PMRI Terminology – Questions for Testifiers

The purpose of the NCVHS Subcommittee on Standards and Security hearings for August 28, 2002 will be to:

(1) define the scope of PMRI terminologies
(2) determine the criteria for selection of PMRI terminologies.

SCOPE OF PMRI TERMINOLOGIES

1. *Is there a better way to group or organize the subsets of PMRI terminologies?*
   a. *What new groups or subsets do you suggest?*

   I believe that there are two additional types of categorizations, which should be endorsed by NCVHS. The first is a categorization by level of granularity and the second is a categorization by Purpose and Scope.

   With the former I recommend broad strokes. I would divide the categorizations into the following three categories. The most granular would be comprehensive compositional systems that allow aggregation of compositional expressions. These terminologies can represent conceptually billions of different representation. This level of detail is most appropriate for clinical data representation (e.g. For direct patient care). Examples of this type of terminology would be SNOMED-CT or GALEN. The second level would be administrative classifications that are aggregations of detailed clinical data for a coding purpose. These classifications are usually not compositional but when they are they have fairly limited and strict rules of composition associated with their usage. Here the prototype example would be ICD9-CM. A third type of coding scheme is a high level aggregation (e.g. for reimbursement) such as is used in the DRG codes for hospital inpatient reimbursement. These in some cases can be directly aggregated from administrative classifications such as ICD9-CM.

   The second type of categorization in by Scope and Purpose. 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 CMS mortality model.*

b. *Should any new PMRI terminologies or groups be added?* Yes.
    1. Foundation Model of Anatomy: This is probably the most comprehensive representational system for Anatomical knowledge. Cornelius Rosse and the Structural Informatics Lab at the University of Washington in Seattle have worked tirelessly to build this compendium. As anatomy is a core discipline within medicine, I suggest adding this terminology to the list of NCVHS recommended vocabularies for clinical data representation.
    2. Dermatology Lexicon This NIH contracted development effort at the University of Rochester is working on the development of a publicly available terminology aimed at the representation of core concepts needed to represent skin diseases. The lexicon is slated to become available in 2003. I suggest this also be included as a clinical terminology.
    3. Patient Safety Coding System This AHRQ funded effort is designed to develop a set of codes that would be suitable for inclusion in ICD9. This classification should also be useful to developers interested in building expert systems directed at improving patient safety. Given the utility of separate utilization of this code set, I suggest that this classification be added to the list of administrative classifications (here also called "Diagnosis and Procedure" codes).
    4. Gene Ontology I suggest adding the gene ontology (GO) to the list of NCVHS terminologies. The GO classification is developed and maintained by the Gene Ontology Consortium and is an emerging standard for the high level organization of genetic concepts. As genomics and proteomics begin to make its

way into the routine practice of medicine, we need to be prepared to represent this type of clinical data. I suggest that it be added to the clinical terminologies as I expect a) that this would be its greatest utility and b) that over time the representation will become increasingly granular.

c. *Should any PMRI terminologies or groups be deleted?* No

2. *Based on the existing graphic, or a new way that you have grouped or organized the subsets of PMRI terminologies:*

   a. *Which categories or subsets should receive the highest priority for the PMRI selection process?* Clinical Terminologies. Please note: That I would also put SNOMED-CT into this category.
   b. *Why did you give these categories or subsets a high priority?* The greatest need if for detailed granular representation of clinical data to fuel decision support, continuing medical education and improved patient safety through guideline-based appropriate patient care.
   c. *What groups or subsets of PMRI terminologies should be a low priority?* The second tier of priorities should be to support the administration of the practice.

