Chapter 8
Public Health Information Standards

J.A. Magnuson, Riki Merrick, and James T. Case

Abstract Standards are one of the most efficient ways to prevent data silos, achieve system interoperability, and promote the value of data. Public health's growing use of electronic data interchange lends increasing urgency to the need to adopt and promote standards, and to participate in standards development as a fully-engaged partner.

However, public health and its many partners must agree upon both the selection and value of standards in order to overcome the significant barriers and challenges to standards adoption. Implementation of standards is complex and resource-intensive, sometimes unevenly more so for one of the partners involved in data interchange.

In this chapter, standards are categorized into process standards and data or content standards. After reviewing a number of the most common standards utilized in public health, we focus in more depth upon three of the most important – HL7®, LOINC®, and SNOMED CT®.

Keywords Electronic data interchange • Clinical decision support • Clinical Document Architecture • Code systems • Confidentiality • Continuity of Care

J.A. Magnuson, PhD (✉)
Department of Medical Informatics & Clinical Epidemiology,
Oregon Health & Science University, 5th Floor, Biomedical
Information Communication Center (BICC), 3181 S.W. Sam Jackson Park Rd.,
Portland, OR 97239, USA
e-mail: jamagnuson@gmail.com

R. Merrick, MPH
Consultant, Public Health Informatics, 2541, Carmichael, CA 95609, USA
e-mail: rikimerrick@ucdavis-alumni.com

J.T. Case, MS, DVM, PhD
California Animal Health and Food Safety Laboratory,
University of California, Davis,
1 Shields Ave, Davis, CA 95616, USA
e-mail: drjtcase@sbcglobal.net

Document • Data standards • Electronic laboratory reporting • Fully specified name • Functional interoperability • Granular • Implementation guide • Integrity • Interoperability • Laboratory Information Management System • Laboratory Information System • Optional • Privacy • Process standards • Required • Semantic interoperability • Standards development organization • Structure • Syntax • Vocabulary • HL7® • LOINC® • SNOMED CT®

Learning Objectives
1. Learn about standards categories, including the two main categories of standards used in this chapter, process standards and data or content standards.
2. Evaluate the anticipated benefits of standards use and the obstacles to adoption of standards.
3. Identify some of the main standards used in public health.
4. Describe the general process for standards development.
5. Review details about three of the commonly-used standards in public health, HL7®, LOINC®, and SNOMED CT®.

Overview

Standards are one of the most efficient ways to prevent data silos, achieve system interoperability, and promote the value of data. Public health's growing use of electronic data interchange lends increasing urgency to the need to adopt and promote standards, and to participate in standards development as a fully-engaged partner.

However, public health and its many partners must agree upon both the selection and value of standards in order to overcome the significant barriers and challenges to standards adoption. Implementation of standards is complex and resource-intensive, sometimes unevenly more so for one of the partners involved in data interchange.

In this chapter, standards are categorized into process standards and data or content standards. After reviewing a number of the most common standards utilized in public health, we focus in more depth upon three of the most important – HL7®, LOINC®, and SNOMED CT®.

Introduction

With a little imagination, one can picture many systems that must communicate over distances, "speak" different languages, and coordinate time-sensitive materials and actions, and that are often critical to the health and safety of individuals or

populations; examples might include systems for air traffic controllers, police, and hospitals. Public health systems may not seem as obvious a choice, but they also fit into this category.

As public health continues its enthusiastic rush into the arena of *electronic data interchange* (*EDI*), interoperability, or the capacity to exchange and utilize data between systems, becomes increasingly critical. Examples of EDI in public health are many and varied, such as:

- *Communicable disease reporting* from laboratories (Electronic Laboratory Reporting or ELR)
- *Wide-ranging surveillance* of sources such as emergency department (ED), emergency medical services (EMS), pharmacy, over-the-counter (OTC), poison control, and absenteeism data
- *Meaningful Use* (*MU*) *objectives* (identified by the Centers for Medicare & Medicaid Services as part of Incentive Programs to promote the adoption of Electronic Health Record Systems), including public health choices for immunizations, syndromic surveillance, and ELR for Stage 1 MU [1].
- *Data sharing* within-state and between state partners such as other states, local or regional health departments, and federal agencies

Jernigan et al. [2] list three main causes of non-communicating or silo-ed public health systems:

- *Functional requirements*: design differences may be based on function, for example the function of case management vs. the function of population surveillance.
- *Policy requirements*: policy restrictions upon systems, such as those that could restrict choices of software
- *External restrictions* imposed by federal funding. Many of the silo-ed systems in wide use today are actually required. One such current example is the Enhanced HIV/AIDS Reporting System (eHARS), a browser-based HIV surveillance system used by state and local health departments to submit de-identified data electronically to the Centers for Disease Control and Prevention (CDC) national database [3].

In this chapter, we are adding two additional causes of public health silos, both closely tied to standards:

- *System architecture* or more specifically, lack of system architecture. If the importance of system integration and architecture is either unknown or discounted when building a new system, then the outcome will be an isolated, non-integrated system. Harmonization, or at least accommodation, of standards is a critical factor in system integration.
- *Exchange partner variations*. There is a wide variety of exchange partners inherent in public health matters, and the corresponding barrier of asking all these partners to agree upon and incorporate any chosen standard can be formidable. Public health exchange partners include local, regional, state, and federal public health agencies; the public and its personal health records; laboratories;

hospitals; and other data generating entities. In the future, these partners should expand to include new data *receiving* entities, such as bi-directional exchange with laboratories or with clinical Electronic Health Record systems, as public health becomes more adept at sharing its wealth of data and information.

The Value of Standards

One of the most efficient ways to prevent data silos, achieve system interoperability, and promote the value of data is through the utilization of standards. Establishing and gaining consensus for standards is not an easy task, however, and to date public health has lagged industry (though not healthcare in general) in agreeing upon and utilizing standards. But to keep perspective on the difficulty of such an endeavor, we need only consider that although the metric system was introduced in France in 1799, the United States is today the only industrialized country that does not utilize it as its official standard of measurement (the metric system is certainly accepted in the US, but it is not yet the official standard). In a nutshell – standards are hard.

