MEDINFO 2001
V. Patel et al. (Eds)
Amsterdam: IOS Press

Guideline for Health Informatics: Controlled Health Vocabularies - Vocabulary Structure and High-level Indicators

Peter L. Elkin[a], Steven H. Brown[b], Christopher G. Chute[a]

a) Mayo Foundation for Medical Education and Research
b) Vanderbilt University and Nashville VA Medical Center

Abstract

Developers and purchasers of controlled health terminologies require valid mechanisms for comparing terminological systems. By Controlled Health Vocabularies we refer to terminologies and terminological systems designed to represent clinical data at a granularity consistent with the practice of today's healthcare delivery. Comprehensive criterion for the evaluation of such systems are lacking and the known criteria are inconsistently applied. Although there are many papers, which describe specific desirable features of a controlled health vocabulary, to date there is not a consistent guide for evaluators of terminologies to reference, which will help them compare implementations of terminological systems on an equal footing[1,2] This guideline serves to fill the gap between academic enumeration of desirable terminological characteristics and the practical implementation or rigorous evaluations which will yield comparable data regarding the quality of one or more controlled health vocabularies.

Keywords:

Controlled Health Vocabularies; Terminology; Quality; Standards

Introduction

Over the past ten or more years, Medical Informatics researchers have been studying controlled vocabulary issues directly. They have examined the structure and content of existing vocabularies 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.[3] As we prepare to enter the twenty-first century, it seems appropriate to pause to reflect on this experience, and publish a guideline for the development of comparable, reusable, multipurpose, and maintainable controlled health vocabularies.

Guideline (See Figure #1)

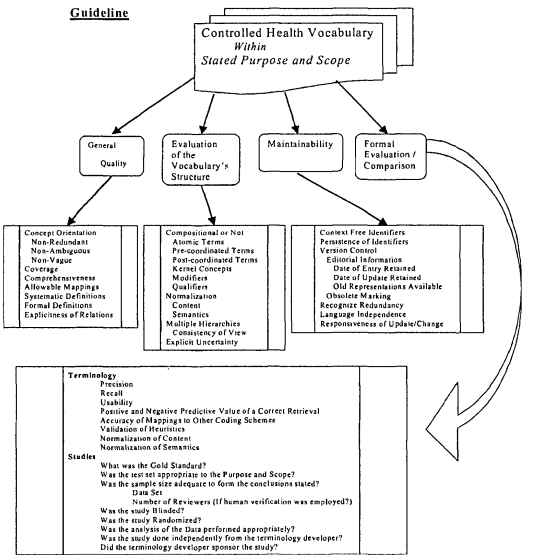

Figure 1 - Guideline

General Quality Metrics

Basic characteristics of a terminology influence its utility and appropriateness in clinical applications. The basic unit of a vocabulary must be a concept, which is the embodiment of some specific meaning and not a code or a character string. Identifiers of a concept must correspond to one and only one meaning and in a well-ordered vocabulary only one concept may have that same meaning. However, multiple terms (linguistic representations) may have the same meaning if they are explicit representations of the same concept. This implies non-vagueness, non-ambiguity and non-redundancy. Terminologies must be internally consistent. There must not be more than one concept in the terminology with the same meaning. This does not exclude synonymy, rather it requires that this be explicitly represented. No Concept should have two or more meanings. However a colloquial term (some authors have referred to this as an "interface term" or an "entry term")

can point to more than one Concept (e.g. MI as a Myocardial Infarction and Mitral Insufficiency). Concept names must be context free (some authors have referred to this as "context laden"). For example "diabetes mellitus" should not have the child concept "well controlled", instead the child concept's name should be "diabetes mellitus, well controlled."

Any controlled vocabulary must have its purpose and scope clearly stated in operational terms so that it 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 – *e.g.* a vocabulary 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 vocabulary sufficiently detailed to cope with cardiology and cardiothoracic surgery might be totally impractical in general practice.

Each segment of the health care process must have explicit in-depth coverage, and not rely on broad leaf node categories that lump specific clinical concepts together. For example, it is often important to distinguish specific diagnosis from categories presently labeled Not Elsewhere Classified (NEC), or to differentiate disease severity such as indolent prostate cancer from widely metastatic disease. The extent to which the depth of coverage is incomplete must be explicitly specified for each domain (scope), and purpose.