Quality Criteria for NCVHS Terminologies

I recommend using a more formal evaluation method for the quality criteria. This criterion is intended to be used in evaluating clinically relevant terminologies. The criteria for classifications are currently being constructed within ASTM E31.01. As an example of the criteria that we suggest I have included this excerpt from ASTM E2087. The implementation guide is similar to your criteria list but provides a somewhat more complete set of questions. The definitions that follow provide supporting material so that you can better understand the meaning of the criteria. I suggest that the terminologies not be ANSI accredited, but that they meet the current ANSI standard for quality of controlled vocabularies (ASTM E2087).

Annex A
(Normative)

Implementation Guide
1 General

Basic Characteristics of a terminology influence its utility and appropriateness in clinical applications.

1.1
Concept Orientation

Is the terminology concept oriented? To how many meanings can one identifier correspond? This must be the case.[iii]

1.1.1
Non-Redundancy

Can concepts be redundantly instantiated within the terminology? This must not be the case.

1.1.2
Non-Ambiguity

Can concepts be ambiguous? This must not be the case.

1.1.3
Non-Vagueness

Are concept definitions independent of their context? This must be the case.

1.1.4
Internal Consistency

Are the relationships used in the terminology applied consistently? This must be the case.

1.2
Purpose and Scope

What is the purpose of the terminology? What is the scope of the terminology? Please state these in operational terms (what functions is the terminology intended to serve?).

1.2.1
Coverage

What is the intended coverage of the terminology?[i]

1.2.2
Comprehensiveness

What is the degree of comprehensiveness (expressed in percent completion) of the terminology within the intended area of coverage? What studies can be referenced to support this assertion (use the criteria under section #4 for assess the validity and generalizability of the study referenced)?[ii]

1.3
Mapping

Is the terminology mappable to classifications or other terminologies? If so, which ones? If it is partially mappable to some classifications or other terminologies, to what extent is this true (expressed in percent completion)? Use the criteria under section #4 for assess the validity and generalizability of the study referenced?[iii]

1.4
Systematic Definitions

Are the meanings of each specific concept within the terminology made available for the users? These should be provided.

1.5
Formal Definitions

Does your terminology support formal definitions? If so, to what extent (expressed in percent completion) is it fully defined? What studies can be referenced to support this assertion (use the criteria under section #4 for assess the validity and generalizability of the study referenced)? It is essential that reference terminologies support formal definitions.

1.6
Explicitness of Relations

Does your terminology support formal subsumption? To what extent are the hierarchies automatically generated by the description logic (expressed as a percentage of all the concepts contained in the terminology)? This is a desirable characteristic.

1.7
Reference Terminologies

Is the terminology intended to be used as a reference terminology?

1.8
Atomic Reference Terminologies

Is there an explicit mechanism for identifying the atomic portion of the reference terminology? Is it intended that pre-coordinated terms can be used within compositional expressions? This should be a goal of all reference terminologies.

1.9
Colloquial Terminologies

Specifically, what is the association between the colloquial terms and the reference terminology? How are these two terminologies maintained so as not to create ambiguous or redundant instantiation of data? This is necessary for all reference terminologies intended to be used clinically.

2 Structure of the Terminology Model

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

For Compositional Terminologies:

2.1
Compositionality

Does your terminology support the creation of compositional expressions? How is a compositional expression created? If this is governed by rules please elaborate them. If so, can you identify equivalence between arbitrary compositional expressions? If so, by what method?

2.1.1
Atomic Concept

Do you make explicit which of your concepts are atomic?

2.1.2
Composite Concept

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

2.1.2.1
Pre-Coordinated Concept

Does your terminology make explicit which concepts are pre-coordinated? This must be true for all compositional terminologies.

2.1.2.2
Post-coordinated Concept

Does your terminology support the creation of post-coordinated expressions?

2.1.3
Types of Atomic and Pre-coordinated Concepts

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.

2.1.3.1
Kernel Concept

Does your terminology identify separately kernel concepts? This should be identified by compositional terminologies.

2.1.3.2
Terms Which Refine the Meaning of a Kernel Concept

Does your terminology identify modifiers and qualifiers within the terminology? If so, how are they used? This should be identified by compositional terminologies.