Obstacles to Adoption

In order for a standard to be both useful and accepted by the community, there must be agreement among the affected industries or groups on the goals to be accomplished through adoption of standards. This agreement may be a challenging objective in itself, especially if the industries and groups are fragmented. Additionally, even the experts often disagree on details. The difficulties and costs inherent in the implementation of standards within any organization must be justified by stated objectives for the exchange of data or the utilization of aggregate data from multiple institutions.

Ideally, standards are developed by a panel of experts and formally approved by a *standards development organization* (*SDO*) such as the International Organization for Standardization (ISO) [4] or the American National Standards Institute (ANSI) [5]. In practice, many "standards" are the product of legacy use within an industry or group. Such *de facto* standards can be extremely useful when no formal standards are available.

The process of developing a standard differs somewhat between SDOs, but there remains a basic similarity, illustrated here by the ISO process. The International Organization for Standardization follows a six-step process [6] when developing a standard. The process begins with (step 1) a proposal to the appropriate technical committee (TC), and then (step 2) a working draft is developed by a group of experts and (step 3) shared with the TC. Next, the draft is (step 4) released for comment by all ISO national members and (step 5) the final draft, after reconciliation of comments, is sent to all ISO members for a vote. If approved by the vote, the draft finally becomes (step 6) an official ISO International Standard.

Agreeing upon a standard is only the beginning. Despite the critical advantages and benefits of standards, it is important to remember that not all partners in a data exchange may share equally in those benefits. The implementation and utilization of standards is often resource-intensive, and many times the essential costs of standards implementations are borne by partners who may not share in the benefits. A good illustration of this situation involves *electronic laboratory reporting* or *ELR*. Around 2000, public health began asking the laboratories legally required to send reportable condition data to public health (including hospital, private, and public health laboratories) to report electronically through the new institution of ELR. Implementation of ELR systems creates a potential for faster disease reporting [7]. This new reporting path, however, requires that the data submitted be standardized in both format and content. These new requirements for standardization often create significant expense for laboratories, especially those with multi-jurisdictional clients requiring multi-jurisdictional reporting. The laboratory must be able to retrieve, format, and transmit data from the *Laboratory Information System* (*LIS*) or *Laboratory Information Management System* (*LIMS*, originally signaling industrial settings, though that distinction is fading and the terms are becoming interchangeable), as well as apply standard codes for laboratory tests and results. This resulting standardization is highly valuable to public health, but confers little practical return on investment for the submitting laboratory. Some funding has been made available to assist laboratories, either directly by public health entities or tangentially by measures such as the Meaningful Use incentives. Nevertheless, in most cases the efforts have been funded largely by the laboratories themselves.

As demonstrated in the previous example, decisions to develop, select, implement, or require standards should not be reached without careful consideration. Figure 8.1 illustrates a decision process flow that represents effective contemplation of such standards issues.
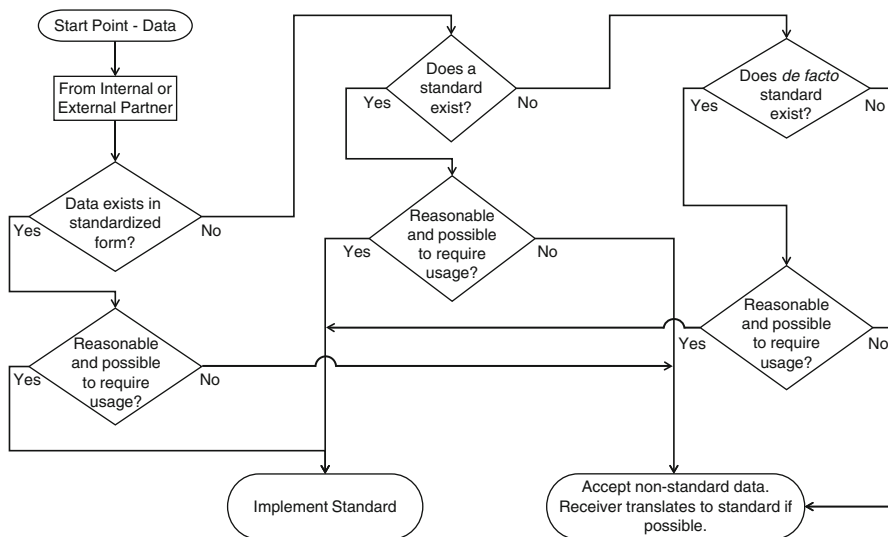

Fig. 8.1 Sample decision process flow for standards utilization

Standards Categories

Categorization of public health standards can help to simplify the subject, but there are a number of different categorical schemas from which to choose. For example, in February 2006, the Health Information Technology Standards Panel (HITSP) [8] separated health information technology standards into the categories and corresponding examples shown in Table 8.1.

For this chapter, we will use a slightly different categorization, and divide public health informatics standards into two fundamental categories, *process* standards and *data* or *content* standards. *Process standards* include procedure and policy standards. Examples of process standards include security policies, data use agreements, workflow, architectural, and metadata standards (creating some overlap with data standards). *Data* (*or content*) *standards* address common terms and methods, and increase the ability to share data between systems, i.e., interoperability and integration. The theoretical components of data standards are (a) vocabulary, (b) format, and (c) transmission. Transmission standards include privacy and confidentiality components, and so tend to overlap somewhat with process standards.

Process Standards

As mentioned earlier, transmission standards and process standards have some degree of overlap in the areas of privacy and confidentiality. *Privacy* generally refers to a 'people' context, a state of being free from unauthorized intrusion or invasion. This concept is as applicable to medical records as it is to your own house. *Confidentiality* is viewed more in the context of information, usually dealing with accessing and sharing information or data.

