion to both DoD and VA patients. The DoD began the transition to the next-generation system, CHCS II, five years ago.

CHCS II aims to store data from all DoD facilities in an enterprise-wide Clinical Data Repository (CDR). Each patient is expected to have a single electronic medical record, available at any DoD facility. Furthermore, 25 months of historical data is expected to be accessible in real time.

Approach

Because of the decade long and DoD-wide use of CHCS, the DoD has a large collection of very valuable historical as well as current patient data. The knowledge gained from analyzing the population data can greatly benefit care delivery and outcome, not only to DoD patients, but also to health care at large. Replacing the master files at each DoD host site with a set of standard codes, without mapping to the legacy codes, would result in the loss of the DoD patient data to computable clinical or administrative use. Essentially, the system will behave as if patient data collection is only starting now. Paper records or text printouts will be needed for past medical data (e.g. for a follow up visit). This has very serious implications for patient care and population management for the DoD, which has worldwide facilities and deployment. In transitioning to CHCS II, it is critical to preserve the historical patient data from the legacy CHCS for continuity of patient care, quality of care delivery, population health management and outcomes research.

For these reasons, the DoD has elected to standardize its master file codes to a single reference set, and to map to standard terminologies where applicable. Historical codes as well as currently active master file codes have been mapped, so that legacy patient data can be interoperable across all DoD facilities and across time. To date, a total of 3 million CHCS master file items from the domain areas of Demographics and Encounters, Laboratory, Microbiology, Pharmacy and Radiology Text Reports have been mapped for all 101 host systems supporting all DoD facilities. For laboratory results, the DoD data is expected to be translated to Logical Observation Identifier Names and Codes (LOINC) for external communication.

System

The 3M Healthcare Data Dictionary (HDD) is a Vocabulary Server application designed to support the integration of coded data in the CDR [1]. The content of the HDD is cross-referenced to standard vocabularies, e.g., LOINC; reference sources, e.g., the Unified Medical Language System (UMLS); and classification schemes, e.g., the International Classification of Diseases, 9th Edition, Clinical Modification (ICD9CM). External vocabularies are not loaded as disparate islands of code sets, but are mapped to unique concepts in the HDD, and linked by the appropriate relationships. Each HDD concept is identified by a meaningless Numerical Concept Identifier (NCID), and the identifiers from external terminologies are mapped to it. For instance, the concept of chickenpox would have the UMLS Concept Unique Identifier (CUI) of C0008049 as one of its mapped representations. It would also have an "ICD9CM Code" relationship to "052.9". As another example, the HDD NCID of 20411 would identify the laboratory result of random serum/plasma sodium reported in Substance Concentration units (e.g. mMol/L). The LOINC code 2951-2 and the LOINC name "SODIUM:SCNC:PT:SER/PLAS:QN:" are two of the mapped representations for NCID 20411, and any non-standard, legacy codes for this laboratory result would all be mapped to this NCID.

NCIDs are used to encode the data in the 3M CDR in CHCS II. Data enters the CDR through incoming transactions such as Health Level 7 (HL7) messages, or via data entry programs such as a clinical workstation. The external code in the incoming HL7 message is translated by the HDD to the corresponding NCID for storage in the CDR. A standard code such as LOINC would already be in the HDD and can be sent in incoming transactions without any prerequisite mapping. If the sending system uses non-standard codes, such as DoD's IENs, then these legacy codes must first be mapped into the HDD so that when they are received in incoming messages, they can be translated to NCIDs. Thus, through mapping, the HDD can translate between one standard and another, between legacy systems, and between a legacy system and a standard. For DoD to exchange data with external systems, such as the VA, the HDD can translate the DoD data from NCIDs to LOINC.

The LOINC mapping process employed by the HDD has been described previously [2]. Put simply, we

developed a knowledge base that maintains a set of six attribute relationships for each LOINC laboratory result. For example, NCID 20411 – random serum/plasma sodium – LOINC code 2951-2, LOINC name "SODIUM:SCNC:PT:SER/PLAS:QN:", would have the following relationships:

- Has Analyte Sodium
- Has Property Substance Concentration
- Has Timing Point in Time
- Has Specimen Serum/Plasma
- Has Scale Quantitative
- Has Method Null

For each of the six LOINC attribute, the knowledge base maintains a comprehensive set of synonyms as well as rules to derive the attribute from other knowledge. For instance, the unit the laboratory result is reported in (e.g. mMol/L) is used to derive the property. Using this knowledge base, an automated LOINC mapping tool has been developed. The output of the tool has been validated through manual, expert review by HDD personnel with laboratory technician background and extensive experience in LOINC mapping.

