



THE CLINICAL TERMS VERSION 3 (THE READ CODES)

MAIN FILE STRUCTURE: OVERVIEW and TECHNICAL DESCRIPTION

APRIL 2008

Purpose of this document

This document is one of a series that, taken together, describe the contents, structure and function of Clinical Terms Version 3 (The Read Codes).

This introduction is intended to provide information on Clinical Terms Version 3. It is also a guide to the other available documents each of which is updated independently. For this reason, different chapters may have different version numbers.

INFORMATION

Distribution	On request
Author	UK Terminology Centre
Further Copies From	UK Terminology Centre Helpdesk Service Support Unit Prospect House, Floor 2 Fishing Line Road Redditch Worcestershire B97 6EW Tel: +44 (0) 1392 206248 Fax: +44 (0) 1392 206945 E-mail: datastandards@nhs.net Internet: http://www.connectingforhealth.nhs.uk/systemsandservices/data/readcodes
Date of Issue	April 2008
Reference Number	168v1.0

© Crown Copyright 2008
Published by the UK Terminology Centre

Table of Contents

Table of Contents	3
1. Introduction	4
2. Background	6
3. Summary of Clinical Terms Version 3	7
4. Principles behind the CTV3 File Structure	9
5. Overview of the File Structure	13
5.1 Describing Terms (Term file)	16
5.2 Linking Terms to Read Codes (Description file)	17
5.3 Describing Concepts (Concepts file)	18
5.4 Describing a Hierarchy	19
5.5 Accessing Terms and Concepts (Key File)	20
5.6 Describing a Speciality's Interest in Clinical Concepts (Specialty Files)	20
5.7 Cross-Mapping Files	21
Technical Annex	21
T.1 Conceptual Model of the Read Codes	21
T.2 Description of the File Layouts in CTV3	24
T.3 Analysis of the Differences of CTV3 in comparison to Earlier Read Code Versions	30
T. 4 Other Available Documents	33
Appendix A	34
Appendix B	35

1. Introduction

Clinical Terms Version 3 (The Read Codes), abbreviated as CTV3, is designed to satisfy the terminology requirements of all health care professional groups including nurses, physiotherapists, general practitioners and the many different medical and surgical specialists. In addition it is also designed to support accurate mapping to other coding schemes particularly statistical classifications including the WHO International Statistical Classification of Diseases – Tenth Revision (ICD-10) for diagnoses and the UK Classification for Surgical Procedures (OPCS-4) for procedures.

The structure and contents of CTV3 was developed in collaboration with professional representative bodies in the UK during a series of “Terms Projects” (The Medical Clinical Terms Project; the Nursing Midwifery and Health Visiting Terms Projects; and the Professions Allied to Medicine Terms Project). The product of this initiative was a detailed corpus of terms and concepts that was too large and complex to be incorporated into the earlier versions of the Read Codes (the 4-Byte Set and the 5-Byte Set). This led to the evolutionary development of the CTV3 file structure which was designed to:

- Satisfy the expression of both general and complex concepts (allowing the use of qualifiers for detail);
- Allow optimal and accurate placement of concepts in a consistent framework;
- Provide multiple perspectives (e.g. nursing, GP, central returns, specialists) of the same concepts;
- Produce more accurate and complete mappings to other coding schemes;
- Permit future evolutionary change.

The new structure allows:

- Concepts to be accurately placed in relation to each other according to their meaning (semantics) allowing the production of an electronic thesaurus;
- All earlier 4-Byte Set and 5-Byte Set Read Codes and terms to be incorporated;
- New terms and concepts to be added (the full capacity is 656 million);
- The addition of detail to concepts by the use of qualifying statements using a formal structure;

- The use of a common formal structure to represent the intrinsic meaning of concepts (semantic definitions);
- More efficient disk storage of the Read Code thesaurus by splitting the Read release into several small files and avoiding duplication of terms;
- On-screen display of restricted subsets of terms;
- Multiple-classification of concepts in all relevant places e.g. Pulmonary tuberculosis is classified both as an infection and as a lung disease;
- Clear and separate identification of terms and concepts, allowing the links between terms and concepts to be altered in a controlled manner as health related knowledge shifts;
- Robust mechanisms for change control (once a Read Code is allocated or a term added they are never deleted from the thesaurus and any changes can be accurately tracked).