The extent to which the degree of comprehensiveness is incomplete must be explicitly specified for each domain (scope), and purpose of the terminology. Within the scope and purpose all aspects of the health care process must 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 include existing AHRQ guidelines, and the HCFA mortality model.*

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 be able to be mapped to classifications, such as ICD-9-CM. This need for multiple granularities is needed for clinical healthcare as well. For example an endocrinologist may specify more detail about a patient's Diabetes Mellitus than a generalist working in an Urgent Care setting, even though both specialties may be caring for the same patient. The degree to which the terminology is isolated from other classifications must be explicitly stated.

Evaluation of the Vocabulary's Structure

Terminology structures determine the ease with which practical and useful interfaces, for term navigation, entry, or retrieval can be supported.[4]

Compositional Terminologies

Composite concepts are created from two or more atomic or pre-coordinated concepts. Atomic concepts (Note: The term "Concept" in this document is used to refer to the Representation of a Concept rather than the thought itself) must be able to be combined to create composite concepts[5]. A concept is a notion represented by language, which identifies one idea. For 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: the 'kernel concept', 'qualifier (also called "status") concept', and 'modifier concepts'.

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 the UMLS Metathesaurus, Colon is a synonym for Large Bowel and Cancer is a synonym for Neoplasm, Malignant. Whereas Colon Cancer is non-atomic as it can be broken down into "Large Bowel" and "Neoplasm, malignant". Each of these two more atomic terms has a separate and unique Concept Unique Identifier (CUI). A composite concept is a concept composed as an expression made up of atomic concepts linked by Semantic Representations (such as Roles, Attributes or Links).

Pre-coordinated concepts are entities, which 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 vocabulary (that is the vocabulary should be Normalized for Content). *Example*: Colon Cancer is non-atomic, however it has a single CUI, which means to the Metathesaurus that it represents a "single" concept. It has the same status in the vocabulary as the site "Large Bowel" and the diagnosis "Neoplasm, malignant."

Post-coordinated concepts are composite concepts, which are not pre-coordinated and therefore must 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

vocabulary 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.

We can classify unique concept representations within a vocabulary 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. A 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. Terms which refine the meaning of a Kernel Concept are constituents of a composite concept which refine the meaning of a Kernel concept, *e.g.* 'stage 1a' in 'having colon cancer *stage 1a*', or 'brittle, poorly controlled', in '*Brittle, poorly controlled* diabetes mellitus'. In general, these concepts are expressed as a link plus a value ('attribute-value pair'). Terminologies must support a logical structure that can support temporal duration and trend. Attributes must be themselves elements of a terminology, and fit into a practical model that extends a terminology. For example, cancers may be further defined by their stage and histology, have been symptomatic for a specifiable time, and may progress over a given interval. Attributes are required to capture important data features for structured data entry and pertinent to secondary data uses such as aggregation and retrieval. Kernel concepts can be refined in many ways including a clinical sense, a temporal sense, and by status terms (e.g. "Recurrent").

Normalization is the process of supporting and mapping alternative words and shorthand terms for composite concepts. All pre-coordinated concepts must be mapped to or logically recognizable by all possible equivalent post-coordinated concepts. There should be mechanisms for identifying this synonymy for user created ("New") post-coordinated concepts as well (i.e. when there is no pre-coordinated concept for this notion in the vocabulary). This functionality is critical to define explicitly equivalent meaning, and to accommodate personal, regional, and discipline specific preferences. Additionally, the incorporation of non-English terms as synonyms can achieve a simple form of multilingual support.

Normalization of semantics is required to insure comparable data in post-coordinated expressions. In compositional systems, there exists the possibility of representing the same concept with multiple potential sets of atoms which may be linked by different semantic links. In this case the vocabulary needs to be able to recognize this redundancy / synonymy (depending on your perspective). The extent to which normalization can be performed formally by the system should be clearly indicated. For example the concept represented by the term "Laparoscopic Cholecystectomy" might be represented in the following two dissections:

"Surgical Procedure: Excision"{Has Site Gallbladder}, {Has Method Endoscopic} and

"Surgical Procedure: Excision"{Has Site Gallbladder}, {Using Device Endoscope}.

A compositional system should contain formal definitions for non-atomic concepts and formal rules for inferring subsumption from the definitions. The logical definition of subsumption should be defined. The formal behavior of all links/relations/attributes should be explicitly defined. The primary hierarchical relation should be subsumption ('kind-of') as defined by logical implication: 'B is a kind of A' means 'All Bs are As'. If a looser meaning such as 'broader than/narrower than' is used, it should be explicitly stated.

