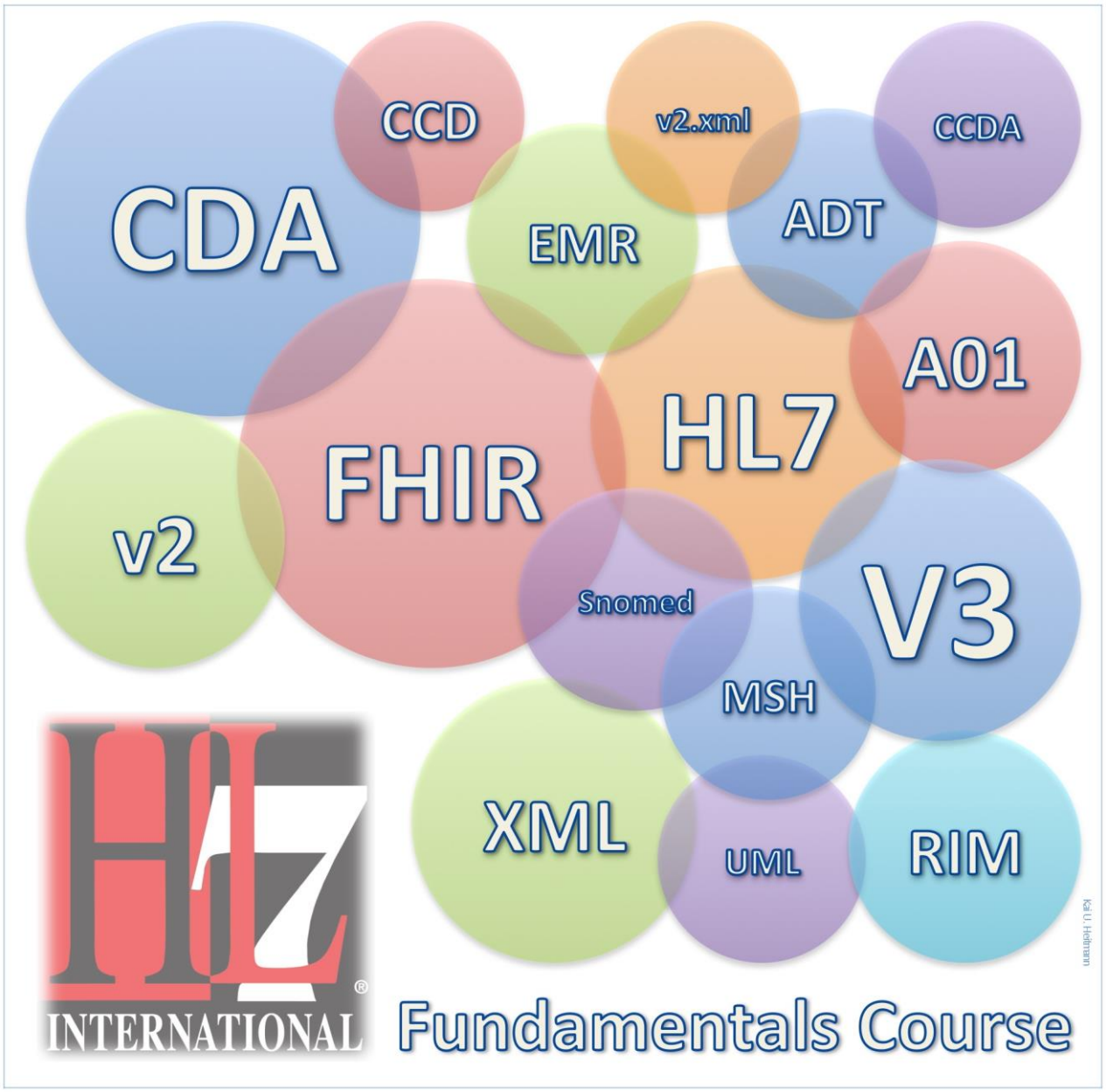

Introduction Module, Unit 1:

Introduction to Healthcare Interoperability

Reading Material

Table of Contents

Unit Content and Learning Objectives

The purpose of this Course is to provide in-depth information about healthcare information standards and to present examples from a variety of implementations that show how organizations are making progress towards interoperability; achieving greater availability of electronic medical records, reducing medical errors and improving the quality of care.

In this first part ("Unit 01") of the 'HL7 Fundamentals' Course, we will establish why reliable and good-quality healthcare data is vital for providing the best possible care to patients. We will then look at ways that communications between computers can improve the reliability and quality of healthcare data and how standards can make these communications affordable and reliable. We will also learn how standards are categorized and developed.

The other Units of this Introductory Module will cover how standard coding systems can also improve healthcare and how the UML and XML tools can improve the modeling and markup of healthcare data.

1. Introduction

The practice of medicine has been evolving continuously since the age of the Greeks. In the past 100 years, more and more complex technological advances have come about (Figure 1). Health organizations have been quick to adopt these new technologies as they have become available - except in the case of medical records.

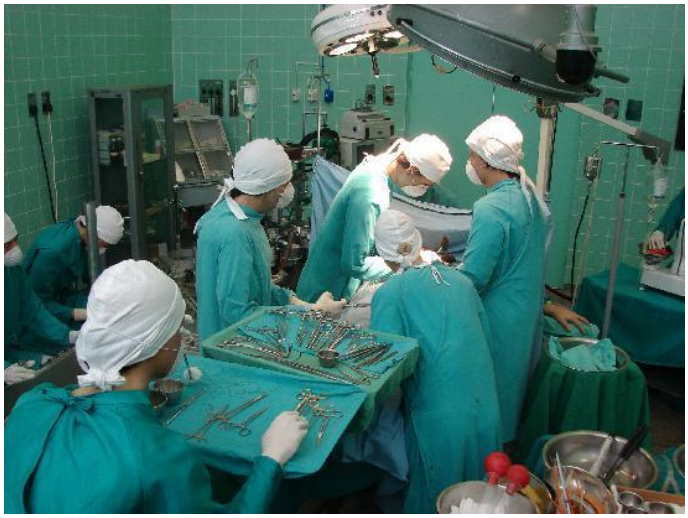

Figure 1: The evolution of medical practice

Today most healthcare organizations still have multiple paper-based medical records and physical archives in different departments of their organization. In many cases, the health-related information is still nearly as fragmented as it was 150 years ago.

The following request reflects this reality:

"I need to bring attention to the urgent need to adopt… some uniform publication system of hospital statistic records. There exists a growing belief that in every hospital, even in those with the best work conditions, there is a large, unnecessary waste of life … In an attempt to reach the truth, I have sought information everywhere, but in few cases was I able to get hospital records adequate for any comparison purpose... If used smartly, these statistics will tell us the actual relative value of some current measures and forms of treatment."

Question: Who said this and when?

The request above could well have been made by any practicing health-care professional today. In fact it was said in 1863 by Florence Nightingale, a famous nurse in England who pioneered epidemiology and the use of health statistics..

A century and a half ago, the need for reliable usable data and the need to enhance the quality of information for strategic decision-making and improvement of the quality of patient care was already a concern.

Nevertheless, integrated Health Information Systems (HIS) have been a topic of discussion for over four decades now. In the 1960's and 70's, such systems were more focused on financial and administrative functions. Gradually, interest began to shift toward patient care, with the objective of integrating the information contained in the various medical records kept in one or several

healthcare sites where the patient was treated. These concepts were implemented in some HISs in the late 1970's and early 1980's, but often failed due to inadequate hardware and software.

An examination of how information technologies have penetrated healthcare organizations shows that relatively little integration is targeted at patient care with the majority allocated to logistics e.g. administrative support for billing and lab operations), as shown in Figure 2:

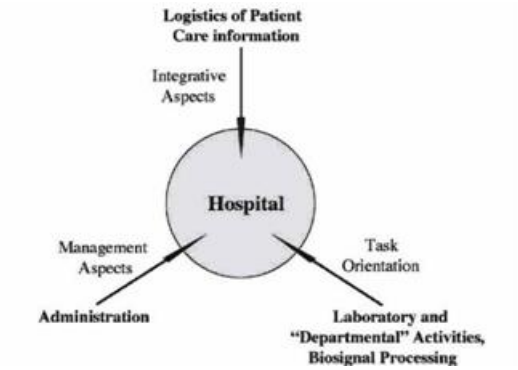

Figure 2: Penetration of information technology in hospital environments

The architecture of the HIS of the 1970's consisted of a monolithic core, made up of a main data-base, a communication system between users and applications and an ADT system ("Admission-Discharge-Transfer")[1] which is the administrative portal for all hospitalization episodes. See Figure 3 below.

