TECHNICAL SPECIFICATION

**ISO/TS 17117**

First edition
2002-02-15

# Health informatics — Controlled health terminology — Structure and high-level indicators

*Informatique de santé — Terminologie contrôlée relative à la santé — Structure et indicateurs de haut niveau*



Reference number
ISO/TS 17117:2002(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.ch
Web www.iso.ch

Printed in Switzerland

# Contents

Page

# Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

In other circumstances, particularly when there is an urgent market requirement for such documents, a technical committee may decide to publish other types of normative document:

— an ISO Publicly Available Specification (ISO/PAS) represents an agreement between technical experts in an ISO working group and is accepted for publication if it is approved by more than 50 % of the members of the parent committee casting a vote;

— an ISO Technical Specification (ISO/TS) represents an agreement between the members of a technical committee and is accepted for publication if it is approved by 2/3 of the members of the committee casting a vote.



An ISO/PAS or ISO/TS is reviewed after three years with a view to deciding whether it should be confirmed for a further three years, revised to become an International Standard, or withdrawn. In the case of a confirmed ISO/PAS or ISO/TS, it is reviewed again after six years at which time it has to be either transposed into an International Standard or withdrawn.

Attention is drawn to the possibility that some of the elements of this Technical Specification may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO/TS 17117 was prepared by Technical Committee ISO/TC 215, *Health informatics*.

Annex B forms a normative part of this Technical Specification. Annex A is for information only.

# Introduction

In 1839, William Farr stated in his First Annual Report of the Registrar-General of Births, Deaths and Marriages in England, “The nomenclature is of as much importance in this department of inquiry, as weights and measures in the physical sciences, and should be settled without delay.” Since that time, this theme has been heard resounding from an increasingly large group of scientists (see Annex A). Today, the need for controlled terminologies to support health record systems has been widely recognized (E-1238, E-1239, E-1384, E-1633, ENV 12017). Controlled terminologies provide systems with the means to aggregate data. This aggregation of data can be carried out at multiple levels of granularity and therefore can enhance the clinical retrieval of a problem-oriented record, data pertaining to a classification for billing purposes, or outcomes data for a given population. Maintenance of large-scale terminologies has become a burdensome problem, as the size of term sets has escalated. Without a well-structured backbone, large-scale terminologies cannot scale to provide the level of interoperability required by today’s complex electronic health record applications.

The solution rests with standards [7]. Over the past ten or more years, medical informatics researchers have been studying controlled terminology issues directly. They have examined the structure and content of existing terminologies to determine why they seem unsuitable for particular needs, and they have proposed solutions. In some cases, proposed solutions have been carried forward into practice and new experience has been gained.[8] As we have entered the twenty-first century, it seems appropriate to pause to reflect on this experience, and to publish a standard set of goals for the development of comparable, reusable, multipurpose and maintainable controlled health terminologies (ISO 12200, ISO 12620).



This Technical Specification is the first deliverable for the ISO/TC 215 *Health informatics*, Working group 3, *Health concept representation*, that is also working on an International Standard to be the basis for future standards in this area. It will serve as a guide for governments, funding agencies, terminology developers, terminology integration organizations and the purchasers and users of controlled health terminology systems toward improved terminological development and recognition of value in a controlled health terminology. This ISO/TS 17117 on quality indicators of controlled health terminologies is based on previous work in ASTM that naturally could not be harmonized with ISO work already in progress. The present work is therefore published as a Technical Specification at this time with the intent to revise it to be compatible with the planned basic terminology standard and converted to a full International Standard after a maximum of three years.

# Health informatics — Controlled health terminology — Structure and high-level indicators

## 1 Scope

This Technical Specification specifies the principal ideas which are necessary and sufficient to assign value to a controlled health terminology. It is applicable to all areas of healthcare about which information is kept or utilized.

## 2 Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this Technical Specification. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this Technical Specification are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.



