Development and Role of the Read Codes

by Ann Harding and Charlie Stuart-Buttle

Produced and maintained by Britain's National Health Service, the Read Codes are a comprehensive, controlled clinical vocabulary. Here's a look at how the codes evolved, their use in the NHS, and the continual process of aligning the system with the needs of its users.

The Read Codes are an alphanumerically coded, controlled clinical vocabulary produced at the National Health Service Centre for Coding and Classification (NHSCCC). This organization is part of the Information Management Group (IMG) of the NHS Executive. There are three versions of the codes: the Four Byte Set, Version 2, and Version 3. This article will describe all the versions, but will concentrate mainly on Version 3 as the most recent development and will outline the development, structure, content, role, and philosophy of the thesaurus.

History of the Read Codes

The Read Codes were developed in the early 1980s by Dr. James Read, then a general practitioner in Loughborough, England, to record clinical and administrative data for general practice. Four-character alphanumeric codes determine the position of a term within the hierarchy, so this version is known as the Four Byte Set. This structure offers computational simplicity, but also restricts hierarchy depth to four levels, which in turn constrains the content of the codes. Synonyms for some concepts are available. In 1988, the Read Codes became the crown copyright.

The limitations of a four-level hierarchy led to the development of a Five Byte Set in 1990. This was initially limited to the primary care domain, but later expanded to include additional concepts to support secondary and tertiary care. There are two version of this set, and both include surgical procedures and a more structured mechanism for representing synonyms. Version 1 has shorter terms and keys, but like Version 2, it has mappings to other systems, including ICD-9-CM and ICD-10-CM and the United Kingdom Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures, Fourth Revision (OPCS4). In addition, the pharmacy concepts also cross-map to the British National Formulary (BNF) and Anatomic and Therapeutic Chemical classification index (ATC). Version 2 is the most widely used format of the Five Byte Set.

Version 3 of the Read Codes was developed during the Clinical Terms Projects (1992 to 1995), a series of major collaborations between the NHS Executive and the Conference Information Group of the Conference of Medical Royal Colleges and their faculties in the UK, the Nursing Professions and Professions Allied to Medicine, including physiotherapy, podiatry, speech and language therapy, dietetics, and occupational therapy. The aim was to provide enhanced functionality and greater specialist detail and to encompass the wider domain of healthcare. Widespread representation of clinical professional interest was promoted by the establishment of more than 50 specialty working groups (SWGs). The initial proposal for a set of clinical terms for use by clinicians came from the clinical professions, and Version 3 was set up based on the existing versions of the Read Codes and the lists supplied by the SWG clinicians. From the outset, the initiative had professional ownership and leadership. During the course of these projects, more than 2000 healthcare professionals were involved in the development and quality assurance of Version 3. The project was managed by the NHSCCC, which also provided training, terminological expertise, computer equipment, and financial control. The differences between the three versions of the Read Codes are summarized in Table 1.

table 1—Properties and Roles of the Read Code Versions

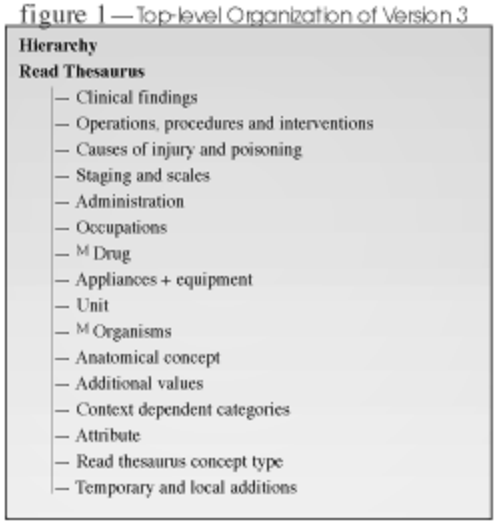
	4 Byte	Version 2	Version 3
Hierarchy representation	Code-dependent	Code-dependent	Link-based
User domains	General practice	General practice and secondary care	All clinical professions and sectors
Multiple parents	No	No	Yes
Hierarchy depth	Four levels	Five levels	Unlimited
Hierarchy relationships	Mixed	Mixed	Subtype
Compositionality	No	No	Yes
Cross-maps	BNF and ATC	OPCS4, ICD-9, ICD-10, BNF, and ATC	OPCS4, ICD-9, ICD-10, BNF, ATC, and EAN
Number of concepts*	20,824	69,470	228,509
Number of terms*	29,031	80,631	264,855

