

HEALTHCARE DATA STANDARDS

Andrei ACHIMAȘ, Ramona Gălățuș

Data standards refer to the elements of information used in healthcare systems. It includes data used for intra-organisational and inter-organisational communications. The appropriate standards cover data models, definitions of data elements and structural relationships.

Coding and Terminology is a traditional central issue of health informatics but also it is closely related with practical issues of the management of healthcare organisations. The goal of standardisation efforts concerning healthcare terminology is to arrive at a consensus on the most appropriate set of terms and the way that they should be structured. There is no a unique standard for healthcare data coding, neither for patient data.

IMPORTANCE OF DATA STANDARDS IN HEALTHCARE

Healthcare activities produce, use and move an enormous amount of data. This applies to all organisations at all levels, from the delivery of care to patients at home, from the top to the lower management, from specialised doctors to administrative personnel.

The use of a common set of terms with standardised definitions is essential to use computers for managing healthcare data and to allow electronic communication between computerised systems.

The use of data standards produces benefits for all the agents including patients, managers, researchers, epidemiologists, etc. Standardised health data make possible to produce more utilisable and communicable health records. Data standards allow to process data producing meaningful comparisons, aggregations, combinations, and statistical analysis. This capacity is very useful for care planning and epidemiological research. Also common data definitions enable managers to gain knowledge needed for operational purposes and quality assessment. Mandatory classifications are being required by governments for healthcare statistics, and by payment agencies for reimbursement purposes.

CODING PRINCIPLES

Codification is a basic discipline linked to information theory. The principles of Information and Communication Technologies are founded on coding data. All digital electronic devices, including computers, are based on binary coding of the data, as well as the instructions and control signals that perform all the operations.

A code can be defined as "a fixed sequence of signs or symbols, serving to designate an object". A code represents an unambiguous way to designate an object, being physical or abstract. In mathematical terms, a codification scheme is an application of a set of objects onto a set of codes.

Barcodes constitute a good example of a codification system widely used for goods. The example of the ASCII code for alphanumeric characters used in computer systems (for example - keyboard encoding) or data transmission, illustrates the importance and use of coding systems within electronic systems.

CODING SYSTEMS IN HEALTH INFORMATICS

Codification is required in healthcare to identify persons, by ciphers such as the social security number, but also it is used for identifying medical devices, pharmaceuticals and the like, in addition to describing illness and therapeutic procedures.

Two classes of coding are typically described in health telematics. One class refers to coding objective or factual data used to identify clearly recognisable objects, such as patients, doctors, nurses, hospital, services, suppliers, medical devices, pharmaceuticals, as well as physical or biochemical magnitudes.

A second class of coding corresponds to subjective or interpreted data. That is used to identify such concepts as abstract entities like diseases, medical procedures, findings, concepts.

There is a great number of coding systems in Health Informatics. Practical realisations of coding and terminological tools are of different categories, such as classifications, nomenclatures, dictionaries, and thesauri.

Some confusion exists between coding and classification. According to ISO (ISO/TC 9789) a classification is "a systematic arrangement of entities or attributes in groups or categories based on the similarity of pre-determined characteristics". Therefore, classification refers to the operation of assigning membership of an entity (object) to a set. This set is named with a "term". The object/or the set can be coded.

Table 1.- Displays different categories of coding systems used in Health Informatics. This table do not intent to constitute an exhaustive description of all existing coding system. It is presented only for illustrative purposes .

Table 1.- Reference list of healthcare terminological systems

| Category | Purpose | System | Author | Scope |
|------------------|--------------------------|----------------|----------|------------------------|
| Nomenclature | Medical Records | SNOMED | CAP | General |
| | | Read Codes | NHS-CCC | General |
| Classification | Statistics | ICD-9-CM | AMA | Diseases |
| | | ICD-10-CM | | and |
| | | ICPC | WONCA | Procedures |
| | | CPT-4 | AMA | Procedures |
| Thesaurus | Information Retrieval | MeSH | NLM | General |
| | Medical Records | THESAM | AP Paris | General |
| | Epidemiology | ADICAP | ADICAP | Pathology |
| Glossary | Comprehensive definition | DSM III | APA | Psychiatry |
| | | CMIT | AMA | Diseases |
| Knowledge Base | Decision aid | DXPLAIN QMR | | Diagnosis Diagnosis |
| Groupers stay | Reimbursement | DRG | Medicare | Hospital |

ISSUES ABOUT TERMINOLOGICAL SYSTEMS AND DATA CODING

Health coding system depends on coding objectives (such as identification, classification) and the application scope (general, surgery). Methodological approaches vary producing different types of codes. Most of the developed coding systems are very useful. One consequence is the need of conversion of coded data when communicated from one system to another. One important characteristic of a coding system is its coverage capability related with the total set of medical terms required by the intended application. Other important issues are the coding rules applied and the mapping into existing codes.

