The Practice of Informatics

JAMIA

White Paper ■

Evaluation of Vocabularies for Electronic Laboratory Reporting to Public Health Agencies

Mark D. White, Linda M. Kolar, DVM, MPH, Steven J. Steindel, PhD

Abstract Clinical laboratories and clinicians transmit certain laboratory test results to public health agencies as required by state or local law. Most of these surveillance data are currently received by conventional mail or facsimile transmission. The Centers for Disease Control and Prevention (CDC), Council of State and Territorial Epidemiologists, and Association of Public Health Laboratories are preparing to implement surveillance systems that will use existing laboratory information systems to transmit electronic laboratory results to appropriate public health agencies. The authors anticipate that this will improve the reporting efficiency for these laboratories, reduce manual data entry, and greatly increase the timeliness and utility of the data. The vocabulary and messaging standards used should encourage participation in these new electronic reporting systems by minimizing the cost and inconvenience to laboratories while providing for accurate and complete communication of needed data. This article describes public health data requirements and the influence of vocabulary and messaging standards on implementation.

■ **JAMIA.** 1999;6:185–194.

Widespread use of clinical laboratory information systems (LISs) and development of electronic data interchange standards provide the nation's public health agencies with the opportunity to supplement or replace current mechanisms for reporting data[1] with far more efficient and timely electronic systems. The current paper systems are inefficient for participating laboratories and limit use of these data by public health agencies. In particular, minimizing the delay between completion of reportable laboratory test results and transmission of these data to public health agencies would improve our ability to rapidly detect and identify events of public health significance, including outbreaks of infectious diseases and changes in antimicrobial resistance patterns. We discuss here one of the key steps toward implementing electronic reporting systems: the evaluation of common nomenclatures and data transmission standards.

These standards must be sufficiently flexible and comprehensive to accommodate the broad and varied reporting needs of both the laboratory community and

Affiliation of the authors: Centers for Disease Control and Prevention, Atlanta, Georgia.

This research was supported in part by an appointment to the Public Health Informatics Fellowship Program at the Centers for Disease Control and Prevention (CDC), administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and CDC.

Correspondence and reprints: Steven J. Steindel, PhD, Centers for Disease Control and Prevention, 4770 Buford Highway, NE (MS G-23), Atlanta, GA 30341. e-mail: ⟨sns6@cdc.gov⟩.

Received for publication: 8/11/98; accepted for publication: 1/12/99.

public health. The types of health care facilities that transmit laboratory data are diverse, ranging from small hospitals and public health clinics to large national reference laboratories. The business needs of these facilities and the information systems that support them are also diverse. To permit nationwide participation, it will be necessary to select one or more vocabularies that enable most reporting laboratories to accurately translate data from their existing LIS.

The basic data requirements for these reportable laboratory test results include such items as test or procedure identifier, specimen type, anatomic collection site (when relevant), reference range, unit of measure, result, and patient demographics. Demographic information is of particular importance for positive results associated with infectious organisms, environmental agents, or patients with repetitive events. In such cases, both demographic and clinical information is often needed in order to understand the source of the problem and formulate strategies for preventing further transmission or exposure. These data must be timely, accurate, and accessible to meet the needs of the agencies responsible for monitoring and protecting the public's health. Identifying infectious disease outbreaks and environmental hazards are but two examples of public health activities that rely on the timeliness, quality, and accessibility of surveillance data.

Effective surveillance often relies on both the test results reported by the laboratory and clinical data provided by the physician. Public health agencies do not all require the same level of detail from the information derived from surveillance data (e.g., local investigation of an infectious disease outbreak vs. national breast cancer rates). For accommodating the data needs of the various public health agencies conducting disease surveillance, a vocabulary capable of describing concepts in varying levels of detail is required. Ideally, the subset of terms selected for reporting laboratory results should be part of a more comprehensive multipurpose vocabulary that is also capable of describing clinical data. This would permit clinical data from physicians to be included in reports without requiring health care facilities to support yet another specialized vocabulary.

This paper provides a brief overview of public health needs and data security concerns, followed by more in-depth discussions of vocabulary characteristics and their potential impact on the ability of laboratories to implement electronic reporting of test results to public health agencies. The choice of a particular vocabulary may profoundly affect a laboratory's ability to participate in this type of reporting and influence the laboratory's accuracy in mapping terms. Representation of concepts, concept identifiers, and vocabulary evolution are addressed with regard to the needs of public health agencies. Also included is a preliminary assessment of the ability of U.S. LIS vendors to comply with the current recommendations of the Centers for Disease Control and Prevention (CDC) for electronic laboratory reporting.