(March 1998 release) - including pharmacy

Structure of Version 3

The requirement to support different clinical disciplines and views identified a need for a more expressive, flexible structure than previous versions. There was also a need to be able to more precisely analyze stored data using the codes. This led to the adoption of a sub-type hierarchy, in which each child is a type of its parent. Concepts are arranged in so-called chapters (Figure 1) which include disorders, history and observations, operations and procedures (including investigations), administration (in some areas specific to the UK), and a number of socially relevant concepts such as occupations. In addition, there are hierarchies of other essential terms such as organisms, chemicals, drugs, physical agents, and environments, including geographical locations.

figure 1—Top-level Organization of Version 3



Version 3 makes the distinction between a concept—which can be, for example, a state, a procedure, an event—and a term, which is a descriptive label for a concept. A separate set of codes are used to identify each concept and each term. In order to identify a particular description of a concept by a particular term the combination of the codes for the concept and the term must therefore be used. This separation allows synonymy, and each concept has a unique, unambiguous "preferred" term but can have any number of synonyms. Where potential ambiguity exists, the more natural terms may need to be relegated to the role of synonyms, and a less intuitive but unambiguous term is used for the preferred term. For example, "Turkey" may refer a

type of meat or a country in Asia Minor, and both are clinically useful terms. Each concept has "Turkey" as a synonym, but there are two unequivocal preferred terms of Turkey—meat and Turkey—country.

The model used to design Version 3 was a directed acyclic graph. This graph is implemented by a link-based hierarchy table, which allows (theoretically) unlimited hierarchy levels, multiple parents, and moving of concepts. For example, the concept "Viral meningitis," which is both a neurological and an infective disorder, is positioned in both hierarchies (see Figure 2). Version 3 also uses qualifiers to define additional detail to concepts. The example of "cholecystectomy" can be qualified by approach, extent, using, method, priority, and site (see Figure 3).

figure 2—
Multiple Parentage of Viral Meningitis

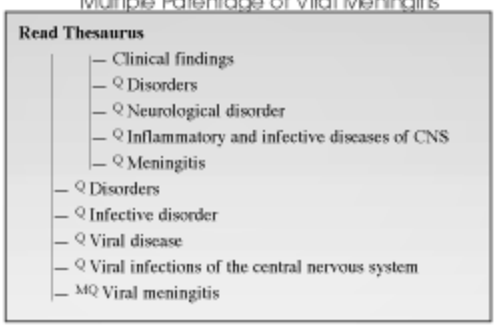
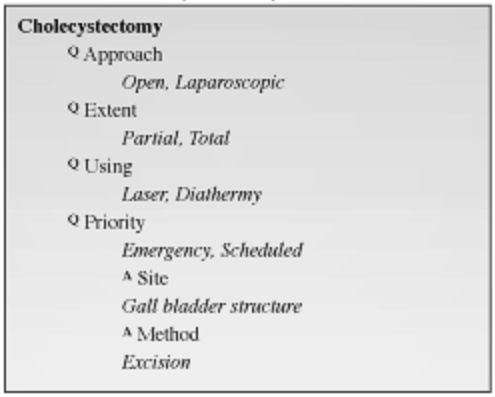


figure 3—
Qualifiers (Q) and Atoms (A) Allowing Additional Detail and Semantic Definition for Cholecystectomy



In this way, the thesaurus is partially compositional in that complex concepts can be constructed from more simple concepts. The site qualifier allows access to laterality (left, right, bilateral) where this is appropriate. A similar approach will semantically define all the concepts within Version 3. This is achieved by identifying intrinsic component attributes of these concepts as "atoms," as opposed to qualifiers.