2.2
Normalization of Content

Is the content of the terminology normalized? What studies can be referenced to support this assertion (use the criteria under section #4 for assess the validity and generalizability of the study referenced)? This must be accomplished for all compositional terminologies.

2.3
Normalization of Semantics

Are the semantics of the terminology normalized? What studies can be referenced to support this assertion (use the criteria under section #4 for assess the validity and generalizability of the study referenced)? For compositional expressions, is it possible to represent the same concept with different semantics? This must be accomplished for all compositional terminologies.

2.4
Multiple Hierarchies

Are multiple hierarchies supported? Are they present within the current version of the terminology?[iv]

2.5
Consistency of View

Is a consistency of views into the terminology maintained? This must be the case for terminologies that support multiple hierarchies.[v]

2.6
Explicit Uncertainty

Does your terminology support the input of explicit uncertainty and incomplete syndromes? This should be a feature of compositional terminologies.

2.7
Representational Form

Does the representational form of the concept identifier place restrictions on the terminology? If so, what are the restrictions? This must not be the case.

3 Maintenance

Technical choices can impact the capacity of a terminology to evolve, change, and remain usable over time.

3.1
Context Free Identifiers

Does the terminology support context free identifiers? This must be the case.[vi]

3.2
Persistence of Identifiers

Are codes ever reused for different concepts? If so, when can this occur? This must be the case.

3.3
Version Control

Are your codes tied explicitly to the version of the terminology? This must be the case.[vii]

3.3.1
Editorial Information

When the terminology is revised, do you record the date of the update and the source or authority of the information leading to the update? This must be the case.

3.3.2
Obsolete Marking

Have you included obsolete marking in your entries? This must be the case

3.4
Recognize Redundancy

Does your terminology recognize redundancy? If so, how is this accomplished? This must be the case.

3.5
Language Independence

Is your terminology presently multilingual? If not, does it have the capacity to become multilingual? If so, please explain. This should be the case.

3.6
Responsiveness

What is the frequency of updates to the terminology? Is it less than or equal to 12 weeks? This should be the case.

4 Evaluation

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 must reflect and specifically identify the purpose and scope of the vocabulary being evaluated.[viii] These criteria stipulate the methods for evaluating studies, which make claims regarding controlled terminologies. These criteria are also useful as a guide to individuals or organizations who wish to perform valid and useful evaluations of one or more controlled health terminologies.

4.1
Purpose and Scope

Important dimensions along which scope should be defined include:

4.1.1
Clinical Area

What is the clinical area of use of the terminology, the disease area of patients addressed and / or the expected profession of users. Within what parts of healthcare is it intended to be used and by whom?

4.1.2
Primary Use

What is the primary intended usage of the terminology? Examples include: reporting for remuneration, management planning, epidemiological research, indexing for bibliographic, Web-based retrieval, recording of clinical details for direct patient care, use for decision support, linking of record to decision support, etc.

4.1.3
Persistence and Extent of Use

Is the intent of the terminology to persist and evolve?. If intended to be persistent, what are the means of updating or change management, etc?

4.1.4
Degree of Automatic Inferencing

Is the terminology intended to support automated classification? Is it is intended that validation on input be possible, and within what limits? Whether post-coordinated expressions are to be accepted and if so what can be inferred about them and what restrictions must be placed on them?

4.1.5
Transformations (Mappings) to Other Vocabularies

What transformations / mappings are supported for what intended purpose (e.g. transformation for purposes of bibliographic retrieval may require less precision than transformation for clinical usage)? What is the sensitivity and specificity of the mappings?

4.1.6
User / Developer Extensibility

Is it intended that the vocabulary be extended by users or application developers? If so, within what limits? If not, what mechanisms are available for meeting new needs as they arise?

4.1.7
Natural Language

Is natural language input or output supported (for analysis or input)? To what level of accuracy?