Findings

As a result of mapping laboratory result master file codes for multiple health care institutions with a variety of laboratory systems, the HDD has developed a comprehensive domain of over 43,000 laboratory result concepts. The size of the laboratory master file from the DoD ranged from a few hundred to over 18,000 rows, averaging about 4,000 rows. However, the items in each file are not unique, and many are mapped to the same HDD concept. The number of unique concepts mapped in each master file ranged from the smallest at 152 to the largest at 4,544, averaging at 1,596 (see Table 1). All together, approximately half a million laboratory master file codes (IENs) from all 101 DoD host systems have been mapped to just over 21,000 unique laboratory result concepts in the HDD (see Table 3).

As of the beginning of February 2005, over half of the laboratory result concepts (metadata) mapped from the DoD master files did not have a LOINC code (see Table 3). Site by site, the proportion of DoD laboratory master file with a LOINC code ranged from 63.7% to 93.4%. The average is 78.5% (see Table 2). Because the DoD master files are widely disparate in size, there would be a lack of significant overlap between files, leading to the lower percentage of LOINC codes when aggregated.

The DoD results are not too different from that found in our commercial mappings. For instance, of the 13,400 unique laboratory result concepts used by the commercial health care organizations that have been mapped in the HDD, just above half have a LOINC code (see Table 2). Site by site, the proportion of commercial laboratory result master files with a LOINC code ranged from 51.1% to 74.0%. The average is 63.4% (see Table 2). File size from commercial systems vary far less than for the DoD. The smallest was mapped to 1,058 unique laboratory results, and the largest to 2,382, with the average being 1,700 (see Table 1).

# of Unique Lab Results	Smallest Site	Largest Site	Average
Used by DoD	152	4,544	1,596
Used by Commercial Sites	1,058	2,382	1,700

Table 1. The spread of the number of unique HDD laboratory result concepts used by the DoD host sites and commercial health care organizations.

Proportion of Unique Lab Results with LOINC Code	Smallest Proportion	Largest Proportion	Average Proportion
Used by DoD	63.7%	93.4%	78.5%
Used by Commercial Sites	51.1%	74.0%	63.4%

Table 2. The spread of the proportion of unique HDD laboratory result concepts with a LOINC code used by the DoD host sites and commercial health care organizations.

Lab Results	# of Unique Lab Results	# of Unique Lab Results With LOINC Code	Proportion of Unique Lab Results With LOINC Code
HDD	43,664	27,509	63.0%
Used by DoD	21,171	9,925	46.9%
Used by Commercial Sites	13,400	6,752	50.4%

Table 3. The number and proportion of unique HDD laboratory result concepts with LOINC codes used by the DoD and commercial health care organizations.

At first glance, the low percentage of DoD master file laboratory results that have a LOINC code would

appear to imply that LOINC alone would not be sufficient as the standard code for data exchange. However, one must bear in mind that the CHCS master file contains all laboratory result codes that are used over the last decade, whether currently active or not. Of the active laboratory result codes, some are used very rarely. All non-reportable as well as reportable laboratory result codes are included. Thus, the proportion of CHCS master file codes that has a LOINC code may not accurately reflect the laboratory result data that may need to participate in data interchange with external organizations such as the VA.

