Published in final edited form as:
Methods Inf Med. 1998 November ; 37(4-5): 394–403.

Desiderata for Controlled Medical Vocabularies in the Twenty-First Century

James.J. Cimino, M.D.
Department of Medical Informatics, Columbia University College of Physicians and Surgeons, 161 Fort Washington Avenue, New York, New York 10032, USA

Abstract

Builders of medical informatics applications need controlled medical vocabularies to support their applications and it is to their advantage to use available standards. In order to do so, however, these standards need to address the requirements of their intended users. Over the past decade, medical informatics researchers have begun to articulate some of these requirements. This paper brings together some of the common themes which have been described, including: vocabulary content, concept orientation, concept permanence, nonsemantic concept identifiers, poly-hierarchy, formal definitions, rejection of “not elsewhere classified” terms, multiple granularities, multiple consistent views, context representation, graceful evolution, and recognized redundancy. Standards developers are beginning to recognize and address these desiderata and adapt their offerings to meet them.

Keywords

Controlled Medical Terminology; Vocabulary; Standards; Review

1. Introduction

The need for controlled vocabularies in medical computing systems is widely recognized. Even systems which deal with narrative text and images provide enhanced capabilities through coding of their data with controlled vocabularies. Over the past four decades, system developers have dealt with this need by creating ad hoc sets of controlled terms for use in their applications. When the sets were small, their creation was a simple matter, but as applications have grown in function and complexity, the effort needed to create and maintain the controlled vocabularies became substantial. With each new system, new efforts were required, because previous vocabularies were deemed unsuitable for adoption in or adaptation to new applications. Furthermore, information in one system could not be recognized by other systems, hindering the ability to integrate component applications into larger systems.

Consider, for example, how a computer-based medical record system might work with a diagnostic expert system to improve patient care. In order to achieve optimal integration of the two, transfer of patient information from the record to the expert would need to be automated. In one attempt to do so, the differences between the controlled vocabularies of the two systems was found to be the major obstacle – even when both systems were created by the same developers [1].

James.Cimino@columbia.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

The solution seems obvious: standards [2]. In fact, many standards have been proposed, but their adoption has been slow. Why? System developers generally indicate that, while they would like to make use of standards, they can't find one that meets their needs. What are those needs? The answers to this question are less clear. The simple answer is, “It doesn't have what I want to say.” Standards developers have taken this to mean that the solution is equally simple: keep adding terms to the vocabulary until it does say what's needed. However, systems developers, as users of controlled vocabularies, are like users everywhere: they may not always articulate their true needs. Vocabulary developers have labored to increase their offerings, but have continued to be confronted with ambivalence. A number of vocabularies have been put forth as standards [3] but they have been found wanting in some recent evaluations [4–6].

Over the past ten years or so, 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. As we prepare to enter the twenty-first century, it seems appropriate to pause to reflect on this additional experience, to rethink the directions we should pursue, and to identify the next set of goals for the development of standard, reusable, multipurpose controlled medical vocabularies.

2. Desiderata

The task of enumeration of general desiderata for controlled vocabularies is hampered in two ways. First, the desired characteristics of a vocabulary will vary with the intended purpose of that vocabulary and there are many possible intended purposes. I address this issue by stating that the desired vocabulary must be multipurpose. Some of the obvious purposes include: capturing clinical findings, natural language processing, indexing medical records, indexing medical literature, and representing medical knowledge. Each reader can add his or her own favorite purpose. A vocabulary intended for any of these can, and often has been, created. But the demands placed on a vocabulary become very different when it must meet several purposes.

A second obstacle to summarizing general desiderata is the difficulty teasing out individual opinions from the literature and unifying them. The need for controlled vocabularies for medical computing is almost as old as computing itself [7–9]. However, it is only in the past ten years or so that researchers have gone past talking about the content of a vocabulary and started to talk about deeper representational aspects. Before then, the literature contains many implications and half-stated assertions. Since then, authors have become more explicit, but the “terminology of terminology” has not yet settled down to a level of general understanding about what each of us means when we discuss vocabulary characteristics [10]. It is a foregone conclusion, then, that the summary I present here is bound to misrepresent some opinions and overlook others.

With these disclaimers, then. I will attempt to enumerate some of the characteristics that seem to be emerging from recent vocabulary research. Some of them may seem obvious, but they are listed formally in order that they not be overlooked.

2.1 Content, Content, and Content

Like the three most important factors in assessing the value of real estate (location, location, and location), the importance of vocabulary content can not be over stressed. The first criticisms of vocabularies were almost universally for more content. The need for expanded term coverage continues to be a problem, as can be seen in numerous studies which evaluate available standards for coverage of a particular domains. For example, recent publications examining the domain of nursing terminology are almost completely focused on the issue of what can be said [11–13]. Issues such as how things can be said, or how the vocabulary is organized are apparently less urgent, although sometimes solutions may need to go beyond the simple addition of more terms [14, 15].

One approach to increasing content is to add terms as they are encountered, responding as quickly as possible to needs as they arise [16]. In this approach, one adds complex expressions as needed rather than attempting a systematic, anticipatory solution. For example, rather than try to anticipate every kind of fracture (simple vs. compression, etc.) of every bone, one would add terms for the most common and add more as needed. This avoids the large numbers of terms occurring through combinatorial explosion and the enumeration of nonsensical combinations (such as “compound greenstick fracture of the stapes”, an anatomically implausible occurrence for a small bone in the middle ear).

An alternative approach is to enumerate all the aloms of a terminology and allow users to combine them into necessary coded terms [17], allowing compositional extensibility [18]. The tradeoff is that, while domain coverage may become easier to achieve, use of the vocabulary becomes more complex. Even with this atomic approach, the identification of all the aloms is nontrivial. The atoms must be substantial enough to convey intended meaning and to preserve their meaning when combined with others. They must be more than simply the words used in medicine. For example, the atom “White” could be used for creating terms like “White Conjunctiva” but would be inappropriate to use in constructing terms such as “Wolff-Parkinson-White Syndrome”. The word “White” needs to be more than a collection of letters – after all, we could represent all medical concepts with just the letters of the alphabet (26, more or less), but this would hardly advance the field of medical informatics. The atomic approach must also consider how to differentiate between atoms and molecules. “White” and “Conjunctiva” are almost certainly atoms, but what about “Wolff-Parkinson-White Syndrome”?

No matter what approach is taken, the need for adding content remains. This occurs because users will demand additions as usage expands and because the field of medicine (with its attendant terminology) expands. The real issue to address in considering the “content desideratum” is this: a formal methodology is needed for expanding content. A haphazard, onesy-twosy approach usually fails to keep up with the needs of users and is difficult to apply consistently. The result can be a patchwork of terms with inconsistent granularity and organization. Instead, we need formal, explicit, reproducible methods for recognizing and filling gaps in content. For example, Musen et al. applied a systematic approach (negotiation of goals, anticipation of use, accommodation of a user community, and evaluation) to creation of a vocabulary for use in a progress note system [19]. Methods of similar rigor need to be developed which can be used for content discovery and expansion in large, multipurpose vocabularies. More attention must be focused on how representations are developed, rather than what representations are produced [20].