Public Health Reporting Characteristics

Public health reporting needs represent both a subset and an extension of requirements for the computerized patient records. They include a variety of infectious diseases, chronic diseases such as lung cancer, and environmentally induced conditions such as lead toxicity. In many cases, the reports may contain epidemiologic information such as risk factors (e.g., smoking history for lung cancer or age of housing for lead reporting).

Public health agencies receive reports from both clinical and nonclinical sources. Information on trauma or accidents may be received from an emergency department, police department, or workplace. Other reports may provide data on the quality of the water supply from the local (public or private) water treatment plant or beach. Information reported on animal or insect vectors can be used to trace the sources of infectious disease outbreaks. Results from tests of soil, air, or building samples are used to help locate the source of environmentally induced outbreaks.

Describing public health concepts through defined vocabularies requires that the above examples, and many similar ones, be considered. Fortunately, the actual information that the vocabularies need to express for electronic laboratory reporting to public health agencies is but a small subset of the information that would be required for a computerized patient record. As we shall describe, information models for reporting public health data are being adapted to meet electronic reporting requirements. In choosing vocabularies for public health reporting, we must consider the diversity of data required and the sometimes rapidly changing informational content of the data. Within this construct, public health agencies need to develop a system that is compatible with the information systems found in the private sector, from which much of the data will be derived.

Security Concerns

The CDC recognizes that the issue of data and transmission security is paramount to the implementation of any electronic public health reporting system and is developing both methods and procedures to ad-

dress this issue. Programs used by CDC have transmitted sensitive information electronically using encryption and dial-up modems for many years. Two examples of such programs are the Laboratory Information Tracking System (National Center for Infectious Diseases) and the HIV/AIDS Reporting System (National Center for HIV, STD and TB Prevention), both widely used by public health departments. Recently implemented was a policy document applying to all programs sponsored by CDC that, in principle, allows sensitive data transmission via Internet using an appropriate encryption method of at least 128 bits, message authentication, and message nonrepudiation (CDC Internet Standard 98.1). In process is the establishment of a secure data facility consisting of a certification server and a public/private key encryption system with digital signatures based on the X.509 standard (ITU-T X.509, version 3). A similar system is used by New York State for the Internet transmission of sensitive data (Ivan Gotham, NY State Health Department, private communication). Lastly, we are engaged in extensive policy discussions with our public health partners concerning the decisions and agreements that must be in place before any transmission system can be widely implemented. Further discussion of this important issue is beyond the scope of this article.[2]

Vocabulary Characteristics

The concepts that a laboratory system can express and report are controlled not only by the richness of a vocabulary's content but also by its structure. The manner in which a vocabulary combines concepts, discriminates between terms, and evolves will affect the clarity, accuracy, flexibility, and level of detail with which it may express the important clinical and laboratory data needed for public health surveillance.

Representation of Concepts

A vocabulary term can represent a single atomic concept or an aggregate concept. For example, ***Mycobacterium tuberculosis*** is a single atomic concept, whereas ***Mycobacterium tuberculosis*** detected in **sputum** by **DNA probe** is an aggregate concept consisting of three atomic concepts (organism + specimen type + testing methodology). Precoordinated vocabularies assign unique concept identifiers (codes) to predefined aggregate concepts (terms) and can therefore convey complex information without ambiguity.[3] This eliminates nonsensical or undesirable combinations of individual concepts, since the aggregate concepts are precoded in the only manner permitted. For example,

Table 1 ■

Detail of Concepts Embedded in LOINC Term 5289-4

Concept/ Dimension	Value	Meaning
Component	Reagin AB Antibody	Nontreponemal screening test for syphilis
Property of measurement	ACNC	Analyte concentration
Time aspect	PT	Point in time
System	CSF Cerebrospinal Fluid	Cerebrospinal fluid
Scale type	QN	Quantitative
Method type	FLOC	Flocculation

the contemporary precoordinated vocabulary known as the Logical Observation Identifier Names and Codes (LOINC) uses the concept identifier 5289-4 to represent an aggregate concept (term) that consists of six individual concepts (Table 1). This fully specified term explicitly describes a reagin antibody analyte concentration conducted on a cerebrospinal fluid specimen collected at a single point in time and measured quantitatively by a flocculation method. Such precoordinated vocabularies are an unambiguous and precise means of representing detailed information as a simple code. This clarity, however, does not come without tradeoffs.

