TECHNICAL REPORT

ISO/TR 12300

First edition
2014-11-15

Health informatics — Principles of mapping between terminological systems

Informatique de santé — Cartographie des terminologies de classifications

Reference number
ISO/TR 12300:2014(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.org
Web www.iso.org

Published 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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT), see the following URL: Foreword — Supplementary information.

The committee responsible for this document is ISO/TC 215, *Health informatics*.

Introduction

The benefits of data sharing and reuse are well known. One of the key principles underpinning health informatics is that data should be collected once and reused to the greatest extent possible.

Mapping is the process of associating concepts from one terminological resource to concepts in another terminological resource and defining their equivalence in accordance with a documented rationale and a given purpose. The terminological resources can be related (different versions of the same resource) or completely different resources. The process of mapping identifies whether there is a relationship between the concepts and, if so, the level of meaning expressed by that relationship. It is a way to integrate different terminological resources used for different purposes and where a bridge between them is required for interoperability and that bridge can be built through mapping. Thus, different data sources can be compared and linked to enable the data to be exchanged between information systems, compared over time, or aggregated for different purposes. The end product (deliverable) of the process is a set of individual maps (relationships) between two terminological resources that defines the cardinality and degree of equivalence between concepts and rule set structures and enables the automated translation between the terminological resources.

As an example in health care, data collected for communicating information about direct patient care (using clinical terminologies) can be reused for statistical and administrative reporting of morbidity data (using clinical classifications) by transforming the terminological representations into classification representations.

Terminological resources include all mechanisms for representation of data including terminologies, classifications, and code systems.

Quality maps are always built for a purpose. Skilled mapping personnel are required to ensure the quality and integrity of map development and mapping rules. The development of rules (either paper-based or computer algorithms) that support conversion of data are crucial to standardize the process and create logical maps that a computer can use repeatedly to consistently convert data from one form to another.

This Technical Report provides guidance for organizations charged with creating or applying maps to meet their business needs. It identifies issues and discusses both the potential in and the limitations of applying the map. This Technical Report also establishes and harmonizes the basic principles for developing, maintaining, and using maps and gives guidelines for good practice that underpin the mapping process. This Technical Report does not provide information or guidance on processes required to produce a map in any given situation nor the intellectual property rights of those who own the various terminologies or classifications.

There is a lack of common understanding of the need for mapping between terminological resources, the process of mapping, and requirements for computational functionality in the mapped relationships between the different terminological resources used in health care. Thus, documenting the general principles that underpin the mapping process are essential to good decision making and governance. These will provide guidance about good practice, will support convergence of international knowledge, standardize processes, structure, and approach to the development of infrastructure and tools supporting the mapping process.

There are broadly three core reasons to map data from one code system to another through a map. These include the following:

— support interoperability (information sharing between systems and organizations);

— reuse of data collected for one purpose to meet another purpose (secondary use);

— convert from an older, no longer relevant to purpose terminological resource to a new alternative representation.

Information sharing might require information collected in the local system to be converted to a “common language” such as that represented by international standard terminological resources such

as SNOMED CT or ICD. The common language should be agreed upon in order for computer systems to communicate effectively. Any information in local systems not in the common language should be translated (mapped) into the common language and when information is received from others, it should be converted from the common language into the language of the local system.

The increasing use of terminologies to collect data supporting direct patient care has enabled the reuse of this data for other purposes. Data collected for secondary use are generally aggregated and collected through classifications. Secondary use includes, but is not limited to, reusing the information for the following purposes:

a) funding;

b) statistical aggregation and reporting (morbidity and mortality);

c) providing a research basis for evidence-based medicine;

d) measuring quality and safety of care;

e) health planning or setting health policy;

f) monitoring resource utilization;

g) public health surveillance.

Reusing the data through mapping reduces the need for recollection of data, thereby simplifying the administrative burden of data collection, although it should be understood that the administrative burden might increase overall due to the maintenance of mapping when continuing to use multiple code systems and maps. Facilitating the automation between various terminological systems used in health care reduces the costs of providing care and improves the quality of the data and the timeliness (availability).

Decisions on whether or not to map or whether to move from, for example, a classification of clinical information to a more precise clinical terminology needs to be based upon a wide range of factors including the ability to accurately represent meaning, the need to represent information in a manner suited to purpose in each use environment, including the need to aggregate and compare data over time. There are also significant costs and skills associated with mapping. The difference between a "once off" map table to meet a singular, conversion process and the decision to use maps as a long-term mechanism to support reporting and analysis need to be understood by those making decisions on these infrastructure approaches recognizing all of the benefits, requirements, and costs which might include the following:

— decision makers in government, healthcare authorities, and healthcare facilities;

— developers, implementers and managers of health information systems, clinical information systems and clinical decision support systems;

— classification and terminology communities of practice;

— all users of clinical data, such as health statisticians, researchers, public health agencies, health insurance providers, health risk organizations, data analysts, and data managers;

In this Technical Report,

— mapping refers to the establishment of semantic comparability between terminological resources (these resources include terminologies, classifications and other code systems),

— the term "concept" is applied throughout this Technical Report to represent a "unit of thought" expressed in a terminology (it should also be noted that some terminological systems do not explicitly represent concepts, but rather terms, i.e. meaning cannot be assumed explicitly by the code or terms used), and

— the term “terminological resource” is applied throughout this Technical Report to collectively mean either a classification or a terminology used to classify or encode data in healthcare.