This document is written with both health care professionals and system developers in mind. The first half includes a discussion of the principles behind the structure and gives an introduction to the elements and use of the structure, based on examples. The second half is devoted to a technical description of the relationship between files and the file layouts.

This document is one of a series that describe the CTV3 file structure. The complementary documents “Clinical Terms Version 3 – Template File” and “Clinical Terms Version 3 – Cross Mapping File” cover aspects that are only briefly mentioned in this document.

Included in this document are:

Part 1

- Background to the changes
- Summary of the CTV3 structure
- Principles behind CTV3
- Overview of the file structure
- Summary of the differences between Read Code Version 2 and CTV3

Part 2

- Conceptual model of the representation

- Description of the layout of each file
- Analysis of the proposed changes

2. Background

The Read Codes were initially developed in the 1980's as a terminology for describing the clinical process of medical general practice. Coverage in the original 4-Byte Set was wide, including symptoms, signs, diseases, procedures, investigations, laboratory medicine and administration. Subsequently they were adapted for use in hospitals in an expanded structure known as the 5-Byte Set which was produced in two file formats (Version 1 and Version 2 file structures). The 5-Byte Set was larger and was based upon the contents of ICD-9 and OPCS-4 providing more codes and terms for diagnoses and surgical procedures which were simply cross-mapped to the ICD-9 and OPCS-4 classifications.

Their primary use has remained that of underpinning a computerised medical record that can subsequently be used for viewing a patient's record from different perspectives, clinical audit, producing reports, meeting central returns and research purposes. After being in widespread use in many different implementations they were adopted as a national standard in the UK for recording the clinical process in General Practice.

The complexity of the file structure in these earlier versions of the Read Codes was restricted by the hardware that a typical end-user owned. Thus a simple hierarchy was constructed using the Read Code both to identify a concept and to represent the hierarchy (removing the last digit of a child left the code of its parent).

As a result of this relatively simple structure, the codes were forced to adopt an uncomplicated perspective of medicine. As central returns were required in ICD-9 and OPCS-4, these were included wholesale in Read Version 2. The clinical view and clinical terms sometimes had to take second place in this view of the world.

More recently, the demands placed on the codes have become heavier and more difficult to satisfy. Specialists have become more interested in using the codes and have found that OPCS-4 and less frequently, ICD terms do not capture aspects of care they are interested in and may accent other aspects deemed of little worth. At the same time ICD-9 was replaced by ICD-10 which abandons many older terms, introduces new terms and restructures both old and new terms in a new hierarchy (notably in mental health).

The structure of the 5-Byte Set of the Read Codes does not allow a flexible approach to these problems. This is because:

- A concept cannot be placed in two or more parts of a hierarchy without being given two (or more) Read Codes. If this is done, then retrieval of concepts from a clinical database becomes more difficult.
- A concept cannot be moved to a different position in the hierarchy without changing its code, but changing a code makes it difficult to perform searches of data over a period of time.
- Complex concepts (e.g. *Open reduction and fixation of fracture of the neck of femur using a locked, reamed, intramedullary nail*) cannot be captured without a huge increase in the number of terms to provide for every combination of the separate characteristics included in such terms.

Changing any coding scheme is a traumatic experience for system developers and users who have invested time, energy and financial resources in building and using systems based on the old framework. The CTV3 file structure takes as a starting point the requirement to produce a coding framework that allows a natural evolution to cope with the new perspectives, detail and functions demanded of the codes.

3. Summary of Clinical Terms Version 3

- The Read Code in CTV3 serves only as a unique identifier for each concept and does not identify a concept's position in the hierarchy (i.e. it is meaningless in portraying its position in the hierarchy)
- A unique Read Code represents each concept. In circumstances, inherited from an earlier Read Codes Version (4-Byte Set and 5-Byte Set), where a concept was previously associated with more than one code they are represented by just one (persistent code). The redundant code(s) are not re-used, but are mapped onto the remaining persistent code.
- The display hierarchy is represented by a database of links, each between one parent and one child concept. This representation theoretically permits any number of levels in the hierarchy. (Other types of hierarchy can be represented with this mechanism by introducing a new file for each type of hierarchy)
- Concepts are arranged according to the principle of "is_a" or sub-type hierarchy, in which a "child concept" is always a "type of" its "parent concept"
- Clinical concepts may be assigned more than one parent concept in any hierarchy. For example *Pulmonary tuberculosis* is placed both

under *Infective disorder* and *Respiratory disorder* (the principle of multiple classification)