In practice these terms are so highly specific and complex that laboratory staff may not have sufficient expertise or knowledge to successfully translate tests to LOINC.[4] Where such expertise does exist, disagreements among local experts can still introduce variance in how identical tests are coded at different institutions. Delays in development or distribution of an updated version of the vocabulary may result in some laboratories temporarily reverting to paper reporting systems or mapping new aggregate concepts to existing but inaccurate terms. Public health agencies often need to aggregate the data received from reporting laboratories and will need laboratories to code reportable information accurately and consistently. If some laboratories map to the fully specified codes while others map using a "best fit" technique, then the data cannot be combined or compared and lose much of their utility.

Precoordination of fully specified terms also requires that assumptions be made about health care business practices and data structures. In facilities that deviate from the assumed model, it may be difficult or impossible to map the individual concepts embedded in the aggregate concept. Continuing with the example above, laboratory systems may not define each test in

terms of mass concentration (MCNC), substance concentration (SCNC), or analyte concentration (ACNC). Likewise, many LISs may not be able to discriminate on the basis of scale type, methodology, or time. Some may not even store specimen type in the LIS, relying instead on the specimen type identified in the collection order. In such cases, substantial effort and resources may be required to add the new fields and data necessary to accurately map all the individual concepts contained in the precoordinated terms. Developing and maintaining complex cross-references may also be required. As a result, the reporting facility could conclude that it is more accurate or cost effective to maintain the current paper reporting system.

Since terms must be provided for all the potentially useful combinations of individual concepts, these concepts occur redundantly throughout the vocabulary. Precoordinated vocabularies are therefore large relative to the number of unique concepts that they describe. If such a vocabulary contains 250 terms representing all the potentially useful combinations of test name and sample type for a specific testing methodology, and if a new or revised testing methodology is subsequently developed, it may be necessary to create as many as 250 new terms. In addition, the original terms must be maintained to preserve the integrity of previously coded data, permit laboratories a gradual transition to the new methodology, and accommodate those who continue using the old method. It is uncertain how well a precoordinated vocabulary can handle the inevitable combinatorial explosion of tests, methodologies, specimens, and results.

Postcoordinated vocabularies, in contrast, assign codes to individual concepts that the laboratory system can use to compose any number of aggregate concepts.[5] For example, six individual concept codes could be combined to compose an aggregate concept equivalent to the fully specified term from the example given in Table 1 (LOINC term 5289-4). Such a vocabulary is generally more flexible in representing new or revised concepts with existing codes. It is also less redundant and thus potentially smaller. Laboratory systems need only map to the individual concepts that fit their business practices and data structures. The Systematized Nomenclature of Human and Veterinary Medicine (SNOMED)[6] is an example of a contemporary postcoordinated vocabulary. When new aggregate concepts are required, the laboratory system may often be able to combine existing individual concept codes to compose the new terms. Similarly, developing a new or revised testing methodology, as in the example above, would require adding a single code that would then be combined with existing test name codes and specimen codes to represent the new term, rather than adding 250 new terms. Since terms consist of discrete (individually) coded concepts, it is also a relatively simple matter to parse them into their constituent concepts. This in turn provides greater flexibility for sorting and querying the data in ways that are meaningful for the various public health agencies. While precoordinated terms can also be disassembled, the resulting individual concepts may not translate to those commonly used in other vocabularies and may not be suitable for complex queries.

Vast expressive power, concept richness, and flexibility also do not come without tradeoffs. Since postcoordinated vocabularies often allow aggregate concepts to be coded in more than one way, it may be necessary to develop guidelines that identify preferred combinations and discourage the use of undesirable or nonsensical terms. Several agencies currently meet on a regular basis to discuss common concepts, methods, and terminology for use in public health (Table 2). These agencies could be asked to develop and distribute appropriate guidelines for creating new public health terms from existing concepts. Until such compositional guidelines can be developed and distributed, public health agencies must carefully weigh the expressive nature of postcoordinated vocabularies against the potential for users to transmit undesirable combinations of concepts.