4.1.8
Other Functions

What other functions are intended? For example, linkage to specific decision support systems, linkage to post-marketing surveillance, etc.

4.1.9
Current Status

To what extent is the system intended to be "finished" or work in progress? If different components of the terminology are at different stages of completion how is this indicated?

4.2
Measures of Quality - Terminological Tools

4.2.1
Interconnectivity (Mapping)

4.2.1.1
Vocabulary and Other Coding Systems

To what extent is the vocabulary mappable to other coding systems or reference terminologies?

4.2.1.2
Vocabulary and Terminological Enhancements

To what extent can the vocabulary accommodate local terminological enhancements?

4.2.1.3
Vocabulary and Networking

Can the vocabulary server respond to queries sent over a network (LAN, WAN)?

4.2.2
Precision and Recall

4.2.2.1
Vocabulary

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, using a standard well-principled method? This should be evaluated only within the intended scope and purpose of the vocabulary system.

4.2.2.2
Search Engine

Is a standard search engine used in the mapping process?

4.2.3
Usability

4.2.3.1
Validation of Usability

Has the usability of the vocabulary been verified?

4.2.3.2
Interface Considerations

How have interface considerations been separated from vocabulary evaluation?

4.2.3.3
Prototypes

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? Evidence for speed of entry, accuracy, comprehensiveness in practice etc. with different approaches? If not, is there a proof of concept?

4.2.3.4
Application Programmer Interfaces

Is there support for computer interfaces and system implementers? 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?

4.2.4
Feasibility

If it is intended for use in an Electronic Patient Record (EPR), what are the options for information storage? Has feasibility been demonstrated?

4.3
Measures of Quality

The generalizability (applicability) of any Study Design reported (Evaluating Reported Evaluations) should be able to be evaluated.

4.3.1
Healthcare / Clinical Relevance

What is the vocabulary's Healthcare / Clinical Relevance?

4.3.2
Gold Standard

What was the Gold Standard used in the evaluation?

4.3.3
Study Population

If published population rates are used for comparison, was the study population comparable to the population from which the rates were derived?

4.3.4
Specific Aims

Were the Specific Aims clear?

4.3.5
Blinding

Was the study appropriately blinded?

4.3.6
Randomization

Was the Test Set Selection randomized or shown in some sense to be a representative sample of the end user population?

4.3.7
Test Location

4.3.7.1
Independence

Was it different from the developer's location?

4.3.7.2
Appropriate for Study Design

How was the test site suited to the study design (tools, resources, etc.)?

4.3.7.3
Principal Investigator Associations

Was the Principal Investigator associated with:

4.3.7.3.1
University

4.3.7.3.2
Academic Medical Center

4.3.7.3.3
Corporation

4.3.7.3.4
Hospital

4.3.7.3.5
Government Agency

4.3.7.3.6
HMO

4.3.7.3.7
Private Practice

4.3.7.3.8
Academic Organization

4.3.7.4
Principal Investigator

4.3.7.4.1
Was the Principal Investigator independent of the vocabulary being evaluated?

4.3.7.4.2
Does the Principal investigator have a track record of publication in this field of study?

4.3.7.4.3
Have there been any conflicts of interest in performing this research?

4.3.8
Project Completion
Was the project completed in a reasonable period of time?

4.3.9
Sample Size

4.3.9.1
Power

Was the sample size of sufficient size to show the anticipated effect, should one exist?

4.3.9.2
Statistics

Who reviewed the Statistical Methods?

4.3.10
Personnel

4.3.10.1
Training Level

What is the average level of training of the study personnel?

4.3.10.2
Reviewers

4.3.10.2.1
Variability

What is the inter-reviewer variability?

4.3.10.2.2
Type

What was the type of reviewer (physician, nurse, other clinician, coder, knowledge engineer) used in the study?

4.3.10.2.3
Independence

Were the reviewers blinded to the other reviewers' judgments (i.e. reviewer independence)?