- Useful clinical concepts are marked as “current”. Legacy classification concepts incorporated from earlier versions of the Read Codes such as those containing NOS (Not Otherwise Specified) terms are marked as “optional”. Where the meaning of earlier version concepts are ambiguous, unclear, or complex they are marked as “extinct” allowing them to be placed in their “best fit” in the CTV3 hierarchy for analysis but excluding them being chosen in the future
- During usage where no specific child concept exists (and the user of an earlier version would previously have selected a NOS term), the user of the codes is now advised to select the more general parent code. Therefore a user who does not wish or need to specify whether a patient had *Extrinsic asthma* or *Intrinsic asthma* would choose simply *Asthma*, rather than *Asthma NOS*
- The CTV3 file structure is rationalised and normalised by comparison with the earlier versions to clarify the meaning of terms and to reduce redundant duplication of information. As a result, the information is distributed in many more files than in the 4-Byte Set or the 5-Byte Set Version 2
- When cross-mapping to other coding schemes (such as ICD-10 and OPCS-4) a default mapping is provided as in Version 2 of the Read Codes. However, this system is now augmented by alternative mappings, where applicable, which incorporate some of the classification ‘rules’ in these systems. This is designed to aid the process of cross-mapping by providing a consistent and comprehensive choice of applicable alternative codes based upon the index and rules of the classification. This aspect of CTV3 is discussed separately in the document titled “Clinical Terms Version 3 – Cross Mapping File”
- The CTV3 file structure includes a mechanism for adding additional modifying detail by providing a list of potential qualifying terms using a “template file”. For example, the most detailed available applicable ‘core concept’ might be *Cholecystectomy* and further detail such as whether the approach was *Endoscopic* or *Open* can be added using a qualifier
- The template file is also used as a common formal mechanism for the expression of the intrinsic characteristics (atoms) of a concept e.g. allowing the definition of *Cholecystectomy* as being an operation performed on the site: Gallbladder. This common template mechanism of expressing both the “atoms” of concepts and qualified detailed allows the detection of equivalence between a complex concept created by different constructs

- The template file is described in the document “Clinical Terms Version 3 – Template File”. It includes a system for determining which qualifiers may be used with which ‘core terms’, partly to make selection of qualifiers easy, partly to allow detail stored in qualifiers to be shared and communicated between users with different information systems and in order to support information retrieval and detection of equivalence

The rest of this document concentrates on the main CTV3 file structure and excludes discussion of the template file and of cross-mapping. Both of these are major topics that are dealt with in other documents.

4. Principles behind the CTV3 File Structure

The platform for the CTV3 file structure is a set of principles that both clarifies the activity of identifying terms and concepts and solves many of the problems of the previous representation schemes.

i) **Any one clinical concept may be described by several terms**

An important feature of the Read Codes is that they represent clinical **concepts**, to which a variety of labels (or **terms**) may be attached. For example, the terms *Furuncle* and *Boil* each refer to the same thing or concept. These labels therefore are the terms we use for describing objects, states and processes (concepts) in the world

ii) **The meaning of a concept is given by its preferred term**

In an ideal world, the meaning of a concept (labelled by a Read Code) would be given by a definition. In the absence of any definitions, we need a single unambiguous term to do the job. Such a **preferred term** (more accurately the term in the preferred description for a concept) must therefore have only a single unambiguous meaning and cannot be a synonym for another term

If the possibility of ambiguity arises with a suggested preferred term, then this is annotated to ensure only one sense can be read into the term. *Cervical pain* is a synonym both of *Cervical neck pain* and *Uterine cervix pain*, the latter being an invented term to clarify the second meaning

The chapters containing values (e.g. *Severe*, *Skin of thigh*) which are only intended to be used as qualifiers (e.g. Skin graft: Site: *Skin of thigh*) may contain ambiguity. For example, *Skin of thigh* may refer to the site of placement of the graft or the site of harvesting. However, ambiguity is resolved by the attribute (e.g. Site of harvest: *Skin of thigh*). Users are strongly advised not to use values other than in the

way specified in the template file (see the document “Clinical Terms Version 3 – Template File”)

iii) A synonym may have several meanings

When any one term has several meanings or senses (in dictionary terminology), it refers to several concepts. For example, the term *Cervical pain* is ambiguous as it may refer to pain in the cervical region of the neck or the cervical region of the uterus