In summary, the unambiguous nature of terms in a precoordinated vocabulary offers the clarity and precision of reporting that public health agencies need to aggregate and analyze the data, but such terms may be difficult or impossible for laboratories to integrate with their current information systems and are more likely to be negatively affected by the inevitable combinatorial explosion of terms. The vast expressive

Table 2 ■

Potential Participants in a Public Health Vocabulary Consensus Organization

Organization (Acronym)	Internet Address
Association of State and Territorial Health Officials (ASTHO)	www.astho.org
Association of Public Health Laboratories (APHL)	www.aphl.org
Centers for Disease Control and Prevention (CDC)	www.cdc.gov
Council of State and Territorial Epidemiologists (CSTE)	www.cste.org
National Association of County and City Health Officials (NACCHO)	www.naccho.org
National Association for Public Health Statistics and Information Systems (NAPHSIS)	www.naphsis.org

power, concept richness, and flexibility of a postcoordinated vocabulary seem better suited to public health surveillance requirements and the diverse information system capabilities of laboratories, but the postcoordinated vocabulary may require additional guidelines for the composition of appropriate and consistent terms.

Hierarchic versus Context-free Concept Identifiers

Hierarchic concept identifiers are codes that indicate the ordered position of a concept in the vocabulary.[7] For example, the SNOMED hierarchic identifier for *Escherichia coli*, serotype O157:H7, is L-15611. The ''L'' in the concept identifier indicates that the code represents a living organism. Concept identifiers beginning with ''L1'' describe Bacteria and *Rickettsiae*. Entries beginning with ''L15'' are *Enterobacteriaceae*. Entries beginning with ''L156'' describe members of the genus *Escherichia*. Thus, ''L-15601'' identifies *Escherichia coli* and ''L-15611'' identifies *Escherichia coli*, serotype O157:H7. This offers users a means of understanding the relationships and differences between concepts and can enhance the ability to map terms accurately. A disadvantage of vocabularies using these hierarchic identifiers is that often only a finite number of terms can be added within each level of the hierarchy, and so expansion is limited.[8] Reclassification of coded concepts can also be problematic, since it may require changes to both the concept and its identifier.

Vocabularies that use context-free concept identifiers are preferred to those that use hierarchic concept identifiers, because context-free identifiers do not restrict the number of terms that can be added and permit additional flexibility in reclassifying terms.[9]

Vocabularies that separate the hierarchic structure from the concept identifier offer public health and laboratory communities the best alternative. These vocabularies still assist users in accurately discriminating between concepts while using context-free identifiers that do not restrict the addition and reclassification of terms.

Vocabulary Evolution[10]

A controlled vocabulary used for reporting laboratory results to public health agencies must be able to maintain order and integrity as it evolves, i.e., the rules governing change must be applied in a consistent manner and the vocabulary must retain its compatibility with previous versions. Many changes will occur in the identification and description of etiologic agents and in the public health requirements for reporting of disease. Our expanding knowledge base will cause researchers to reorder classification schemes, rename known agents, and allow them to identify agents of both new diseases and diseases currently described as being ''of unknown origin.'' Laboratory methods will continue to be developed, refined, and discarded. Public health agencies will revise reportable disease and event lists to reflect these changes and to meet tomorrow's challenges. It is thus vital that the vocabulary used for reporting laboratory test results be capable of evolving at a rate sufficient to meet these needs.

Yet these changes must be made in a careful, well-documented fashion. Users must not only be aware of additions to, deletions from, and name changes in the vocabulary, but also be cognizant of the reasons for the changes and the impact they may have on applying codes in the future and on interpreting old and new data. The creators of the vocabulary should provide both a formal syntax of changes, to convey the surface differences, and a semantics of the changes, to describe how the meaning of a term is or is not altered during the process.[11] This will ensure that the vocabulary evolves in a logical manner and will allow backward and forward compatibility in collected data. In addition, new releases of the vocabulary must be clearly distinguished, so that users can identify which version they are using and can track by date the changes that are made. Without documented evidence of controlled evolution, it will be difficult to combine data from various sources with assurances of compatibility and data comparability.

LOINC was introduced in 1995 and experienced an initial period of rapid growth. Since then updates have been distributed via Internet or diskette about twice a year. LOINC is distributed free of charge and is maintained through grant support. The Regenstrief Institute and the LOINC Committee have indicated that they will maintain the database while grant support is available (at least until October 1999). SNOMED was introduced in 1976 and has been adding new concepts as required and distributing them as part of an annual update on diskette or compact disc. SNOMED is distributed to licensed users and is professionally maintained through license fees.

