The Practice of Informatics JAMIA

*Application of Technology* ■

# Symbolic Anatomic Knowledge Representation in the Read Codes Version 3: Structure and Application

ERICH B. SCHULZ, MB, BS, COLIN PRICE, FRCS, PHILIP J. B. BROWN, MRCGP

**Abstract** The Read Thesaurus (Version 3 of the Read Codes) is a controlled medical vocabulary produced during the Clinical Terms Projects with the involvement of over 2,000 health care professionals from all United Kingdom specialties. In addition to allowing the transfer of clinical information in a meaningful way, it supports analysis of this information and provides a basis for the development of shareable medical knowledge bases. The thesaurus includes a comprehensive, dynamic set of over 7,000 gross anatomic concepts richly linked in a network with over 16,000 operative procedures and 40,000 disorders. The representation of anatomic concepts aims to balance the requirements for expressivity, clearness, and simplicity. The underlying directed acyclic graph hierarchy is independent of the alphanumeric code and enables continued refinement and expansion. A template table allows semantic definition, qualification, and linkage of concepts.

■ **J Am Med Inform Assoc.** 1997;4:38–48.

Musen[1] has pointed out that, before the full potential of clinical information systems can be realized, computers must be able to share clinical knowledge at different levels of abstraction in a structured manner. The crucial role of medical concept representation and the building of shareable ontologies are seen to be increasingly important[2,3] and are presently receiving significant attention from several groups, including the Canon Group,[4] the National Library of Medicine,[5] and the GALEN consortium.[6] There is a requirement to share not only the vocabulary but also the relationships between vocabulary items. The representation of such relationships is facilitated by creating a structured thesaurus.

This paper outlines the development of the Read Thesaurus[7] during the Clinical Terms Project (CTP),[8–11] with emphasis on the structure of the Anatomy chapter and its relationship to other sections of the Thesaurus. In structuring the principal hierarchies for disorders and procedures, we chose a mainly anatomic axis because we felt that anatomy was relatively incontrovertible and could be coherently applied across a wide domain of clinical practice.

## Background to the Read Codes

The Read Codes were devised in the early 1980s by Dr. James Read, then a general practitioner, to support the recording of clinical data by general practitioners using personal computers. The Read Codes have since been widely used to record primary health care data; in 1988, they were endorsed by the United Kingdom (UK) Royal College of General Practitioners for this

Affiliation of the authors: NHS Centre for Coding and Classification, Loughborough, United Kingdom.

Correspondence and reprints: Dr. Erich B. Schulz, NHS Centre for Coding and Classification, Woodgate, Loughborough, Leicestershire, UK. LE11 2TG.

Received for publication: 6/24/96; accepted for publication: 9/18/96.

purpose. In 1990, they were purchased by the Department of Health and became Crown Copyright,[12] and the UK National Health Service Centre for Coding and Classification (NHS CCC) was established to maintain and develop the codes.

The two-year CTP, which ended in April 1994, was a large-scale collaboration between the NHS Executive and the medical profession through the Conference Information Group of the medical Royal Colleges. Forty-three specialty working groups were funded to produce enhancements to relevant sections of the then-current Version 2 Read Codes. The aim was to produce a detailed and comprehensive thesaurus of coded clinical terms to support the requirements of the computerized health care record and to underpin the NHS's Information Management and Technology Strategy.[13] By the end of the project, over 2000 clinicians had been involved in various capacities, including authoring, participating in working groups, and providing independent quality assurance for the work.[14] In parallel with this project, the nursing profession and the professions allied to medicine (PAMs) undertook similar projects of their own.

The CTP generated a requirement for representing increasingly complex clinical concepts, and a revision of the Read Code file structure was therefore needed in order to accommodate these.[7]

## Authoring Methodology

In the CTP, the need for unambiguous representation of anatomic concepts was recognized at an early stage of planning. Consequently, an Anatomy Panel was established, including both clinicians and anatomists, to oversee the development of the anatomy content. Anatomic knowledge can be represented symbolically or graphically.[15] By providing a standard, shareable set of coded concepts, the Read Thesaurus supports sharing of anatomic information at the symbolic rather than the graphic level. The authoring methodology was guided by some of the general principles related to thesaurus construction.[16]

A thesaurus is a structured set of concepts that links similar terms to provide easy navigation and also enables an understanding of the concepts represented by the terms and their interrelationships. Ideally, it should be coherent and should reflect the underlying body of knowledge about the domain.[17] In a thesaurus, the user attempts to find the appropriate term for a given meaning, whereas in a dictionary the aim is