Supporting Definitions:

1 Terms and Definitions

For the purposes of this ASTM E-2087, the following terms and definitions apply:

1.1
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.

1.2
Controlled Health Vocabulary

A terminology intended for clinical use. 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.

1.3
Classification

A terminology, which aggregates data at a prescribed level of abstraction for a particular domain. This fixing of the level of abstraction that can be expressed using the classification system is often fixed to enhance consistency when being the classification is to be applied across a diverse user group, such as is the case with some of the current billing classification schemes.

1.4
Ontology

An organization of concepts for which one can make a rational argument. Colloquially, this term is used to describe a hierarchy constructed for a specific purpose. For example a hierarchy of qualifiers would be a Qualifier Ontology.

1.5
Qualifier

A String which when added to a term changes the meaning of the term in a Temporal or Administrative sense. For example: “History of” or “Recurrent”.

1.6
Modifier

A string which when added to a term changes the meaning of the term in the Clinical sense. For example: clinical stage or severity of illness.

1.7
Canonical Term

A preferred atomic or pre-coordinated term for a particular medical concept.

1.8
Term

A word or words corresponding to one or more concepts.

2 General

2.1
Basics

Basic characteristics of a terminology influence its utility and appropriateness in clinical applications.

2.2
Concept Orientation

The basic unit of a terminology must be a concept, which is the embodiment of some specific meaning and not a code or 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 (DIS 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.

2.2.1
Non-Redundancy

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

2.2.2
Non-Ambiguity

No concept identifier should have more than one meaning. However, an entry term (some authors have referred to this as an "interface terminology") can point to more than on concept e.g. MI as Myocardial Infarction and Mitral Insufficiency).

2.2.3
Non-Vagueness

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."

2.2.4
Internal Consistency

Relationships between concepts should be uniform across parallel domains within the terminology. For example, if heart valve structures are specified anatomically the diagnosis related to each structure should also be specified using the same relationships. (Note Schultz reference)

2.3
Purpose and Scope

Any controlled vocabulary must have its purpose and scope clearly stated in operational terms so that its fitness for particular purposes can be assessed and evaluated (ISO 15188). 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. For example, 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.

2.3.1
Coverage

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 as indicated in section 4.3.[ix]

2.3.2
Comprehensiveness

The extent to which the degree of comprehensiveness is incomplete must be explicitly specified for each domain (scope), and purpose as indicated in section 4.3. 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 Agency for Healthcare Research and Quality guide-lines, and the Health Care Finance Administration (HCFA) mortality model.[x]

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 be able to be mapped to those classifications, such as ICD-9. This need for multiple granularities is needed for clinical healthcare as well (ISO TR 9789). 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 mappable to other classifications must be explicitly stated.[xi]

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 vocabulary have assigned these definitions will need to be explicit and available to the users. Further as relationships are built into vocabularies multiple authors will need these definitions to ensure consistency in authorship. For example, the concept "Hypertension" might be defined as a consistently elevated Blood Pressure and not "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 (E-1712).

4.7
Explicitness of Relations

The logical definition of subsumption should be defined. The formal behavior 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. For 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. These features define the core of the controlled health terminology.

4.9
Atomic Reference Terminologies

A Reference Terminology consisting of only Atomic concepts and their systematic definitions. In this type of reference terminology, no two or more concepts can be combined to create a composite expression as the same meaning as any other single concept contained in the Atomic Reference Terminology.

4.10
Colloquial Terminologies

The set of terms, which consist of commonly used entry points, which map to one or more canonical terms within the vocabulary. These have been called "entry terms" or "interface terminologies" by different authors.

5 Structure of the Terminology Model

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

For Compositional Terminologies:

5.1
Compositionality

Composite concepts are created from Atomic concepts (Note: The term "Concept" in this document is used to refer to the Representation of a Concept rather than the thought itself; also see definition below) must be able to be combined to create composite concepts[xii]. 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".