Security Policies

Data *integrity* (freedom from errors or flaws) and confidentiality are often the prime focus of security concerns. Data integrity must be maintained during any

Table 8.1 One example of Public Health Standards Categorization, based on work done by Public Health Data Standards Consortium [9]

Standards categories	Examples
Data standards	Vocabularies and terminologies
Information content standards	Reference information models (RIM)
Information exchange standards	Message-based and structured document-based
Identifier standards	Identifiers, such as the National Provider Identifier (NPI) [10]
Privacy and security standards	Access control, audit, electronic consent
Functional standards	Work processes, workflow and dataflow models
Other standards	Internet standards, transport mechanisms

transaction; for example, when reporting data, the data received must be exactly the same as the data that were sent. Data confidentiality is a critical concern in public health, which frequently requires the exchange of clinical or laboratory data containing patient identifiable information. Almost all public health agencies are concerned with confidentiality, since they routinely deal with sensitive data that are their legal responsibility to safeguard. A breach in security that allowed patient identifiable data to be made public would jeopardize the ability of a public health agency to perform its data gathering duties, as well as damaging its public reputation as a trustworthy government agency.

The confidentiality of an institution's data depends in large part upon enterprise security – the administrative, physical, and technical security measures enacted by the institution to safeguard its systems. Physical security measures (e.g., locked doors and security patrols), administrative measures (e.g., limiting access rights of employees, providing management and financial support for security policies, prohibiting downloading/playing of music on computers), and technical measures (e.g., firewalls, encryption, digital certificates) – all must be part of an effective enterprise security solution. Effective security policies will address these issues, and may be authored locally or involve collaboration between entities or jurisdictions. HIPAA, the Health Insurance Portability and Accountability Act of 1996 [11], includes efforts to improve health data security nationally; HIPAA is discussed in detail in other parts of this book.

Data Use Agreements (DUA)

Data use agreements are legal agreements between entities that are intended to ensure appropriate safeguarding and use of shared information or data. DUAs will include details of the agreed-upon security measures and confidentiality requirements, such as the conditions under which data may be accessed and disclosed. An effective DUA will also include measures to ensure tracking of data and data use, to enforce compliance with the DUA and provide evidence in the case of a security breach or unauthorized use.

Metadata

Metadata is often described as "Data about Data," and entails structured information that facilitates usage and management of an information resource [12]. Metadata not only makes it easier to generate value from a resource, it enables continued usage of the resource by providing vital descriptive and identifying information for future users. For this discussion, we will review three important divisions of metadata:

- *Descriptive metadata* – generally used for discovery and identification, e.g., title, abstract, author, and keywords

- *Structural metadata* – describes the parts of compound objects, e.g., sections of a document
- *Administrative metadata* – information for resource management, e.g., a database creation date and development platform. May be considered to contain the concepts of rights management metadata (intellectual property rights) and preservation metadata (archival information).

Data or Content Standards

As discussed earlier in this chapter, data or content standards are divided into three categories, (a) vocabulary, (b) format, and (c) transmission. In the following sections, we will discuss the vocabulary and format standards in greater detail. In order for systems to successfully communicate or interface, there must be both functional and semantic interoperability. *Functional interoperability* occurs when systems are able to physically communicate or share data, whereas *semantic interoperability* involves interpretation of data via a common language or *vocabulary*. Interoperability is facilitated by standards of all categories.

Data Format Standards

Information exchange standards define *structure* (parts) and *syntax* (arrangement), including to some extent the vocabulary, of the electronic communication and are referred to as the standard ways of sending and receiving information [13]. These standards can be compared to the grammar requirements in a language.

Health Level Seven (HL7®) [14] is an international standard that is the most widely used formatting standard for health data. Created by developers in the 1980s, it is present in most hospital systems and has been adopted by public health as a data format standard. The term 'Health Level Seven' refers to the Open Systems Interconnection (OSI) standard developed by the International Organization for Standardization (ISO) [15] in 1984. The OSI Reference Model defines the different stages that data must go through to travel over a network, and the seventh level (level 7) is the Application Level, which includes definition and structure of data.

HL7® is a complex and flexible set of format protocols that can encompass a staggering array of data requirements. The flexibility of HL7®can be a 'good news/ bad news' attribute – while it can accommodate an enormous variety of data situations (definitely good), users can also create an astonishing number of variations upon the standard, which may lead to confusion and extra effort.

HL7® defines two major ways to exchange data – message-based (sent as a message) and document-based (sent as a structured document). HL7®, like many standards, was developed over time based on additional requirements for different settings, so there are many versions of HL7® – notably the group of Version 2

messages (also referred to as v2.x), developed first and initially growing in an *ad hoc* or needs-based fashion. There are multiple versions of the international HL7®v2.x standards in use, the latest, v2.8 was balloted in 2012. These versions are backwards compatible with each other, i.e., a system updated to a newer version can still receive data from systems using any previous version [16]. After some experience was gained with v2.x, HL7® developed a formal data model, the Reference Information Model (RIM), that forms the core for all Version 3 (v3) artifacts, to explicitly retain the context in which the exchanged information is used; they can be message- or document-based exchange standards.

HL7®version 2.x messages are identified by message type and trigger event code. For example, a commonly used message for public health laboratory reporting is the ORU^R01 message, which is identified as message type ORU (Observation result unsolicited), and trigger event R01, signifying unsolicited transmission of an observation message. There are many other commonly used message types and trigger events. HL7® tables 0076 and 0003 contain, respectively, 84 message types and 184 event types. A few examples of HL7® message types are: ACK, General acknowledgement, used to let the sender know when the message was received, either successfully or unsuccessfully; ADT, Admit discharge transfer, used in the hospital setting, for example, to exchange information about the patient with the different systems inside a hospital; OSQ, Order status query, used to find out what's going on with an ordered diagnostic test; RAS, Pharmacy administration message, used to report when a specific medication has been given to the patient; VXQ, Query for vaccination record, used to find out if a patient has been immunized against a specific disease.