ISO 704, *Terminology work — Principles and methods*

ISO 860, *Terminology work — Harmonization of concepts and terms*



ISO 1087-1, *Terminology work — Vocabulary — Part 1: Theory and application*

ISO 1087-2, *Terminology work — Vocabulary — Part 2: Computer applications*

ISO/IEC 11179-3, *Information technology — Specification and standardization of data elements — Part 3: Basic attributes of data elements*

ISO 12620, *Computer applications in terminology — Data categories*

ISO/IEC 2382-4, *Information technology — Vocabulary — Part 4: Organization of data*

ISO/IEC/TR 9789, *Information technology — Guidelines for the organization and representation of data elements for data interchange — Coding methods and principles*

E-1284, *Standard Guide for Construction of a Clinical Nomenclature for Support of Electronic Health Records*

E-1712, *Standard Specification for Representing Clinical Laboratory Test and Analyte Names*

ENV 12264, *Health Informatics — Categorical Structures of Systems of Concepts — Model for Representation of Semantics*

## 3 Terms and definitions

For the purposes of this Technical Specification, the following terms and definitions apply.

**3.1**
**terminology**
set of terms representing a system of concepts within a specified domain

NOTE This implies a published purpose and scope from which one can determine the degree to which this representation adequately covers the domain specified.

**3.2**
**controlled health terminology**
set of terms intended for clinical use

NOTE This implies enough content and structure to provide a representation capable of encoding comparable data, at a granularity consistent with that generated by the practice within the domain being represented, within the purpose and scope of the terminology.

**3.3**
**classification**
terminology which aggregates data at a prescribed level of abstraction for a particular domain

NOTE 1 This fixing of the level of abstraction that can be expressed using the classification system is often done to enhance consistency when the classification is to be applied across a diverse user group, such as is the case with some of the current billing classification schemes.



NOTE 2 See Annex A for the history of classification.



**3.4**
**ontology**
organization of concepts for which a rational argument can be made



EXAMPLE A hierarchy of qualifiers would be a qualifier ontology.

NOTE Colloquially, this term is used to describe a hierarchy constructed for a specific purpose.

**3.5**
**qualifier**
string which, when added to a term, changes the meaning of the term in a temporal or administrative sense

EXAMPLES “History of” or “recurrent”.

**3.6**
**modifier**
string which, when added to a term, changes the meaning of the term in the clinical sense

EXAMPLES “Clinical stage” or “severity of illness”.

**3.7**
**canonical term**
preferred atomic or pre-coordinated term for a particular medical concept

**3.8**
**term**
word or words corresponding to one or more concepts

# 4 General

## 4.1 Basics

The basic characteristics of a terminology influence its utility and appropriateness in clinical applications. Terminologies should be evaluated within the context of their stated scope and purpose and are intended to complement and utilize those notions already identified by other national and international standards bodies.

This Technical Specification explicitly refers only to terminologies that are primarily designed to be used for clinical concept representation or to the aspect of a terminology designed to be used for clinical concept representation. This Technical Specification will also provide terminology developers and authors with the quality guidelines needed to construct useful and maintainable controlled health terminologies. These tenets do not attempt to specify all the richness which can be incorporated into a health terminology. However, this Technical Specification does specify the minimal requirements, which, if not adhered to, will assure that the terminology will have only limited generalizability and will be very difficult, if not impossible, to maintain. Terminologies which do not currently meet these criteria, can be in compliance with this Technical Specification by putting in place mechanisms to move toward these goals. Principles for implementation are specified in Annex B.

This Technical Specification will provide terminology developers with a sturdy starting point for the development of controlled health terminologies. This foundation serves as the basis from which terminology developers will build robust, large-scale, reliable and maintainable terminologies.

## 4.2 Concept orientation

The basic unit of a terminology shall be a concept, which is the embodiment of some specific meaning and not a code or character string. Identifiers of a concept shall correspond to one and only one meaning and, in a well-ordered terminology, only one concept may have that same meaning, as specified in ISO 860. However, multiple terms (linguistic representations) may have the same meaning if they are explicit representations of the same concept. This implies non-redundancy, non-ambiguity, non-vagueness and internal consistency.



### 4.2.1 Non-redundancy

Terminologies shall be internally normalized. There shall not be more than one concept identifier in the terminology with the same meaning, as specified in ISO 704 and E-1284. This does not exclude synonymy, rather it requires that this be explicitly represented.

### 4.2.2 Non-ambiguity

No concept identifier should have more than one meaning. However, an entry term can point to more than one concept.

EXAMPLE MI as myocardial infarction and mitral insufficiency.