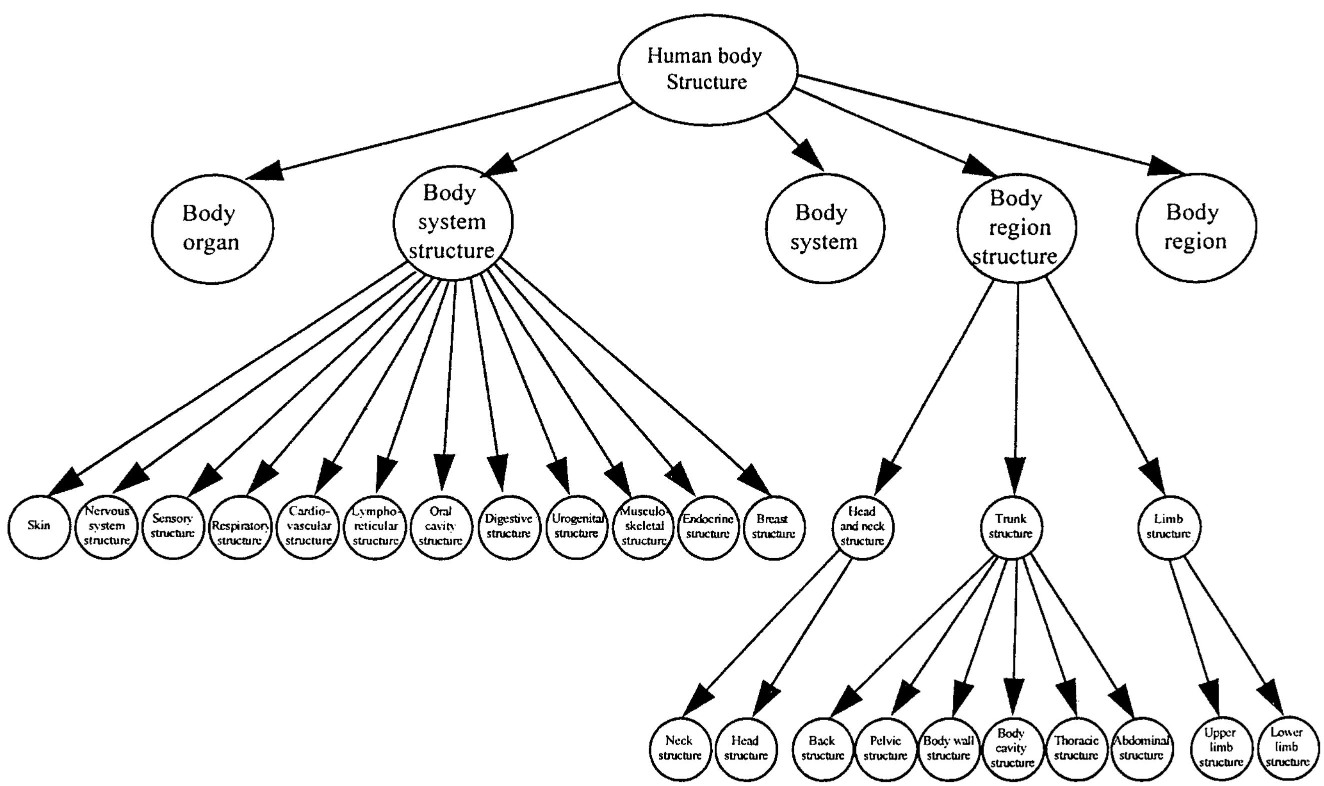


**Figure 1** Upper level head-ings of Version 3 anatomy hierarchy.

to find the meaning of a term. In a dictionary, navigation depends on orthography (spelling), and it is sufficient to arrange the entries in alphabetical order; in a thesaurus, however, navigation is assisted by semantic locality,[18] that is, the user's expectation that concepts are arranged according to their meanings. Further support to navigation in an electronic thesaurus can be provided by preconstructed keys and picking lists.

The Anatomy Panel undertook preliminary reviews of assembled lists of anatomic concepts, which were subsequently forwarded to specialty working groups for comment before submission to the NHS CCC. Since that time, and using the work from the Anatomy Panel as our main resource, a gradual process of integration into a workable hierarchy structure has been undertaken (Fig. 1). The present content of the Anatomy chapter, approximately 7,000 unique concepts, has been largely determined by the pragmatic need to support the clinical record. During the CTP, specialist clinicians stipulated the anatomic elements needed to enhance their procedure and disorder concepts, and the level of detail therefore varies throughout the sections.

In addition to terms from the panel, further submissions from clinical working groups have been included. So, too, has a basic anatomy set from Version 2 of the Read Codes, which had been originally developed from the Site of Operation chapter (Chapter Z) of the Office of Population Censuses and Surveys Classification of Operations, 4th Revision.[19] This anatomy set was subsequently extended by the addition of further concepts, particularly in support of orthopedic surgery. A number of organizational headings have been added by the NHS CCC to underpin the hierarchical representation of heterogeneous concept types (e.g., anatomic regions and structures).

## Concepts and Terms

The Read Thesaurus explicitly separates the notion of a concept from the terms that act as the linguistic labels for that concept. This separation allows a robust approach to synonymous and homonymous terms. For each concept, one unique *preferred term* is identified, in addition to any number of synonymous terms. Within the Anatomy chapter, this mechanism has allowed commonly used alternatives, including eponyms and Latin names, to be incorporated. Sometimes, the decision as to which synonym should be the preferred term is necessarily an arbitrary one, but clear, natural language is the guiding principle. In general, no attempt is made to attach definitions, although, for most anatomic concepts, the unambiguous preferred term is sufficient. In practice, the preferred term is merely the default concept label, but its use is not mandatory; a clinician is free to select a synonym to enter in the electronic record.

Terms that in common usage are homonyms or that have multiple meanings, such as *ventricle* or *knee*, may be linked to multiple concepts. A homonym, however, can never be a preferred term because it is by definition ambiguous. In the case of *knee*, for example, *knee region* and *knee joint* are the two preferred terms; the simple form, *knee*, is provided as a natural language synonym for both.

## Hierarchical Structure

Previous versions of the Read Codes used the code itself to support the hierarchical structure of the thesaurus, enabling the creation of a limited, fixed hierarchical tree. Version 2 was limited to five levels of detail because each of the five characters of the code allowed progression to a further level. As the highest level concept for anatomy was 7N. . ., there were only three further levels to represent additional detail. Figure 2 shows an extract of the Version 2 nervous system hierarchy.

A similar strategy is employed to varying degrees by many clinical terminologies, including the International Statistical Classification of Diseases and Health-Related Problems, 10th Revision (ICD-10),[20] OPCS4,[19] and SNOMED.[21] These traditional hierarchical coding systems, although easy to implement within computer systems, suffer from several limitations. It is often necessary to place concepts arbitrarily within single categories, even when they naturally belong in several. In order to minimize the effect of this problem, designers of these coding schemes frequently need to create and strictly follow fixed axes of classification, even when these axes become unnatural to the system's users. Additionally, a concept cannot be moved or inserted within the hierarchy without altering existing codes. This is illustrated by the inability, within the Version 2 hierarchy shown in Figure 2, to insert the concept of *Central nervous system* between *Nervous system* and *Tissue of brain*.