Think of the messages as information vehicles – put together from a library of building blocks that define information about specific topics. Some of these building blocks (segments) are used in every message, while others are only used when their information needs to be part of the message. Every message has an MSH segment – the message header, which establishes the foundation information. It contains metadata about the message that systems need in order to properly understand the content. Other commonly used segments are PID, patient information; PV1, patient visit; NTE, notes, containing additional information in unstructured text format for clarification; OBR, detailed order request information; and OBX, result information.

A version 3 artifact that has been adopted by several clinical and public health programs in the US is the *Clinical Document Architecture* (*CDA*). The CDA was derived from the HL7® Reference Information Model (RIM) to enable semantic consistency across platforms for the purpose of exchange and re-use of clinical documents [17]. CDA allows representation of clinical or public health information in a structured format, using CDA templates that are similar or identical to the formats of the paper forms [18]. Thus, the CDA standard closely mirrors traditional paper-based reporting workflows, and information is exchanged as documents instead of repackaged into discrete data elements (as is done in messages). The HL7® CDA standard incorporates the concepts of human readability, persistence, stewardship, and wholeness; it allows for authentication and ensures semantic

interoperability through use of the RIM structure and associated controlled vocabulary. It is implemented in Extensible Markup Language (XML). A CDA document has a header and a two-part body, containing the human readable part and the structured data part. The header contains information about the patient, the encounter, and document authors. The body contains the respective clinical content [19].

Vocabulary Standards

Vocabulary standards are often explained using the metaphor of language. If people are speaking different languages, it will be difficult for them to communicate effectively. Similarly, if systems are using different vocabularies to refer to data content, it will be difficult for them to interoperate. Vocabulary standards can be considered to be either local or 'universal', depending upon the partners involved and how widely accepted the standard in question may be. It should be remembered that there are both advantages and disadvantages for local or universal standards. For example, local code sets may be more easily updated or changed by the source institution, but may make sharing data with other institutions much more difficult. 'Universal' codes enhance data sharing between systems and across regions, but may require specialized training to use, and may not be as flexible as local codes in adapting to local circumstances. As with different languages, one can also translate between the local and the universal codes.

One of the most important components of data standards is the consistent representation of clinical concepts or terms through the use of unique codes or identifiers. These are commonly referred to as *code systems*. Some of the areas where code systems are used in public health EDI include:

- Laboratory Tests
- Laboratory Results
- Other subjects, such as diagnoses and clinical findings, administration, or demographics

Laboratory Test and Result Code Standards

Logical Observation Identifiers Names and Codes (LOINC®)

The most widely adopted code system for laboratory observation coding is the Logical Observation Identifier Names and Codes (LOINC®) system [20]; LOINC® can be used to represent the name of both ordered and related performed tests. This code system is owned and maintained by the Regenstrief Institute. Current LOINC®

codes are 3–7 characters in length, and will expand as the code set continues to increase in content. LOINC® codes are constructed as the combination of a simple integer sequence number (beginning with "1"), a "dash" delimiter, followed by a Mod-10 check digit. LOINC® terms are composed of six major parts:

- Component/analyte – The substance or entity that is being measured or observed.
- Kind of property – The kinds of quantities or qualities relating to the same substance.
- Time aspect – measurement relates to either a point in time or a specified time interval. The vast majority of laboratory measurements are "point in time".
- System type – For laboratory observations, this is equivalent to the sample type being analyzed.
- Scale – Specifies the scale of measurement. The most common scales used in laboratory analyses are quantitative (QN), qualitative (QL), nominal (NOM), ordinal (ORD), and narrative (NAR).
- Method – This reflects the technique or procedure used to obtain the result.

Of these LOINC® parts, the code, analyte, property, timing, and scale are required. Both system and method are able to be specified in other parts of an HL7® message, which is the primary vehicle for using LOINC® coded terms.

An example of a fully specified LOINC® term and its component parts is shown below:

13203-5: *Borrelia burgdorferi* AB.IGM:ACNC:PT:CSF:ORD:IB

- COMPONENT – *Borrelia burgdorferi* AB.IGM. The specific immunoglobulin subclass IgM stimulated in response to the presence of *Borrelia burgdorferi* (the organism that causes Lyme disease) antigen.
- PROPERTY – ACNC. Arbitrary Concentration, or an arbitrary number of units in a volume.
- TIME ASPECT – PT. Point or moment in time, i.e., the time the sample was collected.
- SYSTEM – CSF. Sample type, cerebrospinal fluid.
- SCALE – ORD. Ordinal, a qualitative ordered list of values such as "Detected," "Not detected," "Positive," or "Negative."
- METHOD – IB. Measurement method, "Immune blot."

Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT®)

Standardized laboratory result coding for non-numeric values is increasingly being represented using SNOMED CT®. SNOMED CT® was initially produced by the College of American Pathologists (CAP), which entered into an agreement with the US National Library of Medicine (NLM), funded by the Department of Health and

Human Services, to offer open access to the US for the International release content of SNOMED CT®. Since 2007, the code system has been owned and managed by International Health Terminology Standards Development Organisation (IHTSDO®) in Denmark [21].