The combination of a concept (Read Code) and a term is therefore needed to provide an unambiguous way of talking about the world in a health carer's own vocabulary and is called a **description**

iv) Each code represents a single clinical concept

Each Read Code corresponds to one concept. Whenever clinically significant variants are discovered, these are assigned to a different Read Code

v) Each concept distinguished is represented by a unique identifier

In CTV3, the Read Code is a unique identifier for a concept. The CTV3 file structure allows a single code to be placed in more than one place within the hierarchy. (cf. in the 5-Byte Set where concepts had to be placed in more than one part of the hierarchy, they were assigned more than one code)

Where more than one code has previously been assigned to a concept in an earlier version of the Read Codes, only one of these is actively used in the releases accompanying CTV3. The other codes are not removed, but marked as redundant and mapped onto their current Read Code in the **Redundant codes mapping file** (a bonus file described in the Technical Annex to this chapter. Section T.2.6). These redundant codes should not be used in future medical databases. This is described more fully in the document “Clinical Terms Version 3 – Incorporation of Earlier Versions of the Read Codes (The Superset)”.

Importantly this principle does not state that every complex medical concept should be represented by a single code or identifier as there are simply too many of these (e.g. *Severe epigastric pain secondary to a duodenal ulcer by naproxen*). Where such complexity or detail is required, the approach is (1) to rely on qualifying the basic Read Code with other descriptors (see the document “Clinical Terms Version 3 – Template File”) and (2) to split the concept into two or more simpler concepts, e.g. A disease and its symptoms such as *Epigastric pain* – Severity: *Severe*; Associated finding: *Duodenal ulcer*

vi) A concept's code will not change over time

Once G30.. is associated with the concept of an *Acute myocardial infarction* it will remain so, however the term is classified. If new research showed myocardial infarcts to be an auto-immune disease then the disease would be positioned elsewhere in the hierarchy, but retain the same identifying code.

vii) Each concept is understandable without reference to parent or child concepts

There may be many ways of placing a concept in a hierarchy of other concepts and it is dangerous to rely on its parent and child relationships within the structure to provide meaning or definition because any change in the way it is classified potentially alters its meaning. The meaning of a description should therefore be 'context-free'

Classification concepts may break this rule. For example, 'Otherwise Specified' and 'Not Otherwise Specified' concepts (e.g. *Asthma NOS*). The only way to understand what such a term means is to look at what has been specified (often its siblings in the hierarchy). These terms will therefore be marked as optional (not current). These classification concepts are not a design feature of CTV3 per se, but are incorporated from the earlier versions of the Read Codes to ensure forward compatibility. See the document "Clinical Terms Version 3 – Incorporation of Earlier Versions of the Read Codes (The Superset)"

viii) Concepts are arranged according to the principle of being sub-types

Concepts are arranged according to the principle of "is_a" or a sub-type hierarchy, in which a "child concept" is always a "type of" its "parent concept". For example *Tuberculous meningitis* is placed as a child of *Bacterial meningitis* (NB it is also placed as a child of *Tuberculosis*).

ix) No attempt is being made to disallow recognised clinical terms from the thesaurus

Although ICD-10 prefers to classify diabetics as insulin-dependent or non insulin-dependent rather than of juvenile or adult-onset, the latter two terms are not disallowed from the thesaurus as it is not the role of the thesaurus to prescribe what people should or should not record. However, every attempt has been made to place such concepts in the correct place within the hierarchy in relation to others according to the rule maintaining a sub-type hierarchy. Similarly, some health carers prefer to use *Boil* where others prefer *Furuncle*. Clinicians should have access to any recognised clinical term for describing health care.

Where adherence to standards is **required**, then it is up to local policy or software to enforce or encourage this

x) The means for classifying a concept is made independent of the means for identifying it

Dissociating identification and classification of concepts has the major advantage of allowing restructuring of concepts without changing their identifying codes. It also allows for the possibility of multiple (concurrent) ways of classifying concepts. *Pulmonary tuberculosis* can therefore be classed both as an *Infective disorder* and a *Respiratory disorder*. This kind of retrieval structure fits clinical analyses well. A clinician may wish to ask to recover all patients with a *Respiratory disorder* who are on a *Beta-blocker* or all patients with an *Infective disorder* who are on an *Antimicrobial drug* and would expect to include patients with *Pulmonary tuberculosis* in both groups

xi) Linguistic variations of concepts are represented

Any one concept may be represented by different synonyms. Linguistic variations include:

True synonyms

Furuncles and Boils; *Tinea pedis* and *Ringworm of the feet*. Synonyms are attached to the same Read Code (unique identifiers). One term of a set of synonyms, along with a Read Code, is marked as the **preferred term** (the term in the preferred description). The others are marked as alternative or synonymous descriptions

Linguistic variants

Chest pain and *Pain in the chest* are linguistic variants of the same concepts. Some of these are included though it seems unnecessary and is probably impossible, to represent every possible linguistic variant. Linguistic variants, like true synonyms, are attached to the same Read Code

Abbreviations

MI is an abbreviation for the terms *Myocardial infarct* and *Mitral incompetence*. In order to allow access to terms using common abbreviations the synonyms for a concept include both a form with the abbreviation (*MI – Myocardial infarct*) and without the abbreviation (*Myocardial infarct*). This mechanism thus forces a user to clarify what concept the abbreviation they have used relates to, thus avoiding ambiguity.

Natural synonyms

A natural term may be in common use by one specialty that is ambiguous in other contexts. For example *Cord compression* is commonly used by some specialty areas to refer to *Spinal cord compression* and to express *Umbilical cord compression* by others. The CTV3 file structure accommodates this circumstance by allowing a single term (in the case *Cord compression*) to be attached to more than one concept, each with their own unambiguous preferred term. The occurrence of a synonym in common use for more than one discrete concept is described as a “natural synonym”

Detailed variants

In some rare circumstances some synonyms exist, which are strictly not of the same meaning as described by the preferred term and are detailed variants of that concept. These in previous versions of the Read Codes have sometimes occurred in relation to subtly different techniques in performing a procedure that were no thought to be of sufficient significance to merit a separate concept (e.g. Variants of total hip replacements using different named prostheses). These “detailed variant synonyms” (formally known as hypernyms) have largely been given new concepts in CTV3, although some residual detailed variants remain

xii) A Read concept may have multiple ancestors in the display tree

There is no reason to assume that a concept such as *Pulmonary tuberculosis* should be assigned to only one position in the hierarchy (such as *Respiratory disorder*) and in CTV3, the concept may be placed at several points (*Pulmonary tuberculosis* is assigned to both *Respiratory disorder* and *Infective disorder*). Information describing the hierarchy structure is held separately from the unique identifier (Read Code) in the **hierarchy file**