NOTE Some authors have referred to entry terms as an interface terminology.

### 4.2.3 Non-vagueness

Concept names shall be context free.

EXAMPLE “Diabetes mellitus” should not have the child concept “well controlled”, instead the child concept's name should be “diabetes mellitus, well controlled”.

NOTE Some authors have referred to context free as context laden.

### 4.2.4 Internal consistency

Relationships between concepts should be uniform across parallel domains within the terminology.

EXAMPLE If heart valve structures are specified anatomically, the diagnosis related to each structure is also specified using the same relationships.

## 4.3 Purpose and scope

Any controlled terminology shall have its purpose and scope clearly stated in operational terms so that its fitness for particular purposes can be assessed and evaluated. Where appropriate, it may be useful to illustrate the scope by examples or “use cases” as in database models and other specification tools. Criteria, such as coverage and comprehensiveness, can only be judged relative to the intended use and scope.

EXAMPLE A terminology might be comprehensive and detailed enough for general practice with respect to cardiovascular signs, symptoms and disorders, but inadequate to a specialist cardiology or cardiothoracic surgery unit. Conversely, a terminology sufficiently detailed to cope with cardiology and cardiothoracic surgery might be totally impractical in general practice.

### 4.3.1 Coverage

Each segment of the healthcare process shall have explicit in-depth coverage, and not rely on broad leaf node categories that place specific clinical concepts together. The extent to which the depth of coverage is incomplete shall be explicitly specified for each domain (scope) and purpose as indicated in 4.3.[9]

EXAMPLE It is often important to distinguish specific diagnosis from categories presently labelled “not elsewhere classified” (NEC), or to differentiate disease severity such as indolent prostate cancer from widely metastatic disease.

### 4.3.2 Comprehensiveness

The extent to which the degree of comprehensiveness is incomplete shall be explicitly specified for each domain (scope), and purpose as indicated in 4.3. Within the scope and purpose, all aspects of the healthcare process shall be addressed for all related disciplines, such as physical findings, risk factors, or functional status — across the breadth of medicine, surgery, nursing and dentistry. This criterion applies because decision support, risk adjustment, outcomes research and useful guidelines require more than diagnoses and procedures.



EXAMPLES The existing Agency for Healthcare Research and Quality Guidelines, and the Healthcare Finance Administration (HCFA) mortality model.[10]



## 4.4 Mapping

Government and payers mandate the form and classification schema for much clinical data exchange. Thus, comprehensive and detailed representations of patient data within computer-based patient records should have the ability to be mapped to those classifications, such as ICD-9. This need for multiple granularities is required for clinical healthcare, as well as is specified in ISO/IEC/TR 9789. The degree to which the terminology is mappable to other classifications shall be explicitly stated.[11]

EXAMPLE An endocrinologist may specify more detail about a patient’s diabetes mellitus than a generalist working in a primary care setting, even though both specialities may be caring for the same patient.

## 4.5 Systematic definitions

In order for users of the terminology to be certain that the meaning that they assign to concepts is identical to the meaning which the authors of the terminology have assigned to these definitions will need to be explicit and available to the users. Further, as relationships are built into terminologies, multiple authors will need these definitions to ensure consistency in authorship.

EXAMPLE The clinical concept “hypertension” might be defined as a consistently elevated blood pressure and needs to be distinguished from a single “BP > 140/85”.

## 4.6 Formal definitions

A compositional system should contain formal definitions for non-atomic concepts and formal rules for inferring subsumption from the definitions, as specified in E-1712.

### 4.7 Explicitness of relations

The logical definition of subsumption should be defined. The formal behaviour of all links/relations/attributes should be explicitly defined. If a looser meaning such as “broader than/narrower than” is used, it should be explicitly stated.

EXAMPLE The primary hierarchical relation should be subsumption as exemplified by logical implication: B is a kind of A means all Bs are As.

### 4.8 Reference terminologies

The set of canonical concepts, their structure, relationships and, if present, their systematic and formal definitions define the core of the controlled health terminology.

### 4.9 Atomic reference terminologies

In a reference terminology consisting of only atomic concepts and their systematic definitions, no two or more concepts can be combined to create a composite expression which has the same meaning as any other single concept contained in the atomic reference terminology.