Other Data Content Coding Standards

Laboratory test and result standards are of course not the only coding standards of importance to public health informatics. A sampling of other important code systems includes:

- *Procedural codes*: – The Current Procedural Terminology (CPT) code system contains content developed (and copyrighted) by the American Medical Association. CPT codes are five-digit alphanumerics that classify medical service and are used for insurance billing [22]. As an example, the 2013 CPT codes and Medicare payment information show that "Application of short leg cast (below knee to toes)" has the assigned CPT code of 29405, and a cost (facility: in hospital) of US$68.65. It is of note that fee-based code systems, especially those that are generally accepted for reimbursement, are often far more advanced in their acceptance, adoption, and implementations. This reflects the popular adage, "Money talks and people listen."
- *Geographic codes*: In 2006, the Geographic Names Information System (GNIS) Feature ID became the official federal reference to named geographic entities [23]. Using this system in 2013, the White House in Washington DC has an ID of 531723. However, public health often uses the legacy standard, the Federal Information Processing Standards (FIPS), to identify geographic areas such as states and counties [24]. Using FIPS, the Washington, DC code is 11001. Another option for geographic coding is the US Census Bureau coding INCITS 38:200x, "Codes for the Identification of the States, the District of Columbia, Puerto Rico, and the Insular Areas of the United States" [25]. Using this resource, the District of Columbia is identified as ANSI State Code "11"; Official United States Postal Service (USPS) Code "DC"; Name "District of Columbia"; and Geographic Names Information System Identifier (GNISID) "01702382."
- *Industry and Occupation codes*: These code systems may be used by public health programs, such as programs tracking environmental issues like lead exposure. The Standard Occupational Classification (SOC) system, Bureau of Labor Statistics, classifies workers into occupational categories [26]. The 2010 dataset includes 840 detailed occupations, which are also grouped into broader categories. For example, Carpenters Assistant is classified as 47-3012 Helpers – Carpenters; Broad Occupation is 47-3010 Helpers, Construction Trades; Minor Group is 47-3000 Helpers, Construction Trades; and Major Group is 47-0000 Construction and Extraction Occupations. The North American Industry Classification System (NAICS) is the standard used by Federal statistical

agencies in classifying business establishments for the purpose of collecting, analyzing, and publishing statistical data [27]. As an example, the 2012 NAICS Definition assigns code 238350 to "Finish Carpentry Contractors."

- *Demographic codes*: Information on demographic or population variables, such as gender, race, ethnicity, and age, are crucial to public health. Demographic codes used for public health data include the Race Value Set developed by HITSP [28]. As an example, this value set assigns a code of 2010-7 to the concept "Aleutian Islander."

Format Standards Paradigm – HL7®

HL7® Version 2.x Artifacts

Reporting Observations (ORU)

Let's look in detail at the contents of a v2.x HL7® message using the Observation message – for instance, to report a laboratory result. The way the segments are arranged within a message creates a hierarchy of information. There are two types of observation messages used in laboratory reports, (a) the ORU, which is patient centric, meaning the information about the patient comes before anything else, and (b) the OUL, which is sample centric, allowing for grouping under the sample rather than a patient. As an example, the OUL could be used because you want to report a result from testing a water sample for contamination; water usually is not considered a patient. For this illustration, we want to look at the ORU because in healthcare we most often are interested in patient-related observations. The ORU message structure in the standard is defined by its required and optional segments, assembled in a specific order. *Required* means the segment must be sent, while *optional* means that you don't have to send it – but if you send it, it has to follow the rules of the standard. The standard also specifies (in an *implementation guide*, a document that contains the specifications for the message) whether a segment or a group of segments can be repeated. The underlying segment order must be maintained, but changes can be made in the optionality (whether segments are required or optional) or the number of times a segment or a group of segments can repeat.

Each segment has a specified number of fields that carry specific information related to the general topic of the segment (Fig. 8.2). For example the PID or patient information segment will have fields for name, date of birth, birthplace, address, gender, etc. Each field has a specific format called a data type; data types can be a *string of characters* (*ST*, *string*), while in other cases the format can be more complex and have several components. One such complex data type is *Extended Person Name* (*XPN*), which can contain last name, first name, other given names, suffix, etc. Data types follow a precise order that has meaning, such as in the *Date/Time* (*DTM*) data type, used for values like date of birth; DTM values are listed as

Fig. 8.2 Generic ORU message structure with the most commonly used segments (Published with kind permission of © Riki Merrick 2013. All Rights Reserved)
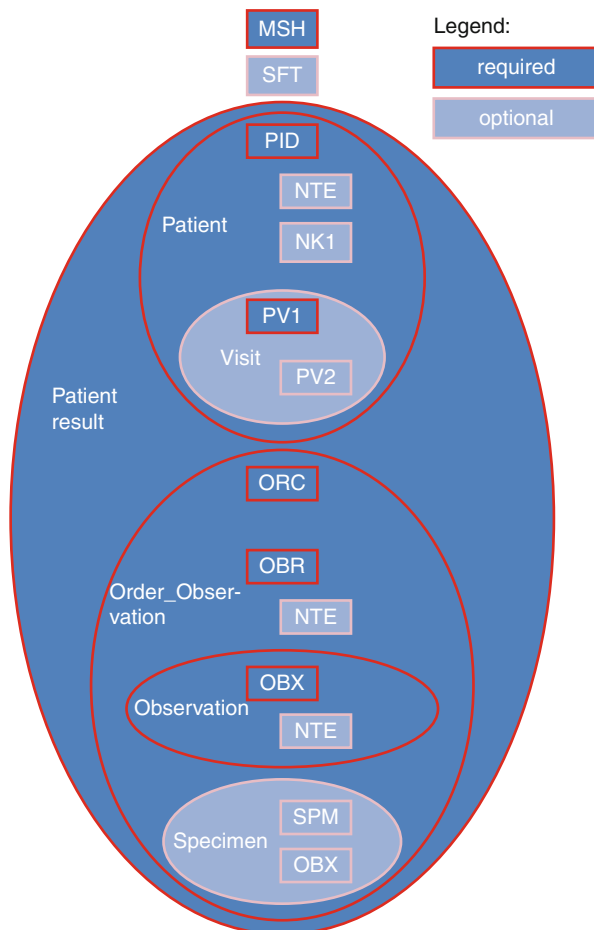
four-digit year, two-digit month, two-digit day, and, if available, two-digit hours, etc. Another complex datatype very important for reporting observations is the *Coded with Exceptions* (*CWE*) datatype, which is used to carry the codes describing the ordered tests, performed tests, sample types, and results.