Version 2 of the Read Codes has provided a simple and stable data structure to support the requirements of primary care and also the aggregation of data in the hospital sector. However, its inflexible structure has led to increasing difficulty in accommodating requests for expansion. Also, it became apparent at a very early stage of the CTP that the existing structure could not meet specialist requirements for the representation of clinical data.

To overcome these problems, the Version 3 alphanumeric Read Code acts solely as the symbol to uniquely identify each concept but no longer determines its hierarchy position. The hierarchical structure is carried instead in an additional table of superordinate-to-subordinate relationships. The separate hierarchy table allows adjustments and extensions to be made to the hierarchy without altering the identifying code.

In Version 3, there is no theoretical limit to the number of levels of detail. A single concept may be placed in multiple positions within the structure (multiple classification) by adding extra subordinate-to-superordinate links within the hierarchy table. If the concepts in the thesaurus are considered as *nodes*, subordinate to a single root node, and the hierarchical links between concepts are considered as *arcs*, the resulting data structure has the mathematical properties of a *directed acyclic graph*.[22] Acyclicity is the property of being unable to return to a node by traversing any number of arcs in a given direction; that is, it is impossible to go in a circle. The *tree* allowed by code-dependent hierarchies is a specific type of directed acyclic graph in which each node may only be reached by one path from the root node.

Within a controlled vocabulary, concepts will have defined relationships; in Read Version 3, these are specified in both hierarchy and template tables. In the hierarchy, a *generic relation* exists between two concepts when the intention of a superordinate concept is contained in its subordinate concept. In a generic hierarchy, such as that of Version 3, a subordinate concept should have a *type_of* (or *is_a*) relationship to its superordinate, and the integrity of the generic relation can be tested by a linguistic construction:

Mitral valve **is_a_type_of** Heart valve.

A *partitive relation* exists between two concepts when the subordinate concept refers to a part of the superordinate concept:

**Mitral valve**
  Mitral valve annulus
  Mitral valve leaflet

Applying the above test to this hierarchy demonstrates that the subordinate concepts (annulus and leaflet) are not subtypes of mitral valve but rather its parts. Occasionally, undetected homonomy is exposed by the apparent ability to describe a concept both as a subpart and a subtype of another concept. The relationship between the terms for *Nervous system* and *Central nervous system* illustrates this phenomenon. In this case, depending on how it is interpreted, *Nervous system* may refer to **a** nervous system (i.e. central, peripheral, autonomic) or **the** nervous system. This ambiguity has, in general, been avoided through the use of adequately specific terms.

| | |
|---|---|
| 7N... | Site of operation |
| 7N0.. | Nervous system |
| 7N00. | Tissue of brain |
| 7N004 | Tissue of cerebellum |

**Figure 2** Version 2 Read Code hierarchy.

A traditional view of anatomy places concepts in a mixed partitive and subtype arrangement, but the Read Version 3 hierarchy table holds only subtype relationships. This neutral structure supports the multiple views required of an anatomic thesaurus and also maintains consistency with the subtype arrangement of the rest of Read Version 3. Nevertheless, because it is necessary to present a more familiar partitive view to users, a mechanism has been developed to allow the creation of subsidiary meronymic links. Organizational concepts in the hierarchy act as indirect partitive links between an anatomic structure and its substructures. In most cases, the word *structure* has been used to denote these organizational concepts. Thus, *Aortic structure* may alternatively be read as a heading term for *Parts of the aorta* (Fig. 3).

A meronymic hierarchy structure (meronomy) is derived from the subtype hierarchy by explicitly identifying the links between the named body part and the organizational node. In Figure 3, these links are shown as dotted lines, and the derivable meronomy is illustrated in Figure 4.

Separating meronymic from subtype links enables the construction of simple, clinically useful alternative hierarchies of basic anatomic types, such as *Organs*, *Bones*, and *Joints*, that contain the named structures only and not their subparts. If these meronymic links and the labeled links within the template table discussed below are considered as additional arcs overlaid on the hierarchically directed acyclic graph, a large *associative* or *semantic network* results.

## Multiple Classification

Multiple classification enables both the appropriate placement of concepts along intersecting axes and also flexibility when a concept overlaps physically. Within