Implementation

To take advantage of the potential benefits of electronic laboratory reporting, CDC, in consultation with its partners, has elected to evaluate Health Level Seven (HL7), version 2.3,[12] as the messaging standard for pilot testing the transmission of reportable labo-

ratory test results to public health agencies.[13] These pilot studies will assist in identifying the obstacles and problems that must be overcome before widespread deployment. As part of this test implementation, these public health agencies will use the unsolicited transmission of an observation (ORU) transaction set. The HL7 v2.3 documentation strongly encourages the use of universal identifiers in the observation identifier (OBX-3) segment of this transaction set.* It also specifies LOINC as one of the possible universal identifiers that could be used in this segment. Based on this documentation and suggestions from HL7 members, CDC has chosen to evaluate the use of LOINC as a universal identifier in OBX-3 during our HL7 pilot studies. To minimize difficulties in aggregation for observation (result) values that are non-numeric, such as organism names, we will require laboratories participating in the pilot studies to use a coded element (SNOMED codes) rather than text in OBX-5.

CDC is currently considering several HL7-related activities that will allow the agency to evaluate and develop the potential to use LIS to transmit HL7 messages of results of laboratory tests for infectious diseases or detection of incident cancer cases. Such studies will provide valuable information for assessing the future applicability of direct reporting to appropriate public health agencies. Among the proposed activities are a survey of U.S. LIS vendors to determine their ability to implement electronic laboratory reporting of clinical and anatomic laboratory data to public health agencies using the HL7 messaging standards, and development of an implementation specification based on the CDC's HL7 electronic laboratory reporting message recommendations and several pilot projects. These pilot studies will provide an assessment of the effectiveness of both the implementation specification and the transmission system and will assist public health agencies in developing a national electronic laboratory reporting system. To optimize our data collection efforts, CDC has been conducting a nationwide retrospective survey to gather objective information regarding the type and volume of laboratory testing that was performed at representative testing locations in 1996.[14]

The pilot studies will evaluate the ability of laboratory information systems to use fully specified and precoordinated observation identifiers in OBX-3. They will also identify and evaluate the differences between the business practices and data structures of laboratories and the model used to precoordinate the LOINC terms.

Laboratory Information Systems

Large U.S. laboratories often have adequate resources and technical expertise to program their own information systems. These laboratories should be capable of implementing the vocabularies that public health agencies recommend for electronic laboratory reporting. However, many U.S. laboratories rely on LIS vendors to provide the software, hardware, and programming to meet their information system needs, and our recommendation may have a much greater impact on their ability to participate. To gauge the ability of these laboratories to participate in our pilot studies, we conducted unstructured telephone interviews with 11 HL7-capable LIS vendors who agreed to participate in this informal survey. These vendors represent approximately 52 percent of hospital laboratory, 25 percent of independent laboratory, 29 percent of clinic or group practice laboratory, and 56 percent of other LIS installations as reported by 67 LIS vendors in a 1995 review.[15]

These companies were experienced in traditional electronic data interchange relationships, where they define the meaning of every test code with each partner with whom they exchange information. They were generally less familiar with the differences between precoordinated and postcoordinated vocabularies. All the vendors welcomed the concept of a universal test identifier. While 10 of the 11 vendors were aware of the existence of the LOINC codes, only two were aware that each LOINC code incorporates concepts other than test or procedure name. These two were also the only vendors who indicated they currently store the information required to complement all the individual concepts (dimensions) of a fully specified LOINC code (Table 3). Of the remaining vendor systems that included fields that could be mapped to two or more dimensions of the LOINC code, these fields occurred in two or more tables in their systems. Most indicated that fields identifying specimen type are located in separate tables from test identifications or test descriptions. Nine vendors believed that the additional fields and complex cross-references necessary to implement fully specified observation identifiers such as LOINC codes would be both difficult and expensive. These nine vendors also indicated that they could probably provide laboratories with the ability to transmit reportable test results to public health

*HL7 messages consist of variable-length data fields divided by a field separator character. The data fields are combined into logical groupings called ''segments.'' The OBX segment contains information related to observations (such as laboratory test results). OBX-3 contains the observation (test) identifier.

Table 3 ■

LOINC Dimensions Compared with Data Structures of the HL7-capable LIS Vendors Surveyed

	Component (Analyte)	Property of Measurement	Time Aspect	System (Specimen)	Scale Type	Method Type
A	+	−	+	+	±	+
B	+	−	+	+	±	+
C	+	−	±	±	−	−
D	+	+	+	+	+	+
E	+	−	±	+	−	−
F	+	−	+	+	−	−
G	+	−	±	+	−	−
H	+	−	±	−	±	−
I	+	+	+	+	+	+
J	+	−	+	+	+	−
K	+	−	−	+	−	+

NOTE: HL7 indicates Health Level Seven, version 2.3; LIS, laboratory information system. A plus sign (+) indicates that the vendor could identify a field in the data structure that corresponds to this LOINC dimension; a minus sign (−), that the vendor could not identify such a field; a plus-or-minus sign (±), that the vendor was uncertain whether their system could accommodate this dimension (concept).

agencies in the near future if they used a less complex mapping for OBX-3, such as that permitted by a post-coordinated vocabulary or a simplified universal test identifier.