Experience has shown that it is not a realistic solution to search for a single comprehensive international medical coding system. A practical approach proposed by some authors has been to define a registration scheme for the

coding systems. Following that a standard identifier could be assigned to each registered coding system.

In the short term, administration and healthcare agencies keen to obtain aggregate clinical data are driven to adopting existing systems, even if they are imperfect. This has led to much debate among those supporting particular systems about their relative merits over competing ones.

The longer term need will be to introduce more maintainable and extensible systems as the cost of supporting existing systems becomes harder to sustain.

Since any general medical terminology will cover only a small part of the specific vocabulary of any medical specialty, separate systems may need to be developed for use between and within specialties. Communication standards on set of terms necessary to pass information between specialties have been detected as a need to cover in the future.

OVERVIEW OF CODING SYSTEMS OF PATIENT DATA

Patient data needs to be represented in a suitable coded form for computing and communication purposes. The typical approach is to encode the information using a set of standard terms from a controlled vocabulary. Coded patient data is used for a variety of applications related to direct patient care, clinical research, statistical reports and automated decision support. All these applications require to record and code data in a standard way.

The coding of patient information had been performed for a long time before the presence of computers in healthcare. The principal interest has been to reduce the complexity of original records by abstracting patient data in order to facilitate their management. Medical record abstraction has been traditionally performed for producing health statistics (for example - diseases incidence, mortality of surgical procedures) and more recently for costs assessment and reimbursement purposes.

A number of coding schemes have been designed and used for abstracting patient data. The following types correspond to this category - the classical International Classification of Diseases (ICD) of the WHO; the International Classification of Primary Care (ICPC) from WONCA; and Diagnostic Related Groups (DRGs).

Professional specialty groups have found some limitations in using general coding systems and they have developed their own coding system for medical record abstraction. Examples are the Current Procedural Terminology (CPT) of the American Medical Association (AMA); the Diagnostic and Statistical Manual of Mental Disorders of the American Psychiatric Association; Current Dental Terminology (CDT) of the American Dental Association (ADA) and the Micro-glossary of SNOMED for Dentistry

from the Advisory Committee on Dental Electronic Nomenclature Indexing and Classification (ACODENIC).

Nurses have also been active in the field with numerous contributions. One basic reference is the classification of Nursing Practice of the International Council of Nurses.

Another system of interest, concerning medical record abstraction, is the Medical Subject Headings (MeSH) maintained by the US National Library of Medicine.

The development of computer-based healthcare information systems, requiring more extensive representation of patient data than the needs of mere abstraction for statistical and reimbursement purposes, has raised further vocabulary capabilities. Some developers have expressed the inadequacy of traditional vocabularies to deal with coding requirements of the electronic medical record. Thus diverse initiatives have produced systems aimed to support coding of detailed patient data for usage in Electronic Health Records and other clinical applications.

More recently, efforts directed to develop comprehensive clinical coding schemes are seeking to serve both purposes. Examples of this category of coding systems are SNOMED, Read Clinical Codes, the Gabrieli (ASTM) Medical Nomenclature, and the Unified Medical Language System (UMLS) of the US National Library of Medicine.

Also worthy of mention is the coding schema for laboratory Logical Observation Identifier Names and Codes (LOINC) developed by the Logical Observation Identifier Names and Codes (LOINC) consortium.

Healthcare data standards are compiled in Data Dictionaries. Data Manuals are often provided for facilitating handling the content in Data Dictionaries. The following paragraphs outline the standard and non-standards characteristics of the above mentioned coding and classification systems.

INTERNATIONAL CLASSIFICATION OF DISEASES (ICD)

The International Classification of Diseases (ICD) is the archetypal coding system for medical record abstraction. Its origin is traced to the International List of Causes of Deaths, adopted in 1893 by the International Statistical Institute. The classification has been revised at (approximately) ten-yearly intervals. The World Health Organisation (WHO) has undertaken their maintenance and updating since 1948. Its scope was then extended to include non-fatal conditions. Subsequent revisions have enhanced its usefulness for morbidity applications.