### 4.10 Colloquial terminologies

The set of terms, which consists of commonly used entry points, maps to one or more canonical terms within the terminology.



NOTE These have been called “entry terms” or “interface terminologies” by different authors.



## 5 Structure of the terminology model



### 5.1 Terminology structures

Terminology structures determine the ease with which practical and useful interfaces, for term navigation, entry, or retrieval can be supported, as specified in ISO 704, ISO 1087-1 and ENV 12264.

### 5.2 Compositional terminologies

#### 5.2.1 Compositionality

Composite concepts are created from atomic and pre-coordinated concepts and shall have the ability to be combined to create compositional expressions[12].

EXAMPLE “Colon cancer” comprises “malignant neoplasm” and “large bowel” as atomic components. In a compositional system, concept representations can be divided into atomic and composite concept representations.

Composite concept representations can be further divided into named “pre-coordinated concept representations” and “post-coordinated representation” expressions. Within a composite concept, it may be possible to separate the constituents into three categories: “kernel concept”, “qualifier (also called “status”) concept”, and “modifier concept”.

NOTE A concept is a notion represented by language, which identifies one idea. However, the term “concept” in this Technical Specification is used to refer to the representation of a concept rather than to the thought itself.

#### 5.2.1.1 Atomic concept

An atomic concept is a representation of a concept that is not composed of other simpler concept representations within a particular terminology. In many cases, atomic concepts will correspond to what philosophers call “natural kinds”. Such an entity cannot be meaningfully decomposed. Concepts should be separable into their constituent components, to the extent practical. These should form the root basis of all concepts.

EXAMPLE In SNOMED-RT, “colon” is a synonym for “large bowel” and “cancer” is a synonym for “neoplasm, malignant”. Therefore, the term “colon cancer” is non-atomic as it can be broken down into “large bowel” and “neoplasm, malignant”. Each of these two atomic terms has a separate and unique concept identifier, as does the pre-coordinated term “colon cancer”.

#### 5.2.1.2 Composite concept

A composite concept is composed as an expression made up of atomic concepts linked by semantic relations (such as roles, attributes or links).

##### 5.2.1.2.1 Pre-coordinated concept

Such an entity can be broken into parts without loss of meaning (can be meaningfully decomposed), when the atomic concepts are examined in aggregate. These are representations, which are considered single concepts within the host vocabulary. Ideally, these concepts should have their equivalent composite concepts explicitly defined within the terminology (that is the terminology should be normalized for content, as indicated in 5.2.2).

EXAMPLE The term “colon cancer” is non-atomic, however it has a single unique identifier, which means to the SNOMED-RT that it represents a “single” concept. It has the same status in the terminology as the site “large bowel” and the diagnosis “neoplasm, malignant”.



##### 5.2.1.2.2 Post-coordinated concept

A post-coordinated concept is a composite concept, which is not pre-coordinated and therefore shall be represented as an expression of multiple concepts using the representation language. This is the attempt of a system to construct a set of concepts from within a controlled terminology to more completely represent a user’s query.



EXAMPLE The concept “bacterial effusion, left knee” is not a unique term within the SNOMED-RT terminology. It represents a clinical concept that some patient has an infected left knee joint. As it cannot be represented by a single concept identifier, to fully capture the intended meaning a system would need to build a representation from multiple concept identifiers or lose information to free text.

#### 5.2.1.3 Types of atomic and pre-coordinated concepts

Unique concept representations can be classified within a terminology into at least three distinct types: kernel concepts, modifiers and qualifiers (which contain status concepts). This separation allows user interfaces to provide more readable and therefore more useful presentations of composite concepts.

##### 5.2.1.3.1 Kernel concept

This is an atomic or pre-coordinated concept, which represents one of the one or more main concepts within a pre-coordinated or post-coordinated composition.

##### 5.2.1.3.2 Modifiers and qualifiers

Constituents of a composite concept that refine the meaning of a kernel concept are known as modifiers or qualifiers.

EXAMPLE 1 “Stage 1a” in the expression “having colon cancer stage 1a” or “brittle, poorly controlled” in the expression “brittle, poorly controlled diabetes mellitus” are examples of qualifiers and modifiers.