HL7® Version 3 Artifacts

Clinical Document Architecture (CDA)

Along with several HL7® v2.x artifacts, the Health IT Standards Federal Advisory Committee, in their September 2011 rule about Meaningful Use, named the HL7® *Clinical Document Architecture* (*CDA*) standard for use in data exchanges between
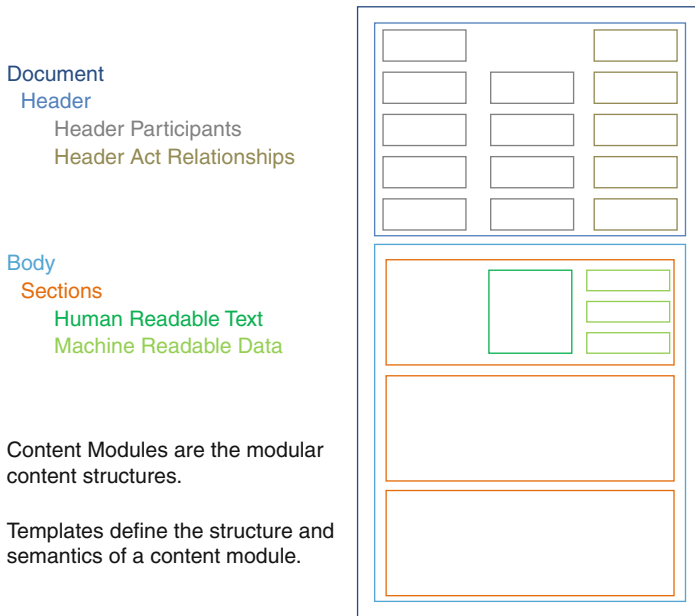
Document
  Header
      Header Participants
      Header Act Relationships

Body
  Sections
      Human Readable Text
      Machine Readable Data

Content Modules are the modular
content structures.

Templates define the structure and
semantics of a content module.

Fig. 8.3 CDA building blocks (© 2012 Lisa R. Nelson, used with permission)

clinical *Electronic Health Record systems* (*EHRs*), as well as from EHRs to public health information systems [29]. In the US, Meaningful Use requires EHRs to create, transmit, receive, and display the *Continuity of Care Document* (*CCD*) [30], which uses the CDA as a framework. CDA serves as the basis for the creation of closely-related documents within a document-based health information exchange. CDA defines templates at different levels; documents, which have a header and a body, which in turn has its human readable and structured part, comprised of template-based sections and entries. This structure is illustrated in Fig. 8.3, and can also be organized to construct valid public health reports.

MU requires the use of CCD for exchange of discharge information from hospitals to the patient's primary care provider, for example, core clinical information about the hospital stay and instructions on what to do next. It must also be used to summarize clinical information when a specialist needs to be involved in the patient care. The CCD described in MU is a collection of CDA templates, which will be combined in a specific order for each specific purpose, but the core information in each template is pre-defined [31].

An example of a CDA-based PH report is that sent for the group of reports about Healthcare-Associated Infections, such as bloodstream infections, surgical site infections, urinary tract infections, etc., to the National Healthcare Safety Network at the CDC. CDA parts and specific vocabulary (LOINC®, SNOMED CT®, demographic standards, etc.) are defined in the implementation guide to ensure all

required data for this reporting purpose are included. A different implementation guide, also CDA-based, is used to report about cancer patients to cancer registries in the US. Both the way the sections are put together and the vocabulary that is required are specialized to the needs of cancer reporting. In order to reduce the many variations imposed on the data providers (i.e., the EHR systems) the Public Health Reporting Initiative (PHRI) has convened many public health programs at the local, state, and federal level to collaborate and harmonize the format and vocabulary used for data that is needed across many different programs. These harmonized "Common Core" data elements have been incorporated into another information model to retain the context, by defining format and vocabulary binding in the Federal Health Information Model (FHIM) [32], which is also linked to the underlying HL7® RIM. PHRI has also created the Reference Implementation Framework document, intended to be a "one-stop shop" to access all currently available standards for public health reporting, regardless of the format used for exchange (can be HL7® v2.x messages or HL7® v3 messages or CDA based) [33].

Vocabulary Standards – SNOMED CT®

As mentioned earlier in this chapter, SNOMED CT® is a comprehensive reference terminology that encompasses all areas of healthcare. Its primary use is within EHRs, for the purposes of both meaning-based retrieval and use in *clinical decision support* (assistance to health professionals in making choices). The by-product of consistent representation of clinical data is the ability to perform broad-ranging data aggregation, reporting, and analysis. SNOMED CT®, at its highest level, is based on three primary structures - concepts, descriptions, and relationships. A brief outline of the content and structure of SNOMED CT® is provided below.

Concepts

Within the SNOMED® terminology, a concept is a unit of content that is assigned a unique "meaningless" identifier in numeric format. These identifiers are meaningless in that it is not possible by simply looking at the identifier to deduce any meaning of the associated term or its position within the SNOMED CT® hierarchy. Each concept is represented by a description called the *fully specified name* (FSN) that uniquely represents the concept; this is accomplished through the combination of the description string and a semantic tag, which represents the top-level category to which the concept belongs. To illustrate this, let's examine the term "swab," which has multiple meanings within the healthcare environment. It may represent a physical object, a unit of a product, or a specimen type. Within SNOMED CT®, these are represented by uniquely identified concepts: 408098004 identifies swab

as a physical object; 420401004, as a unit of product usage (qualifier value); and 257261003, as a specimen.

Concepts are arranged hierarchically within SNOMED CT®, such that less granular (more general) concepts are assigned as "parents" to more granular (more detailed) "children" through explicitly defined "is a" relationships (i.e., a granular concept "is a" child to the more general parent concept). In some cases, concepts may have many parents depending on the types of defining relationships assigned to them. There are a number of important considerations when assessing these parent–child relationships:

1. For a concept to be a child of another concept, all of the defining attributes for the purported parent must be always and necessarily true for the child.
2. One cannot, by looking at the children of a concept, deduce the meaning of a parent concept: i.e., a parent defines the children, children do not define the parent.
3. Not all levels of intermediate granularity of meaning are represented by the terminology: i.e., there may be perceived "gaps" in the hierarchies.
4. It may not be possible to ascertain the full meaning of a concept without looking at all of the parents.