2.2 Concept Orientation

Careful reading of medical informatics research will show that most systems that report using controlled vocabulary are actually dealing with the notion of concepts. Authors are becoming more explicit now in stating that they need vocabularies in which the unit of symbolic processing is the concept – an embodiment of a particular meaning [21–25]. Concept orientation means that terms must correspond to at least one meaning (“nonvagueness”) and no more than one meaning (“nonambiguity”), and that meanings correspond to no more than one term (“nonredundancy”) [26, 27].

Review of the literature suggests that there is some argument around the issue of ambiguity. Blois argues that while low-level concepts (such as protons) may be precisely defined, high-level concepts, including clinical concepts like myocardial infarction, are defined not by necessary attributes but by contingent ones (e.g., the presence of chest pain in myocardial infarction) [28]. Moorman et al. suggest that ambiguity can be allowed in the vocabulary as long as it is reduced to unequivocal meaning, based on context, when actually used (e.g., stored in a clinical record) [29].

However, a distinction must be made between ambiguity of the meaning of a concept and ambiguity of its usage [26, 30]. It is unfair, for example, to say that the concept “Myocardial Infarction” is ambiguous because it could mean “Right Ventricular Infarction”, “Left Ventricular Infarction” and so on. Any concept, no matter how fine-grained, will always subsume some finer-grained concepts. But “Myocardial Infarction” has a meaning which can be expressed in terms of a particular pathophysiologic process which affects a particular anatomic site. Now, if we use this concept to encode patient data, the meaning of the data will vary with the context (“Myocardial Infarction”, “Rule Out Myocardial Infarction”, “History of Myocardial Infarction”, “Family History of Myocardial Infarction”, “No Myocardial Infarction”, etc.).

This context-sensitive ambiguity is a different phenomenon from context-independent ambiguity that might be found in a controlled vocabulary [31]. For example, the term “Diabetes” does not subsume “Diabetes Mellitus” and “Diabetes Insipidus”; it has no useful medical meaning (vague). The concept “Ml” might mean “Myocardial Infarction”, “Mitral Insufficiency”, or “Medical Informatics”; before it even appears in a context, it has multiple meanings (ambiguous). Concept orientation, therefore, dictates that each concept in the vocabulary has a single, coherent meaning, although its meaning might vary, depending on its appearance in a context (such as a medical record) [29].

2.3 Concept Permanence

The corollary of concept orientation is concept permanence: the meaning of a concept, once created, is inviolate. Its preferred name may evolve, and it may be flagged inactive or archaic, but its meaning must remain. This is important, for example, when data coded under an older version of the vocabulary need to be interpreted in view of a current conceptual framework. For example, the old concept “pacemaker” can be renamed “implantable pacemaker” without changing its meaning (as we add the concept “percutaneous pacemaker”). But, the name for the old concept “non-A-non-B hepatitis” can not be changed to “Hepatitis C” because the two concepts arc not exactly synonymous (that is, we can't infer that someone diagnosed in 1980 as having non-A-non-B hepatitis actually had hepatitis C). Nor can we delete the old concept, even though we might no longer code patient data with it.

2.4 Nonsemantic Concept Identifier

If each term in the vocabulary is to be associated with a concept, the concept must have a unique identifier. The simplest approach is to give each concept a unique name and use this for the identifier. Now that computer storage costs are dropping, the need for the compactness provided by a code (such as an integer) has become less compelling. If a concept may have several different names, one could be chosen as the preferred name and the remainder included as synonyms. However, using a name as a unique identifier for a concept limits our ability to alter the preferred name when necessary. Such changes can occur for a number of reasons without implying that the associated meaning of the concept has changed [32].

Because many vocabularies are organized into strict hierarchies, there has been an irresistible temptation to make the unique identifier a hierarchical code which reflects the concept's position in the hierarchy. For example, a concept with the code 1000 might be the parent of the concept with the code 1200 which, in turn might be the parent of the concept 1280, and so on. One advantage to this approach is that, with some familiarity, the codes become some-what readable to a human and their hierarchical relationships can be understood. With today's computer inter-faces, however, there is little reason why humans need to have readable codes or, for that matter, why they even need to see the codes at all. Another advantage of hierarchical codes is that querying a database for members of a class becomes easier (e.g., searching for “all codes beginning with 1” will retrieve codes 1000, 1200, 1280, and so on). However, this advantage is lost if the concepts can appear in multiple places in the hierarchy (see “Polyhierarchy”, below); fortunately, there are other ways to perform “class-based” queries to a database which will work even when concepts can be in multiple classes [32].

There are several problems with using the concept identifier to convey hierarchical information. First, it is possible for the coding system to run out of room. A decimal code, such as the one described above, will only allow ten concepts at any level in the hierarchy and only allow a depth of four [34]. Coding systems can be designed to avoid this problem, but other problems remain. For example, once assigned a code, a concept can never be reclassified without breaking the hierarchical coding scheme. Even more problematic, if a concept belongs in more than one location in the hierarchy (see “Polyhierarchy”, below), a convenient single hierarchical identifier is no longer possible. It is desirable, therefore, to have the unique identifiers for the concepts which are free of hierarchical or other implicit meaning (i.e., nonsemantic concept identifiers); such information should instead be included as attributes of the concepts [14].

2.5 Polyhierarchy

There seems to be almost universal agreement that controlled medical vocabularies should have hierarchical arrangements. This is helpful for locating concepts (through “tree walking”), grouping similar concepts, and conveying meaning (for example, if we see the concept “cell” under the concept “anatomic entity” we will understand the intended meaning as different than if it appeared under the concepts “room” or “power source”). There is some disagreement, however, as to whether concepts should be classified according to a single taxonomy (strict hierarchy) or if multiple classifications (polyhierarchy) can be allowed. Most available standard vocabularies are strict hierarchies.

Different vocabulary users may demand different, equally valid, arrangements of concepts. It seems unlikely that there can ever be agreement on a single arrangement that will satisfy all; hence the popular demand for multiple hierarchies [31, 34–36]. Zweigenbaum and his colleagues believe that concept classification should be based on the essence of the concepts, rather than arbitrary descriptive knowledge [37]. They argue quite rightly that arbitrary, user-specific ad hoc classes can still be available using additional semantic information. However, unless there can be agreement on what the essence of concepts should be, there can never be agreement on what the appropriate hierarchy should be. Furthermore, if the essence of a concept is defined by its being the union of the essence of two other concepts, its classification becomes problematic. For example, until medical knowledge advances to provide a better definition, we must define the essence of “hepatorenal syndrome” as the occurrence of renal failure in patients with severe liver disease. If our vocabulary has the concepts “liver disease” and “renal disease” (which seem desirable or at least not unreasonable), “hepatorenal syndrome” must be a descendant of both.