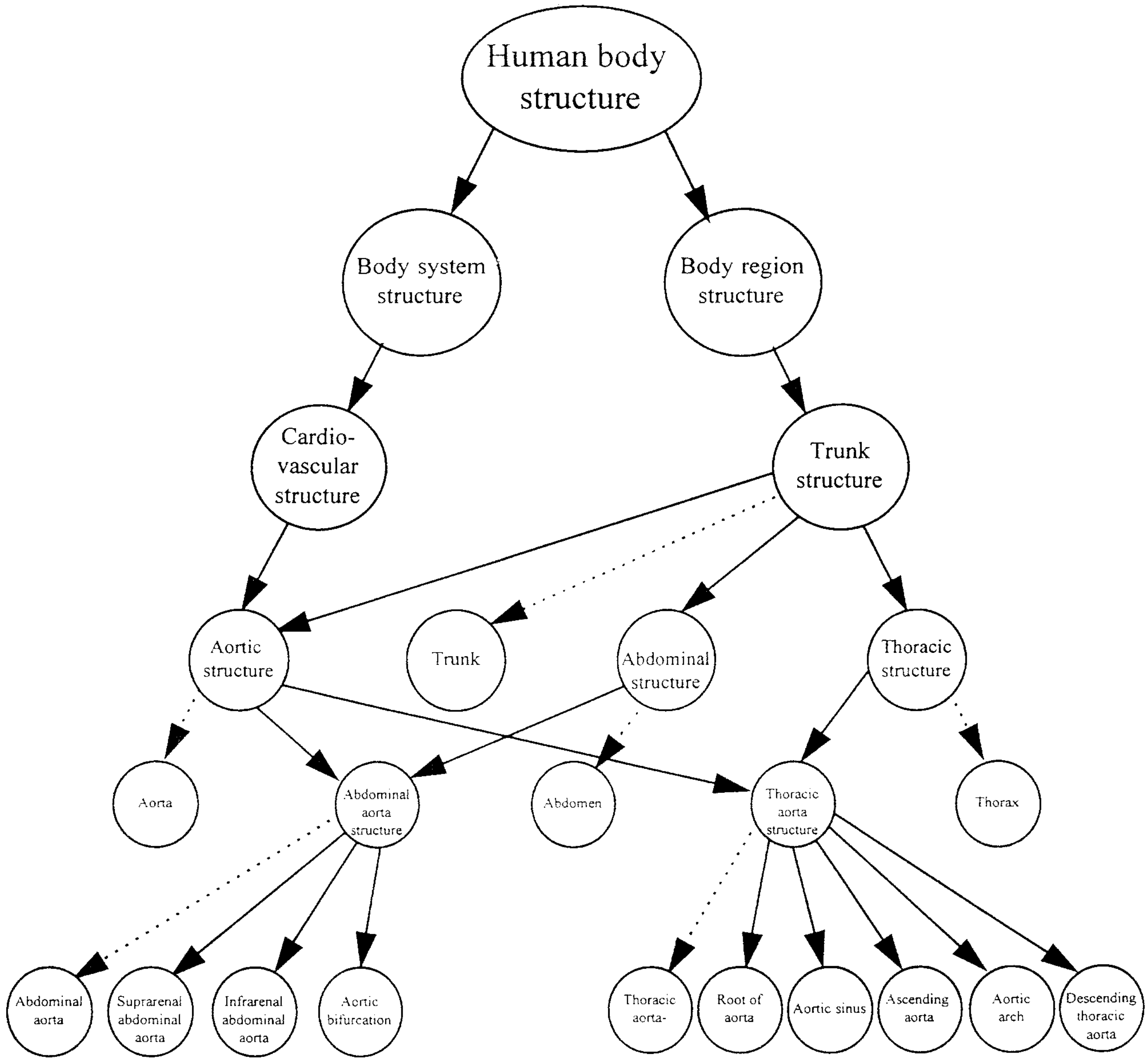


**Figure 3** Aortic struc-ture sub-type hierarchy.

a traditional coding scheme, it may be impossible to describe the site of pathology as being at the junction of two structures; if it is possible, these junctional locations would be placed arbitrarily into one category or another. In Version 3, junctional concepts are placed within the classes of both superstructures. For example, the *ureterovesical junction* is classified as a type of bladder structure and as a type of ureteric structure.

Furthermore, multiple classification allows additional axes to be used. While the predominant organizing principle of the Read Thesaurus is *body system*, additional axes have been added to increase usability and to facilitate translation to other classification systems. For example, within the Disorders chapter, in parallel with ICD-10, there are additional subchapters based on disease process (infection, neoplasm, trauma) and also period of disease onset (neonatal, puerperal). The major alternative axis within the Anatomy chapter of Version 3 is that of *body region*, so that both a regional and a systemic view of anatomy are presented (Fig. 1). The *aortic structure* hierarchy (Fig. 3) illustrates this dual classification axis and also a more complex example of the handling of overlapping structures. The provision of multiple axes coupled with cross-referencing provides a powerful resource for analysis.

When data is being analyzed, multiple classifications can potentially lead to multiple counting of cases. Once recognized, however, this situation can be readily dealt with and ultimately leads to greater flexibility. For example, a urologist analyzing a clinical database may wish to retrieve patients with a history of urologic pathology and then categorize the patients on the basis of site into the categories (A) *kidney*, (B) *ureter*, and (C) *bladder* (Fig. 5). An additional two categories could be created for the overlapping zones so that five non-overlapping categories existed. The five new categories (1 to 5) could then be treated in one of several ways, depending on the nature of the analysis required.

The advantages of multiple classification may be outweighed by the difficulties associated with authoring and implementing a directed acyclic graph. Advocates of restricting hierarchical representation to a tree propose the use of non-subtype semantic links as the mechanism for specifying additional characteristics. Deciding which characteristics to represent in the tree and which single axes of classification to choose at each level of the tree, and the problem of devising non-overlapping categories, may render this strategy somewhat difficult to implement across a wide domain.[23]

## Supporting Detail and Expressivity

The inability to cope with the level of detail required by clinical practice is a well-recognized limitation of traditional coding schemes. The encoding of complex medical concepts using controlled terminologies or coding systems may be supported by following either an *enumerative* or a *compositional* strategy. Purely compositional terminologies seek to construct all clinical concepts from their most basic elements, whereas enumerative terminologies seek to preassemble all relevant concepts as single, stand-alone entities.