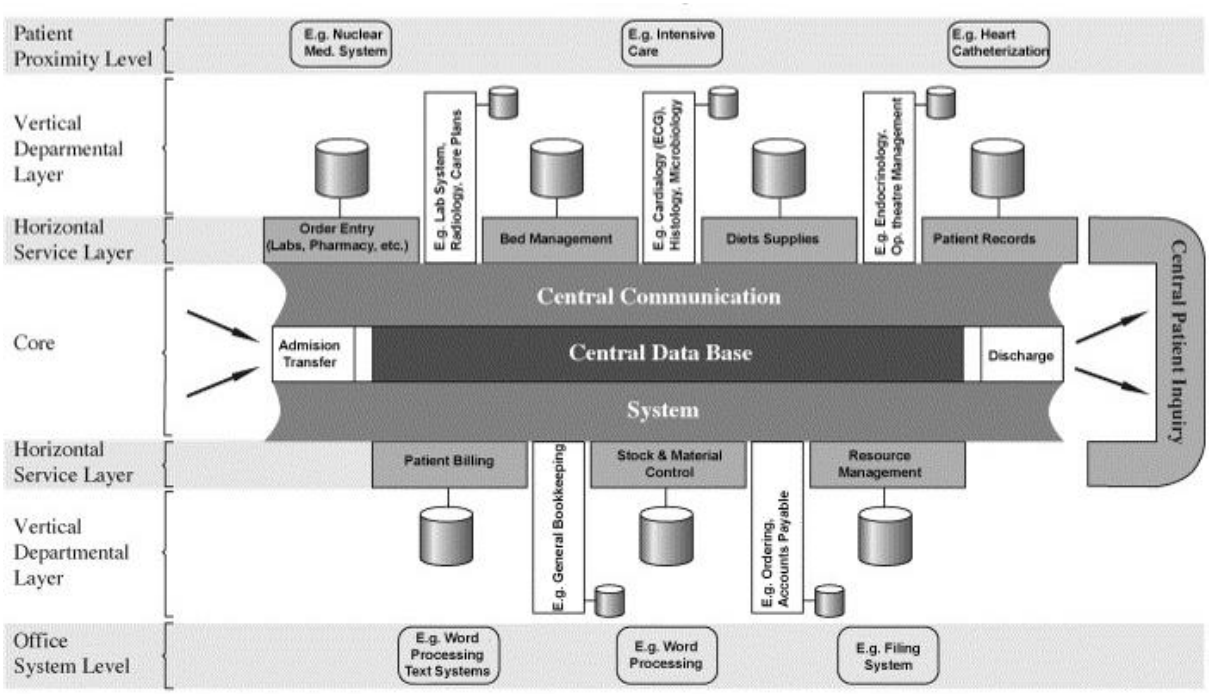

Figure 3: Typical Architecture of a Health Information System in the 1970's

[1] ADT systems are in some countries called ATD systems ("Admission-Transfer-Discharge")

There is a horizontal layer providing such services as pharmacy, bed administration, clinical charts, billing, supplies and human resources.

Finally, there are vertical departmental layers representing different independent subsystems (e.g. Intensive Care, Nuclear Medicine, Patient Billing, Stores Management, etc.) that communicate with the main core for administrative purposes.

There are also independent systems that have survived without any communication with the main system - systems that, with the advent of personal computers ("PCs"), gradually offered more and more diverse applications without connectivity (Figure 4).

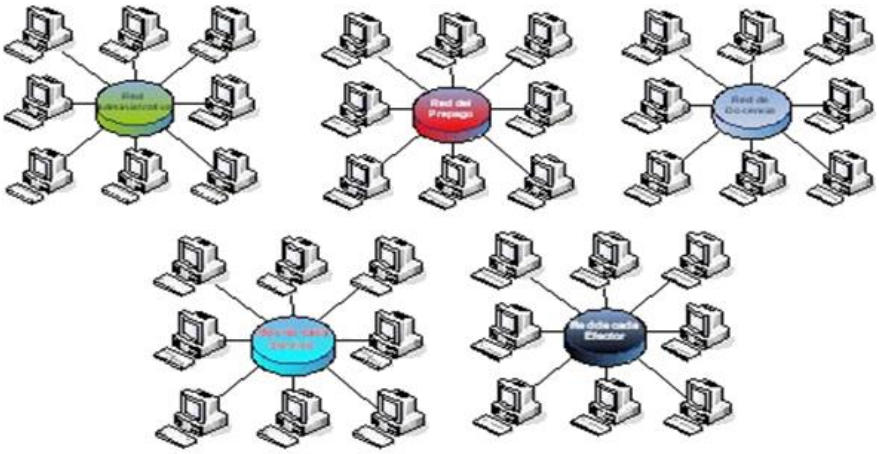

Figure 4: Independent systems

These departmental or layered systems helped to meet specific objectives and needs. At present the focus on outpatient care with care decentralization, geographical dispersion, new consultation interrelations, penetration of information networks (Internet) and rules by payers (what and how medical care is covered) govern the need for clinical management.

The trend today is to eliminate non-interconnected systems, as they are "clinical information silos" that prevent a comprehensive overview of a patient's healthcare data. Interoperability standards such as Health Level 7 (HL7) play a major role in connecting these legacy independent systems.

Managing Clinical Information

The Clinical Management concept is broad and attempts to achieve population-wide cost-effective medical care in outpatient, hospital and rehabilitation environments, while at the same time striving to maintain quality results.

This model has an indispensable requirement for **data interchange among health applications** in order to have high quality, contextual, up-to-date clinical information. Such a requirement implies new challenges for the design and development of information systems, such as:

1. Common, single-source person identification services
2. A person-oriented integrated lifetime electronic clinical record
3. Physical, semantic and syntactic interoperability among the systems in the different organizations
4. Information exchange standards (such as HL7)

5. Medical knowledge terminology and representation services (such as SNOMED, LOINC, etc.)
6. Clinical events monitoring services and rules to prevent medical mistakes (such as CPOE[2], etc.)

[2] Computerized Physician Order Entry

2. CASE STUDY: Implementing a Hospital Information System

The need to interchange healthcare data in a healthcare organization - e.g. a hospital - results in new challenges. Distributed systems **do not always** have compatible data architectures, which creates the need for interfaces that provide the communications between the systems. This in turn creates the need to provide resources for the development and maintenance of each interface. The following real-life experience shows how these issues can be solved.

Overview

When the 'Hospital Italiano de Buenos Aires' in Argentina started to plan its new medical information management system, it found the following environment:

- Several independent, non-coordinated development groups
- Multiple platforms, databases and development tools, including but not limited to:
    - AS400/DB2/COOL:Plex
    - Solaris/Sybase/PowerBuilder
    - Windows NT/SQL Server/Visual Basic
- Multiple proprietary systems with no interrelation and ad-hoc dictionary use

After evaluating several commercial electronic medical record systems, the hospital decided to develop its own in-house system, which took into account the following considerations and needs:

- Strategic and political aspects
- Organizational aspects
- High availability ("24/7" e.g. 24 hours each day of the week)
- Differential support ("24/7")
- Application interoperability
- Semantic interoperability
- Unequivocal patient identification
- Electronic Medical Record (EMR)

Strategic and Political Aspects

Firstly, it was necessary to have political support at the highest institutional levels, to identify leaders of both medical and administrative areas, to set a budget and to estimate realistic but flexible annual deadlines.

The Hospital Information Department had areas such as administrative development, biomedical development, technology and communications, medical computing, rules and procedures administration and an epidemiology, biostatistics and quality area to support clinical management.

Organizational Aspects

New work teams were built and classified as:

- A strategic project follow-up group made up of representatives from hospital administration, medical services and information technology areas.
- A group of representatives of the healthcare area that defined the characteristics of vocabularies, EMR and medical rules.
- A group of representatives of the computing area that addressed technology and networks, messaging (HL7) and middleware, and data security and privacy.

High Availability and Support

It was deemed necessary to implement and support a high-availability environment that would ensure that information technology services would be available 24 hours a day, 7 days a week ("24/7") - the same hours during which medical care is provided by the hospital.

A help desk was also needed which was available anytime to assist all information management system users.

Unambiguous Patient Identification

An electronic medical record (EMR) represents a change in the concept of medical record management. The medical record can no longer be a disjointed collection of fragmented records of patient encounters, but must rather become part of an integrated clinical information system. In order to achieve this, all the information related to a patient must be collated into one single repository, regardless of the place where the medical care was provided. This repository is the EMR.

Data integration is difficult, if not impossible, unless patients are identified unambiguously. This critical issue must be resolved in any health information system - and in any healthcare organization. The problem is not simply to use an existing patient identifier, since no single identifier has proved infallible. The Argentinean National Identification Document, for example, could provide an identifier, but many people have a passport instead, and other patients may have no document that can unambiguously identify them. The solution is therefore not to find suitable patient identifiers, but rather in creating an identification service that ensures the unambiguous identification of patients.

A widely implemented solution to this problem is the creation and maintenance of a Master Patient Index ("MPI" or "PMI") with an associated patient identification service supporting and maintaining it.

Centralized control, maintenance and auditing of the MPI is essential to eliminate duplicated patient records (e.g. one patient who has multiple IDs) and duplicated IDs that belong to different patients.

The true magnitude of this problem is often underestimated as many patient-related administrative processes take place after the patient has left the hospital and therefore errors are often not identified.

An even bigger problem may occur in care-provision scenarios. Incorrect patient identification during an encounter may not only result in serious errors in the patient record but can have dangerous effects on the patient's health - for example, if an allergy to a specific drug (e.g. penicillin) is

incorrectly recorded or if a positive HIV virus laboratory test result is associated with the wrong patient.

Electronic Medical Record (EMR)

The medical record is a **repository of data** collected during a patient's interactions with healthcare professionals. It is a medical document that allows, among other things, a complete understanding of the patient's health situation over many years (e.g. it is "longitudinal" and is therefore invaluable when making a complicated diagnosis or determining a long-term care plan.

The medical record must be included as a source document in an integrated information system to provide comprehensive information leading to more optimal health care.

The electronic medical record (EMR) is a computer system that records patients' demographic and medical data for long-term storage and analysis.

An EMR provides many advantages. It allows quick access to the patient's medical information and supports recording daily evaluations, requesting procedures and medical interventions, accessing results, and even, with the more advanced systems, provides a clinical decision support system - all at the point of care, e.g. next to the patient's bed. It also allows for centralized and safe storage, distributed information access and epidemiological analysis, thus becoming a very useful tool for managing the health of not only the individual but also the population.

3. Interoperability

Interoperability is defined as:

"The ability of two or more systems or components to exchange information and use the information that has been exchanged."[3]

Information interchange is known as **functional interoperability**, whereas the capacity to understand and use shared information is called **semantic interoperability**.

For different computer systems to share a patient's information, the information needs to be transferred from one system to another. Before interoperability standards such as HL7 were created, this transfer was generally done through custom interfaces.