The Ninth Edition (ICD-9) was published in 1977 and the Tenth Edition in 1992.

After the publication of ICD-9, some problems were detected. The United States National Centre for Health Statistics published a set of "clinical modifications" to ICD-9 known as ICD-9 CM. Many countries for a number of purposes have adopted the ICD-9-CM, widely accepted and used in the healthcare industry: data collection, quality of care analyses, resource utilisation, research and reimbursement, and statistical reporting.

ICD-10-CM is currently under development, with a planned implementation date of October 1, 2000.

DIAGNOSIS RELATED GROUPS (DRGS)

The Diagnostic Related Groups (DRGs) were originated at Yale University in US for use in prospective payment within the Medicare Program.

This coding system is an abstraction of an abstraction. It is applied to lists of ICD-9-CM codes that are themselves derived from patient medical records. The purpose of DRG coding is to provide a relatively small number of codes for classifying patient hospitalisations, while at the same time providing some separations of cases based on severity of illness. The principal motivations for the groupings are factors that affect cost and length of stay.

INTERNATIONAL CLASSIFICATION OF PRIMARY CARE (ICPC) FROM WONCA

This coding systems was originated in response to some perceived limitations of ICD-9 from the World Organisation of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians (WONCA). IPC provides seven axes of terms and a structure to combine them to represent clinical encounters.

PHYSICIANS' CURRENT PROCEDURAL TERMINOLOGY (CPT) OF THE AMERICAN MEDICAL ASSOCIATION (AMA)

Physicians' Current Procedural Terminology, Fourth Edition (CPT-4) is a listing of descriptive terms and identifying codes for reporting medical services and procedures. The purpose of the terminology is to provide a uniform language that will accurately describe medical, surgical and diagnostic services, and will thereby provide an effective means for reliable communication among health professionals, patients and third parties. The first edition of CPT appeared in 1966.

CPT is used as the reporting mechanism by physicians and many other health professionals in the United States. It is the coding system used by Medicare and virtually all third party payers, including workers compensation and Medicaid. Hospitals use CPT codes to report outpatient service to Medicare.

CPT contains over 7,300 codes to describe medical and surgical procedures. CPT is divided into 6 major sections, including Evaluation and Management Services, Anaesthesia, Surgery, Radiology, Pathology and Medicine. The section heads, subheads, and titles provide an implicit hierarchy. The Surgery section is subdivided anatomically. The Medicine Section is divided by medical sub-specialties. A series of two digit modifiers are also included to make the coding system more specific.

INTERNATIONAL MEDICAL TERMINOLOGY (IMT)

The International Medical Terminology (IMT) is a medical terminology designed to support the classification, retrieval, presentation, and communication of medical information throughout the medical product regulatory cycle. The foundation of the IMT is the Medical Dictionary for Drug Regulatory Affairs (MEDDRA) developed by the UK Medicines Control Agency (MCA) in its Adverse Drug Reactions On-line Information Tracking System (ADROIT).

Included in the IMT are terms describing diseases, diagnoses, signs, symptoms, therapeutic indication names, and qualitative results of investigations (such as laboratory tests, radiological studies), medical and surgical procedures, and terms describing medical, social, and family history. The IMT consists of a five level hierarchy, starting with 26 System Organ Classes (SOCs), that represent the highest level groupings of the terminology. Including all levels, it contains approximately 40,000 terms. The Preferred Term (PT) is the internationally agreed upon level at which regulatory information is to be exchanged. The IMT contains approximately 8,800 Preferred Terms.

NURSING INTERVENTIONS CLASSIFICATION (NIC)

The Nursing Interventions Classification (NIC) is a comprehensive, standardised language describing treatments that nurses perform in all settings and in all specialties.

NIC groups approximately 13,000 nurse activities into 433 standardised intervention terms, each with a unique code. NIC has numerous uses including care planning, documentation, standards construction, critical paths, competency evaluation, job descriptions, and curriculum and course syllabus construction. The use of NIC in nursing information systems allows for the collection of standardised data to be used in effectiveness research and in determining the costs of nursing.

NIC interventions have been linked with NANDA nursing diagnoses and the Omaha System problems and are in the process of being linked with Nursing Outcomes Classification (NOC) patient outcomes. The classification work is

part of the Centre for Nursing Classification at the University of Iowa College of Nursing.