A compositional approach allows a very expressive system to be constructed from a relatively small set of basic concepts. Repetitive anatomic structures such as

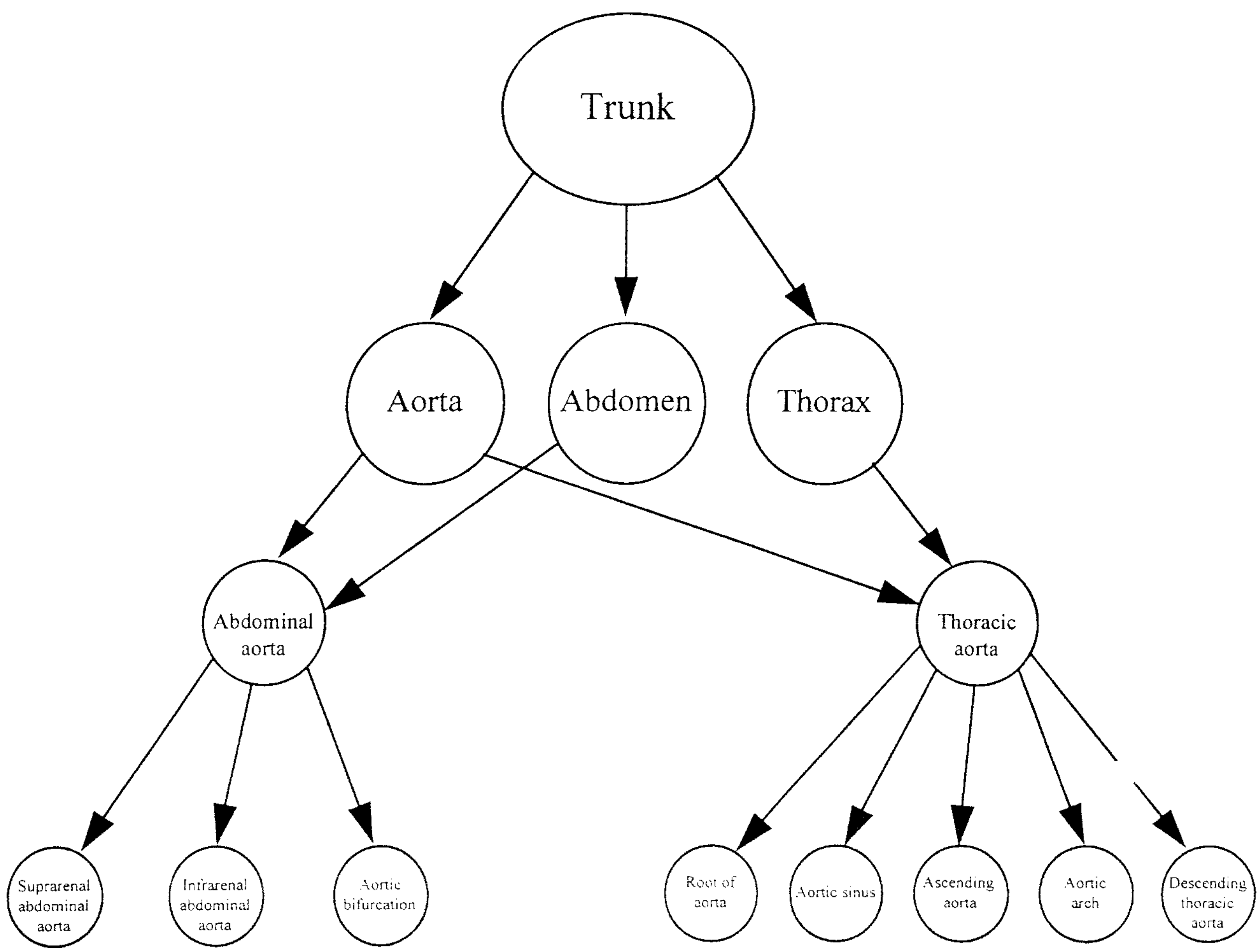


**Figure 4** Meronomy for aortic structure.

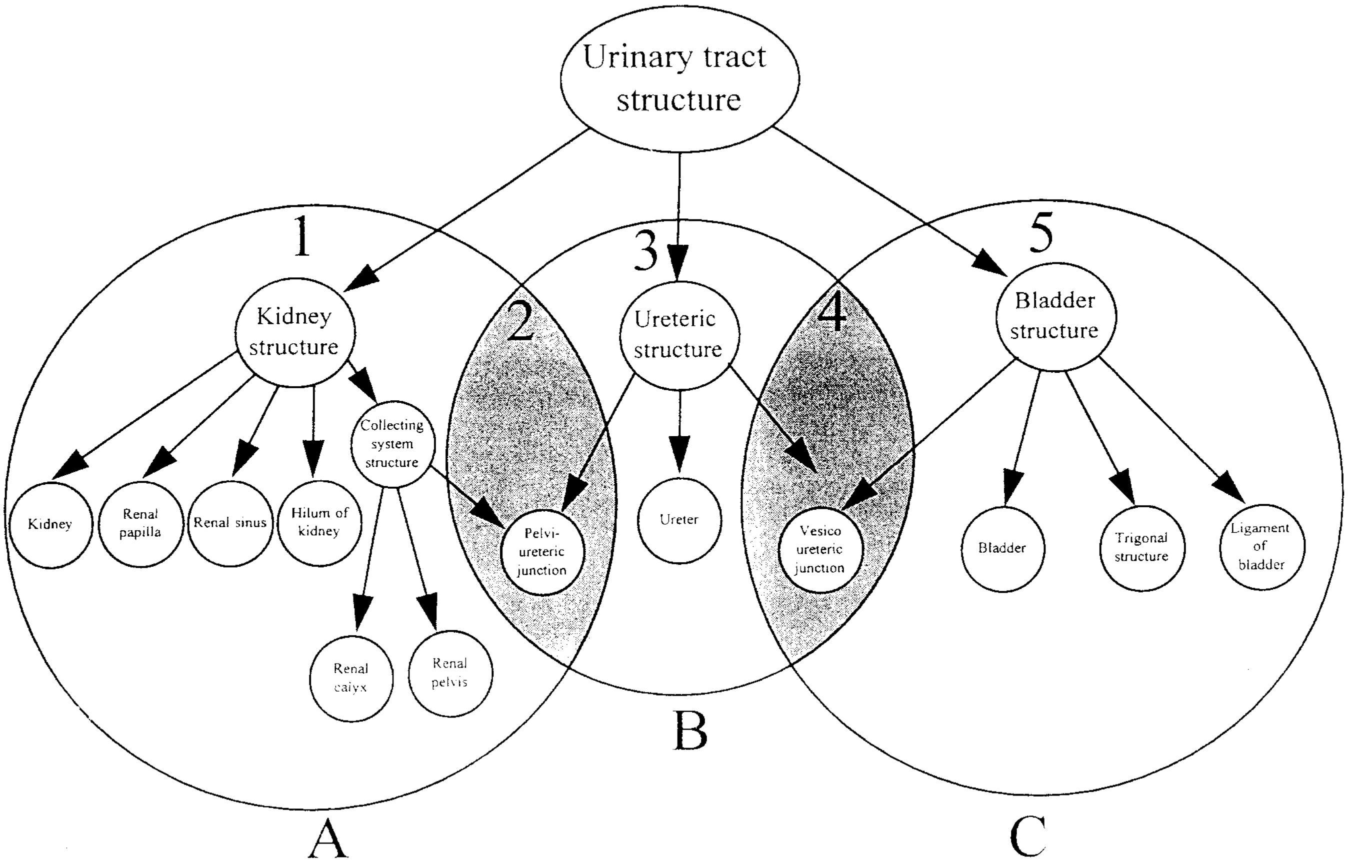


**Figure 5** Overlapping categories for analysis.

digits and vertebral elements illustrate this. Thus, from the notion of the 7 cervical, 12 thoracic, and 5 lumbar vertebral levels and the notions of *vertebra*, *vertebral body*, *lamina*, *pedicle*, and *spinous process* (29 concepts in total), it is possible to allow construction of 120 distinct concepts. The addition of *left* and *right* to *pedicle* allows another 48 concepts to be created. Allowing *rib* to be combined with *left* and *right* and a thoracic vertebral level provides a further 24 concepts. So, from 32 primitive concepts a compositional terminology could enable the representation of 192 new concepts. This exponentially increasing expressivity is the principal attraction of the compositional strategy.

The compositional approach has several drawbacks. First, a potentially complex sets of rules is required to ensure that non-sensible concepts such as *Left spinous process* or *Body of C1* are not assembled. Second, if it is possible to represent the same concept in more than one way, then the equivalence should be detectable automatically. Without this system of constraints, uncontrolled expressivity can seriously compromise subsequent analysis. Research into natural language processing may eventually help overcome these difficulties, but an internal conceptual representation scheme will still be needed. The European-funded GALEN project is a multi-center effort involving the development of *Terminology Servers* that strongly depend on composition and sophisticated sets of constraints.[6]

On the other hand, if sensible, relevant clinical concepts have been preassembled, predominantly enumerative terminologies such as Read Version 2 and SNOMED can bypass this complexity, allowing ready incorporation into existing clinical information systems. Indeed, in its most simple form, an enumerative terminology may be presented as a single table of data. Clearly, it is neither feasible nor sensible to enumerate all possible clinical concepts because, unlike compositional terminologies, enumerative schemes increase in size exponentially with the level of detail represented by a single concept. Very long lists of terms varying by only one or two words result from enumerating along multiple axes simultaneously. These large, repetitive lists present difficulties with both maintenance and the provision of usable interfaces.

The difficulties arising from enumeration, coupled with the requirement to provide a terminology that could be widely and immediately implemented without major modification of existing software, have greatly influenced the evolution of the Read Thesaurus. This problem has resulted in a pragmatic middle ground approach in which Version 3 combines aspects of both enumerative and compositional designs.