There can be little argument that strict hierarchies are more manageable and manipulable, from a computing standpoint, than polyhierarchies. This is small consolation, however, if the vocabulary is unusable. General consensus, seems to favor allowing multiple hierarchies to coexist in a vocabulary without arguing about which particular tree is the essential one. It is certainly possible that if a single hierarchy is needed for computational purposes, one could be so designated with the others treated as nonhierarchical (but nevertheless directed and acyclic) relationships.

2.6 Formal Definitions

Many researchers and developers have indicated a desire for controlled vocabularies to have formal definitions in one form or another [23, 25–27, 36, 38–50]. Usually, these definitions are expressed as some collection of relationships to other concepts in the vocabulary. For example, the concept “Pneumococcal Pneumonia” can be defined with a hierarchical (“is a”) link to the concept “Pneumonia” and a “caused by” link to the concept “Streptococcus pneumoniae”. If “Pneumonia” has a “site” relationship with the concept “Lung”, then “Pneumococcal Pneumonia” will inherit this relationship as well. This information can be expressed in a number of ways, including frame-based semantic networks [40], classification operators [51], categorical structures [52], and conceptual graphs [53–55]. The important thing to realize about these definitions is that they are in a form which can be manipulated symbolically (i.e., with a computer), as opposed to the unstructured narrative text variety, such as those found in a dictionary. Many researchers have included in their requests that the definitional knowledge be made explic-itly separated from assertional knowledge which may also appear in the vocabulary [25, 41, 43, 46, 56]. For example, linking “Pneumococcal Pneumonia”, via the “caused by” relationship, to “Streptococcus pneumoniae” is definitional, while linking it, via a “treated with” relationship, to “Penicillin” would be assertional. Similarly, the inverse relationship (“causes”) from “Streptococcus pneumoniae” to “Pneumococcal Pneumonia” would also be considered assertional, since it is not part of the definition of “Streptococcus pneumoniae”.

The creation of definitions places additional demands on the creators of controlled vocabularies. However, with careful planning and design, these demands need not be onerous. For example, the definition given for “Pneumococcal Pneumonia”, given above, only required one additional “caused by” link to be added, assuming that it would be made a child of “Pneumonia” in any case and that the concept “Streptococcus pneumoniae” was already included in the vocabulary. Many of the required links can be generated through automatic means, either by the processing of the concept names directly [18] or through extraction from medical knowledge bases [57]. Also, the effort required to include definitions may help not only the users of the vocabulary, but the maintainers as well: formal definitions can support automated vocabulary management [58], collaborative vocabulary development [59], and methods for converging distributed development efforts [60, 61].

2.7 Reject “Not Elsewhere Classified”

Since no vocabulary can guarantee domain completeness all of the time, it is tempting to include a catch-all term which can be used to encode information that is not represented by other existing terms. Such terms often appear in vocabularies with the phrase “Not Elsewhere Classified”, or “NEC” (this is not lo be confused with “Not Otherwise Specified”, or “NOS”. which simply means that no modifiers are included with the base concept). The problem with such terms is that they can never have a formal definition other than one of exclusion – that is, the definition can only be based on knowledge of the rest of concepts in the vocabulary. Not only is this awkward, but as the vocabulary evolves, the meaning of NEC concepts will change in subtle ways. Such “semantic drift” will lead to problems, such as the proper interpretation of historical data. Controlled vocabularies should therefore reject the use of "not elsewhere classified" terms.

2.8 Multiple Granularities

Each author who expresses a need for a controlled vocabulary, does so with a particular purpose in mind. Associated with that purpose, usually implicitly, is some preconception of a level of granularity at which the concepts must be expressed. For example, the concepts associated with a diabetic patient might be (with increasingly finer granularity): "Diabetes Mellitus", "Type II Diabetes Mellitus", and "Insulin-Dependent Type II Diabetes Mellitus" (note that the simpler term "Diabetes" is so coarse-grained as to be vague). A general practitioner might balk at being required to select a diagnosis from the fine-grained end of this spectrum of concepts, while an endocrinologist might demand nothing less.

In reviewing the various writings on the subject, it becomes clear that multiple granularities are needed for multipurpose vocabularies. Vocabularies which attempt to operate at one level of granularity will be deemed inadequate for application where finer grain is needed and will be deemed cumbersome where coarse grain is needed. Insistence on a single level of detail within vocabularies may explain why they often are not reusable [62]. It also conflicts with a very basic attribute of medical information: the more macroscopic the level of discourse, the coarser the granularity of the concepts [63].

It is essential that medical vocabularies be capable of handling concepts as fine-grained as "insulin molecule" and as general as "insulin resistance". However, we must differentiate between the precision in medical knowledge and the precision in creating controlled concepts to represent that knowledge. While uncertainty in medical language is inevitable [64], we must strive to represent that uncertainty with precision.

2.9 Multiple Consistent Views

If a vocabulary is intended to serve multiple functions, each requiring a different level of granularity, there will be a need for providing multiple views of the vocabulary, suitable for different purposes [30]. For example, if an application restricts coding of patient diagnoses to coarse-grained concepts (such as "Diabetes Mellitus"), the more fine-grained concepts (such as "Insulin-Dependent Type II Diabetes Mellitus") could be collapsed into the coarse concept and appear in this view as synonyms (see Figs. 1a and 1b). Alternatively, an application may wish to hide some intermediate classes in a hierarchy if they are deemed irrelevant (see Fig. 1c). Similarly, although the vocabulary may support multiple hierarchies, a particular application may wish to limit the user to a single, strict hierarchy (sec Fig. 1d).

We must be careful to confine the ability to provide multiple consistent views, such that inconsistent views do not result. For example, if we create a view in which concepts with multiple parents appear in several places in a single hierarchy, care must be taken that each concept has an identical appearance within the view (see Figs, 1e and 1f) [31].

2.10 Beyond Medical Concepts: Representing Context

Part of the difficulty with using a standard controlled vocabulary is that the vocabulary was created independent of the specific contexts in which it is to be used. This helps prevent the vocabulary from including too many implicit assumptions about the meanings of concepts and allows it to stand on its own. However, it can lead to confusion when concepts are to be recorded in some specific context, for example, in an electronic patient record. Many researchers have expressed a need for their controlled vocabulary to contain context representation through formal, explicit information about how concepts are used [21, 65, 66].