All Terminologies

Concepts should be accessible through all reasonable hierarchical paths (i.e. they must allow multiple semantic parents), e.g. stomach cancer can be viewed as a neoplasm or as a gastrointestinal disease. A balance between number of parents (as siblings) and number of children in a hierarchy should be maintained. This feature assumes obvious advantages for natural navigation of terms (for retrieval and analysis), as a concept of interest can be found by following intuitive paths (i.e. users should not have to guess where a particular concept was instantiated).

A concept in multiple hierarchies must be the same concept in each case. Our example of stomach cancer must not have changes in nuance or structure when arrived at via the cancer hierarchy as opposed to GI diseases. Inconsistent views could have catastrophic consequences for retrieval and decision support, by inadvertently introducing variations in meaning which may be unrecognized and therefore be misleading to users of the system.

Uncertainty should be represented explicitly. Notions of "probable", "suspected", "history of" or differential possibilities (i.e. a Differential Diagnosis list) must be supported. The impact of certain versus very uncertain information has obvious impact on decision support and other secondary data uses. Similarly, in the case of incomplete syndromes clinicians should be able to record the partial criteria consistent with the patient's presentation. This criterion is listed separately as many current terminological systems fail to address this adequately.

Computer coding of concept identifiers must *not* place arbitrary restrictions on the terminology, such as numbers of digits, attributes, or composite elements. To do so subverts meaning and content of a terminology to the limitations of format, which in turn often results in the assignment of concepts to the wrong location because it might no longer "fit" where it belongs in an hierarchy. These reorganizations confuse people and machines alike, as intelligent navigation agents are led astray for arbitrary reasons. The long, sequential, alphanumeric tags used as concept identifiers in the UMLS project of the National

Library of Medicine exemplify well this principal (i.e. A meaningless identifier).

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

Maintainability of a Terminology

Technical choices can impact the capacity of a terminology to evolve, change, and remain usable over time. Unique codes attached to concepts must not be tied to hierarchical position or other contexts; their format must not carry meaning. This is known as a terminology with "concept free identifiers." Because health knowledge is being constantly updated, how we categorize health concepts is likely to change (e.g. Peptic Ulcer Disease is now understood as an infectious disease, but this was not always so.) For this reason, the "code" assigned to a concept must not be inextricably bound to a hierarchy position in the terminology, so that we need not change the code as we update our understanding of, in this case, the disease. Changing the code may make historical patient data confusing or erroneous. This notion is the same as using Non-Semantic Identifiers. Codes must not be re-used when a term is obsolete or superseded. Consistency of patient description over time is not possible when concepts change codes; the problem is worse when codes can change meaning. This practice not only disrupts historical analyses of aggregate data, but also can be dangerous to the management of individual patients whose data might be subsequently misinterpreted. This notion is often designated by the term "Concept Permanence."

Updates and modifications must be referable to consistent version identifiers. Usage in patient records should carry this version information. Because the interpretation of coded patient data is a function of terminologies that exist at a point in time[6] (e.g. AIDS patients were coded inconsistently before the introduction of the term AIDS). Terminology representations should specify the state of the terminology system at the time a term is used; version information most easily accomplishes this, and may be hidden from ordinary review. The frequency of updates, or sub-versions, should be sufficiently short to accommodate new codes and repairs quickly, ideally on the order of weeks.

New and revised terms, concepts, and synonyms must have their date of entry or effect in the system, along with pointers to their source and / or authority. Previous ways of representing a new entry should be recorded for historical retrieval purposes. Superseded entries should be so marked, together with their preferred successor. Because data may still exist in historical patient records using obsolete terms, their future interpretation and aggregation are dependent upon that term being carried and cross-referenced to subsequent terms (e.g. HTLV III to HIV).

Authors of these large-scale vocabularies will need mechanisms to identify redundancy when it occurs. This is essential for the safe evolution of any such vocabulary. This implies Normalization of Concepts and Semantics, and specifically addresses the need for vocabulary systems to provide the tools and resources necessary to accomplish this task.

It would be desirable for terminologies to support multilingual presentations. As healthcare confronts the global economy and multiethnic practice environments, routine terminology maintenance should evolve toward supporting multilingual representations. While substantially lacking the power and utility of machine translation linguistics, this simplistic addition will enhance understanding and use in non-English speaking areas. Evaluators of terminologies should inquire whether or not there have been translations, and if not what is the expected cost of translation?