The goal has been to develop a thesaurus capable of representing most clinical concepts while avoiding complexities that result in only a minimal increase in expressiveness. For this reason, Version 3 has enumerated many anatomic concepts that could otherwise have been constructed compositionally; to date, within the Anatomy chapter, laterality is the only applicable qualifying statement. These anatomic terms are then available within the thesaurus to be used to construct detailed disorder or procedure concepts compositionally (Fig. 6A, B).

The existence of a trade-off between expressivity and ease of subsequent analysis should be recognized. The developers of any coding system should be aware that users seeking unfettered expressivity without a desire to perform analysis may always fall back on the power of natural language. For this reason, in the development of Read Version 3, expressivity has sometimes been sacrificed in order to ensure analyzability.

The decision to enumerate rather than provide mechanisms to support compositional construction of anatomic concepts may be justified on several grounds. First, the level of anatomic detail recorded in present clinical information systems seems adequately supported by the existing 7,000 anatomic concepts. Second, the Read Thesaurus is already being incorporated into clinical information systems in the NHS; in order to enable this implementation of the Thesaurus across many platforms, there must be a stable product. Therefore, despite the inherent structural flexibility of the Thesaurus, a relatively conservative stance has been adopted in the proposed model. If current research activity[24,25] confirms that a usable system of anatomic composition is indeed possible, then this type of model may be incorporated into the structure of subsequent releases of Version 3.

Composition is supported in Version 3 by a *template* table (Fig. 6). This table allows the linkage of an *object* concept with allowable *values* for a given *attribute* and may thus be used to attach potential qualifying statements. The characterization of the anatomic component of disorder and procedure concepts is achieved by defining anatomic values linked to the concept by the attribute *Site* (Fig. 6A). In addition, all non-unilateral anatomic structures within Version 3 have values for the attribute of *Laterality* appropriately specified (Fig. 6B).

The specification of *Artery* as the site of *Arterial cannula insertion* allows the construction of interfaces that provide users with a choice of arteries to add detail to the core concept. Because all arteries occur as descendents of *Artery* in the hierarchy table, the user is able to choose *Radial artery*, which can then itself be qualified with laterality.