A decade ago, Huff and colleagues argued that a vocabulary could never be truly flexible, extensible and comprehensive without a grammar to define how it should be used [67]. Campbell and Musen stated that, in order to provide systematic domain coverage, they would need both a patient-description vocabulary and rules for manipulation of the vocabulary [68]. Rector et al. add an additional requirement: not only is there a grammar for manipulation, but there is concept-specific information about "what is sensible to say" that further limits how concepts can be arranged [43]. Such limitations are needed in order for the vocabulary to support operations such as predictive data entry, natural language processing, and aggregation of patient records: Rector (and others in the Galen Project) simply request that such information be included as part of the vocabulary, in the form of constraints and sanctions [69].

If drawing the line between concept and context can become difficult [41], drawing the line between the vocabulary and the application becomes even more so. After all, the ultimate context for controlled medical vocabulary concepts is some external form such as a patient record. Coping with such contexts may be easier if such contexts are modeled in the vocabulary [70]. A schematic of how such contexts fit together is shown in Fig. 2. The figure differentiates between levels of concept interaction: what's needed to define the concepts, what's desired to show expressivity of the vocabulary, and how such expressiveness is channeled for recording purposes (e.g., in a patient record).

Of course, patient records vary a great deal from institution to institution and, if we have difficulty standardizing on a vocabulary, what hope is there for standardizing on a record structure? One possible solution is to view the recording of patient information from an "event" standpoint, where each event is constitutes some action, including the recording of data, occurring during an episode of care which, in turn occurs as part of a patient encounter [71, 72]. These add more levels to the organization of concepts in contexts, but can be easily modeled in the vocabulary, as in Fig. 2.

2.11 Evolve Gracefully

It is an inescapable fact that controlled vocabularies need to change with time. Even if there were a perfect vocabulary that "got it right the first time", the vocabulary would have to change with the evolution of medical knowledge. All too often, however, vocabularies change in ways that are for the convenience of the creators but wreak havoc with the users [32]. For example, if the name of a concept is changed in such a way as to alter its meaning, what happens to the ability to aggregate patient data that are coded before and after the change? An important desideratum is that those charged with maintaining the vocabulary must accommodate graceful evolution of their content and structure. This can be accomplished through clear, detailed descriptions of what changes occur and why [73], so that good reasons for change (such as simple addition, refinement, precoordination, disambiguation, obsolescence, discovered redundancy, and minor name changes) can be understood and bad reasons (such as redundancy, major name changes, code reuse, and changed codes) can be avoided.[74]

2.12 Recognize Redundancy

In controlled vocabulary parlance, redundancy is the condition in which the same information can be stated in two different ways. Synonymy is a type of redundancy which is desirable: it helps people recognize the terms they associate with a particular concept and, since the synonyms map to the same concept (by definition), then the coding of the information is not redundant. On the other hand, the ability to code information in multiple ways is generally to be avoided. However, such redundancy may be inevitable in a good, expressive vocabulary.

Consider an application in which the user records a coded problem list. For any given concept the user might wish to record, there is always the possibility that the user desires a more specific form than is available in the vocabulary. A good application will allow the user to add more detail to the coded problem, either through the addition of a coded modifier, through the use of unconstrained text, or perhaps a combination of both. For example, if a patient has a pneumonia in the lower lobe of the left lung, but the vocabulary does not have such a concept, the user might select the coded concept “Pneumonia” and add the modifier “Left Lower Lobe”. Suppose that, a year later, the vocabulary adds the concept “Left Lower Lobe Pneumonia”. Now, there are two ways to code the concept – the old and the new. Even if we were to somehow prevent the old method from being used, we still have old data coded that way.

As vocabularies evolve, gracefully or not, they will begin to include this kind of redundancy. Rather than pretend it does not happen, we should embrace the diversity it represents while, at the same time, provide a mechanism by which we can recognize redundancy and perhaps render it transparent. In the example above, if I were to ask for all patients with “Left Lower Lobe Pneumonia”, I could retrieve the ones coded with the specific concept and those coded with a combination of concepts. Such recognition is possible if we have paid sufficient attention to two other desiderata: formal definitions and context representation. If we know, from context representation, that the disease concept “Pneumonia” can appear in a medical record together with an anatomical concept, such as “Left Lower Lobe”, and the definition of “Left Lower Lobe Pneumonia” includes named relationships lo the concepts “Pneumonia” and “Left Lower Lobe” (“is a” and “site of” relationships, respectively), sufficient information exists to allow us to determine that the representation of the new concept in the vocabulary is equivalent to the collection of concepts appearing in the patient database (see Fig. 3).

3. Discussion

The intense focus previously directed at such issues as medical knowledge representation and patient care data models is now being redirected to the issue of developing and maintaining shareable, multipurpose, high-quality vocabularies. Shareabilily of vocabulary has become important as system builders realize they must rely on vocabulary builders to help them meet the needs of representing large sets of clinical terms. The multipurpose nature of vocabularies refers to their ability to be used to record data for one purpose (such as direct patient care) and then be used for reasoning about the data (such as automated decision support), usually through a variety of views or abstractions of the specific codes used in data capture. Even the ability to support browsing of vocabularies remains problematic [75]. “High quality” has been difficult to define, but generally means that the vocabulary approaches completeness, is well organized, and has terms whose meanings are clear. The above list of desiderata for shareable, multipurpose controlled vocabularies reflect one person's view of the necessary priorities; however, they are based on personal experience with attempts to adopt vocabularies [76–79] and gleaned from the reported experiences of others. The solutions necessary to meet the above list of desiderata vary from technical to political, from simple adoption to basic shifts in philosophy, and from those currently in use to areas ripe for research.

Developers of controlled vocabularies are recognizing that their products are in demand for multiple purposes and, as such, they must address a variety of needs that go beyond those included for the vocabulary's original purpose [80]. The simple solution of “add more terms until they're happy” is not satisfying vocabulary users; they want content, but they want more. They want information about the terms, so they know what they mean and how to use them. They also want this information to supplement the knowledge they create for their own purposes. These purposes are as diverse as natural language processing, predictive data entry, automated decision support, indexing, clinical research, and even the maintenance of vocabularies themselves.

Simple, technical solutions are at hand for some characteristics, and are already being adopted. For example, using nonsemantic concept identifiers and allowing polyhierarchies are straightforward. The systematic solution for some others, such as multiple granularities and multiple consistent views will require more thought, but generally should be tractable. Allowing graceful evolution and recognized redundancy are still areas for research, with some promising findings. For example, systematic approaches for vocabulary updates are being discussed to support evolution [73], while conceptual graphs provide a mechanism for transforming between different synonymous (i.e., redundant) arrangements of associated concepts [54].