However, as can be seen below, the number of interfaces grows exponentially as the number of connected systems increases:

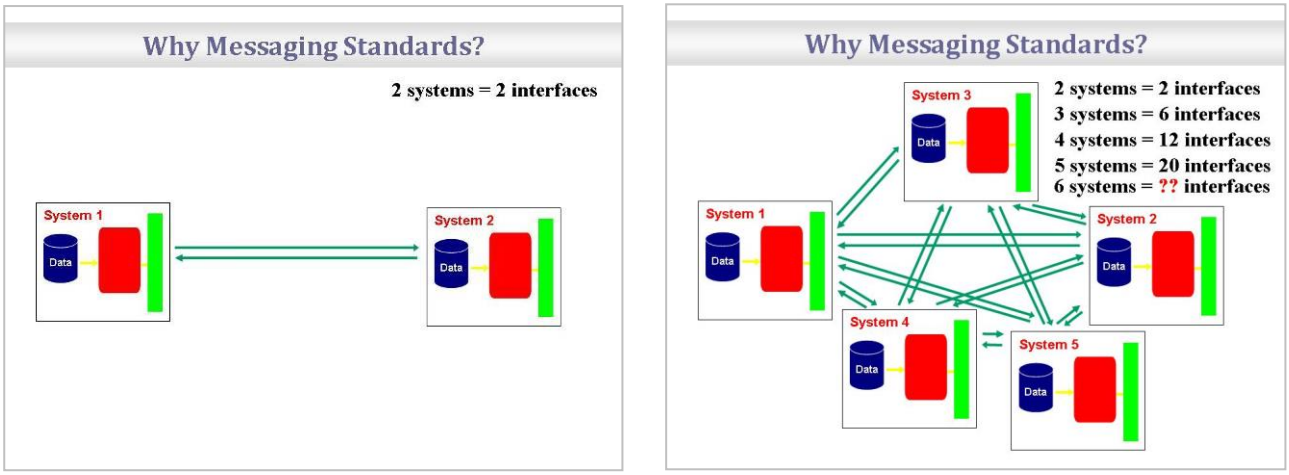

Figure 5: As more systems are interconnected, the number of point-to-point interfaces increases exponentially!

If all computer systems in a healthcare organization are to exchange data, the increase of the number of interfaces follows the following formula: **Interfaces = (n x (n-1))** Therefore the number of healthcare data interfaces in a regional or national health service could be more than 10,000!

[3] IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries, 1990

How can this Problem be Solved?

Using a single "hub" interface with a standardized message syntax, such as HL7 messaging, is one way to solve this multiple interface issue:

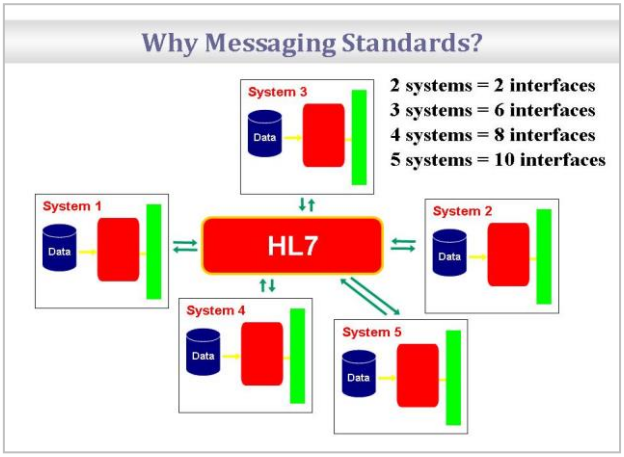

Figure 6: A common interface "hub" using a common communications protocol, e.g. HL7

These messages, in turn, use standardized data (semantics), such as patient identification and lab data. To maximize the semantic interoperability, it is common to define coded values for many fields based on standard controlled vocabularies, such as those discussed in the following Units of this Course.

HL7 and related protocols for electronic health data interchange can allow clinical applications to communicate with each other regardless of their technology platforms or development languages.

4. Why is use of Standards so Important?

Information communication is a key component in any system. In the health area, information is transferred among health-care professionals, institutions, decision support systems, etc.

Effective communication requires that information senders and receivers share a common "reference framework" that enables all interactions to be unambiguously understood. Standards provide this common framework, promoting uniformity in the definition and identification of health system components, whether they are objects, diagnosis, people, interventions, etc.

Developing an individual ("ad hoc") interface solution for each problem as it arises can appear to be relatively quick and inexpensive in the short term. However, these custom interfaces have a very limited use and can be difficult or impossible to adapt to new requirements that arise as the organization and its systems and processes grow.

Standard solutions may initially appear to be more complex to implement - but they can be adapted and scaled to many different scenarios, thus providing an essential framework for growth.

The need for interoperating systems is evident in each part of the healthcare organization. For example, a physician treating a patient in the Emergency Room (ER) needs to have access to the patient's encounter history, a list of current medications, and recent lab results in order to be able to best provide optimum treatment. In order to ensure proper follow-up, the ER physician also needs to be able to communicate treatment and outcomes to the patient's general practitioner. It would not be practical to expect hospital staff or the patient to collect each piece of information from multiple information systems that are not connected, or to wait for a medical briefing, which may lack much of the linked information mentioned above. Interoperable systems should share information continuously and automatically through the participating institutions and display the information in a useful way.

Interoperability requires creation, acceptance and implementation of standards to ensure that data is available and retains its meaning and context through the various processes of clinical care, in any part of the health system.

Health-care standards implement rules that govern the way patient information is electronically stored and interchanged. Ideally, a single set of standards would provide efficient access to text, numeric and image data, allowing information to be shared appropriately by health professionals, payers, administrators and consumers.

Certain industries, such as banking and telecommunications, have developed and implemented international standards for electronic data interchange. In healthcare, the development of national and international has faced formidable challenges. One of the reasons appears to be that patient records are typically accumulations of interactions involving health professionals, patients, insurance companies and governmental agencies. The data they contain is often not uniformly categorized and may contain large amounts of free text and images.

Therefore, it is not surprising that clinical data standards are sometimes seen as a complex collection of different vocabularies and obscure technical details.

It is important for managers, general practitioners and healthcare policymakers to understand the basics of standards. Key policy decisions are routinely made regarding how and when standards are to be implemented to ensure the optimal provision of health care. If these decisions are based on incorrect information, mandating standards that are unsuitable for the purpose, or require implementation approaches that are flawed, then serious degradations of healthcare outcomes are possible!

Categorizing Standards

Interoperability depends on two significant components:

1. Syntactic (functional) interoperability
2. Semantic interoperability

"Syntactic" or "functional" refers to the structure of a communication; it can be thought of as equivalent to grammar rules in spoken languages.

Semantics hold the meaning of a communication, the equivalent of a dictionary or thesaurus. Terminologies such as SNOMED CT, LOINC, the ISO code descriptor lists of countries, spoken languages and currencies, the list of US States (CA, TX, DC, etc.) are examples of semantic standards.

Without semantic interoperability, information may be interchanged but there is no certainty that the receiver can use or understand it.

The healthcare standards currently available address both components of interoperability and are grouped into the following six categories:

1. **Messaging and data interchange standards**: These allow for automatic health data exchanges between systems and organizations: specifying format, data elements and structure. Common standards include HL7 for administrative and clinical care data, DICOM for diagnostic imaging (X-ray, radiology, ultrasound, etc), NCPDP (only in the USA) for retail prescriptions, X12N (only in the USA) for insurance eligibility and claiming, etc.
2. **Terminology standards:** These vocabularies provide specific codes for clinical concepts such as diseases, problem lists, allergies, medications and diagnoses. Examples of terminology standards are LOINC (Logical Observation Identifiers, Names and Codes) for lab results, SNOMED (Systematized Nomenclature of Medicine) for clinical terms, ICD (International Classification of Diseases) for diseases, etc.
3. **Document standards**: These indicate the types of information that may be included in documents and where information can be found in documents. A common standard in paper medical records is the SOAP format (Subjective, Objective, Assessment, Plan)[4]. The Continuity of Care Record (CCR - a healthcare ASTM standard mainly used in the USA) provides a standard data set for electronic referral among healthcare professionals that includes patient identification information, encounter and treatment records, medications, allergies and recommendations for the care plan. The HL7 Clinical Document Architecture (CDA) is an interchange standard for any

[4] This should not be confused with the Simple Object Access Protocol (SOAP) used in information technology

clinical document, such as progress notes, surgery reports, discharge summaries, etc. The Continuity of Care Document (CCD - an HL7/ASTM joint standard used mainly in the USA) enables representation of the CCR data set as a CDA document using XML encoding.

4. **Conceptual standards**: These provide a framework for understanding the concept of clinical data and how it can be exchanged between systems without losing meaning or context. For example, the HL7 Reference Information Model (RIM) provides a framework to model and describe clinical data and the surrounding context - e.g. "who, what, when, where and how".
5. **Application standards**: These determine the way business rules are implemented and how users interact with software systems. Examples include user authorization (openID, OAuth, etc.), single sign-on, which allows users to access multiple applications from the same desktop environment, as well as standards produced by the HL7 Clinical Content Object Workgroup (CCOW) that provide for a comprehensive desktop view of information from multiple, non-integrated clinical systems.
6. **Architecture standards**: These define the logistical processes involved in data storage and distribution. Examples include the Public Health Information Networks (PHIN) of the US Centers for Diseases Control and Prevention (CDC)[5] and the US National Disease Electronic Surveillance System[6] that both use HL7 standards.

[5] More information at www.cdc.gov/phin/resources/PHINguides.html

[6] See www.cdc.gov/nndss/

5. Standards Development

How are Standards Developed?

There are four basic standards development mechanisms:

1. When a group informally agrees to use common processes whose details are not generally published, these processes are called "**ad hoc standards**".
2. **De facto standards**, such as those for computer operating systems, are those imposed by their popular use or market acceptance.
3. Governments and jurisdictions can determine and impose standards to be used in particular scenarios; these are called "**de jure standards**".
4. **Consensus-based standards** such as those developed by HL7, which result from all parties interested in using a standard meeting in an open and democratic environment to discuss and reach consensus on the definition of the standard.

Healthcare information standards are typically developed by volunteers working in committees or work groups organized around interest communities. Interested parties may include general practitioners, researchers, health informaticians, chief information officers (CIOs), database managers, information system analysts, project directors, managers, etc. Organizations with special interests in public health, patient safety and electronic records may participate to ensure that the standards will be relevant to their areas of concern.