Simplified Observation (Test) Identifiers

In 1997, a group of state epidemiologists and public health officials developed a preliminary table containing approximately 64 reportable entities (infectious organisms and agents) important to public health surveillance and the accepted testing methodologies and procedures used in their identification.[13] Members of the group used a precoordinated vocabulary (LOINC) to provide coded terms for the preferred tests and procedures needed to identify these reportable entities. The group was able to describe most of the 64 reportable entities using 280 LOINC codes. Codes were not yet available for some entities. To permit useful data aggregation and analysis, many of these would require a coded element in OBX-5 for non-numeric results such as the name of the organism identified.

As one means of providing a simplified observation identifier, we attempted to represent the 64 reportable entities, methods, and procedures using existing SNOMED codes (v3.3) from the procedure, living organism, and modifier axes. For this exercise, we chose to use a two-component code, but more complex and specific terms can be composed. This would allow a laboratory system to compose as detailed a term in OBX-3 as their data structures permit without imposing the limitations of fully specified precoordinated terms. We found we could represent most of these concepts with 50 procedure codes, a living organism code for each reportable entity, and 2 modifier codes (121 codes total). Adding approximately a dozen coded concepts (2 living organism codes and 10 procedure codes) would permit all the entities to be identified and would increase the specificity of the coded methodologies. The resulting terms were less detailed than the LOINC codes, containing only information on the test procedure or method and organism identified.[16] However, additional details can be placed elsewhere in the reporting message. For example, the HL7 ORU transaction set contains an observation request (OBR) segment that provides separate fields for the specimen source and anatomic site, and an observation result (OBX) segment that contains fields for units, reference range, and further distinction of the methodology. It should not be necessary to also include this information in the laboratory observation (test) identifier (OBX-3) for public health reporting. Use of a simplified observation identifier in conjunction with the information in these fields can provide all the information required for reportable laboratory test results without the potential mapping limitations that may be associated with a fully specified and precoordinated observation identifier.

The following examples illustrate the differences between using fully specified LOINC codes containing up to six dimensions and simplified two-dimensional SNOMED terms as observation identifiers in OBX-3. When the observation values (results) are non-numeric, such as for organism identification, a SNOMED code is used in OBX-5 to supplement the LOINC code and provide a coded entry for data aggregation and analysis. The SNOMED examples use a two-dimen-

sional observation identifier composed of a procedure code and a living organism code for OBX-3. A living organism code is also used in OBX-5. This may appear redundant in some cases, but where the observation value in OBX-5 is qualitative, semiqualitative, or numeric, it is at times necessary to include the organism name in the observation identifier. For example, when *Yersinia pestis* (the plague) is reported, the preferred antibody test is the enzyme-linked immunosorbant assay (ELISA). A reportable result for this assay would be numeric (≥1:64). In this case we would use the codes P3-70200 | L-1E401 in OBX-3 to represent ELISA for *Yersinia pestis* and place the numeric result (titer) in OBX-5. The following examples utilize the specimen source information coded in the OBR-15 segment, which is required regardless of the vocabulary used in OBX-3.

EXAMPLE 1: REPORTABLE RESULTS DATA TO BE CODED

Rabies virus identified by antibody neutralization in serum or cerebrospinal fluid

Rabies virus identified by direct fluorescent antigen detection in tissue or other (unspecified specimens)

Rabies viral culture in saliva, cerebrospinal fluid, central nervous system, tissue, or other (unspecified) tissue

LOINC can represent this information with eight multidimensional codes (Table 4).

SNOMED can represent this information by combining a procedure code with a living organism code in OBX-3. The specimen source information is obtained from the OBR segment (OBR-15), where it can be coded using the standard HL7 specimen table (0070) or terms from the SNOMED topography table (Table 5).

EXAMPLE 2: DATA TO BE CODED

Microbial culture of *Streptococcus pyogenes* in blood, cerebrospinal fluid, pleural fluid, peritoneal fluid/ascites, wound, or other

LOINC can describe most of these microbial cultures with six multidimensional codes. For representing the organism that was identified (*Streptococcus pyogenes*), a coded element (SNOMED) code is required in OBX-5 to permit data aggregation and analysis (Table 4).