This mechanism enables identification of equivalence—for example, "Osteoarthritis" qualified by site "Hip joint" is identified as being equivalent to "Osteoarthritis of the hip" with atom site "Hip joint." This is especially important for accurate retrieval and analysis of stored data, but it can also be used to validate hierarchical placements, which facilitates accuracy of editing and maintenance.

Version 3 is produced in a relational database and released as flat text files in relational format. The terms are intended to be used by clinicians, with emphasis on natural clinical terms as found in written records. This has meant the replacement of classification categories, such as "Asthma not otherwise specified," with an "optional" status so that they do not appear on initial picking lists for data entry (although they can be accessed if the user group will benefit from being able to use these terms). Like Version 2, the thesaurus has cross-mappings to OPCS4, ICD-9 and ICD-10, and the BNF and ATC classifications. In addition, there is a map to the European Article Number (EAN).

Maintenance and Refinement of the Read Codes

Creation of Version 3 was a logistically complex and time-consuming procedure. It is recognized that maintenance of such a product requires continued clinical knowledge, resources, and effort. There are a number of reasons for this:

- The thesaurus is updated every three to six months (monthly for drugs and appliances). This allows both ongoing refinement and changes to be made at the request of users who expect a rapid response. The resource implications exist not only in clinical authoring time but also in the provision of appropriate software to implement, record, and control changes to the thesaurus

- Semantic definitions are being applied to concepts to support retrieval and placement of new terms. As more of the thesaurus is semantically defined there will be greater shift of responsibility for the thesaurus to the NHSCCC clinical authors, with relatively little involvement by the professions who created it. The process of semantic definition is not always straightforward, especially in the case of multisystem disorders such as scleroderma or poorly understood disorders of obscure etiology, and judgmental clinical decisions may have to be made. Where such a decision is outside the knowledge base of the clinician, the issue may be referred to the relevant SWG

- Forward compatibility between earlier versions is maintained. There are many clinicians and coders, both in primary and secondary care, who have stored data using earlier versions. These users need to be able to carry through and analyze such data. Users of different versions also must be able to communicate with each other. To assist with these processes, Version 3 has been made a superset of both the Four Byte Set and Version 2, so that all concepts from these versions exist within Version 3

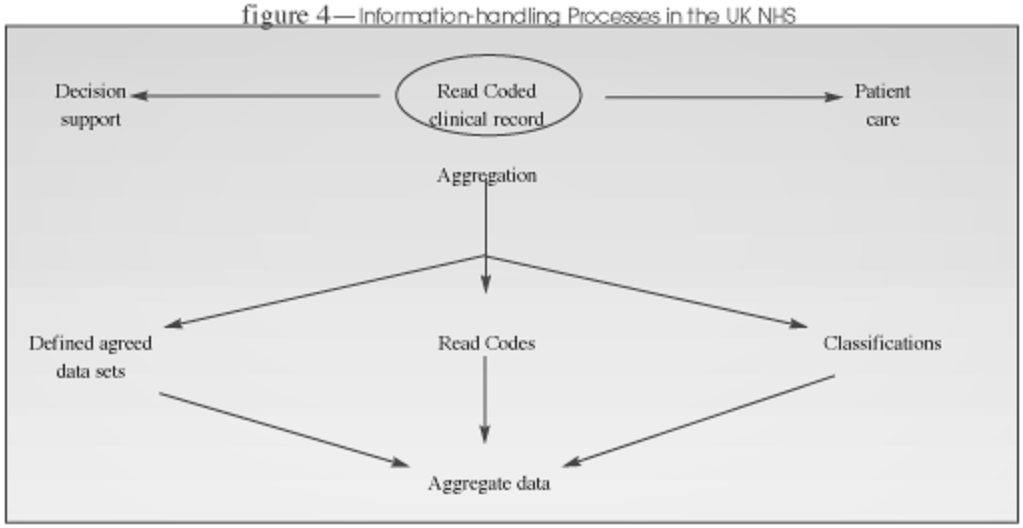
- Maps to the classifications require constant maintenance. New concepts will of course require a cross-map, but subtle changes to concepts or their qualifiers can also engender a need for refinement of the mappings

This work is done by a team of clinically and scientifically trained authors and classification experts who require knowledge of the structure and philosophy of the thesaurus in addition to that of their specific domain.