The success of any standard depends on the credibility of the organization developing the standard, the competence and diligence of its committees and the ability of the organization to achieve industry adoption. It is also important that the committees have enough members in each applicable sector of the industry to provide a representative cross-section of domain experts.

Early adopters generally come from within the standards development community. They validate the adequacy and efficacy of the standard and may also be seen as leaders ("evangelists") communicating the standard to the wider audience of users. Ultimately, the standard may be accredited or otherwise approved by an external body such as the International Standards Organisation (ISO), the American National Standards Institute (ANSI) for the USA, DIN for Germany, etc.

Who creates Standards?

Standards have been created by a variety of healthcare organizations, including service provider entities, management staff, vendors, and independent advisory bodies.

The leading international standards development and coordination organization is the International Standards Organisation (ISO), whose Technical Committee number 215 ("TC/215") creates and approves health informatics standards. This committee is composed of representatives from the national standards bodies such as ANSI, DIN, BS, CSA, Standards Australia, IRAM, etc.

Technical Committee 251 of the Comité Européen de Normalization (CEN - European Committee for Standardization) is another important participant in health information standards development representing the European Union.

Health Level Seven (HL7) focuses on developing international health information standards. It has offices in the USA and Belgium and holds Working Meetings three times a year in various countries around the globe.

In the United States, the American National Standards Institute (ANSI) creates and approves standards. It also accredits standards development organizations (SDOs) such as HL7 to create standards for America.

The American Society for Testing and Materials (now known as "ASTM International") publishes thousands of standards related to materials and materials testing. It participates in health information standards creation through a single small subcommittee (ASTM E31).

Table 2 summarizes the key standards and the organizations responsible for developing and maintaining them. They are listed in the six categories mentioned above.

Data Exchange/ Messaging

Standard	Acronym	Description	Developer
Health Level Seven Messaging Standards Version 2 and Version 3	HL7 V2.x HL7 V3	Electronic message formats for clinical, financial and administrative data. V2 is commonly used in commercially available software. Publication of V3 began in January 2005.	HL7 International www.HL7.org
Digital Imaging and Communications in Medicine	DICOM	Format for communicating radiology images and data.	US National Electronics Manufacturers Association www.NEMA.org
Clinical Data Interchange Standards Consortium	CDISC	Format for reporting data collected in clinical trials.	Clinical Data Interchange Standards Consortium www.CDISC.org
National Council for Prescription Drug Programs	NCPDP	A full suite of standards supporting pharmacy benefits management and retail pharmacy operations in the USA.	The US National Council for Prescription Drug Programs www.NCPDP.org
X12N Insurance EDI Transaction Sets	X12N	EDI transactions for claims, eligibility and payments. Mandated for use in the USA by HIPAA	US ANSI Accredited Standards Committee X12 www.X12.org
Standard for Medical Device Communications	IEEE1073	Messages for medical device communications.	Institute of Electrical and Electronics Engineers Standards Association www.IEEE.org

Terminologies

Standard	Acronym	Description	Developer
International Classification of Diseases	ICD-9 (1978) ICD-10 (1993) ICD-11 (WIP)	Coding system for Diagnosis/ disease and procedure/operation codes - used for research, statistics and billing/ claims. Various Modifications in use worldwide(CM, AM)	World Health Organization www.WHO.int
Logical Observation Identifiers Names and Codes	LOINC	Concept-based terminology for lab orders and results.	US Regenstrief Institute for Health Care www.LOINC.org

Standard	Acronym	Description	Developer
Systematized Nomenclature of Medicine - Clinical Terms	SNOMED CT	Mapping of clinical Concepts with standard descriptive terms.	International Health Terminology Standards Development Organization www.IHTSDO.org
Unified Medical Language System	UMLS	Database of 100 medical terminologies with concept zapping Tools.	US National Library of Medicine www.NLM.NIH.gov

Documents

Standard	Acronym	Description	Developer
Continuity of Care Record	CCR	Data set that gives a snapshot of a patient's core data and recent encounters (allergies, medications, treatment, care plan) and makes it available to subsequent care providers on referral. Mainly used in the USA.	ASTM International, E31 Committee on Health Informatics www.ASTM.org
Continuity of Care Document	CCD	Representation of the CCR data set as a CDA document encoded in XML.	HL7 International www.HL7.org ASTM www.ASTM.org
Clinical Document Architecture	CDA	Standard exchange model for clinical documents such as discharge summaries and progress notes.	HL7 International www.HL7.org

Conceptual

Standard	Acronym	Description	Developer
HL7 Reference information Model	HL7 Reference Information Model (RIM)	Shared, generic model that facilitates interoperability through standardization of all domain models to a norm. The RIM is referenced by V3 messages, CDA documents and the FHIR standard.	HL7 International www.HL7.org

Applications

Standard	Acronym	Description	Developer
HL7 Clinical Content Object Workgroup Standard	CCOW	Standard for providing comprehensive desktop view and single sign-on capability across systems without integrating databases	HL7 International www.HL7.org

Architecture

Standard	Acronym	Description	Developer
Public Health Information Network	PHIN	Components of an electronic surveillance and management system for bioterrorism and public health preparedness data in the USA. Formerly the National Electronic Disease Surveillance System (NEDSS).	US Centers for Disease Control www.CDC.gov/phin

The Healthcare Standards Development Process

A healthcare interoperability standard is generally developed in five stages.

The first stage is **identification**, where the issues and scope that will be covered by the standard are identified.

During **conceptualization** it is decided how the problem will be solved and how the standard will be created. Then, all the interested parties discuss whether and how the resulting model should adjust to their needs.

At this point, the standard **specification** is created and a draft standard is published. Then, in the early implementation stage, industry leaders begin using the standard to see how it performs in real life. The experience gained from early implementation is used to produce the final specification.

In the **adoption** stage, the rest of the industry implements the standard.

Finally, in some cases, **certification** processes are created, where a certifying entity verifies that a software program or process complies with the standard.

6. Introduction to Vocabularies in Healthcare.

"If we cannot name it, we cannot control it, finance it, research it, teach it or develop public policies..."

Norma M Lang PhD FAAN FRCN RN

Professor and Dean Emerita, University of Pennsylvania School of Nursing

Semantic interoperability - as outlined in the previous Unit - is the way in which, once data has been collected, information can be meaningfully interpreted and incorporated into the receiving system. In order to achieve this type of interoperability for any healthcare record, we need to use the same "language", the same **vocabulary**.

The use of computers in managing healthcare information has served to highlight the complexity of language handling, especially medical vocabularies. Therefore, it is necessary to use vocabulary control strategies so that either the clinical information stored in Health Information Systems can be shared, for administrative purposes or in making clinical decisions (perhaps incorporating the use of automated decision-support tools) in ways that maximizes the quality and safety of patient care.

*In order to understand the **problems of language,** please do the following exercise: take a piece of paper, imagine a **screen** and make a simple sketch of what the screen you imagine looks like.*

When you have finished your sketch, review the following definitions:

Definitions

Concept: An idea, action or thing with one single meaning. Example: 

Word: A set of characters expressing a concept that can be found in a dictionary. Example: "S-C-R-E-E-N"

Term: One or more words used as a unit. Examples: "Computer Screen", "LCD Screen", etc."

Synonyms: Two or more terms for the same concept. Examples: "Computer Screen", "Computer Display Screen", "Screen Display", "LCD Screen", etc.

Homonyms: Words or terms that are spelled and pronounced alike but have different meanings

Code: An expression of a term or concept in a formal notation or classification

• Natural Language Problems

A few paragraphs ago we asked you to imagine and sketch a "screen". What does your sketch show?

Each of you may have come up with a different sketch because many interpretations of the word "screen" are possible! As we see in Figure 1 below, there are computer screens, classroom projection screens, TV screens, movie screens, etc. - and even sunscreens!

Figure 7: What is a "screen"?

What is the reason for these different interpretations of the word "screen"? Without the contextual information, it is not clear what type of screen we wanted you to sketch.

So if you are in a computer store and ask for a "screen", the salesperson will probably show you the latest models of LCD monitors and if you are in a home entertainment section of a large store you will most likely be demonstrated state-of-the-art TVs. However, if the context is that you are in a drugstore or pharmacy and you ask for a "screen", the salesperson will likely offer you different brands of sunscreen or sun lotions!

The purpose of this exercise was to show you that people communicate and understand each other not only because they use the same language and words, but also because there is a commonly understood context surrounding these words.

Figure 2 below represents the communication process among people. Unlike people, computers cannot know the context in which messages are sent. That is why it is vitally important to explicitly state the context.

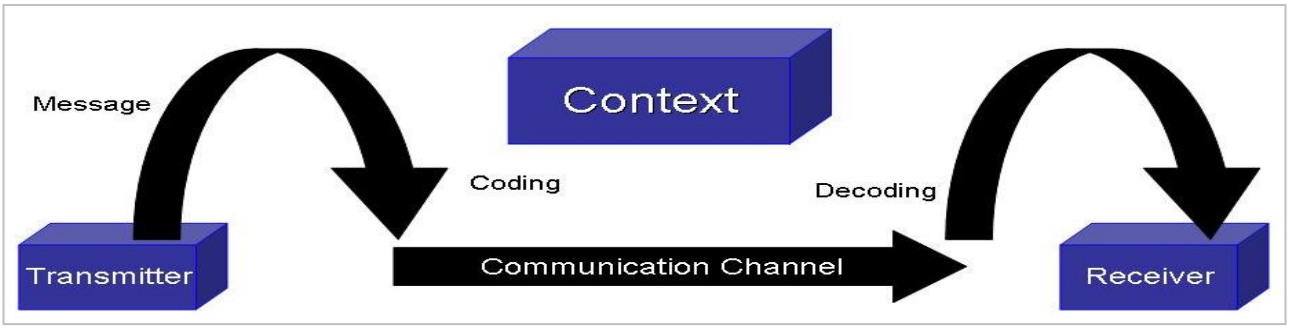

Figure 8: Successful Communication process between people

Both the sender and receiver must be familiar with both the language (encoding) and the context so that the content of a message can be interpreted the same way by both parties. Therefore, knowing the environment and the audience is essential in the communication process.