SNOMED can represent the same organism identification by combining a procedure code (Microbial Culture) and a living organism code (*Streptococcus pyogenes*) in OBX-3. The living organism code (*Streptococcus pyogenes*) is also used in OBX-5. The specimen source information is coded in the OBR segment as above (Table 5).

Conclusions

For widespread implementation of electronic laboratory reporting, public health agencies must first ensure that the electronic transmission, storage, and use of this information is at least as confidential and secure as current systems. Next we must ensure the most complete surveillance data possible by selecting coding schemes, vocabularies, and messaging standards that allow reporting laboratories to participate and accurately code their results.

While the HL7 messaging standard does not currently address confidentiality or security issues, it is becoming widely accepted in the health care industry, and public health agencies are investigating its potential for transmitting surveillance data. Projects sponsored by CDC will use encryption, message authentication, and message nonrepudiation via a secure data facility to evaluate the ability to send and receive HL7 messages that meet or exceed current standards for the confidentiality and security of patient information.

The vocabulary selected for encoding laboratory information must also ensure that most reporting laboratories are able to participate and can accurately and efficiently code their results. This vocabulary will need to be unambiguous, expressive, comprehensive, and flexible in accommodating varying health care business practices and data structures. It should also eliminate redundancy, minimize maintenance, simplify mapping, and permit useful data aggregation and analysis.

LOINC has been frequently publicized and recommended by the HL7 community, yet it has not been widely implemented. The CDC is initiating several projects to evaluate the potential to use LOINC to encode test results data from laboratory information systems in various settings. To facilitate useful aggregation and analysis of the data, a second vocabulary or table is required to code non-numeric observation values such as organism names in OBX-5. The CDC will evaluate the potential to utilize SNOMED, which laboratories have used primarily to describe anatomic pathology data, to encode organism names and other non-numeric observation values in OBX-5.

Much of the information precoordinated in each LOINC code may be difficult or impossible to obtain within existing LISs and may be retrieved or inferred from other portions of the HL7 message. If the data structures or business rules of laboratories prevent them from implementing electronic laboratory reporting using LOINC as the observation identifier in OBX-3, alternative coding systems will be necessary. The

Table 4 ■

Examples of LOINC Codes for Reporting in OBX-3

OBX-3 LOINC Code	Concept						OBX-5 SNOMED Code	Concept
	Component	Property of Measurement	Time Aspect	System	Scale Type	Method Type		
Rabies virus:								
6523-5	Rabies virus AB	ACNC	PT	CSF	QN	NEUT	L-33301	Rabies virus
6524-3	Rabies virus AB	ACNC	PT	SER	QN	NEUT	L-33301	Rabies virus
6528-4	Rabies virus AG	ACNC	PT	TISS	ORD	IF	L-33301	Rabies virus
6529-2	Rabies virus AG	ACNC	PT	TISS	QN	IF	L-33301	Rabies virus
6532-6	Rabies virus AG	ACNC	PT	XXX	ORD	IF	L-33301	Rabies virus
6533-4	Rabies virus AG	ACNC	PT	XXX	QN	IF	L-33301	Rabies virus
6536-7	Rabies virus identified	PRID	PT	TISS	NOM	Organism-specific culture	L-33301	Rabies virus
6539-1	Rabies virus identified	PRID	PT	XXX	NOM	Organism-specific culture	L-33301	Rabies virus
Streptococcus pyogenes:								
600-7	Microorganism identified	PRID	PT	BLD	NOM	Blood culture	L-25102	*Streptococcus pyogenes*
606-4	Microorganism identified	PRID	PT	CSF	NOM	Sterile body fluid culture	L-25102	*Streptococcus pyogenes*
618-9	Microorganism identified	PRID	PT	PLR	NOM	Sterile body fluid culture	L-25102	*Streptococcus pyogenes*
619-7	Microorganism identified	PRID	PT	PRT	NOM	Sterile body fluid culture	L-25102	*Streptococcus pyogenes*
363-1	Microorganism identified	PRID	PT	XXX	NOM	Sterile body fluid culture	L-25102	*Streptococcus pyogenes*
6462-6	Microorganism identified	PRID	PT	WND	NOM	Routine bacterial culture	L-25102	*Streptococcus pyogenes*

NOTE: ACNC indicates analyte concentration; PT, point in time; CSF, cerebrospinal fluid; QN, quantitative; NEUT, neutralization; SER, serum; TISS, tissue; ORD, ordinal; IF, immunofluorescence; XXX, system not specified; PRID, presence/identity/existence; NOM, nominal; BLD, blood; PLR, pleural fluid; PRT, peritoneal fluid; WND, wound.