5.1.1
Atomic Concept

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. For 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), as does the pre-coordinated term "Colon Cancer."

5.1.2
Composite Concept

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

5.1.2.1
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 vocabulary (that is the vocabulary should be Normalized for Content). For 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."

5.1.2.2
Post-coordinated Concept

A composite concept, which is 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. For 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.

5.1.3
Types of Atomic and Pre-coordinated Concepts

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.

5.1.3.1
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.

5.1.3.2
Terms Which Refine the Meaning of a Kernel Concept

Constituents of a composite concept which refine the meaning of a Kernel concept. For example, "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").

5.2
Normalization of Content

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.

5.3
Normalization of Semantics

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:

5.3.1
"Surgical Procedure: Excision"{Has Site Gallbladder}, {Has Method Endoscopic}

and

5.3.2
"Surgical Procedure: Excision"{Has Site Gallbladder}, {Using Device Endoscope}.

5.4
Multiple Hierarchies

Concepts should be accessible through all reasonable hierarchical paths (i.e. they must allow multiple semantic parents). For example, 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).[xiii]

5.5
Consistency of View

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.[xiv]

5.6
Explicit Uncertainty

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.

5.7
Representational Form

The representational form of the identifiers within the terminology should be meaningless. 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 principle.

6 Maintenance

Technical choices can impact the capacity of a terminology to evolve, change, and remain usable over time.

6.1
Context Free Identifiers

Unique codes attached to concepts must not be tied to hierarchical position or other contexts; their format must not carry meaning. 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 Non-Semantic Identifiers.[xv]

6.2
Persistence of Identifiers

Codes must not be re-used when a concept 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 can be dangerous to the management of individual patients whose data might be subsequently misinterpreted. This encompasses the notion of Concept Permanence.

6.3
Version Control

Updates and modifications must be referable to consistent version identifiers. Usage in patient records should carry this version information. This is true because the interpretation of coded patient data is a function of terminologies that exist at a point in time (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 (IS 15188, IS 12620, IS 1087-2, IS 11179-3, IS 2382/4).[xvi, xvii]

6.3.1
Editorial Information

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.

6.3.2
Obsolete Marking

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).

6.4
Recognize Redundancy

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, but specifically addresses the need for vocabulary systems to provide the tools and resources necessary to accomplish this task.

6.5
Language Independence

It would be desirable for terminologies to support multi-lingual presentations. As healthcare confronts the global economy and multiethnic practice environments, routine terminology maintenance must incorporate multilingual support. While substantially lacking the power and utility of machine translation linguistics, this simplistic addition will enhance understanding and use globally. Have there been translations? What is the expected cost of translation?

6.6
Responsiveness

The frequency of updates, or sub-versions, should be sufficiently short to accommodate new codes and repairs quickly, ideally on the order of weeks.

7 Evaluation

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 must reflect and specifically identify the purpose and scope of the vocabulary being evaluated.[xviii]

7.1
Purpose and Scope

Important dimensions along which scope should be defined include:

7.1.1
Clinical Area

What is the clinical area of use of the terminology, the disease area of patients addressed and / or the expected profession of users. Within what parts of healthcare is it intended to be used and by whom?

7.1.2
Primary Use

What is the primary intended usage of the terminology? Examples include: reporting for remuneration, management planning, epidemiological research, indexing for bibliographic, Web-based retrieval, recording of clinical details for direct patient care, use for decision support, linking of record to decision support, etc.

7.1.3
Persistence and Extent of Use

While some vocabularies are intended, at least initially, primarily for a specific study or a specific site, others are not. If intended to be persistent, what are the means of updating or change management, etc?

7.1.4
Degree of Automatic Inferencing

Developers should define whether or not and to what degree automatic inferencing is intended. Developers should define whether or not classification is intended to be automatic. Developers should define whether or not it is intended that validation on input be possible and within what limits? Developers should define whether or not post-coordinated expressions are to be accepted and if so, what can be inferred about them and what restrictions must be placed on them (is formal sanctioning required)?