When the communication channel cannot fully depend on external context, as is the case with computerized information systems, the message sender must be as specific enough with both data and context so make sure that the recipient can understand the message without ambiguity.

This is also critical in patient identification. For example, a search for a name such as "John Smith" may refer to thousands of different people, but if an identification number and the date of birth are added to the search, there is a much higher chance of identifying the right person. Even so, when trying to be specific in health-care communications, ambiguity problems inherent to natural language can be encountered that must be resolved when dealing with medical language.

Ambiguity occurs when a language or vocabulary allows for more than one meaning for a single word or expression of a concept. Often, resolution is only possible when the context or situation is known.

This ambiguity arises from any one or more of the following three effects:

Synonymy: Relationship of similarity in the meaning of certain words called ***synonyms***; for example "fever" and "pyrexia".

Polysemy: The capacity of a single term to express many different meanings. For example, the word "mouth" may refer to the opening in a person's face', the opening of a cave or where a river flows into the ocean. Similarly, the term "Paget's Disease" refers to two completely different ailments, one affecting the breast and the other the bones (and has additional meanings besides).

Homonymy: Two terms with the same pronunciation and/or spelling that mean difference things. For example, the noun "rose" (flower) has a different source than the verb "rose" (past tense of "rise").

The difference between homonymy and polysemy is that in homonymy there are two or more unrelated source words whereas in polysemy the source is a single source word. Therefore, in order to identify a homonym the origin of a word (etymology) must be studied.[7]

Here is an example of synonymy in clinical practice: If a doctor wants to communicate to a nurse that a patient's body temperature is too high, the doctor may also use any of the following alternative expressions:

"The patient has hyperthermia"
"The patient is pyretic"
"The patient is in a febrile state"
"The patient has a temperature above 38° C"
"The patient has a temperature above 100° F"
etc …

Despite the synonymy, the two healthcare professionals understand each other - because they both work in a healthcare context!

Representing the same concept with different words may not be a problem, as long as there is a known context facilitating the communication process. As was said above, communication be-

[7] The Online Etymology Dictionary describes the origins of English-language words: www.etymonline.com

tween professionals and computers requires additional information about the context that may be omitted in communication among people.

Here are' several examples of ways that the blood type/group of a patient could be communicated between two computers:

Blood Type A+	Blood Group A+	Blood Type A positive
ABO Group A+	BLDTYP A+	ABOTYPE Type Apos

So even without detailed knowledge of how a sender expresses the blood group/type information, a clinician reading the information received can infer that each of the expressions above indicates a patient's blood group/type of "A positive". The computer however, may not unambiguously understand the information, because **what is obvious for a human is unclear for a computer!**

• Knowledge Representation

Knowledge representation is an area of artificial intelligence (AI) that represents information about the world in a form that a computer system can utilize to solve complex tasks such as diagnosing a medical condition or having a dialog in a natural language. Knowledge representation incorporates findings from psychology about how humans solve problems and represent knowledge in order to design formalisms that will make complex systems easier to design and build.[8]

One of the biggest challenges for medical information systems is the reliable representation of medical knowledge. In some systems, decision support modules interpret patient data entered through the electronic medical record and access knowledge bases (such as pharmacology databases, the Cochrane library, electronic books and other sources of information with scientific evidence levels) in order to generate reminders, warnings and in some cases therapeutic recommendations. Therefore, any vocabulary of medical words (or "terms") must be strictly controlled, because even a minor natural language ambiguity can result in errors dangerously impacting patient safety.

To understand the features and components of a controlled vocabulary, it is necessary to clearly define the types of controlled vocabularies shown in Figure 3:

Natural Language: The universe of expressions used to communicate ideas. It may include words from native languages (English, Spanish, Portuguese, Hindi, Mandarin, etc). Generally, the narrative text of medical records is expressed in natural language.

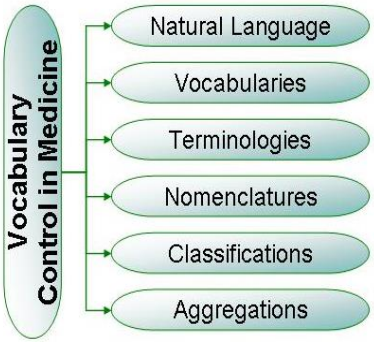

Figure 9: Vocabulary Control

Vocabulary: A set of terms used or available for use by an individual or group, or within a specific type of work or knowledge field ("domain"). The vocabulary of physicians is different from the vocabularies of astronauts or Olympic athletes.

Similarly, we see many different types of vocabularies in medical records: laboratory data, medication lists, surgery reports, dietary requirements, care plans, etc as well as natural language narrative text. Because patients do not generally communicate using a medical vocabulary, the scope of the medical communications must extend to the natural language vocabulary.

Controlled Vocabulary: An approved set of words or terms limited to those used in a specific environment. In an electronic user interface, a controlled vocabulary may provide a limited "pick-list" of terms to choose from in order to enter an item of medical data. Numerous controlled vocabularies have been developed for use in clinical applications.

Controlled medical vocabularies are classified according to their characteristics or organization:

[8] See http://en.wikipedia.org/wiki/Knowledge_representation_and_reasoning

Terminology: a set of concepts, designations and relationships for a specific subject area or special reference within a discipline, i.e. a collection of words or phrases called terms, aggregated in a systematic fashion.

Taxonomy: a terminology ordered according to the logical relations of terms regarding a particular point of view. For example, if the terms in the terminology of a specific diagnosis are ordered according to their meaning, a taxonomy is created.

Nomenclature: a sub-set of a terminology for a specific subject area. It is made up of the terms (or groups of terms) from a terminology and their relationships. A nomenclature does not occur naturally as a result of use or custom, but is created by some official body that organizes (and/or standardizes) the terminology.

Classification: A classification is an orderly system of concepts belonging to a specific subject area, with implicit and explicit order principles. Their definition depends on their expected use.

Depending on its application in an information system, a controlled vocabulary may also be characterized as one of the following:

Reference Vocabulary or Reference Terminology: A controlled vocabulary used for a more detailed representation of data in an information system. The reference terminology is used to store data in the database. Multiple interface terminologies converge into a single reference terminology. Generally, a nomenclature is used that must incorporate all the domains in which a set of interface terminologies applies.

Interface Vocabulary or Interface Terminology: A controlled vocabulary is a list from which the user can choose a term to enter into a system. Such a list may include all the lexical varieties, acronyms, abbreviations and jargon that are employed by a system's users, each with an implicit meaning drawn from the context they are used in. Within the same system, several interface terminologies can be used, having been adjusted to a specific type of user.

Output Vocabulary: A Terminology or classification used for information analysis. Terminology information systems must provide tools to recover information represented with the reference terminology and transform it into the desired output vocabulary.

Vocabulary control is the strategy by which automated systems implement solutions to the natural language problem. This strategy is applied by limiting vocabularies to a strict set of terms (therefore called "controlled vocabulary") and involves agreeing on the exclusive use of those terms for information expression.

A Controlled Vocabulary consists of a combination of restricted terms and grammar rules.
A Controlled Vocabulary is a key component in achieving the interoperability of health information systems!

The main characteristics of a controlled vocabulary are:
A strict set of terms, which are unambiguous and accurate
The set of terms is standardized
New concepts require integration into the vocabulary and may not be introduced *ad hoc*
Users may require training before using the vocabulary

In healthcare information technology, controlled vocabularies not only facilitate system interoperability, but also enable statistical and epidemiological analysis, reports for computer-assisted decision-making, planning of care and follow-up strategies, etc.

Natural Language	Controlled Vocabulary
Ambiguous	Unambiguous
Many occurrences of synonymy and polysemy	Restricted terms
Highly context dependent	Context is clearly defined
Very expressive and flexible	With restriction in the different levels
New concepts are easy to express and add	New concepts require integration into the vocabulary
Not standardized	Standardized
Does not require specialized training	May require training before being used
	Rigid, accurate
	Sub-group of natural language:
	May include grammar rules

Figure 10: Natural Language vs Controlled Vocabulary

• Why do we Need Controlled Vocabularies?

Controlled vocabularies are essential to system interoperability. What else makes these vocabularies essential?

They can be used for:
Standardizing free text or structured content of the medical record
Representing clinical observations and evaluations
Coding tests and results
Identifying drugs
Interchanging clinical data in real time
Representing syntactic and semantic aspects of medical concepts
Recovering and analyzing data, and supporting the decision-making process

The choice of vocabulary is influenced by each vocabulary's characteristics. It is important to remember that each vocabulary is created with a particular purpose. For example, it is generally not advisable to use a vocabulary that was created for medical practice billing for epidemiology purposes.

• Examples of Controlled Vocabularies

The following list contains the main controlled vocabularies used in healthcare and outlines their content and purpose.

SNOMED Clinical Terms (SNOMED CT)

The Systematized Nomenclature of Medicine (SNOMED) is a medical nomenclature originally developed by the College of American Pathologists (CAP). It includes terms of all medical domains, including veterinary medicine.

In 2007, SNOMED was transferred from the College of American Pathologists (CAP) to the International Health Terminology Standards Development Organization (IHTSDO). IHTSDO has an international management board whose members have extensive experience in medical information technology.

The current version, SNOMED CT, results from its combination with the Read Codes (the U.K. clinical use nomenclature) to create an extensive and detailed nomenclature that is strongly clinically oriented with international validity.

SNOMED CT is the most comprehensive vocabulary available to describe clinical findings, diseases, procedures, etc. Its main features are:
A compositional focus that allows the combination of simple terms (lung + inflammation), or the addition of modifications to a concept (severe, mild, sudden onset, etc)
Interface and reference vocabulary functionalities

Due to its level of detail and quality of semantic relations, SNOMED CT is highly suitable for use as a reference vocabulary. SNOMED CT also includes interface functionalities - for each concept there are several possible descriptions that can be used as elements of an interface vocabulary in an implementation.