Table 5 ■

Examples of SNOMED Codes for Reporting in OBX-3

OBX-3 SNOMED Code	Concept	OBX-5 SNOMED Code	Concept
Rabies virus:			
P3-50260 | L-33301	Viral culture, NOS | Rabies virus	L-33301	Rabies virus
P3-6A200 | L-33301	Viral neutralization test | Rabies virus	L-33301	Rabies virus
P3-60115 | L-33301	Fluorescent antigen measurement, NOS | Rabies virus	L-33301	Rabies virus
Streptococcus pyogenes:			
P3-50100 | L-25102	Microbial culture, NOS | *Streptococcus pyogenes*	L-25102	*Streptococcus pyogenes*

NOTE: NOS indicates not otherwise specified.

HL7 standard could permit less fully specified observation identifiers in OBX-3, which in turn would permit LOINC to provide a table of precoordinated but simplified observation identifiers. While this would permit those facilities whose data structures or business rules cannot accommodate fully specified precoordinated observation identifiers to participate in public health electronic laboratory reporting, a means of coding non-numeric observation values would still be required. Another potential alternative is to use a postcoordinated vocabulary in OBX-3. This would allow the laboratory to compose as detailed a term in OBX-3 as their data structures permit without imposing the limitations of fully specified and precoordinated terms. Such simplification is OBX-3 could make the use of universal test identifiers genuinely feasible and reduce potential barriers to reporting electronic laboratory results to public health agencies. Some postcoordinated vocabularies such as SNOMED contain concepts that can also be used for observation values in OBX-5 and specimen codes in OBR-15.

References ■

1. Centers for Disease Control and Prevention. Manual of Procedures for the Reporting of Nationally Notifiable Diseases to CDC. Atlanta, GA: CDC, 1995.
2. Gostin L, Lazzarini Z, Neslund V, Osterholm M. The public health information infrastructure: a national review of the law on health information privacy. JAMA. 1996;24:1921–48.
3. Arden WF, McDonald CJ, Demoor G, et al. Logical observation identifier names and codes (LOINC) database: a public use of set codes and names for electronic reporting of clinical laboratory test results: Clin Chem. 1996;42(1):81–90.
4. Baorto DM, Cimino JJ, Parvin CA, Kahn MG. Using Logical Observation Identifier Names and Codes (LOINC) to exchange laboratory data among three academic hospitals: Proc AMIA Annu Fall Symp. 1997:96–100.
5. Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J. Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions and clarity. J Am Med Inform Assoc. 1997;4(3):238–51.
6. Rothwell DJ, Côté RA, Cordeau JP, Boisvert MA. Developing a standard data structure for medical language: the SNOMED proposal. Proc Annu Symp Comput Appl Med Care. 1993:695–9.
7. Rothwell DJ, Côté RA. Managing information with SNOMED: understanding the model. Proc AMIA Annu Fall Symp. 1996:80–3.
8. Schulz EB, Price C, Brown PJ. Symbolic anatomic knowledge representation in the Read Codes version 3: structure and application. J Am Med Inform Assoc. 1997;4:38–48.
9. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med. 1998;37:394–403.
10. Cimino JJ. Formal descriptions and adaptive mechanisms for changes in controlled medical vocabularies. Methods Inf Med. 1996;35:202–10.
11. Cimino JJ, Clayton PD. Coping with changing controlled vocabularies. Proc Annu Symp Comput Appl Med Care. 1994:135–9.
12. Health Level Seven. An Application Protocol for Electronic Data Exchange in Healthcare Environments, version 2.3. Ann Arbor, Mich.: Health Level Seven, 1997.
13. Centers for Disease Conrol and Prevention. Electronic Reporting of Laboratory Data for Public Health: Meeting Report and Recommendations. Atlanta, GA: CDC, 1997. Available at: http://www.cdc.gov/phppo/dls/guidstd.htm. Accessed Mar 5, 1999
14. Centers for Disease Control and Prevention. National Inventory of Clinical Laboratory Testing Services, Final Report. Atlanta, Ga.: CDC, 1999. CDC contract 200-95-0933.
15. Aller R, Weilbert M, Carey K. Some LIS capabilities worth searching for. CAP Today. 1995;9(11):41–58.
16. Rocha RA, Huff SM. Coupling vocabularies and data structures: lessons from LOINC. Proc AMIA Annu Fall Symp. 1996:90–4.