7.1.5
Transformations (Mappings) to Other Vocabularies

What transformations / mappings are supported for what intended purpose? For example, transformation for purposes of bibliographic retrieval may require less precision than transformation for clinical usage? What is the sensitivity and specificity of the mappings?

7.1.6
User / Developer Extensibility

Is it intended that the vocabulary be extended by users or application developers? If so, within what limits? If not, what mechanisms are available for meeting new needs as they arise?

7.1.7
Natural Language

Is natural language input or output supported (for analysis or input)? To what level of accuracy?

7.1.8
Other Functions

What other functions are intended? For example, linkage to specific decision support systems, linkage to post-marketing surveillance, etc.

7.1.9
Current Status

To what extent is the system intended to be "finished" or work in progress? If different components of the terminology are at different stages of completion how is this indicated?

7.2
Measures of Quality - Terminological Tools

7.2.1
Interconnectivity (Mapping)

7.2.1.1
Vocabulary and Other Coding Systems

To what extent is the vocabulary mappable to other coding systems or reference terminologies?

7.2.1.2
Vocabulary and Terminological Enhancements

To what extent can the vocabulary accommodate local terminological enhancements?

7.2.1.3
Vocabulary and Networking

Can the vocabulary server respond to queries sent over a network (LAN, WAN)?

7.2.2
Precision and Recall

7.2.2.1
Vocabulary

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, using a standard well-principled method? This should be evaluated only within the intended scope and purpose of the vocabulary system.

7.2.2.2
Search Engine

Is a standard search engine used in the mapping process?

7.2.3
Usability

7.2.3.1
Validation of Usability

Has the usability of the vocabulary been verified?

7.2.3.2
Interface Considerations

How have interface considerations been separated from vocabulary evaluation?

7.2.3.3
Prototypes

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? Evidence for speed of entry, accuracy, comprehensiveness in practice etc. with different approaches? If not, is there a proof of concept?

7.2.3.4
Application Programmer Interfaces

Is there support for computer interfaces and system implementers? 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?

7.2.4
Feasibility

If it is intended for use in an Electronic Patient Record (EPR), what are the options for information storage? Has feasibility been demonstrated?

7.3
Measures of Quality

The generalizability (applicability) of any Study Design reported (Evaluating Reported Evaluations) should be able to be evaluated.

7.3.1
Healthcare / Clinical Relevance

What is the vocabulary's Healthcare / Clinical Relevance?

7.3.2
Gold Standard

What was the Gold Standard used in the evaluation?

7.3.3
Study Population

If published population rates are used for comparison, was the study population comparable to the population from which the rates were derived?

7.3.4
Specific Aims

Were the Specific Aims clear?

7.3.5
Blinding

Was the study appropriately blinded?

7.3.6
Randomization

Was the Test Set Selection randomized or shown in some sense to be a representative sample of the end user population?

7.3.7
Test Location

7.3.7.1
Independence

Was it different from the developer's location?

7.3.7.2
Appropriate for Study Design

How was the test site suited to the study design (tools, resources, etc.)?

7.3.7.3
Principal Investigator Associations

Was the Principal Investigator associated with:

7.3.7.3.1
University

7.3.7.3.2
Academic Medical Center

7.3.7.3.3
Corporation

7.3.7.3.4
Hospital

7.3.7.3.5
Government Agency

7.3.7.3.6
HMO

7.3.7.3.7
Private Practice

7.3.7.3.8
Academic Organization

7.3.7.4
Principal Investigator

7.3.7.4.1
Was the Principal Investigator independent of the vocabulary being evaluated?

7.3.7.4.2
Does the Principal investigator have a track record of publication in this field of study?

7.3.7.4.3
Have there been any conflicts of interest in performing this research?

7.3.8
Project Completion
Was the project completed in a reasonable period of time?

7.3.9
Sample Size

7.3.9.1
Power
Was the sample size of sufficient size to show the anticipated effect, should one exist?

7.3.9.2
Statistics

Who reviewed the Statistical Methods?