What this means in practice is that in some cases it may be difficult to ascertain the full meaning of the "words" in the concept description without looking at the surrounding content in SNOMED CT® to gain the full context of the term.

Concept Identifiers

Concept identifiers are assigned permanently to any concept that is incorporated into the terminology. This means that once an identifier is assigned, it is never reused. Local extensions to SNOMED CT® are assigned namespace identifiers that allow for the unique assignment of extension concepts, descriptions, or relationships that augment the content of the International Release in order to meet specific needs of the extension owner. In the US, the National Library of Medicine has been assigned the extension namespace identifier for the official US extension to SNOMED CT®. This extension is designed to support the specific needs of US healthcare as designated by legislative mandates such as Meaningful Use. In general, SNOMED CT® identifiers have the general structure demonstrated below.

- SNOMED CT® Identifier (SCTID): 101291009, is comprised of an item identifier [101291], a partition identifier [00], and a check digit [9].
- SCTID: 430261000124101, is comprised of an extension item identifier [43026], a namespace identifier [1000124], a partition identifier [10], and a check digit [4].

SNOMED CT® differs from most other clinical terminologies in that it provides a multi-hierarchical representation of distinct clinical concepts as well as a set of

defining relationships that allow systems to perform reasoning against the terminology. For example, a disease concept in SNOMED CT® may have assigned relationships to a causative agent (e.g., a particular species of bacteria) and a finding site (a specific anatomic structure). This would allow a reasoning system to classify the concept as a bacterial disease, an infectious disease, a disease affecting a particular part of the body, etc. This allows one to analyze SNOMED CT® encoded content from a variety of perspectives, based on the associated defining relationships attached to a concept.

SNOMED CT® Descriptions

In addition to fully specified names, SNOMED CT® allows for a variety of alternative descriptions to represent the intended meaning of the concept. The primary purpose of these alternative descriptions, contrary to the general perception, is not to provide different display terms for the concept but to provide users with assistance in searching the terminology for the proper concept that meets their particular need. While these alternative descriptions have historically been called "synonyms," in many cases they are not true synonyms; the meaning of these terms could be more general, or in some cases, ambiguous. Looking back at the example provided at the beginning of this section, all three of the concepts related to "swabs" have alternate descriptions of "swab." Without the knowledge of the concept's fully specified name, one could not determine the full meaning of the descriptive term "swab." Thus, it might be possible for the term to be used incorrectly if a user did not have access both to the descriptive term and the associated FSN.

This lack of true synonymy, and confusion as to the purpose of SNOMED CT® descriptions, often causes users some frustration because SNOMED CT® "does not have my words." Because the ways in which users might want to have terms displayed by their own EHRs is nearly limitless, SNOMED CT® does not attempt to provide an exhaustive list of potential alternative descriptions.

SNOMED CT® Relationships

The relationships defined by SNOMED CT® are at the heart of the true value of the terminology, to provide enhanced usefulness for a variety of analytical needs. Through the explicit relationships, it is possible to easily select concepts based on particular attributes such as infectious disease, neoplastic disease, location on the body, or clinical manifestation. Because all concepts are related to one or more "parents," it is possible to computationally aggregate highly specific terms into more general categories for trend reporting and analysis. This value allows data recorders to be as specific as possible with their entries, without having to worry about how their entries will be categorized during analysis.

SNOMED CT® Browsers

Generally, users will first become familiar with the content of SNOMED CT® through exposure to one of a growing population of SNOMED CT® specific "browsers." These tools, whether stand-alone or web-based, provide mechanisms to search for specific concept, descriptions, or identifiers within the entire SNOMED CT® terminology and then traverse associated hierarchies to view the terms located within the same "vicinity" of the searched term. Currently, the US National Library of Medicine maintains a listing of available SNOMED CT® browsers [34]. These browsers include the NLM SNOMED CT® Browser, which differs from all the others in that it leverages the NLM Unified Medical Language System (UMLS) to find terms within SNOMED CT®. As mentioned above, SNOMED CT® does not attempt to include all possible descriptions that might be applicable for a particular SNOMED CT® concept. The NLM browser, by utilizing the power of the UMLS® Metathesaurus® (a multi-lingual collection of biomedical and health-related concepts, synonyms, and relationships), can use descriptions that originate from any of its over 150 source terminologies. Of these sources, 15–20 are updated annually. Thus, the NLM SNOMED CT® browser allows users to search for concepts in SNOMED CT® using descriptions that do not actually exist in the full SNOMED CT® terminology. This additional power provides more comprehensive retrieval of concepts than can be accomplished through the use of SNOMED-only browsers.

Summary

There is a well-worn saying among standards aficionados, to the effect that the nice thing about standards is that there are so many from which to choose. Entertaining as that phrasing may (or may not) be, it actually may be true that the variety of standards really is a positive development. The incredible complexity of situations and data inherent in public health and healthcare EDI demands a similar complexity in standards.

Review Questions
1. Discuss the two categories of standards used in this chapter, process standards and data or content standards. What are some examples of these standards that are used in public health?
2. Describe the general process for standards development. How can public health participate in standards development, and why should it do so?
3. Select two of the standards reviewed in this chapter and describe in detail (a) the benefits accrued to public health from use of the standard, and (b) the barriers to implementing this standard in public health.

4. If you had to choose between process standards and data or content standards, which category do you think is most important for public health to utilize?
5. Based on the two standards presented in depth in this chapter, HL7® and SNOMED CT®, compare and contrast their role in public health and their importance to population health as a whole. Use examples to strengthen your key points.