SNOMED CT contains
More than 365,000 concepts
Almost one million descriptions
Nearly one and a half million relationships

SNOMED CT is not merely a diagnostic nomenclature; it attempts to embrace the whole spectrum of controlled vocabularies within the healthcare domain. Its 365,000+ concepts are grouped in hierarchies. The following Figure 5 shows some of the hierarchies with examples:

Hierarchy	Example
Findings	Swelling of arm
Disease	Pneumonia
Procedure/intervention	Biopsy of lung
Observable entity	Tumor stage
Body structure	Structure of thyroid

Hierarchy	Example
Organism	DNA virus
Substance	Gastric acid
Pharmaceutical/biologic product	Tamoxifen
Specimen	Urine specimen
Qualifier value	Bilateral
Physical object	Suture needle
Physical force	Friction
Events	Flash flood
Environments/geographical locations	Intensive care unit
Social context	Organ donor

Figure 11: SNOMED CT Hierarchies with Examples

SNOMED CT is concept-based terminology. Each medical concept is uniquely identified and the concepts are related to each other by hierarchical relationships. The concepts and their relationships can be understood by an information system.

Here are some examples of the possible relationships used to describe diseases or clinical findings:

- place of finding
- associated with
- after its
- causal agent
- due to
- associated morphology
- severity
- course
- episodicity
- pathological process
- etiology
- finding
- followed by

For example, in SNOMED CT there is a relationship between the "diabetes" concept and the "diabetic foot" concept, through a "due to" attribute ("diabetic foot" - "due to" - "diabetes"). These semantic relationships provide information, which decision support systems use to apply rules.

Therefore, it can be very effectively used as a reference clinical terminology to unambiguously record all relevant aspects of medical care in standardized terms.

An interface terminology may incorporate jargons, localisms and lexical variants of each institution and therefore must always be mapped to a reference terminology in order to be able to correctly represent medical knowledge. This reference vocabulary must meet the following characteristics:

Domain scope or coverage: Must include terms for all possible objects or events that may be collected, as the terms in the vocabulary must represent the entire domain.

Non-redundancy: Mechanisms must be established that prevent multiple terms for the same concept from being added to the vocabulary as different concepts.

Synonymy: Non-unique multiple terms must be accommodated for the same concept.

Explicit (not vague): Incomplete meanings must be avoided. E.g.: *ventricle* is not a completely meaningful concept; it means nothing unless it is clarified that it is a *cardiac ventricle* or *brain ventricle*.

Unambiguous: Each concept must have a single meaning; in case of homonymy, each concept must be stripped of its ambiguity. *E.g.*, "Paget's Disease" must be divided into two concepts: "Paget's disease of bone" and "Paget's disease of breast".

Multi-classification or polyhierarchy: The system must not be restricted in such a way that the same concept cannot be assigned to as many classes as necessary. E.g., *pneumonia* is a descendant of both *pulmonary* and *infectious disease*.

Context consistency: Concepts that exist in several classes must be allowed to have the same meaning in each of them. Inconsistencies must be avoided. E.g., "corticosteroid," in the context of both "hormones" and "anti-inflammatory drugs", must have identical attributes, for both itself and its descendants.

Relationship explicitness: The meaning of the relationships among concepts must be clearly stated. E.g., the relationship between "Staphylococcal pneumonia" and "pneumonia" must be differentiated from the relationship between "pneumonia" and "staphylococci"; the former is a class relationship whereas the latter is an etiological relationship. Pneumonia "is a" Staphylococcal pneumonia, pneumonia "is caused by" Staphylococci.

Concept permanence: Old concepts must not be deleted. Replacement of existing concepts by better concepts must also be supported: E.g., when "HIV infection" replaces "AIDS", there must be an explicit link between the two concepts.

The use of "not classified somewhere else" must not be allowed; such terms cannot be used in a reference vocabulary.

In SNOMED, a single concept can map a broad number of terms: myocardial infarction, cardiac infarction, heart attack, myocardial infarct, MI - Myocardial infarction, infarction of heart, etc. all map to concept ID 22298006 in SNOMED CT; see it in a terminology browser below:

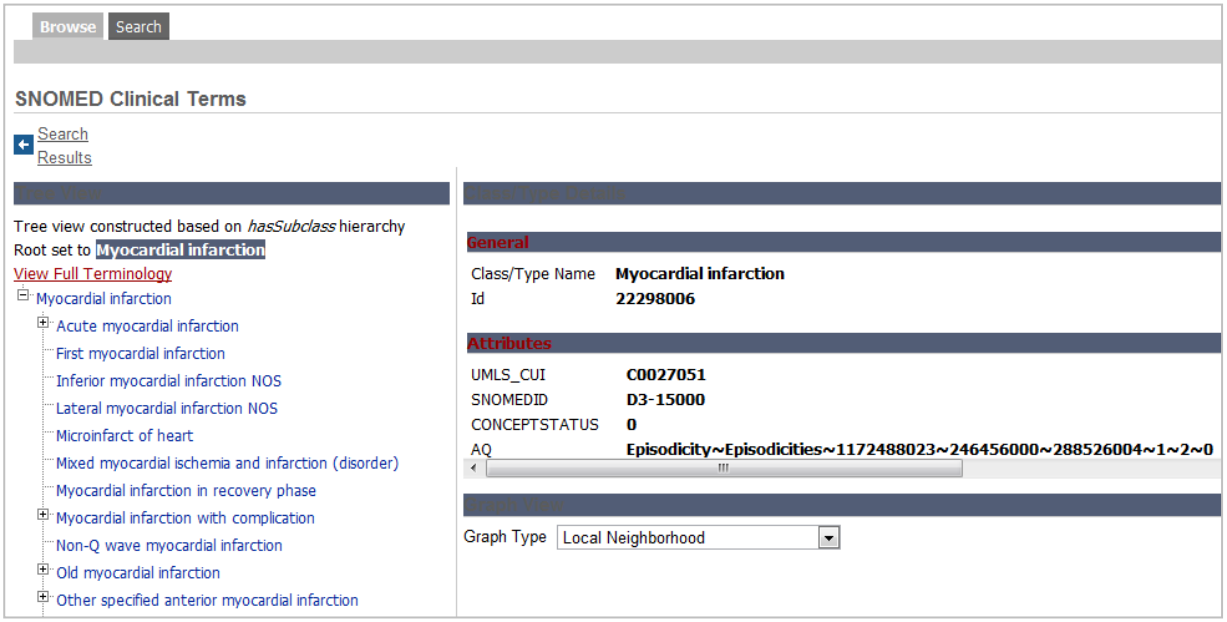

Figure 12: The Term "Myocardial Infarction" in a SNOMED CT Browser

SNOMED Terms can be searched online here: http://phinvads.cdc.gov/vads/SearchHome.action

International Classification of Diseases (ICD)

The World Health Organization (WHO) created the International Classification of Diseases (ICD) and has maintained it for more than 120 years.

Remember that a classification is an orderly system of concepts belonging to a domain with implicit and explicit principles of order and class definition, depending upon expected use.

ICD originated as a classification of causes of death and is used today to report morbidity as well as mortality. It was first created in 1892; its most recent revision in 1992 was the tenth (ICD-10).

In some countries, ICD was adapted ("localized") to better meet the local needs. For example, in the USA, codes were added to ICD to allow its use for billing and the CM (Clinical Modification) variant was created, currently in its 10th edition (ICD-10-CM). In Australia, ICD-10-AM (Australian Modification) was created, which is also used in Ireland, Slovenia, New Zealand and Singapore.[9]

ICD-10 has no procedures, only diagnoses, and has progressed from a numeric organization of chapters to an alphanumeric organization, which can be mapped back to previous ICD editions. ICD-10 is today considered the worldwide standard for mortality and morbidity reporting.

ICPC-2 (International Classification of Primary Care)

ICPC (International Classification of Primary Care) is a classification created by the World Organization of Family Doctors (specifically the WONCA[10] International Classification Committee) to be used exclusively in primary care. First published in 1987, it is the result of more than 25 years of experience in primary care classification.

It contains approximately 700 terms and has sufficient granularity for use by general practitioners in primary care or outpatient environments. It allows representation of reasons for consultations, diagnosis, diagnostic procedures, administrative procedures, etc.

ICPC has a solid mapping to ICD-10 that allows using it as an access methodology to ICD-10.

ICPC is used optionally within a "care episodes" data model. "Care episodes" are made up of one or more consultations; each consultation is codified according to the reason for consultation, the diagnosis and any resulting plan. The episodes are linked to reflect the process of the diagnosis.

For example, a consultation due to a dry cough may initially be recorded with a preliminary diagnosis of "cough" and a thoracic x-ray may be requested. A second consultation is then scheduled for interpreting the x-ray and a tumor is detected, resulting in a new diagnosis of "lung cancer". This episode therefore has two consultations and it demonstrates the progression from the identification of the symptom to a final clinical diagnosis.

Diagnosis Related Groups (DRG)

Diagnosis Related Groups (DRG) were originally developed in the early 1980's by a multidisciplinary team at Yale University in the USA to compare the quality of treatment of groups of inpatients with a similar diagnosis at different hospitals. In 1984, the US Department of Health adopted DRGs for funding Medicare and Medicaid patients in hospital as a response to continually rising medical costs and deep economic deterioration in healthcare in the late 1970s. Rather than simply paying hospitals whatever costs they charged to treat public patients ("fee for service"), the

[9] See http://nccc.uow.edu.au/icd10am-achi-acs/overview/icd10am/index.html

[10] World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians

DRG-based model paid hospitals a predetermined amount based on the patient's diagnosis. This was the most significant change in US health policy since 1965.[11]

A DRG is a single code (eg "411" Cholecystectomy - Removal of Gall Bladder) that combines the principal diagnosis as well as any secondary or additional diagnoses (co-morbidities) or complications, procedures, plus age and sex and length of stay that occurred within an acute hospital episode of care. For some DRGs - birth-weight of newborns, discharge status (e.g. alive), days on mechanical ventilation and some other factors also affect the final allocation of the DRG.