Formal Evaluation of Terminologies

As we seek to understand quality in the controlled vocabularies that we create or use, we need standard criteria for the evaluation of these systems. All evaluations should reflect and specifically identify the purpose and scope of the vocabulary being evaluated.[7]

What are the vocabulary's precision and recall for mapping Diagnoses, Procedures, Manifestations, Anatomy, Organisms, etc., against an established and nationally recognized standard query test set? This should be evaluated only within the intended scope and purpose of the vocabulary system.

Is a standard search engine used in the mapping process? Where different tool sets are employed for the evaluation of competing terminological implementations, bias may be introduced.

Has the usability of the vocabulary been verified? How have interface considerations been separated from vocabulary evaluation? What support exists for the development of user interfaces. Has an effective user interface been built? Has the vocabulary been shown to have an effective user interface for its intended use? If not, what are the questions or issues outstanding? This includes documentation of the speed of entry, accuracy, comprehensiveness in practice etc. with different approaches, and support for computer interfaces and system implementers. Is there a proof of concept implementation? Is there a demonstrated proof of concept implementation in software? Can it be shown to be usable for the primary purpose indicated? Have there been failed implementations?

To what extent is the vocabulary mappable to other coding systems or reference terminologies? To what extent can the vocabulary accommodate local terminological enhancements? Can the vocabulary server respond to queries sent over a network (LAN, WAN)?

For compositional terminologies, to what extent can the terminology claim to be complete with respect to the normalization of content and semantics? As we have already asserted that normalization of a compositional terminology is essential to maintain comparable data, this needs to be a primary measure of a terminology's reliability. We assert that one method of testing this is to run all of the pre-coordinated terms from within the terminology through the compositional engine intended for use with it (excluding the exact match). Then measure to what extent all of the potential equivalent compositions are made explicit within the terminology being evaluated.

For each formal evaluation used to make an opinion regarding the quality of a terminology, it is useful to ascertain whether or not the methodology used in the evaluation is valid and generalizable to your environment.

Conclusion

This guideline is intended to document the principal notions, which are necessary and sufficient to assign value to a controlled health vocabulary. It is applicable to all areas of healthcare about which information is kept or utilized. This guideline explicitly does not refer to classifications or coding systems, which are not designed to be used clinically.

This guideline will provide vocabulary developers and authors with the principles needed to construct useful, maintainable controlled health vocabularies. These tenets do not attempt to specify all of the richness, which can be incorporated into a health terminology. However this guideline does specify the minimal requirements, which if not adhered to will assure that the vocabulary will have limited generalizability and will be very difficult if not impossible to maintain. This guideline will provide terminology developers with a sturdy starting point for the development of controlled health vocabularies. This foundation serves as the basis from which vocabulary developers will build robust large-scale reliable and maintainable terminologies.

For consumers and users who find the sea of terminologies and terminological tool sets over-whelming, we have created a practical guide and an enumerated set of questions to ask to insure their success in comparing alternative systems.

References

[1] Cimino, JJ. Desiderata for controlled medical vocabularies in the Twenty-First Century. *Meth Inform Med* 1998; 37:394-403.

[2] Chute CG, Cohn SP, Campbell JR. A framework for comprehensive health terminology systems in the United States: Development guidelines, criteria for selection, and public policy implications. *JAMIA* 1998; 5(6):503-10.

[3] Solbrig, H. (1998). Final submission to the CorbaMED Request for Proposals on Lexical Query Services (CorbaLex), OMG. http://www.omg.org/cgi-bin/doc?formal/99-3-6.pdf or http://www.omg.org/cgi-bin/doc?formal/99-3-1.pdf

[4] Elkin PL, Chute CG, et al. "The Role of Compositionality in Standardized Problem List Generation." in Safran(ed), Medinfo Proceedings, 1998

[5] Bernauer J, Franz M,Schoop D, Schoop M, Pretschner DP. The Compositional Approach for Representing Medical Concept Systems. Medinfo 95;8 Pt (1):70-4.

[6] Cimino JJ. Formal Descriptions and Adaptive Mechanisms for Changes in Controlled Medical Vocabularies, *Methods of Information in Medicine* 1996; 35(3): 211-217.

[7] Elkin PL, Chute CG. ANSI-HISB Code Set Evaluation Criterion Survey, 1998; Minutes ANSI-HISB meeting 4/98.

Address for correspondence

Peter L. Elkin
Mayo Clinic
200 First Street SW
Rochester, MN 55905
(507) 266-4569
elkin.peter@mayo.edu