The next step is to query the DoD CDR to find the laboratory result codes that have been stored because they were used in the past two years, and look at the proportion that have a LOINC code. The results are summarized in Table 4. Note that Tables 4 and 5 present the count of actual patient data encoded with the laboratory result concept codes, as opposed to Tables 1 to 3 which present the count of metadata (vocabulary concepts). Thus, the “Lab Results” in Tables 4 and 5 represent usage data from the DoD CDR, whereas the “Lab Results” in Tables 1 to 3 refer to the vocabulary concepts in the HDD. The number of unique laboratory results stored in the DoD CDR is only 11,010, just a little over half of all the DoD master file laboratory results mapped in the HDD. Thus, as much as half of the CHCS laboratory master file codes are either inactive or so rarely used that they have not been ordered in two years. Interestingly, the proportion of unique laboratory result concepts with a LOINC code, stored in the CDR, remains around 46% (metadata count). However, looking at all the laboratory results that have been stored in the CDR, we found that almost 90% have a LOINC code (actual patient data use). Therefore, the implication is that most of DoD’s laboratory result concepts that do not have a LOINC code are likely the rarely used ones that have not yet been submitted to the LOINC Committee to be assigned a LOINC code.

# of Lab Results in CDR	252,538,886
# of Lab Results with LOINC Code	226,936,672
Proportion of Lab Results with LOINC Code	89.9%
# of Unique Lab Results in CDR	11,010
Proportion of Unique Lab Results in CDR with LOINC Code	46.0%

Table 4. The number and proportion of laboratory results in the DoD CDR with LOINC codes.

Further analysis showed that almost 84% of all DoD laboratory results belong in five LOINC classes, accounting for only 2,739 LOINC codes, less than 25% of the 11,010 LOINC codes used by the DoD and stored in the CDR (see Table 5). The five LOINC classes are hematology/blood counts, chemistry, urinalysis, drug and toxicology, and serology.

LOINC Class	# of Lab Results in CDR	# of Unique LOINC Codes
HEM/BC	93,841,611	359
CHEM	85,024,991	1,477
UA	30,971,309	104
DRUG/TOX	1,384,939	556
SERO	469,528	243
Total	211,692,378 (83.8% of all lab results in CDR)	2,739

Table 5. The number and proportion of laboratory results in the DoD CDR with LOINC codes.

Discussion

A standard vocabulary may not provide all the codes that correspond to the entire set of data in use at a health care organization. This is particularly true of standard vocabularies that are built via voluntary submission from participating organizations over time (e.g. LOINC). In certain use cases, mandating the use of only the codes from a standard coding scheme may be acceptable (e.g. ICD9CM for reimbursement). In other situations, particularly where clinical care or workflow is concerned, it is important to capture the data accurately according to what really occurred.

The LOINC database for laboratory results was started with the master files from seven US laboratories. It was first released in April 1995 with approximately 6,500 codes, and has since grown through submission from laboratories, hospitals, and other organizations, including 3M. The December 2004 release (version 2.14) contains over 34,000 active codes, of which nearly 28,000 are laboratory codes. In comparison, the October 2003 release contains approximately 20,000 laboratory codes.

LOINC is released periodically. The latest release is June 2005. In 2003 it was released in May and October; in 2002, January, February, August and September; in 2001, January and July; in 2000, February and June. If LOINC codes are used to code

data directly to be stored in the CDR, the clinician would be restricted to ordering only those laboratory tests that have associated LOINC codes at the time. This could have a significant impact upon clinical practice and workflow.

Therefore, the LOINC committee recommends that LOINC codes should be recorded “as attributes of existing test/observation master files” for use in the appropriate message segments to communicate among systems [3]. Internally, each organization needs to use its existing test/observation master files for the necessary laboratory and clinical functions. Many of these internal codes are for organization specific operations or functions and are unlikely to encode data that needs to be sent externally. An example would be the codes for internal quality assurance in the laboratory.

Many laboratory observations would never receive a LOINC code. Reasons include reporting a panel of multiple results through a single result field, or an interpretive data field containing only theoretical information. Some laboratory observations are used for internal system processes, e.g., “DoD DNA Samples”. Yet others have attributes that are not compliant with LOINC definitions or rules, e.g., “ABO Group, Serum or Plasma, Qualitative”. All examples are characteristic of laboratory results used by the DoD. The last one, for instance, is used by over a third of the DoD host systems. The situation where laboratory observations do not have a LOINC code is not unique to the DoD. For instance, the last example is used by commercial health care organizations as well.

For the HDD laboratory results, the commonest reason a LOINC code would not be assigned is that the laboratory result is an interpretation or comment that are meant for text blobs of data or that would not hold the actual result value. Some examples include:

- Autoantibody Profile, Serum Qualitative
- 17-Ketosteroids Fractionated, Urine Qualitative
- Cladosporium Herbarum Ab Interpretation, Serum Qualitative
- Assay Reporting Limits, Other Specimen, Qualitative

There are nearly 2,000 laboratory results of this nature in the HDD.