xiii) There may be an unlimited number of levels in the hierarchy

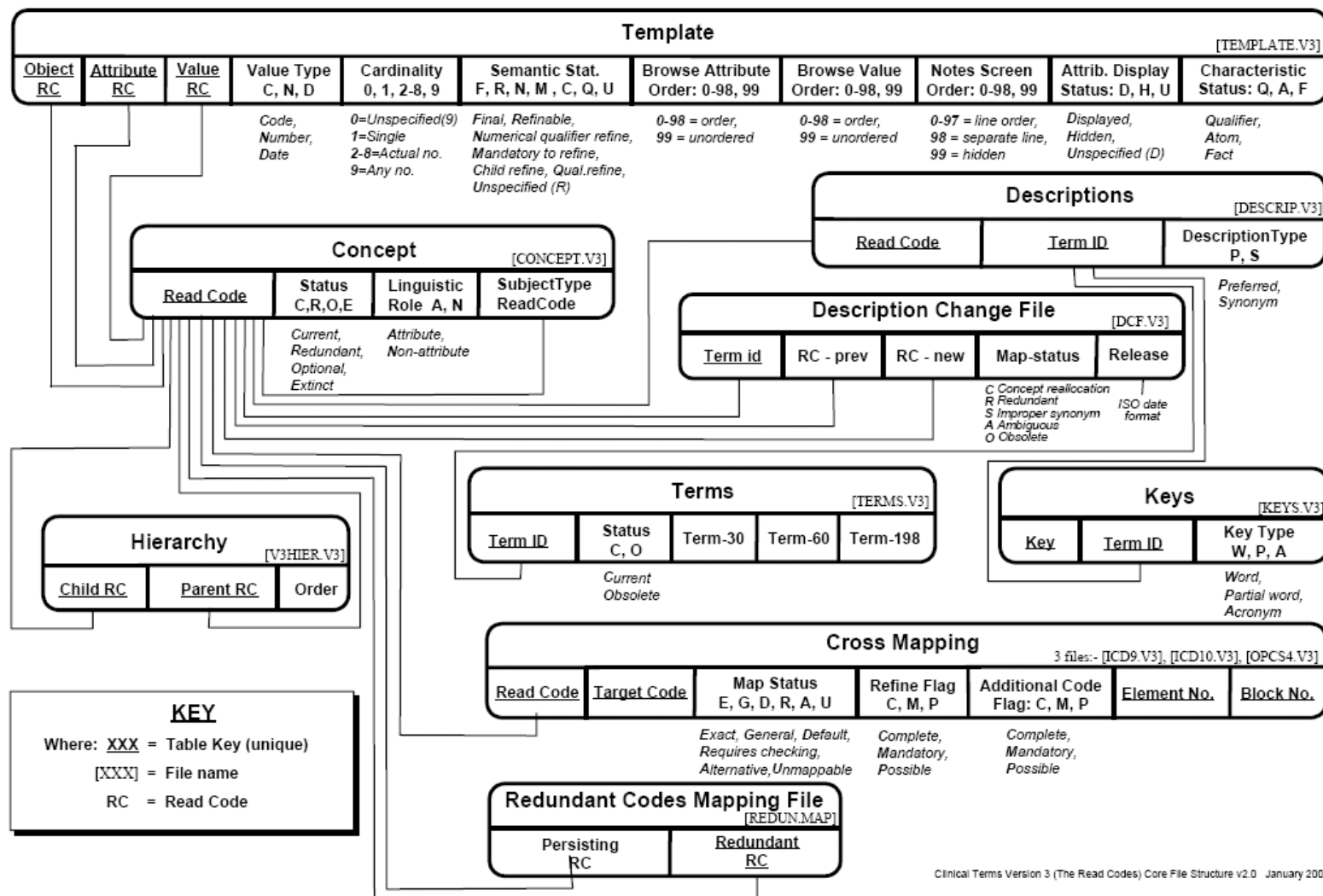
The CTV3 hierarchy is represented using a hierarchy table that has a single entry for each “parent-child” relationship. This scheme therefore theoretically can support an unlimited number of levels in the hierarchy. This structure allows concepts to be placed in their optimum position without the restraint of code length enforced upon them as in earlier schemes

5. Overview of the File Structure

Full details of the file structure are given in the Technical Annex of the document. However, this section describes how the principles are achieved in the file structure. In summary, the files are as follows:

File	Brief description
Term file*	Contains the 30, 60 and 198 character versions of each term
Concept file*	Contains information about Read Codes such as whether the code is current, optional, redundant or extinct
Descriptions file*	Links the terms to the Read Codes and defines whether the term is the preferred term or a synonym
Hierarchy file*	Contains links between parents and children that make up the hierarchy
Key file	Contains 10 character keys and links these to terms
Speciality files	One for each specialty, these contain codes that the specialty is particularly interested in
Cross-mapping files*	There is one file for each classification or coding system that is being mapped to. Described in a separate document "Clinical Terms Version 3 – Cross Mapping File"
Template file*	The mechanism by which "core concepts" are defined (semantic definitions or atoms) and associated with appropriate potential qualifiers and facts. Described in a separate document "Clinical Terms Version 3 – Template File"
Redundant codes mapping file*	Contains the mapping between redundant Read Codes and the persistent equivalent (Read Code) concept
Description change File	Contains the mapping of terms and concepts whose statuses have changed between releases. Described in a separate document "Clinical Terms Version 3 – Managing Change: Description Change File"

= Illustrated in the following diagram

Clinical Terms Version 3 (The Read Codes) Core File Structure

5.1 Describing Terms (Term file)

This file contains a list of all of the terms. The longest acceptable term is 198 characters long, but all such terms where present are also abbreviated so that 30 and 60 character versions are available for systems that can only handle these shorter forms. In the example below, *Tuberculosis of meninges and central nervous system* is greater than 30 characters long, so is placed in the 60 character field, but also shortened to a 30 character form. Each term is given an identifier (term_id) which is used whenever referring to this term elsewhere in the file structure, or in clinical information systems. Each term is also marked as current (C) or obsolete (O)¹. Note that there is no mention of Read Codes in this file (though term identifiers look like a Read Code).

term_id	term_status	term_30	term_60	term_198
Y71DW†	C	<i>Infective disorder</i>		
Y1001	C	<i>Respiratory disorder</i>		
Y71HU	C	<i>Tuberculosis</i>		
Y1040	C	<i>Pulmonary tuberculosis</i>		
Y71lg	C	<i>TB of meninges and CNS</i>	<i>Tuberculosis of meninges and central nervous system</i>	
Y11xs	C	<i>Spinal cord compression</i>		
Y40xs	C	<i>Cord compression</i>		
YaaGm	C	<i>Umbilical cord compression</i>		

† NB the Term-id is case sensitive

¹ Whilst terms may be current or obsolete, concepts may be current, optional, redundant or extinct.