DRGs are initially classified into medical or surgical categories, and then further defined by the presence of minor or major complications or co-morbidities: e.g.:

411 Cholecystectomy with major complications and/or co-morbidities
412 Cholecystectomy with complications and/or co-morbidities
413 Cholecystectomy without complications and/or co-morbidities

A DRG therefore reflects the seriousness of the disease, the complexity of the treatment, the resource consumption, the length of hospital stay and other factors that affect the cost of treating the patient.[12] DRGs provide a statistically meaningful method of comparing the treatment, resource use and length of stay of similar patients in different hospitals.

ATC (Anatomical-Therapeutic-Chemical)

The ATC vocabulary is considered an important controlled vocabulary for drugs. It is part of the World Health Organization's drug dictionary. First published in 1976, ATC classifies drugs according to anatomical, therapeutic and chemical criteria.

The ATC pharmaceutical coding system divides drugs into different groups according to the organ or system on which they act and/or their therapeutic and chemical characteristics. This means that one drug can have more than one code. For example, acetylsalicylic acid ("aspirin") has the code "A01AD05" as a drug for local oral treatment, "B01AC06" as a platelet inhibitor and "N02BA01" as an analgesic and antipyretic. On the other hand, several different brands share the same code if they have the same active pharmaceutical substance and indications.

The ATC vocabulary has international characteristics, combining the clinical experience of more than 34 countries and every year is updated with some 2,000 new drugs.

National Drug Codes (NDC) & RxNorm

The National Drug Codes (NDC)[13] is a drug classification of the US Federal Drug Administration (FDA). Each drug is identified by an 11-digit code that is made up of three parts. NDC has flaws, such as the inability to group certain similar drugs, which has led to the development of a new, more functional vocabulary called RxNorm, which has a much more solid semantic structure.

---

[11] From: Mayes, Rick, "The Origins, Development, and Passage of Medicare's Revolutionary Prospective Payment System" Journal of the History of Medicine and Allied Sciences Volume 62, Number 1, January 2007, pp. 21–55

[12] More information on DRGs at http://en.wikipedia.org/wiki/Diagnosis-related_group

[13] More information at www.fda.gov/drugs/informationondrugs/ucm142438.htm

The US RxNorm provides normalized names for clinical drugs and links its names to many of the drug vocabularies commonly used in pharmacy management and drug interaction software, including those of First Databank, Micromedex, MediSpan, Gold Standard Drug Database and Multum. By providing mappings between these vocabularies, RxNorm can facilitate semantic interoperability between systems not using the same drug vocabularies.[14]

Logical Observation Identifiers, Names and Codes (LOINC)

LOINC[15] is a classification for clinical observations. It is primarily used for lab results, but can also be applicable to aspects of a physical examination or any other clinical observation.

LOINC was developed by the Regenstrief Institute at the University of Indiana.

For each observation, the following is specified:
Properties - type of measure, e.g., concentration, numeric fraction, etc.
Time - point in time
Sample - e.g., blood, cerebrospinal fluid
Method - e.g., qualitative, quantitative, and it sometimes include whether it is automatic or manual, etc.

These aspects are represented within a text system with predefined separators and abbreviations to specify each dimension (see Figure 7 below).

```
Blood glucose GLUCOSE:MCNC:PT:BLD:QN:
Serum glucose GLUCOSE:MCNC:PT:SER:QN:
Urine glucose concentration GLUCOSE:MCNC:PT:UR:QN:
Urine glucose by dip stick GLUCOSE:MCNC:PT:UR:SQ:TEST STRIP
Ionized whole blood calcium CALCIUM.FREE:SCNC:PT:BLD:QN:
24 hour calcium excretion CALCIUM.TOTAL:MRAT:24H:UR:QN:
Automated hematocrit HEMATOCRIT:NFR:PT:BLD:QN:AUTOMATED COUNT
Manual spun hematocrit HEMATOCRIT:NFR:PT:BLD:QN:SPUN
Erythrocyte MCV ERYTHROCYTE MEAN CORPUSCULAR
VOLUME:ENTVOL:PT:RBC:QN:AUTOMATED COUNT
ESR by Westergren method ERYTHROCYTE SEDIMENTATION
RATE:VEL:PT:BLD:QN:WESTERGREN
```

Figure 13: LOINC Example Terms

This is an example of LOINC, showing first the description of the observation to be represented, in this case lab observations, and then the encoding in LOINC. Even if it looks confusing at first glance, LOINC is very useful for automated information system processing.

[14] More information at www.nlm.nih.gov/research/umls/rxnorm/

[15] See www.LOINC.org

Current Procedural Terminology (CPT-4)

The Current Procedural Terminology (CPT) was created by the American Medical Association (AMA) to represent procedures and services provided exclusively by physicians. It is used for reporting and reimbursement for medical practices and its use is compulsory in the outpatient environment for government reimbursement in the United States.

HCFA COMMON PROCEDURE CODING SYSTEM (HCPCS)

The HCFA Common Procedure Coding System (HCPCS) is another procedure classification complementary with CPT. It is maintained by the US Government. The first level consists of CPT-4 codes; the second level includes the codes corresponding to procedures performed by personnel other than doctors and for billable supplies such as prostheses.

This classification is used for hospital inpatient scenarios.

Controlled Vocabularies for Nursing

Traditionally controlled vocabularies have focused on pathologies and symptoms, but have failed to contemplate the needs of nurses, who must address concepts from different functional evaluations such as "activity intolerance".

This has led to several US initiatives to introduce controlled vocabularies in the nursing domain, resulting in the creation of several non-traditional classifications:
North American Nursing Diagnosis Association (NANDA) Taxonomy II
Nursing Intervention/Outcomes Classification (NIC/NOC)
Omaha System

Other Considerations on Controlled Vocabularies

Controlled vocabularies present several limitations, namely:
The terms included in the vocabularies often do not relate to the terms used naturally by physicians, making their adoption and use difficult.
In the natural expression of diagnoses or clinical conditions, modifiers are used which often cannot be expressed in the controlled vocabulary (mild, severe, acute, recurrent, etc.)
Another frequent problem is that billing needs distort the categories of a classification, unifying different clinical concepts that have the same significance for billing within the same category.

Unified Medical Language System (UMLS)

As the number of controlled vocabularies increased, there arose several initiatives to provide a unification of such vocabularies. The most important of these is the Unified Medical Language System (UMLS), a US National Library of Medicine (NLM) project.

UMLS consists of three parts:

a) **Metathesaurus**: A combined repository of all the vocabularies, with interconnections. Here the most widely known vocabularies can be found, including their Spanish translations.

b) **UMLS semantic network:** Relationships that provide information on the meaning of concepts.

c) **Specialist lexicon:** An application targeted at facilitating the association of natural language terms with the words included in the Metathesaurus (only available in English).

There are some limitations to the use of UMLS:
It only includes one-to-one relationships.
It does not provide for the addition of terms, but only for the use of terms included in the source vocabularies.
It does not have a hierarchy that unifies all the concepts, only those hierarchies (where they exist) that apply to each source vocabulary.
It is not extensible (as are SNOMED and LOINC).

These characteristics have limited the utility of UMLS, which mainly functions as a repository of controlled medical vocabularies.

UMLS access is free of charge for academic and research purposes, but to use any vocabulary included in the metathesaurus in clinical practice a license (which may include license fees) must be obtained.

• HL7 and Vocabularies

Overview

As we have seen above, the effective use of vocabularies and terminologies is a vital component of achieving semantic interoperability. The HL7 organization has included vocabularies in its standards since its first implemented standard V2.1 was published in June 1990. Although HL7 aims for every data item in its standards to have a well-specified set of values, it only creates a vocabulary where no other terminology can be used. Therefore, HL7 collaborates and leverages in using vocabularies from other organizations, rather than competing with them.

The HL7 Vocabulary Work Group

The Vocabulary Work Group[16] in the HL7 standards organization focuses on all aspects of the vocabularies and terminologies used in the various HL7 Standards, e.g. V2.x, V3, CDA, FHIR, etc. To achieve this goal, the Work Group works cooperatively with other groups that have an interest in coded vocabularies used in clinical computing. These groups include standards development organizations, creators and maintainers of vocabularies, government agencies and regulatory bodies, clinical professional specialty groups, other HL7 Work Groups constructing models including vocabularies for standardization, vocabulary content providers and vocabulary tool vendors.

Mission: "To identify, organize and maintain coded vocabulary terms used in HL7 information structures, provide clear documented guidelines on the principles of vocabulary content and structure to support the retention of meaning over time and to maintain the HL7 Vocabulary Model and guidance of the use of Vocabulary in HL7 Standards."

HL7's vocabulary development strategy includes:
Reference existing vocabularies: SNOMED CT, LOINC, RxNorm, ISO, FDA identifiers, etc.
Collaborate with other SDOs: DICOM, ISO, CEN, NCPDP, ASTM, X12, etc.
Collaborate with government-sponsored efforts: US NCVHS Patient Medical Record Information (PMRI) standards, Canada Health Infoway's Consolidated Health Informatics (CHI) standards, etc.
Add value by creating linkages between HL7 messages and existing vocabularies
Collaborate with vocabulary developers to add needed content to existing vocabularies
Only create vocabularies that do not already exist
Register code systems used in HL7 standards for conformance purposes

How HL7 uses Vocabularies

HL7's policy is to use existing vocabularies from external sources whenever possible (e.g. "External Tables", "Imported tables", "Referenced Tables"). HL7 does not develop its own vocabularies (e.g. "HL7 Tables") unless no viable external terminology exists.

[16] More information at www.hl7.org/special/Committees/Vocab/index.cfm

What is a Code System?