Use of the Read Codes in the NHS

An agreed clinical thesaurus forms part of the NHS information management strategy for the UK. Information provision in the NHS should be based on operationally gathered clinical data, which can then be used for subsequent aggregation of data for secondary purposes. This can take place directly or indirectly via existing classifications or defined data sets (see Figure 4). The Read Codes are designed to record the everyday care of patients, and provision of cross-maps to ICD-10 and OPCS4 enables a link to these statistical classifications, although the codes can be used directly for aggregation. Such data can then used to generate healthcare resource groups (HRGs) and healthcare benefit groups (HBGs) that group patients requiring similar resources and costs and who experience similar outcomes, thereby underpinning data analysis and subsequent planning of patient care. The Four Byte Set and Version 2 are already in widespread use in the UK. Approximately 90 percent of general practices are now computerized, and of these, 73 percent use the Read Codes in computer systems for the recording and storing of clinical data. The requirements for accreditation for such systems include compatibility with the Read Codes. Version 2 is in use, to one extent or another, in some 46 percent of NHS hospitals. There are also pilot projects using Read Codes for electronic data interchange between hospital pathology laboratories, radiology departments, and general practitioners.

figure 4—Information-handling Processes in the UK NHS



Live Use of Version 3

The first live use of Version 3 commenced in 1995. The codes are now in use across 23 specialties and in 12 sites in the UK and are being used by several hundred individual clinicians. Version 3 is a considerably more comprehensive product than its predecessors, covering a far wider domain in greater detail. As such it has required a greater level of testing. It is not possible to subject a terminology to a test of fitness for purpose outside the sphere of operational use, since testing outside of live implementations can only reflect limited areas of usage, with concomitant restrictions on the scale and validity of data. During the course of the past year, the NHSCCC has been involved in operationally testing the thesaurus in live use.

Feedback from a steadily growing number of operational testing sites has already been extremely valuable and of a significantly different nature to that received from early piloting. This operational testing is currently taking place at sites throughout the UK and is mainly focusing on selected areas of the hierarchy. This collaboration involves the NHSCCC assisting sites that are implementing Version 3 systems. The sites provide feedback on term requirements and hierarchy placements, evaluation studies, cross-mapping feedback, and studies on data analysis. This work has been taking place in a number of specialties, including general medicine; care of the elderly; dentistry; general; ophthalmic; ear, nose, and throat; orthopedic surgery; obstetrics and gynecology; and pediatrics and midwifery. It is hoped that live use will be widened in the next year, including general practice.

The more complex structure of Version 3 has resulted in the development of three commercially available Read Code engines, produced independently of the NHSCCC, which implement an applications programming interface, thus reducing the overhead for software suppliers producing compliant clinical systems. These engines are currently undergoing beta testing by a number of software suppliers at 16 sites.

Summary

The Read Codes have been in use within the NHS for more than 10 years. Development and maintenance of Version 3 has been driven by users. During these processes, significant lessons have been learned, both about the needs of clinicians and how these can be aligned with the needs of clinical coders and managers. The creation of the Read thesaurus was a significant undertaking, supported by the NHS with funding, but also by the goodwill of many members of the clinical professions.

For More Information

For further information about Read Codes or the NHS Centre for Coding and Classification, please contact the NHS Executive Headquarters, Information Management Group, NHS Centre for Coding and Classification, 58-60 Woodgate, Loughborough LE11 2TQ, England. Telephone 0044 1509 211411; fax 0044 1509 211611.

Ann Harding *is executive director of the NHS Centre for Coding and Classification, Loughborough, England.* ***Charlie Stuart-Buttle*** *is the centre's director of Read Code development.*

Article Citation:

Harding, Ann and Charlie Stuart-Buttle. "The Development and Role of the Read Codes." *Journal of AHIMA* 69, no. 5 (1998): 34-38.