Some of the desiderata will require fundamental philosophical shifts. For example, decisions to have a truly concept-oriented vocabulary and avoid the dreaded “NEC” terms are simple ones, but can not be taken lightly. Some of these decisions, such as formal definitions and representing context, will also require significant development effort to make them a reality. Several developers describe commitment to these goals, and one group has actually provided formal, computer-manipulable definitions of their concepts [81, 82]. But the amount of work remains formidable. Finding ways to share the burden of vocabulary design and construction will be challenging [83], but some approaches seem promising [59]. Finding ways lo coordinate content development and maintenance among multiple groups will require sophisticated approaches [60]. Despite their perceived infancy [84], the currently available standards should be the starting point for new efforts [85].

Predictions may not be difficult to make, given the current directions in which standards development is proceeding. It is likely that vocabularies will become concept-oriented, using nonsemantic identifiers and containing semantic information in the form of a semantic network, including multiple hierarchies. Development of a standard notation for the semantic information may take some time, but the conceptual graph seems to be a popular candidate. Maintenance of vocabularies will eventually settle down into some form which is convenient for users and concept permanence will become the norm. Still unclear is whether the semantic, definitional information provided by developers will be minimal, complete, or somewhere in between. Some of the other desiderata, such as context representation, multiple consistent views, and recognition of redundancy will probably be late in coming. However, the knowledge and structure provided with the vocabulary will at least facilitate development of implementation-specific solutions which have not heretofore been possible.

4. Conclusions

This list of desiderata is not intended lo be complete; rather, it is a partial list which can serve to initiate discussion about additional characteristics needed to make controlled vocabularies sharable and reusable. The reader should not infer that vocabulary developers are not addressing these issues. In fact, these same developers were the sources for many of the ideas listed here. As a result, vocabularies are undergoing their next molt. Current trends seem to indicate that this one will be a true metamorphosis, as lists change to multiple hierarchies, informal descriptive information changes to formal definitional and assertional information, and attention is given not just to the expansion of content, but to structural and representational issues.

Acknowledgments

The ideas in this paper have been based on the medical informatics literature and synthesized from years of work and conversations with many researchers. In particular, the author's collaborators on the UMLS project, the Canon group, and the InterMed Collaboratory can be credited with helping to precipitate these thoughts. In addition, since this paper has been prepared as an accompaniment to a presentation at the IMIA Working Group 6 conference, it is hoped that additional desiderata will be put forth at the meeting which will find their way into the final version of this manuscript. This work has been supported in part by the National Library of Medicine.

REFERENCES

1. Wong ET, Pryor TA, Huff SM, Haug PJ, Warner HR. Interfacing a stand-alone diagnostik expert system with a hospital information system. Comput Biomed Res. 1994; 27:116–29. [PubMed: 8033537]
2. Masys DR. Of codes and keywords: standards for biomedical nomenclature. Acad Med. 1990; 65:627–9. [PubMed: 2261036]
3. Cimino, JJ. Coding systems in health care. In: van Bemmel, JH.; McCray, AT., editors. Year-book of Medical Informatics, International Medical Informatics Association, Rotterdam. 1995. p. 71-85.Reprinted in Methods of [Information in Medicine 1996>; 35 (4/5): 273-84
4. Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The Content Coverage of Clinical Classifications. JAMIA. 1996; 3:224–33. [PubMed: 8723613]
5. Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J. Phase II evaluation of clinical coding schemes; completeness, taxonomy, mapping, definitions and clarity. Journal of the American Medical Informatics Association. 1997; 4:238–51. [PubMed: 9147343]
6. Conference Summary Report: Moving Toward International Standards in Primary Care Informatics: Clinical Vocabulary; New Orleans. Nov. 1995 United States Department of Health and Human Service 1996
7. Anderson J. The computer: medical vocabulary and information. British Medical Bulletin. 1968; 24(3):194–8. [PubMed: 5672184]
8. Bates JAV. Preparation of clinical data for computers. British Medical Bulletin. 1968; 24(3):199–205. [PubMed: 5672185]
9. Howell RW, Loy RM. Disease coding by computer: the "fruit machine" method. British Journal of Preventive and Social Medicine. 1968; 22:178–81. [PubMed: 5667313]
10. Cimino JJ. Controlled medical vocabulary construction: methods from the Canon Group. JAMIA. 1994; 1:296–7. [PubMed: 7719811]
11. Ozbolt JG, Fruchtnicht JN, Hayden JR. Toward data standards for nursing information. JAMIA. 1994; 1:175–85. [PubMed: 7719798]
12. McCloskey J, Bulechek G. Letter: Toward data standards for nursing information. JAMIA. 1994; 1:469–71. [PubMed: 7850573]
13. Ozbolt JG, Fruchtnicht JN, Hayden JR. Letter: Toward data standards for nursing information. JAMIA. 1994; 1:471–2.
14. Humphreys BL. Comment: Toward data standards for nursing information. JAMIA. 1994; 1:472–4.
15. Goossen WTF, Epping PJMM, Abraham IL. Classification systems in nursing: formalizing nursing knowledge and implications for nursing information systems. Meth Inform Med. 1996; 35:59–71. [PubMed: 8992226]
16. Gabrieli ER. Computerizing text from office records. MD Comput. 1987; 4:44–9. 56. [PubMed: 3613935]
17. Côté RA, Robboy S. Progress in medical information management – Systematized nomenclature of medicine (SNOMED). JAMA. 1980; 243:756–62. [PubMed: 6986000]
18. Evans DA, Rothwell DJ, Monarch IA, Lefferts RG, Côté RA. Towards representations for medical concepts. Med Decis Making. 1991; 11(suppl):S102–S108. [PubMed: 1770838]
19. Musen MA, Weickert KE, Miller ET, Campbell KE, Fagan LM. Development of a controlled medical terminology: knowledge acquisition and knowledge representation. Meth Inform Med. 1995; 34:85–95. [PubMed: 9082143]

20. Moehr, JR.; Kluge, EH.; Patel, VL. Advanced patient information systems and medical concept representation. In: Greenes, RA.; Peterson, HE.; Protti, DJ., editors. Proceedings of the Eight World Congress on Medical Informatics (MEDINFO '95); Vancouver, British Columbia, Canada: Healthcare Computing & Communications (Canada); 1995. p. 95-9.

21. Evans, DA. Pragmatically-structured, lexical-semantic knowledge bases for a Unified Medical Language System. In: Greenes, RA., editor. Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care (SCAMC); Washington, DC; New York: IEEE Computer Society Press; 1988. p. 169-73.

22. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Meth Inform Med. 1993; 32:281–91. [PubMed: 8412823]

23. Volot, F.; Zweigenbaum, P.; Bachimont, B.; Ben Said, M.; Bouaud, J.; Fieschi, M.; Boisvieux, JF. Structuration and acquisition of medical knowledge using UMLS in the conceptual graph formalism. In: Safran, C., editor. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care (SCAMC); Washington, DC. 1993. p. 710-14.McGraw-Hill, New York; 1994

24. Evans DA, Cimino JJ, Hersh WR, Huff SM, Bell DS. Toward a medical-concept representation language. JAMIA. 1994; 1:207–17. [PubMed: 7719804]

25. Rassinoux, AM.; Miller, RA.; Baud, RH.; Scherrer, JR. Modeling principles for QMR medical findings. In: Cimino, JJ., editor. Proceedings of the AMIA Annual Fall Symposium (Formerly SCAMC); Washington, DC, Philadelphia: Hanley and Belfus; 1996. p. 264-98.

26. Henkind, SJ.; Benis, AM.; Teichhols, LE. Quantification as a means to increase the utility of nomenclature-classification systems. In: Salmon, R.; Blum, B.; Jürgensen, editors. Proceedings of the Seventh World Congress on Medical Informatics (MED-INFO 86); North-Holland (Amsterdam). 1986. p. 858-61.

27. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. JAMIA. 1994; 1:35–50. [PubMed: 7719786]

28. Blois, MS. The effect of hierarchy on the encoding of meaning. In: Orthner, HF., editor. Proceedings of the Tenth Annual Symposium on Computer Applications in Medical Care (SCAMC); Washington, DC: New York: IEEE Computer Society Press; 1986. p. 73

29. Moorman PW, van Ginneken AM, van der Lei J, van Bemmel JH. A model for structured data entry based on explicit descriptional knowledge. Meth Inform Med. 1994; 33:454–63. [PubMed: 7869942]

30. van Ginneken, AM.; van der Lei, J.; Moorman, PW. Towards unambiguous representation of patient data. In: Frisse, ME., editor. Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care (SCAMC); Baltimore, MD. 1992. p. 69-73.McGraw-Hill, New York; 1993

31. Cimino, JJ.; Hripcsak, G.; Johnson, SB.; Clayton, PD. Designing an introspective, controlled medical vocabulary. In: Kingsland, LC., editor. Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care (SCAMC); New York, Washington, DC: IEEE Computer Society Press; 1989. p. 513-8.

32. Cimino JJ. Formal descriptions and adaptive mechanisms for changes in controlled medical vocabularies. Meth Inform Med. 1996; 35:202–10. [PubMed: 8952304]

33. Forman, BH.; Cimino, JJ.; Johnson, SB.; Sengupta, S.; Sideli, R.; Clayton, P. Applying a controlled medical terminology to a distributed production clinical information system. In: Gardner, RM., editor. Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care (SCAMC); Philadelphia, New Orleans, LA: Hanley and Belfus; 1995. p. 421-5.

34. Bernauer, J.; Franz, M.; Schoop, M.; Schoop, D.; Pretschner, DP. The compositional approach for representing medical concept systems. In: Greenes, RA.; Peterson, HE.; Protti, DJ., editors. Proceedings of ihe Eight World Congress on Medical Informatics (MEDINFO '95); Vancouver, British Columbia, Canada: Healthcare Computing & Communications (Canada); 1995. p. 70-4.

35. Dunham GS, Henson DE, Pacak MG. Three solutions to problems of categorized medical nomenclatures. Meth Inform Med. 1984; 23:87–95. [PubMed: 6472122]

36. Baud, R.; Lovis, C.; Alpay, L.; Rassinoux, AM.; Nowlan, A.; Rector, A. Modeling for natural language understanding. In: Safran, C., editor. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care (SCAMC); New York, Washington, DC. 1993. p. 289-93.New York: McGraw-Hill 1994
37. Zweigenbaum P, Bachimont B, Bouaud J, Charlet J, Boisvieux JF. Issues in the structuring and acquisition of an outology for medical language understanding. Meth Inform Med. 1995; 34:15–24. [PubMed: 9082125]
38. Masarie FE Jr, Miller RA, Bouhaddou O, Giuse NB, Warner HR. An interlingua for electronic interchange of medical information: using frames to map between clinical vocabularies. Computers in Biomedical Research. 1991; 24(4):379–400.
39. Gabrieli, ER. Interface problems between medicine and computers. In: Cohen, GS., editor. Proceedings of the Eighth Annual Symposium on Computer Applications in Medical Care (SCAMC); New York, Washington, DC: IEEE Computer Society Press; 1984. p. 93-5.
40. Barr, CE.; Komorowski, HJ.; Pattison-Gordon, E.; Greenes, RA. Conceptual modeling for the Unified Medical Language System. In: Greenes, RA., editor. Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care (SCAMC); New York, Washington, DC: IEEE Computer Society Press; 1988. p. 148-51.
41. Haimowitz, IJ.; Patil, RS.; Szolovitz, P. Representing medical knowledge in a terminological language is difficult. In: Greenes, RA., editor. Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care (SCAMC); New York, Washington, DC: IEEE Computer Society Press; 1988. p. 101-5.
42. Rothwell, DJ.; Côté, RA. Optimizing the structure of a standardized vocabulary – the SNOMED model. In: Miller, RA., editor. Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care (SCAMC); New York, Washington, DC: IEEE Computer Society Press; 1990. p. 181-4.
43. Rector, AL.; Nowlan, WA.; Kay, S. Conceptual knowledge: the core of medical information systems. In: Lun, KC.; Deguolet, P.; Piemme, TE.; Rienhoff, O., editors. Proceedings of the Seventh World Congress on Medical Informatics (MEDINFO '92), Geneva; North-Holland (Amsterdam). 1992. p. 1420-6.
44. Campbell KE, Das AK, Musen MA. A logical foundation for representation of clinical data. JAMIA. 1994; 1:218–32. [PubMed: 7719805]
45. Bernauer, J. Subsumption principles underlying medical concept systems and their formal reconstruction. In: Ozbolt, JG., editor. Proceedings of the Eighteenth Annual Symposium on Compute Applications in Medical Care (SCAMC); Washington, DC: Philadelphia: Hanley and Belfus; 1994. p. 140-4.
46. Smart JF, Roux M. A model for medical knowledge of representation application to the analysis of descriptive pathology reports. Meth Inform Med. 1995; 34:352–60. [PubMed: 7476466]
47. Kiuchi T, Ohashi Y, Sato H, Kaihara S. Methodology for the construction of a system for clinical use. Meth Inform Med. 1995; 34:511–7. [PubMed: 8713767]
48. Rector, AL. Thesauri and formal classifications: terminologies for people and machines. In: Chute, CG., editor. IMIA Working Group 6 Conference on Natural Language and Medical Concept Representation; Jacksonville, Florida. Jan. 1997 p. 183-95.
49. Brown, PJB.; O'Neil, M.; Price, C. Semantic representation of disorders in Version 3 of the Read Codes. In: Chute, CG., editor. IMIA Working Group 6 Conference on Natural Language and Medical Concept Representation; Jacksonville, Florida. Jan. 1997 p. 209-14.
50. Price, C.; O'Neil, M.; Bentley, TE.; Brown, PJB. Exploring the ontology of surgical procedures in the Read Thesaurus. In: Chute, CG., editor. IMIA Working Group 6 Conference on Natural Language and Medical Concept Representation; Jacksonville, Florida. Jan. 1997 p. 215-21.
51. Bernauer, J. Formal classification of medical concept descriptions: graph oriented operators. In: Chute, CG., editor. IMIA Working Group 6 Conference on Natural Language and Medical Concept Representation; Jacksonville, Florida. Jan. 1997 p. 109-15.
52. Rossi Mori, A.; Consorti, F.; Galeazzi, E. Standards to support development of terminological systems for healthcare telematics. In: Chute, CG., editor. IMIA Working Group 6 Conference on