A code system is a set of unique codes that have associated designations and meaning. It contains:
A unique identifier (codes) which when combined with the code system identifier creates an ID that is unique within healthcare
A name or designation
Additional text or annotations to further define the concept,
sometimes synonyms, relationships and hierarchies

Within the HL7 context, codes within a code system must not change meaning and must nor be reused, but codes may be added or retired.

What is a Concept Domain?

This expression is used HL7's CDA and V3 standards meaning a high level grouping for all things possible in a given domain from which value sets will be constructed, e.g. an abstract conceptual space such as "countries of the world," "the gender of a person (for administrative purposes)," "languages of the world," etc.

In the CDA and V3 standards, the external vocabulary domains are described in the External Domains list.

External Code Systems

HL7's policy is to use existing code systems from external sources whenever possible. HL7 does not develop its own code system unless all external possibilities have proven unworkable. For its standards, HL7 references the contents of external code systems (as "External Tables", "Imported tables", "Referenced Tables", etc.) but does not maintain or distribute content (e.g. SNOMED, LOINC, ISO, etc.). In some cases (e.g. ISO Standard 639 for languages, ISO Standard 3166 for country codes, ISO Standard 4217 for currencies, etc.), HL7 maintains a copy of the contents in the standards documents for the convenience of its users.

Internal Code Systems

Where external vocabularies or code systems do not exist or are not viable, internal code systems are code systems developed and maintained within the HL7 organization. In most cases, these internal code systems are either part of the standard (e.g. the V2.x list of version numbers, etc.) or are tightly linked to HL7's models.

What are Common Terminology Services (CTS)?

The HL7 Common Terminology Services (HL7 CTS) is an Application Programming Interface (API) that can be used to access the content of vocabularies. It is intended to specify only the unique

services needed in the implementation, but does not state how the service is to be implemented.[17]
In summary, the HL7 CTS is:

an HL7 ANSI standard defining the minimum set of requirements for interoperability across disparate health-care applications.

a specification for accessing terminology content: The CTS identifies the minimum set of functions a terminology resource must possess for use in HL7.

a functional model: Defining the functional characteristics of vocabulary service as a set of Application Programming Interfaces (APIs)

not a software package, although certain software exists that implements the specification, and it's not a common vocabulary data structure, although the standard can be implemented to interface with varying models (data and terminology.)

a specification for an API to access vocabulary/terminological content. A software client using the API does not have to know the vocabulary's data structure and/or how to access them, but can access its information using the standard API functions.

The CTS Architecture

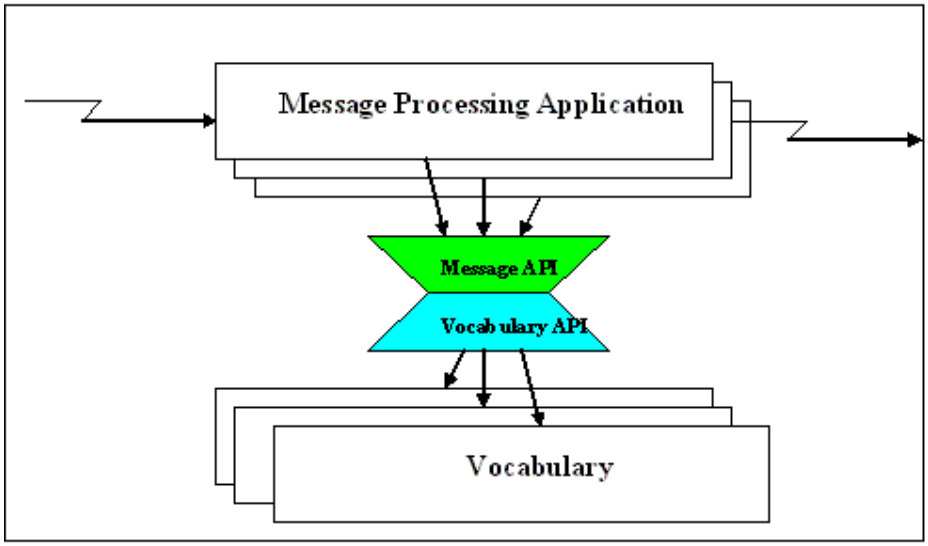

Figure 14: The HL7 Common Terminology Services Architecture

HL7 CTS Versions

HL7 international has published two releases of its Common Terminology Services standard. The original HL7 CTS specification[18] published in 2005 deliberately avoided issues related to terminology distribution, versioning and authoring; it focused on static value sets and did not fully address the definition or resolution of value sets that define post-coordinated expressions - issues that are now in the scope of HL7 CTS 2 (published in 2009).[19]

[17] More details at www.HL7.org/implement/standards/product_brief.cfm?product_id=10

[18] See HL7 CTS - April 2005 - Download at www.hl7.org/documentcenter/private/standards/CTS/R1/HL7_CTS_R1_FINAL.zip

[19] See Health Level 7 Common Terminology Services (HL7 CTS 2) - Draft Standard for Trial Use (DSTU) - Release 2 - October 2009 - Download at www.hl7.org/documentcenter/public/standards/dstu/2009may/V3_CTS_R2_DSTU_2009OCT.pdf

To compliment HL7's CTS high-level functionality and requirements standard, the Object Management Group (OMG) has created precise technical API interface specifications.[20]

Typical CTS API Calls

lookupCodeSystemInfo: Return detailed information about the named code system.
lookupValueSetExpansion: Return a hierarchical list of selectable concepts for the vocabulary domain and context
isConceptIdValid: Determine whether the concept code is valid in the code system.
fillInDetails: Fill in the details for the coded attribute, including all code system names, versions and display names.
validateCode: Determine whether the supplied coded attribute is valid in this vocabulary domain and context.
validateTranslation: Determine whether the translation portion of the coded attribute is valid in this domain and context.
translateCode : Translate the input code into a form that is valid in the target context.

[20] See http://hssp.wikispaces.com/specs-cts2

Unit Summary and Conclusion

This Unit explains why the use of standards in the healthcare area is necessary and discusses some of the standards and the organizations involved in their development and maintenance. Also, we have examined general concepts and characteristics of real-world healthcare information systems, in order to understand the need for standards. We have also reviewed the essential elements of the implementation of an HIS, focusing specifically on interoperability, a key theme in this course.

Without Terminology Standards:

- health data is non-comparable
- data aggregation is difficult, if not impossible
- health systems cannot interchange data
- linkage to decision support resources is not possible
- secondary uses (research, efficacy, etc.) are not possible

Interoperability cannot be accomplished without effective terminologies!

You should now have a clear understanding that a useful information interchange depends upon the existence of an agreed set of semantic and syntactic rules. Controlled vocabularies, especially terminologies, are a key component for achieving interoperability of health information systems.

In the latter part of this Unit you also learnt how Health Level 7 handles and utilizes vocabularies. You were also introduced to the HL7 Common Terminology Services - an API that simplifies interfacing into different vocabularies/terminologies.

Additional Reading Material

HL7 Vocabularies

V2.x: Health Level 7 V2.7, Chapter 2C "Control - Code tables"
CDA: Health Level 7 CDA R2, "HL7 Vocabulary Domains"

HL7 and OMG Common Terminology Services (CTS) Standards

Overview: www.HL7.org/implement/standards/product_brief.cfm?product_id=10

More HL7 information: http://wiki.hl7.org/index.php?title=Common_Terminology_Services_-_Release_2_%28Normative%29

More OMG information at: http://hssp.wikispaces.com/specs-cts2

Health Level 7 Common Terminology Services - April 2005. Download from www.HL7.org/documentcenter/private/standards/CTS/R1/HL7_CTS_R1_FINAL.zip
Health Level 7 Common Terminology Services (HL7 CTS 2) - Draft Standard for Trial Use (DSTU) - Release 2 - October 2009. Download from www.HL7.org/documentcenter/public/standards/dstu/2009may/V3_CTS_R2_DSTU_2009OCT.pdf

Note: You can download the two HL7 CTS standards free from the above URLs on the HL7 International Web site (www.HL7.org) after you create a free log-in account.

HL7 Common Terminology Services (CTA) Implementations

US National Cancer Institute (NCI) LexBIG (part of the caBIG Program)
Use: Vocabulary Service for NCI LexBIG
Platform: Java
Backend: Relational Database
More info at https://wiki.nci.nih.gov/display/LexEVS/LexBIG

VHA Enterprise Terminology Service

Use: Internal vocabulary management
Platform: Java
Middle-tier: Weblogic
Backend: Relational Database (Oracle)
Read "VHA Enterprise Terminology Project" at
http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_028716.pdf

Mayo Clinic

Use: Open Source Reference Implementation
Platform: Java
Backend: SOAP, LDAP, Relational Database, Protégé

Additional Reading Material

Please read the following:

Mandatory Reading

- HL7 on Wikipedia: http://en.wikipedia.org/wiki/Health_Level_7

Further Useful Reading

- "The e-Health Revolution - easier said than done" - Research Paper no. 3 2011-12 at www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/rp1112/12rp03
- "E-health Standards and Interoperability" - This report by the International Telecommunication Union's (ITU) Telecommunication Standardization Bureau outlines how e-health systems can potentially transform healthcare through mobile health delivery, personalized medicine and social media e-health applications. Reaching the potential for advancements in e-health will only be achieved through information and communication technology standards efforts that facilitate interoperability among systems and devices, provide unqualified privacy and security, address the unique needs of the developing world and leverage existing ubiquitous technologies such as social media applications and mobile devices.
  The report concludes by suggesting five standards prerequisites for achieving the promise of e-health: emphasizing greater interoperability, increasing coordination over global e-health standardization, ensuring privacy and security, reducing the standardization gap in the developing world, and leveraging existing technologies like mobile devices and social media applications.
  Download at www.itu.int/dms_pub/itu-t/oth/23/01/T23010000170001PDFE.pdf
- Updates on e-Health in the USA: www.cms.gov/eHealth/
- Updates on e-Health and knowledge management: www.openclinical.org/e-Health.html