In the case of *Radial artery cannula insertion*, the template table entry serves simply to identify anatomic detail inherent in the object concept, rather than allowing the addition of further qualifying detail. This makes possible the automatic detection of equivalence between *Arterial cannula insertion* plus *Site: Radial artery* and *Radial artery cannula insertion*. A set of additional fields within the tables released to implementors supports the process of detecting this equivalence and assists in the construction of clinical interfaces.

The formal characterization of the elemental components of complex concepts blurs the distinction between terminology and knowledge base. This blurring is to a certain extent inevitable, and the incorporation of knowledge of this nature has been advocated as a necessary element in maintaining large, controlled medical vocabularies.[26,27]

In addition to enabling construction of clinical interfaces and detection of equivalence, the template table allows the derivation of the anatomy implicit within a concept and thus provides a valuable resource for data analysis. Furthermore, from the anatomic hierarchy, which has a dual classification based on systems and regions, both the system and the region of the body involved may be deduced. For example, it can be automatically determined that a clinical record entry of *Radial artery cannula insertion* involves the radial artery, which is both an artery and an upper-limb structure.

The process of linking to anatomy the hierarchies of disorders and procedures submitted by the clinical

| Object | Attribute | Value |
|---|---|---|
| Arterial cannula insertion | Site | Artery |
| Radial artery cannula insertion | Site | Radial artery |

**Figure 6A** Template table for *Arterial cannula insertion*.

| Object | Attribute | Value |
|---|---|---|
| Radial artery | Laterality | Right<br>Left<br>Unilateral<br>Bilateral |

**Figure 6B** Template table for *Radial artery*.

specialty working groups has been an invaluable validation of both the content and the structure of the Anatomy chapter. When an appropriate anatomic concept could not be located to specify the site, the Anatomy chapter was augmented. In this way, expansion of the Anatomy chapter has been directed by clinical requirements. As the authoring of the site linkages progressed, it became possible to automatically detect instances when the axis of the Anatomy chapter diverged from that of the Disorder and Procedure chapters, enabling harmonization of the hierarchies.

Arranging concepts into a hierarchy enables the template records of many concepts to be initially inherited by their subordinate concepts. Although a valuable authoring tool, inheritance usually needs to be manually checked; there are frequently exceptions, and the inheritance of qualifiers often forces a view of the hierarchy that is different from that of the original author.

## Lessons Learned, Unresolved Issues, and Direction for Future Work

The CTP was initiated with Version 2 of the Read Codes as its starting point. The limitations of its fixed, code-dependent hierarchy and the lack of formal qualifying mechanisms quickly led to the adoption of the more flexible data structures[27] now implemented in Version 3. The transition from a fixed to a dynamic hierarchy required several significant changes to the authoring practices. The inability to rely on a stable set of hierarchically related concepts and the absence of definitions within the Thesaurus made it essential that each concept be labeled by a stand-alone, unambiguous term. Additionally, residual categories, generally identifiable by the presence of "other" or "not elsewhere classified" in the term became unstable in a dynamic hierarchy as its evolution continually changed their meanings. For this reason, these terms have not been included within the recommended clinical set of Version 3 concepts.

The use of inheritance to assist in generating template table records imposed a discipline on the construction of hierarchies within Version 3 that was not required in the earlier versions. This use has generally resulted in increased consistency but has caused some difficulty in structuring those sections of the Thesaurus that do not lend themselves to sub-type hierarchies. The use of heading terms with a carefully chosen suffix, such as "structure," has been a useful technique for organizing these sections.

An important feature of the Read Codes is their dynamic nature; quarterly updates allow for corrections and for incremental additions to the content. The flexible nature of the Version 3 data structures enables improvements in the scope and detail of the Anatomy chapter to continue. A phased series of enhancements is planned over successive quarterly releases to include additional concepts, particularly in response to requests from the growing number of sites now using Read Version 3. Expansion of the current Anatomy chapter is likely in embryology, dysmorphology, and microscopic anatomy. The more complex data structure of Version 3, although necessary, has imposed a significant overhead on system developers compared with the simpler structure of the early versions.

Within most domains, a model for representing the majority of required concepts can usually be defined quite readily; the stability and general applicability of human anatomy has made it the natural focus around which to harmonize the often conflicting conceptual models of the various specialist groups. Unresolved issues at the boundaries of the Version 3 Anatomy chapter also remain. Examples include dysmorphologic and variant structures, as well as persistent embryonic structures that may all be considered either as structures in their own right or as clinical findings. For example, the *ileum* is, without question, an anatomic structure, but *distal ileitis* is a clinical finding. *Meckel's diverticulum*, however, may be viewed as either.

There are three options for the authoring of these terms. First, these concepts may be placed solely in the Anatomy chapter as structures. The presence of this structure may then be constructed by the linkage of the notion of *presence* with that of the finding. Second, a special category may be created for dysmorphologic and remnant structures, and they may then be allowed to assume a dual role as either clinical findings or anatomic structures, depending on the context. Third, two different concepts may be created: one for the structure and another for the finding that relates to that structure. Which of these options will eventually prove to be the most useful remains to be determined. Finally, developmental work is now taking place in other domains, including surgical actions and disease causative agents, with the intention of underpinning representation of secondary axes in these chapters.

## Evaluation

The design of the CTP included a phase of widespread review of the content by specialist clinicians. This is now being followed by collaborative field testing in cooperation with partnership hospital sites. In addition, closer relationships are being forged with system suppliers and users in the primary care sector. It is the philosophy of the NHS CCC that links with system developers and rapid response to user requirements will enable us to most rapidly achieve our goal of uniform coding of clinical data across the NHS. The evaluation metrics for the Read Thesaurus, therefore, relate to the *extent* of usage within the health service and *user satisfaction*. These will be influenced by a range of factors, including ease of implementation, coverage, and analyzability. Evaluating each of these factors in isolation is of questionable value because of the complex interrelationships among them.