Examples are drawn from the published literature on mapping to illustrate key concepts and enhance understanding. However, full and complete understanding of the principles and guidelines requires some background knowledge of the coding of healthcare data, the various terminological systems used, and the many uses of the coded data.

Health informatics — Principles of mapping between terminological systems

1 Scope

This Technical Report provides guidance for organizations charged with creating or applying maps to meet their business needs. It explains the risks inherent in the mapping process and discusses the issues that need to be considered in the development, maintenance, and use of maps in health care. This Technical Report also identifies variations in process, precision, and administration when mapping for different purposes and in different environments.

Importantly, this Technical Report establishes and harmonizes the basic principles for developing, maintaining, and using maps and gives guidelines for good practice that underpin the mapping process. Terminological resources includes terminologies, classifications, and code systems used in the regulatory environment as it relates to healthcare and reporting requirements in healthcare.

This Technical Report is general in nature and does not describe the specific methods applied in the mapping process nor does it describe maps between databases and data sets, even though many of the principles stated here will apply to those types of maps. This Technical Report does not include consideration of the intellectual property rights and expectations of the owners of terminologies or classifications. It is the responsibility of the mapper and process to ensure that these legal rights are protected and acknowledged as part of the mapping processes.

2 Terms and definitions

Where there are terms used in this document that are not defined in this clause, they are considered to be generic to the English language and not specific to this document. Additional definitions and terms can be found at the international health informatics Standards Knowledge Management Tool and Glossary website www.skmtglossary.org.

Terms are presented alphabetically in logical groups and each definition is best understood through an understanding of the whole family of terms to which it belongs.

2.1 General

2.1.1
auto-matching
computational mapping task undertaken using an algorithm based upon the relationship between concepts

Note 1 to entry: Separate files of concept content from different coding systems are compared using an algorithm to determine whether there are concepts which match each other; that is, whether each coding system has content in common

[SOURCE: National eHealth Transition Authority — Australia (NEHTA), 2005]

2.1.2
cross map
see *map* (2.1.11)

2.1.3
cross map target
see *map target* (2.1.15)

2.1.4
data aggregation
process by which information is collected, manipulated, and expressed in summary form

Note 1 to entry: Data aggregation is primarily performed for reporting purposes, policy development, health service management, research, statistical analysis, and population health studies.

[SOURCE: ISO/TS 18308:2004]

2.1.5
cardinality
number of times a data element can repeat within an individual occurrence or object view

EXAMPLE A person can have one date of birth (cardinality 1) but *n* addresses in a lifetime (cardinality of *n* = many)

2.1.6
custodian
one who guards and protects or maintains property or records

2.1.7
equivalence
condition of being equal or the same in value, worth, or function

Note 1 to entry: In terminological systems, two concepts are (semantically) equivalent if their domain of meaning overlap and their semantic definitions are interpreted as identical.

Note 2 to entry: In the context of terminological resources, equivalence and semantic equivalence are often considered as synonyms.

[SOURCE: Oxford Dictionary (2013), modified]

2.1.8
human mapping
use of human knowledge and skill to build maps between concepts and/or terms in different coding systems

Note 1 to entry: Each map is built singly and individually. The process requires examination of each and every concept in each coding system within the scope of the map. Informed judgements or decisions are made about the shared meaning of concepts. Some electronic or computational tools are used but only in support of work process. The use of tools might still require manual oversight to determine equivalence of meaning.

[SOURCE: National eHealth Transition Authority — Australia (NEHTA): 2005, modified]

2.1.9
individual map
cross map
index from one term to another, sometimes using rules that allow translation from one representation to another indicating degree of equivalence

Note 1 to entry: Entry in a map which indicates how to translate from an individual source concept to a target concept. The term map is often used to indicate a table of individual map entries. It is for this reason that the individual and map tables are being differentiated.

Note 2 to entry: The use of this term is often used in ways which are confusing. It is essential to always make it clear whether you are referring to an individual map or a map table (or set).

Note 3 to entry: In SNOMED CT, each individual map is represented as a row or group of rows in a map Reference Set. It links a single map source concept code (e.g. SNOMED CT Concept ID) to one or more codes in a map target (e.g. ICD Code).

Note 4 to entry: A map is often computable and is the outcome of the mapping process.

2.1.10
map
See *individual map* (2.1.9) or *map table* (2.1.11)

2.1.11
map table
map reference set (in SNOMED CT)
map set
group of individual maps used to convert a range of entries from source to target code system

2.1.12
mapping
process of defining a relationship between concepts in one coding system to concepts in another coding system, in accordance with a documented rationale, for a given purpose

Note 1 to entry: Quality mapping will produce a usable map table, be a reproducible and understandable process.