Our experience supports the LOINC committee’s recommendation. Because it is HDD NCIDs, not

LOINC codes, that are used to encode data in the CDR, there is no obstacle to operation or workflow if a laboratory result does not have a LOINC code. DoD-wide data standardization and interoperability have been accomplished through mapping to NCIDs. A small proportion of laboratory results stored in the DoD CDR does not currently have a LOINC code and thus cannot be translated to LOINC for external data exchange. Some of these laboratory results may be for DoD's internal use only and thus would never require a LOINC code for external data exchange. Others may be valid for external data exchange, but are not commonly used, and thus not yet submitted to LOINC for code assignment. For those laboratory results, we would submit them to LOINC. With each LOINC release the HDD is updated by assigning (mapping) the new LOINC codes to their corresponding, existing NCIDs. There is, therefore, no delay in using the laboratory result at the DoD facility or any interruption to workflow. Since it is the same NCID that is stored whether there is a LOINC code or not, there is no need to do any update or transformation to the data in the CDR.

An alternative LOINC mapping strategy that has been heard in discussions with other organizations is to force the selection of a LOINC code, presumably the closest match. This ensures that every laboratory result would have a LOINC code, even though it may not be the exactly correct match. The risk of this approach is that imprecise information could be captured in the CDR, particularly if different sites picked a different LOINC code for the same item.

One of the cardinal rules of HDD mapping is to be as precise and detailed as the data is. In other words, two items are mapped only if they are identical in every respect. This ensures that no information is lost through mapping, and that the concept is correctly defined. The HDD is designed for "graceful evolution" [4]. If two concepts were later found to be duplicates, one can be inactivated and superceded by the other, with no deletion of NCIDs or change in meaning. The two duplicate concepts and their NCIDs are in fact linked in the HDD, and programs that interact with the HDD will treat them as one. In other words, there is no recoding of data or any work required on the part of the CDR. On the other hand, if two concepts were mistakenly thought to be the same and mapped as one in the HDD, once data has been encoded to a single NCID, the work to "split apart" and recode the data to two different NCIDs would have to be done in the CDR.

Given our HDD design, it is therefore risk-free to create a laboratory result concept with the precise attributes according to the LOINC definition rules, if an exact match is not yet in LOINC, and submit the item to LOINC for code assignment. If LOINC determines that the submission is a duplicate of an existing LOINC code, which means it is a duplicate of a concept already in the HDD, the "newer" concept without the LOINC code can simply be inactivated and superceded by the "older" concept already with the LOINC code.

Conclusion

A high proportion, nearly 90%, of the laboratory results stored in the DoD CDR can be translated to LOINC code for data exchange with the VA. Of the remainder, further analysis would be needed to understand if they are needed for external exchange and why they do not currently have a LOINC code. We plan to review the laboratory results stored in the DoD CDR that currently do not have a LOINC code, starting from the highest count, and prepare them for submission to LOINC if appropriate. We also plan to review the attribute representations and relationships maintained by our laboratory result knowledge base, used for automated mapping to LOINC, to ensure that duplicates are found, inactivated and superceded. This would also enhance the capture rate of our automated laboratory result mapping.

Acknowledgement

We would like to thank LTC (ret.) Joel D. Bales, MS MHA, of Integic Corporation for providing the laboratory LOINC query results for the DoD CDR.

References

1. Rocha R, et al. Designing a Controlled Medical Vocabulary Server: The VOSER Project. *Computers and Biomedical Research*, 27(6): 472-507, 1994.
2. Lau L, et al. A Method for the Automated Mapping of Laboratory Results to LOINC. *JAMIA*, Proceedings Supplement 2000.
3. McDonald C, et al. LOINC, a Universal Standard for Identifying Laboratory Observations: A 5-Year Update. *Clin Chem*, 49(4), 624-633.
4. Cimino J. Desiderata for Controlled Medical vocabularies in the Twenty-First Century. *Proceedings of the IMIA WG6 Conference on Natural Language and Medical Concept Representation*. 1997; 257-267.