7.3.10
Personnel

7.3.10.1
Training Level

What is the average level of training of the study personnel?

7.3.10.2
Reviewers

7.3.10.2.1
Variability

What is the inter-reviewer variability?

7.3.10.2.2
Type

What was the type of reviewer (physician, nurse, other clinician, coder, knowledge engineer) used in the study?

7.3.10.2.3
Independence

Were the reviewers blinded to the other reviewers' judgments (i.e. reviewer independence)?

---

[i] Cimino JJ. Desiderata for Controlled Medical Vocabularies in the Twenty-first Century. *Methods of Information in Medicine* 1998, in press.

[ii] Cote RA, Rothwell DJ. The Classification-nomenclature Issues in Medicine: A Return to Natural Language. *Medical Informatics* 1989;14(1):25-41.

[iii] Rocha RA, Rocha BH, Huff SM. Automated Translation Between medical Vocabularies using a Frame-based Interlingua. *Proceedings of the Annual Symposium of Computer Applications in Medical Care*. 1993:690-694.

[iv] Campbell KE, Musen MA. Representation of Clinical Data Using SNOMED III and Conceptual Graphs. *Proceedings of the Annual Symposium on Computer Applications in Medical Care* 1992:354-8.

[v] Rossi Mori A, Galeazzi E, Gangemi A, Pisanelli DM, Thornton AM. Semantic Standards for the Representation of Medical Records. Medical Decision Making 1991; 4(Suppl): S76-80.

[vi] Tuttle MS, Olson NE, Campbell KE, Sherertz DD, Nelson SJ, Cole WG. Formal Properties of the Metathesaurus. *Proceedings of the Annual Symposium on Computer Applications in Medical Care* 1994:145-9.

[vii] Campbell KE, Cohn SP, Chute CG, Rennels G, Shortliffe EH. Galapagos: Computer-based Support for Evolution of a Convergent Medical Terminology. *Journal of the American Medical Informatics Association* 1996;SympSuppl:269-73.

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

Formatted

[ix] Cimino JJ. Desiderata for Controlled medical Vocabularies in the Twenty-first Century. *Methods of Information in Medicine*. 1998;37(4/5):394-403.

[x] Cote RA, Rothwell DJ. The Classification-nomenclature Issues in Medicine: A Return to Natural Language. *Medical Informatics*. 1989;14(1):25-41.

[xi] Rocha RA, Rocha BH, Huff SM. Automated Translation Between medical Vocabularies using a Frame-based Interlingua. *Proceedings of the Annual Symposium of Computer Applications in Medical Care*. 1993:690-694.

[xii] 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.

[xiii] Campbell KE, Musen MA. Representation of Clinical Data Using SNOMED III and Conceptual Graphs. *Proceedings of the Annual Symposium on Computer Applications in Medical Care* 1992:354-8.

[xiv] Rossi Mori A, Galeazzi E, Gangemi A, Pisanelli DM, Thornton AM. Semantic Standards for the Representation of Medical Records. Medical Decision Making 1991; 4(Suppl): S76-80.

[xv] Tuttle MS, Olson NE, Campbell KE, Sherertz DD, Nelson SJ, Cole WG. Formal Properties of the Metathesaurus. *Proceedings of the Annual Symposium on Computer Applications in Medical Care* 1994:145-9.

[xvi] Campbell KE, Cohn SP, Chute CG, Rennels G, Shortliffe EH. Galapagos: Computer-based Support for Evolution of a Convergent Medical Terminology. *JAMIA* 1996;SympSuppl:269-73.

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

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

Formatted

Formatted

Formatted

Formatted