Nevertheless, external bodies are being actively encouraged to perform and publish their own formal evaluations of coverage and usability. To date, there have been no British studies formally evaluating Version 3, although one American study has been published.[28] Although its findings were generally positive, it lacked comparison with other major controlled clinical vocabularies. A recent, larger scale comparative study by the Computer-Based Patient Record Institute[2] examined only Version 2 of the Read Codes. Unfortunately, the fundamental changes to the data structures and the significantly increased coverage of Version 3 make it impossible to extrapolate the results of a Version 2 evaluation to Version 3. However, a further study by the same group, this time including Version 3, is now in progress.

## Conclusion

Read Version 3 has been developed with widespread specialist clinical and scientific input and will continue to evolve in response to the requirements of clinical care. This paper has examined the structural features that have permitted the construction of a flexible, symbolic representation of anatomic knowledge. In addition, the linkages that enable these anatomic concepts to support the representation of mainstream clinical concepts within the vocabulary, such as disorders and procedures, have been discussed.

It is recognized that objective, external evaluation of the content and structure of Version 3 is presently limited. Ultimately, however, its successful use in the clinical arena will provide the best test.

*References* ■

1. Musen MA. Dimensions of knowledge sharing and reuse. Comp Biomed Res. 1992;25:435–67.
2. Rigby MJ, Roberts R, Williams JG. Objectives and prerequisites to success for integrated patient records. Comp Meth Prog Biomed. 1995;48:121–5.
3. Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content and coverage of clinical classifications. J Am Med Inform Assoc. 1996;3:224–33.
4. Evans DA, Cimino JJ, Hersch WR, Huff SM, Bell DS. The Canon Group: Position statement toward a medical concept representation language. J Am Med Inform Assoc. 1994;1: 207–17.
5. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Meth Inform Med. 1993;32:281–91.
6. Rector AL, Glowinski AJ, Nowlan WA, Rossi-Mori A. Medical concept models and medical records: An Approach based on GALEN and PEN&PAD. J Am Med Inform Assoc. 1995;2:19–35.
7. O'Neil MJ, Payne C, Read JD. Read Codes Version 3: A user led terminology. Meth Inform Med. 1995;34:187–92.
8. Buckland R. The language of health. BMJ. 1993;306:287–8.
9. Williams JG, Morgan JM, Severs MP, Howlett PJ. Let there be light. British Journal of Healthcare Computing. 1993;10: 30–2.
10. Stannard CF. Clinical terms project: a coding system for clinicians. Br J Hosp Med. 1994;52:46–8.
11. Barron SL. The Read clinical classification. Br J Obstet Gynaecol. 1993;100:800.
12. Chisholm J. The Read clinical classification. BMJ. 1990;300: 1092.
13. Leaning MS. The new information management and technology strategy of the NHS. BMJ. 1993;307:217.
14. Tabaqchali MA, Venables CW. The clinical terms project: its potential for computerised surgical audit. Ann R Coll Surg Engl (suppl). 1995;77:124–9.
15. Rosse C. The potential of computerized representations of anatomy in the training of health care providers. Acad Med. 1995;70:499–505.
16. ISO 2788: 1986. Guidelines for the Establishment and Development of Monolingual Thesauri. Geneva: International Standards Organisation, 1986.
17. Rada R, Ghaoui C, Russell J, Taylor M. Approaches to the construction of a medical informatics glossary and thesaurus. Med Inf. 1993;18:69–78.
18. Tuttle MS, Sherertz D, Olson, N, et al. Using Meta-1—The first versions of the UMLS Metathesaurus. In: Miller RA (ed). Proc Annu Symp Comput Appl Med Care. New York: IEEE Computer Society Press, 1990:131–5.
19. Classification of surgical operations and procedures (4th revision). Office of Population Censuses and Surveys. London: HMSO, 1990.
20. International Statistical Classification of Diseases and Related Health Problems. 10th Revision. Geneva: WHO, 1992.
21. Côté RA, Rothwell DJ, Palotay JL, Becket RS, Brochu L. The systematized nomenclature of human and veterinary medicine, SNOMED International (4 volumes). College of American Pathologists, April 1993.
22. Aho AV, Hopcroft JE, Ullman JD. Data structures and algorithms. Reading, Massachusetts. Addison–Wesley, 1983.
23. Zweigenbaum P, Bachimont B, Bouaud J, Charlet J, Boisvieux J-F. Issues in the structuring and acquisition of an ontology for medical language understanding. Meth Inform Med. 1995;34:15–24.
24. Rector AL, Gangemi A, Galeazzi E, Glowinski AJ, Rossi-Mori A. The GALEN CORE Model Schemata for Anatomy: Towards a Re-usable Application-Independent Model of Medical Concepts. In: Barahona P, Veloso M, Bryant J. Proc MIE Lisbon. 1994:229–33.

25. Rosse C, Ben Said M, Eno KR, Brinkley JF. Enhancements of Anatomical Information in UMLS Knowledge Sources. In: Gardner RM (ed). Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care. Philadelphia: Hanley and Belfus, 1995:873–7.
26. Cimino JJ, Hripcsak G, Johnson SB, Clayton PD. Designing an Introspective, Multipurpose, Controlled Medical Vocabulary. In: Kingsland LC (ed). Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care. New York: IEEE Computer Society Press. 1989:513–8.
27. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. J Am Med Inform Assoc. 1994; 1:35–50.
28. Hausam RR, Hahn AW. Representation of Clinical Problem Assessment Phrases in U.S. Family Practice Using Read Version 3.1 Terms: A Preliminary Study. In: Gardner RM (ed). Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care. Philadelphia: Hanley and Belfus, 1995:426–30.