2.1.13
map source
<mapping>
synonym: source
terminology, coding scheme, or classification used as the starting point for map production

Note 1 to entry: Map source is used as a term which can apply to an individual map source (a single code or term), as well as to the coding system. To differentiate between these, individual map source should be used when referring to a single term/concept.

2.1.14
map target
synonym: target (in a map), target scheme
terminology, coding scheme, or classification to which some or all of the concepts in another terminology, coding system, or classification (the map source) are mapped

Note 1 to entry: Map target is used as a term which can apply to an individual map source (a single code or term), as well as to the coding system. To differentiate between these, individual map target should be used when referring to a single term/concept.

Note 2 to entry: In SNOMED CT, some map targets might be derived from two or more associated statements and in these cases; the combination can be expressed as a set of associated rules. Each map target is represented as a row in the map table with each individual map target appearing at least once in the map reference set used to define the map table.

[SOURCE: International Health Terminology Standards Development Organization (IHTSDO) – to be non SNOMED/US specific, modified]

2.1.15
ontology
organization of concept for which a rational argument can be made

[SOURCE: ISO 17117]

2.1.16
scenario
description of high-level business activities defining process and requirements

Note 1 to entry: People often refer to scenarios as use cases, but in the context of this Technical Report, a use case is considered to be a more technical approach as defined in ISO/IEC 19501:2005.

2.1.17
semantic correspondence
measure of similarity between two concepts

Note 1 to entry: This term measures and indicates whether two concepts are semantically equivalent (thereby achieving equivalence) or not.

2.1.18
semantic equivalence
condition of being equal or the same in meaning

Note 1 to entry: In terminological systems, two concepts are semantically equivalent if their domain of meaning overlap and their semantic definitions are interpreted as identical.

Note 2 to entry: In the context of terminological resources, equivalence and semantic equivalence are often considered as synonyms.

2.2 Terminological resources

2.2.1
classification
exhaustive set of mutually exclusive categories to aggregate data at a pre-prescribed level of specialization for a specific purpose

[SOURCE: ISO 17115:2006]

Note 1 to entry: Classifications include a place, though not always specific, for all concepts required for the specific purpose of the classification. They include broad catch all categories and 'unspecified' sections to capture those concepts, where it is not possible or practical, for the purpose to be more specific.

2.2.2
classification

<healthcare>
synonym of statistical classification, statistical healthcare classification, healthcare classification system, and healthcare classification.

EXAMPLE International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10), International Standard Classification of Occupations.

Note 1 to entry: Classifications include a place, though not always specific, for all concepts required for the specific purpose of the classification. They include broad catch all categories and 'unspecified' sections to capture those concepts, where it is not possible or practical, for the purpose to be more specific.

2.2.3
coding system
combination of a set of concepts, a set of code values, and at least one code scheme mapping code values to coded concepts

[SOURCE: ISO 17115:2006]

EXAMPLE Country codes.

Note 1 to entry: Coded concepts are typically represented by terms but can have other representation. Code values are typically numeric or alphanumeric or a mixture of these, e.g. ICD Code J44.5.

2.2.4
coding system
<information technology>
collection of rules that maps the elements in one set, the "coded set" onto the elements in a second set.

[SOURCE: ISO 2382-4]

2.2.5
coding system
<health messaging>
combination of a set of code meanings and a set of code values, based on a coding scheme.

[SOURCE: Health Level 7 (HL7)]

2.2.6
concept
unit of knowledge created by a unique combination of characteristics

[SOURCE: ISO 1087-1:2000]

Note 1 to entry: A concept can be represented using one or more terms, pictures, icons, or sounds.

Note 2 to entry: Informally, the term "concept" is often used when what is meant is "concept representation". However, this leads to confusion when precise meanings are required. Concepts arise out of human individual and social conceptualizations of the world around them. Concept representations are artefacts constructed of symbols (from ISO 17115).

Note 3 to entry: Concept representations are not necessarily bound to particular languages. However, they are influenced by the social or cultural context of use often leading to different categorizations (from ISO 1087-1:2000).

[SOURCE: http://www.nlm.nih.gov/research/umls/meta2.html]

2.2.7
context
related conditions and situations that provide a useful understanding and meaning of a subject

[SOURCE: ISO/TR 17119:2005]

2.2.8
term
linguistic representation of a concept

Note 1 to entry: A term can contain symbols and have variants, e.g. different forms of spelling.

2.2.9
terminology
structured, human readable, and machine-readable representation of concepts

Note 1 to entry: This includes the relationship of the terminology to the specifications for organizing, communicating, and interpreting such a set of concepts.

[SOURCE: ISO 17115:2007, modified (used by IHTSDO)]

2.2.10
terminology
<healthcare> a terminology that is designed for use in computer systems. The term vocabulary or health or medical language is used to indicate the broader idea of linguistic representation without the specification of computability.

[SOURCE: ISO 17115:2007, modified (used by IHTSDO)]

2.2.11
terminological resource
controlled set of terms in healthcare

Note 1 to entry: Usually designed and controlled for use with computers for specific healthcare purpose, such as data entry, aggregation, retrieval, and analysis.