References

1. HealthIT.gov. Policymaking, regulation, & strategy. Cited 23 Mar 2013. Available from: http://www.healthit.gov/policy-researchers-implementers/meaningful-use
2. Jernigan DB, Davies J, Sim A. Data standards in public health informatics. In: O'Carroll PW, Yasnoff WA, Ward ME, Ripp LH, Martin EL, editors. Public health informatics and information systems. New York: Springer; 2003. p. 213–38.
3. Centers for Disease Control and Prevention. HIV/AIDS statistics and surveillance. Cited 23 Mar 2013. Available from: http://www.cdc.gov/hiv/topics/surveillance/resources/qa/generalqa.htm
4. International Organization for Standardization. Cited 23 Mar 2013. Available from: http://www.iso.org/iso/home.htm
5. American National Standards Institute. About ANSI. Cited 27 Mar 2013. Available from: http://www.ansi.org/about_ansi/overview/overview.aspx?menuid=1#.UVNc5lcu-jg
6. International Organization for Standardization. Standards development. Cited 23 Mar 2013. Available from: http://www.iso.org/iso/home/standards_development.htm
7. Panackal AA, M'ikanatha NM, Tsui F, McMahon J, Wagner MM, Dixon BW, et al. Automatic electronic laboratory-based reporting of notifiable infectious diseases at a large health system. Emerg Infect Dis. 2002;8(7):685–91.
8. Health Information Technology Standards Panel (HITSP). Cited 27 Mar 2013. Available from: http://www.hitsp.org
9. Public Health Data Standards Consortium (PHDSC). Health IT standards. Web-Resource Center. Cited 27 Mar 2013. Available from: http://www.phdsc.org/standards/health-information-tech-standards.asp
10. Centers for Medicare and Medicaid Services, US Department of Health and Human Services. National Provider Identifier Standard. Cited 27 Mar 2013. Available from: http://www.cms.hhs.gov/NationalProvIdentstand
11. U.S. Department of Health & Human Services. Summary of the HIPAA Security Rule. Cited 22 Mar 2013. Available from: http://www.hhs.gov/ocr/privacy/hipaa/understanding/srsummary.html.
12. National Information Standards Organization. Understanding metadata. Cited 22 Mar 2013. Available from: http://www.niso.org/publications/press/UnderstandingMetadata.pdf
13. Public Health Data Standards Consortium (PHDSC). Health IT standards. Web-Resource Center. Information Exchanges Standards. Cited 27 Mar 2013. Available from: http://www.phdsc.org/standards/health-information/IE_Standards.asp
14. Health Level Seven International. Cited 23 Mar 2013. Available from: www.hl7.org
15. International Organization for Standardization. Cited 23 Mar 2013. Available from: http://www.iso.org/iso/home.html
16. Public Health Data Standards Consortium (PHDSC). Health IT standards. Web-Resource Center. Information Exchanges Standards. Cited 12 Mar 2013. Available from: http://www.phdsc.org/standards/health-information/IE_Standards.asp

17. Health Level Seven (HL7®). Clinical Document Architecture (CDA). Cited 27 Mar 2013. Available from: http://www.hl7.org/implement/standards/cda.cfm
18. Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, Shabo A. HL7® Clinical Document Architecture, Release 2. J Am Med Inform Assoc. 2006;13(1):30–9.
19. Health Level Seven (HL7®). Frequently asked questions. Cited 12 Mar 2013. Available from: http://www.hl7.org/about/FAQs/index.cfm
20. LOINC® from Regenstrief. Logical Observation Identifiers Names and Codes (LOINC®). Cited 27 Jan 2013. Available from: http://loinc.org/
21. International Health Terminology Standards Development Organisation. Cited 27 Jan 2013. Available from: http://www.ihtsdo.org/
22. American Medical Association. CPT – current procedural terminology. Cited 23 Mar 2013. Available from: http://www.ama-assn.org/ama/pub/category/3113.html
23. USGS: National Hydrography Dataset. Cited 22 Mar 2013. Available from: http://nhd.usgs.gov/gnis.html
24. National Institute of Standards and Technology (NIST). Information Technology Laboratory. Cited 22 Mar 2013. Available from: http://www.nist.gov/itl/fips.cfm
25. US Census Bureau. Cited 22 Mar 2013. Available from: http://www.census.gov/geo/www/ansi/statetables.html
26. Bureau of Labor Statistics. Standard occupational classification. Cited 23 Mar 2013. Available from: http://www.bls.gov/soc/home.htm
27. US Census Bureau. North American Industry Classification System. Cited 22 Mar 2013. Available from: http://www.census.gov/eos/www/naics/
28. US Health Information Knowledgebase. Health Information Technology Standards Panel. Cited 22 Mar 2013. Available from: http://ushik.ahrq.gov/ViewItemDetails?system=hitsp&itemKey=86685000
29. Health IT Standards Federal Advisory Committee. Recommendations from the Public Health Surveillance Summer Camp. 2011. Cited 12 Mar 2013. Available from: http://healthit.hhs.gov/portal/server.pt?open=512&objID=1817&parentname=CommunityPage&parentid=28&mode=2&in_hi_userid=11673&cached=true#09281
30. US Department of Health and Human Services. Health information technology: initial set of standards, implementation specifications, and certification criteria for electronic health record technology. Regulation Identification Number 0991-AB58. Washington, D.C.: Office of the National Coordinator for Health Information Technology (US). 2010.
31. S&I Framework. Transitions of care (ToC) initiative. Cited 26 Mar 2013. Available from: http://wiki.siframework.org/Transitions+of+Care+(ToC)+Initiative
32. FHIMS.org. Background on THE FHIM. Cited 26 Mar 2013. Available from: http://www.fhims.org/background.html
33. S&I Framework. Public Health Reporting Initiative. Cited 26 Mar 2013. Available from: http://wiki.siframework.org/Public+Health+Reporting+Initiative
34. US National Library of Medicine. Unified Medical Language System® (UMLS®). Cited 12 Mar 2013. Available from: http://www.nlm.nih.gov/research/umls/Snomed/snomed_browsers.html