NIC is recognised by the American Nurses' Association and is included in the National Library of Medicine's Meta-thesaurus for a Unified Medical Language. Both the Cumulative Index to Nursing Literature (CINAHL) and Silver Platter have added NIC to their nursing indexes.

NIC is included in the Joint Commission on Accreditation for Healthcare Organisation's (JCAHO) as one nursing classification system that can be used to meet the standard on uniform data.

SNOMED INTERNATIONAL: SYSTEMATISED NOMENCLATURE OF HUMAN AND VETERINARY MEDICINE FROM THE COLLEGE OF AMERICAN PATHOLOGISTS

SNOMED International is a comprehensive, multi-axial nomenclature classification system created for the indexing of the entire medical vocabulary, including signs and symptoms, diagnoses, and procedures.

SNOMED was the first coding system that attempted to provide terms for a broad range of clinical domains. It represented a significant improvement over this predecessor - Systematised Nomenclature of Pathology (SNOP) - originated in the early 1960'.

First published in 1975, and then revised as SNOMED II in 1979, it has been more recently released in a greatly expanded version SNOMED International: Systematised Nomenclature of Human and Veterinary Medicine, in 1993.

It consists of a set of eleven axes, each of which serves as taxonomy for a specific set of concepts (covering such as organisms, disease, procedures). The eleven modules of the current version of SNOMED contain more than 144,000 terms and term codes.

The American Veterinary Medical Association and the American Dental Association have endorsed its use. In addition, the American College of Radiology/National Equipment Manufacturers Association (ACR/NEMA) is using a subset of SNOMED in their Digital Imaging and Communications in Medicine standard (DICOM).

SNOMED has been translated from English into 12 other languages.

READ CODES

The Read Clinical Codes were developed in the early 1980s by Dr. James Read to record clinical summaries in General Practice. The first version of the Read Codes was adopted by the UK National Health Service in 1990. Version 2 was developed to meet the needs of hospitals for cross-mapping their data to ICD-9.

Version 3 has been developed to support not only medical record summarisation, but to support patient care applications directly. The first versions of the Read Codes were organised in a strict hierarchy, Version 3 allowed users to construct acyclic graphs (conceptually analogous to situations where children can be the offspring of many parents).

It is intended that Version 3 will allow the creation of the full electronic patient record, should this be required. As with Version 1 and 2 Read Codes, Version 3 terms are provided, where appropriate, with validated cross-references to ICD-9, ICD-10 and OPCS4.

The NHS Centre for Coding and Classification (NHS CCC), is part of the Information Management Group of the Department of Health in UK and it is their role to develop the Read Codes for use by all healthcare professionals.

THE GABRIELI (ASTM) MEDICAL NOMENCLATURE

This system was first developed at the University of Buffalo. It consists of a single, large hierarchy that contains successively more complex expressions as one moves down through the hierarchy.

Although it was initially available as a commercial product, it is being used as the foundation for nomenclature work under the American Society for Testing and Materials (ASTM) to produce an official standard.

UNIFIED MEDICAL LANGUAGE SYSTEM

The Unified Medical Language System (UMLS) is a long-term research and development project of the National Library of Medicine (NLM) started in 1986. The purpose of the UMLS is to aid the development of systems that help health professionals and researchers retrieve and integrate electronic biomedical information from a variety of sources. The UMLS approach involves the development of machine-readable Knowledge Sources that can be used by a wide variety of applications programs to :

compensate for differences in the way concepts are expressed in different machine-readable sources and by different users

- to identify the information sources most relevant to a user inquiry
- to negotiate the telecommunications and search procedures necessary to retrieve information from these sources

The goal is to make it easy for users to link disparate information systems, including computer-based patient records, bibliographic databases, factual databases and expert systems.

There are four UMLS Knowledge Sources: the Meta-thesaurus, the SPECIALIST Lexicon, a Semantic Network and an Information Sources Map. Most heavily used to date, the Meta-thesaurus provides a uniform, integrated distribution format for more than 30 biomedical vocabularies and classifications, linking many different names for the same concepts. The Lexicon contains syntactic information for many Meta-thesaurus terms, component words, and English words, including verbs that do not appear in the Meta-thesaurus. The Semantic Network contains information about the types or categories (such as "Disease or Syndrome" or "Virus") to which all Meta-thesaurus concepts have been assigned and the permissible relationships among these types (including "Virus" causes "Disease or Syndrome"). The Information Sources Map or directory contains both human-readable and machine-"processable" information about the scope, location, vocabulary, syntax rules and access conditions of biomedical databases of all kinds.