Natural Language and Medical Concept Representation; Jacksonville, Florida. Jan. 1997 p. 131-45.

53. Bernauer, J. Conceptual graphs as a operational model for descriptive findings. In: Clayton, PD., editor. Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care (SCAMC); Washington, DC. 1991. p. 214-8.McGraw-Hill, New York; 1992
54. Campbell, KE.; Musen, MA. Representation of clinical data using SNOMED III and conceptual graphs. In: Frisse, ME., editor. Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care (SCAMC); Baltimore, MD. 1992. p. 354-8.McGraw-Hill, New York; 1993
55. Baud RH, Rassinoux AM, Wagner JC, Lovis C, Juge C, Alpay LL, Michel PA, Degoulet P, Scherrer JR. Representing clinical narratives using conceptual graphs. Meth Inform Med. 1995; 34:176–86. [PubMed: 9082129]
56. Robé, PF.; de, V.; Flier, FJ.; Zanstra, PE. Health classifications and terminological modeling. In: Chute, CG., editor. IMIA Working Group 6 Conference on Natural Language and Medical Concept Representation; Jacksonville, Florida. Jan. 1997 p. 223-4.
57. Cimino JJ, Barnett GO. Automatic knowledge acquisition from MEDLINE. Meth Inform Med. 1993; 32:120–30. [PubMed: 8321130]
58. Cimino, JJ.; Johnson, SB.; Hripcsak, G.; Hill, CL.; Clayton, PD. Managing Vocabulary for a Centralized Clinical System. In: Greenes, RA.; Peterson, HE.; Protti, DJ., editors. Proceedings of the Eight World Congress on Medical Informatics (MEDINFO '95); Vancouver, British Columbia, Canada: Healthcare Computing & Communications (Canada); 1995. p. 117-20.
59. Friedman C, Huff SM, Hersh WR, Pattison-Gordon E, Cimino JJ. The Canon group's effort: working toward a merged model. JAMIA. 1995; 2:4–18. [PubMed: 7895135]
60. Campbell, KE.; Cohn, SP.; Chute, CG.; Rennels, G.; Shortliffe, EH. Gálapagos: computer-based support for evolution of a convergent medical terminology. In: Cimino, JJ., editor. Proceedings of the AMIA Annual Fall Symposium (Formerly SCAMC); Washington, DC: Philadelphia: Hanley and Belfus; 1996. p. 269-73.
61. Campbell, KE.; Cohn, SP.; Chute, CG.; Shortliffe, EH.; Rennels, G. Scaleable methodologies for distributed development of logic-based convergent medical terminology. In: Chute, CG., editor. IMIA Working Group 6 Conference on Natural Language and Medical Concept Representation; Jacksonville, Florida. Jan. 1997 p. 243-56.
62. Musen MA. Knowledge sharing and reuse. Comput Biomed Res. 1992; 25:435–67. [PubMed: 1395522]
63. Blois MS. Medicine and the nature of vertical reasoning. New Engl J Med. 1988; 318:847–51. [PubMed: 3352667]
64. Kong A, Barnett GO, Mosteller F, Youtz C. How medical professionals evaluate expressions of probability. New Engl J Med. 1986; 315:740–4. [PubMed: 3748081]
65. Henry, SB.; Mead, CN. Standardized nursing classification systems; necessary, but not sufficient, for representing what nurses do. In: Cimino, JJ., editor. Proceedings of the AMIA Annual Fall Symposium (Formerly SCAMC); Philadelphia, Washington, DC: Hanley and Belfus; 1996. p. 145-9.
66. Huff, SM.; Rocha, RA.; Solbrig, HR.; Barnes, MW.; Schank, SP.; Smith, M. Linking a medical vocabulary to a clinical data model using Abstract Syntax Notation 1. In: Chute, CG., editor. IMIA Working Group 6 Conference on Natural Language and Medical Concept Representation; Jacksonville, Florida. Jan. 1997 p. 225-41.
67. Huff, SM.; Craig, RB.; Gould, BL.; Castagno, DL.; Smilan, RE. Medical data dictionary for decision support applications. In: Stead, WW., editor. Proceedings of the Eleventh Annual Symposium on Computer Applications in Medical Care (SCAMC); New York, Washington, DC: IEEE Computer Society Press; 1987. p. 310-7.
68. Campbell, KE.; Musen, MA. Creation of a systematic domain for medical care: the need for a comprehensive patient-description vocabulary. In: Lun, KC.; Deguolet, P.; Piemme, TE.; Rienhoff, O., editors. Proceedings of the Seventh World Congress on Medical Informatics (MEDINFO '92), Geneva; Norih-Holland (Amsterdam). 1992. p. 1437-42.