The UMLS Knowledge Sources were designed as multi-purpose tools, to facilitate the development of more effective biomedical information systems. As intended, they have been applied in a wide variety of research and development environments to many different tasks, including vocabulary development, knowledge representation, clinical data capture, linking patient data to knowledge sources, curriculum analysis, natural language processing, automated indexing and information retrieval.

EUROPEAN OFFICIAL STANDARDISATION ACTIVITIES OF CEN TC251

The Working Group 2 of CEN/TC 251 has been devoted to standardisation in the area of Healthcare Terminology, Semantics and Knowledge Bases. The activities of this working group has produced the following outcomes:

- ENV1614:1995 Concept Structure for Nomenclature, Classification and Coding of Properties in Clinical Laboratory Sciences
- ENV1828:1995 Structure for Classification and Coding Surgical Procedures
- ENV1264 Structure of Concept Systems- Vocabulary
- ENV Terminology and Coding Systems of Drugs
- ENV Structure for Coding Systems of Medical Devices
- ENV 12435: 1996_Standards for Notation of Units for Quantities in Clinical Sciences
- ENV 12381: 1996_Time Standards for Healthcare Specific Problems.

ENV 1614: 1995 Concept Structure for Nomenclature, Classification and Coding of properties in Clinical Laboratory Sciences - this pre-standard provides a structure for systematic naming, classification and coding of properties, including quantities (analyses) in laboratory medicine. The system is intended to facilitate the communication of messages about such properties through computing and telecommunication equipment by creating a standard concept structure to which local nomenclatures used in routine laboratory practice may be related.

It is not the purpose of this standard to standardise the language used by healthcare practitioners in requesting or reporting laboratory data.

ENV 1828: 1995 Structure for Classification and Coding of Surgical Procedures - this European pre-standard is applicable to any interventional procedure. It provides a unique reference concept system to represent surgical procedure wording by terminological phrases. This computer-based concept system is confined to the computer to support the exchange of surgical procedure information expressed in natural language between different national standard terminology classifications within Europe regardless of what purpose the underlying classification has been developed for.

This pre-standard is merely the first step in the process of voluntary convergence of surgical procedure coding systems. The process could involve using the pre-standard to complement existing classifications in order to ease international comparisons or as a basis for developing new ones.

The pre-standard can be used to develop new coding systems for surgical procedures, which may be multi-axial in nature or the more traditional, hierarchical ones.

ENV 12264: 1996 Structure of Concept Systems-Vocabulary - the document provides the vocabulary and the guidelines to describe the categorical structure of a concept system. The structure consists in practice of a list of involved categories, their typical relations and the criteria used to build taxonomies.

The standard structures may be used as the basis from which to prepare principled classifications and nomenclatures, multi-hierarchical systems, combinatorial systems, and so on. The customised structure of the implemented coding systems may remain implicit, or may preferably be described in detail in the documentation accompanying the system or may be exploited by computers for advanced processing.

prENV 12610 Terminology and Coding Systems of Drugs - the standard will fit as well medicinal product as herbal remedies, anthroposophic products, nutrients, homeopathic products, and similar

substances. The standard will enable the identification of medical substances (MSI-Medical Substance Identifiers), of medicinal products as such (MPI-Medicinal Product Identifiers) and the packages of medicinal products (MPPI-Medicinal Product Package Identifiers).

prENV 12611 Structure for Coding Systems of Medical Devices

This pre-standard defines a categorical structure for a system of concepts for medical device groups, in such a way that each medical device group may be precisely described. The document is to be used by organisations involved with the development or maintenance of nomenclatures and coding systems for medical devices and by designers of databases or information systems involving medical devices.

ENV 12435: 1996 Standards for Notation of Units for Quantities in Clinical Sciences

This pre-standard provides rules for rounding of numeric values and for statements of uncertainty. An annex will list kinds-of-quantity likely to be encountered in medical data and conversion factors for their units. The pre-standard does not deal with the "grammar" of informal quantitative statements nor with expression of financial information.

ENV 12381: 1996 Time Standards for Healthcare Specific Problems - this document provides a set of basic entities, with precisely defined properties and inter-relationships among them, that is sufficient to allow an unambiguous representation of time-related expressions. In addition, a number of preference rules, facilitating the generation and interpretation of such an unambiguous representation in a communicative context, are proposed.