69. Rector AL, Glowinski AJ, Nowlan WA, Rossi-Mori A. Medical-concept models and medical records: an approach based on GALEN and PEN&PAD. JAMIA. 1995; 2:19–35. [PubMed: 7895133]
70. Prokosh, HU.; Amiri, F.; Krause, D.; Neek, G.; Dudeck, J. A semantic network model for the medical record rheumatology clinic. In: Greenes, RA.; Peterson, HE.; Protti, DJ., editors. Proceedings of the Eight World Congress on Medicai Informatics (MEDINFO '95), Vancouver; British Columbia, Canada: Healthcare Computing & Communications (Canada); 1995. p. 240-4.
71. Huff SM, Rocha RA, Bray BE, Warner HR, Haug PJ. An event model of medical information representation. JAMIA. 1995; 2:116–34. [PubMed: 7743315]
72. Johnson, SB.; Friedman, C.; Cimino, JJ.; Clark, T.; Hripcsak, G.; Clayton, PD. Conceptual data model for a central patient database. In: Clayton, PD., editor. Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care (SCAMC); Washington, DC. 1991. p. 381-5.McGraw-Hill, New York 1992
73. Tuttle MS, Nelson SJ. A poor precedent. Meth Inform Med. 1996; 35:211–7. [PubMed: 8952305]
74. Cimino, JJ.; Clayton, PD. Coping with changing controlled vocabularies. In: Ozbolt, JG., editor. Proceedings of the Eighteenth Annual Symposium on Computer Applications in Medical Care; New York, Washington, DC: McGraw-Hill; Nov. 1994 p. 135-9.
75. Hripcsak G, Allen B, Cimino JJ, Lee R. Access to data: comparing Access Med with Query by Review. JAMIA. 1996; 3:288–99. [PubMed: 8816352]
76. Cimino, JJ. Representation of clinical laboratory terminology in the Unified Medical Language System. In: Clayton, PD., editor. Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care (SCAMC); Washington, DC. 1991. p. 199-203.McGraw-Hill, New York; 1992
77. Cimino JJ, Sideli RV. Using the UMLS to bring the library to the bedside. Medical Decision Making. 1991; 11(Suppl):S116–S120. [PubMed: 1770840]
78. Cimino, JJ.; Barrows, RC.; Allen, B. Adapting ICD9-CM for clinical decision support (abstract). In: Musen, MA., editor. Proceedings of the 1992 Spring Congress of the American Medical Informatics Association; Portland. Oregon. May. 1992 p. 34The Association, Bethesda, MD; 1992
79. Cimino JJ. Use of the Unified Medical Language System in patient care at the Columbia-Presbyterian Medical Center. Meth Inform Med. 1995; 34:158–64. [PubMed: 9082126]
80. Schulz EB, Price CS, Brown PJB. Symbolic anatomic knowledge representation in the Read Codes Version 3: structure and application. Journal of the American Medical Association. 1997; 4:38–48.
81. Forrey AW, McDonald CJ, DeMoor G, Huff SM, Leaville D, Leland D, Fiers T, Charles L, Griffin B, Stalling F, Tullis A, Hutchins K, Baenziger J. Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. Clinical Chemistry. 1996; 42:81–90. [PubMed: 8565239]
82. Rocha, RA.; Huff, SM. Coupling vocabularies and data structures: lessons from LOINC. In: Cimino, JJ., editor. Proceedings of the AMIA Annual Fall Symposium (Formerly SCAMC); Washington, DC: Philadelphia: Hanley and Belfus; 1996. p. 90-4.
83. Tuttle MS. The position of the Canon group: a reality check. JAMIA. 1994; 1:298–9. [PubMed: 7719812]
84. United States General Accounting Office. Automated Medical Records: Leadership Needed to Expedite Standards Development: report to the Chairman/Committee on Governmental Affairs, U.S. Senate. Washington, DC: Apr. 1993 USGAO/IMTEC-93-17
85. Board of Directors of the American Medical Informatics Association. Standards for medical identifiers, codes, and messages needed to create an efficient computer-stored medical record. JAMIA. 1994; 1:1–7. [PubMed: 7719784]

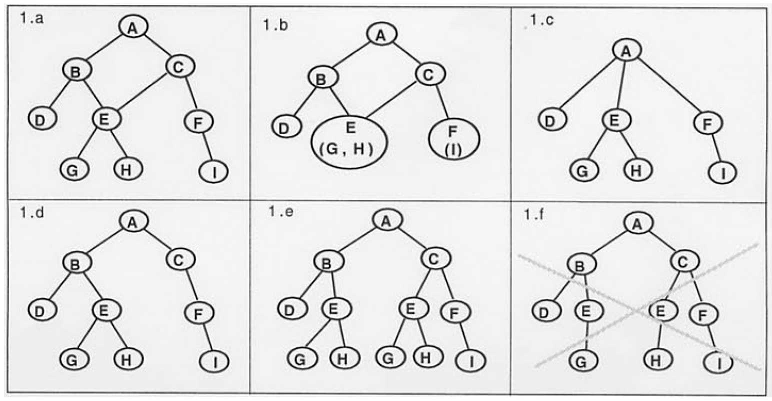

Fig. 1.
Multiple views of a polyhierarchy. a) Internal arrangements of nine concepts in a polyhierarchy, where E has two parents; b) Hierarchy has been collapsed so that specific concepts serve as synonyms of their more general parents; c) Intermediate levels in the hierarchy have been hidden; d) Conversion to a strict hierarchy; e) Strict hierarchy with multiple contexts for term E; f) Multiple contexts for E are shown, but are inconsistent (different children).

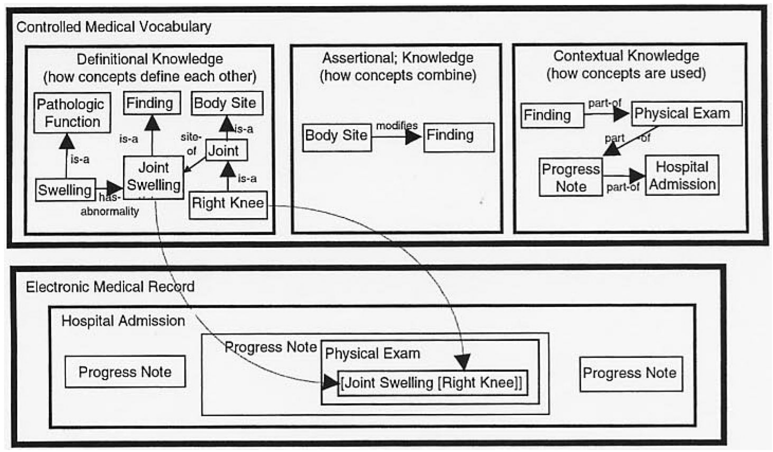

Fig. 2.

Definitional, assertional, and contextual information in the vocabulary showing how concepts can be combined and where they will appear in a clinical record.

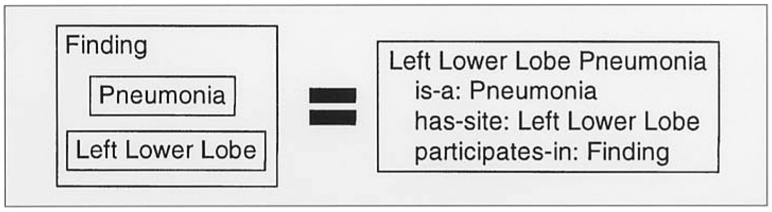

Fig. 3.
Interchangability of redundant data representations. The structure on the left depicts the post coordination of a disease concept (Pneumonia) and a body location (Left Lower Lobe) to create a finding in an electronic medical record. The structure on the right shows a precoordinated term for the same finding (Left Lower Lobe Pneumonia). Because this latter term includes formal, structured definitional information (depicted by the is-a, has-site, and participates-in attributes), it is possible to recognize, in an automated way, that data coded in these two different ways are equivalent.