While term shortening had advantages, users and developers need to be aware of its limitations. The 30 character abbreviations can be so short as to mislead or confuse. The following are two worst case examples

Term – 30	Full term
ESSS – Ep with con s/w in s/w s	ESSS – Epilepsy with continuous spike/wave during slow wave sleep
Infect. After inject./infusion	Infection after injection/infusion/trans-Fusion/vaccination

Developers using 30 character abbreviations are advised to present users with the full clinical term to confirm the true meaning of the concept. It is recommended that the full clinical term be offered to the user, along with the Read Code and the Term ID as stored in the clinical record.

5.2 Linking Terms to Read Codes (Description file)

Two important principles discussed earlier in the document were the requirements for a clinical concept to be referred to by several synonyms, (and in some circumstances for any one clinical term to have several senses, or to label several concepts).

These are achieved in the descriptions files, an example extract of which is shown below, in which terms are linked to their Read Codes. For example, the term for *Spinal cord compression* (its term identifier, as described above, is Ya1XS) is linked to its Read Code XaONk. As this is the preferred term for this Read Code, the description type is P.

Synonyms are flagged with a description type of S, as can be seen from the row in which *Cord compression* (term_id Y40xj) is also linked to the Read Code X40Cc.

read_code	term_id	desc-type
A....	Y71DW	P
H....	Y1001	P
A1...	Y71HU	P
A11..	Y1040	P
A13..	Y71lg	P
Xa0Nk	Ya1XS	P
X40Cc	Y40xj	S
Xa0Nk	Y40xj	S
X40Cc	YaaGm	P

5.3 Describing Concepts (Concepts file)

This file holds information about the concepts (Read Codes), but makes no mention of the terms. It includes information about whether the code is designated for **current use** (concept_status is C); *Tuberculosis*, which has the Read Code A1..., remains a current concept. Alternatively the Read Code may be **redundant** (concept_status is R), when there is a (persistent) code that describe exactly the same concept, or **optional** (concept_status is O), in the case where the concept should be avoided unless no other term can be found. The latter category includes classification concepts such as NOS concepts derived the earlier versions of the Read Codes (e.g. *Asthmas NOS* which has the Read Code XE0YX) as these codes have a meaning which is not clinically useful or consistent and are being abandoned. Redundant concepts are mapped onto the single persistent concept in the Redundant codes mapping file (see Section &.2.6).

A fourth concept status (E) or **Extinct concepts** identify depreciated concepts mostly from earlier versions of the Read Codes whose nature is inherently complex and ambiguous. These concepts should never be offered to or chosen by the user and have been placed in the hierarchy in their optimal position purely to allow retrospective analysis of mixed version analysis to be attempted. More suitable concepts exist elsewhere in CTV3. A new file, the description change file, contains mappings from descriptions containing extinct Read Codes onto current descriptions. This file is described in the document "Clinical Terms Version 3 – Managing Change: Description Change File".

The **linguistic role** that each concept plays is discussed in the document "Clinical Terms Version 3 – Template File". The majority of Read concepts have the linguistic role N (standing for non-attribute). A small minority have a linguistic role A (standing for atttribute) which are a particular class of Read Codes for linking together (non-attribute) concepts.

The **subject type** of a Read concept is a field that describes whether concepts are drugs, procedures, diseases, findings, etc. Each concept may belong only to one of these types – it is either a procedure or a disease, for example. The type is represented by a Read Code. Subject types are only used for programming purposes (for instance in determining whether or not qualifiers may be 'inherited' down the hierarchy: see the document "Clinical Terms Version 3 – Template File").

read code	concept_status	linguistic role	subject_type
A1...	C	N	Clinical State
H33zz	O	N	Clinical State

System developers will also find **medical types** useful as these delineate 'chapters' of codes. A mapping from the first level Read Code onto their Medical types is released in a separate bonus file *contype.v3*. Medical types are described in more detail in Appendix B.

5.4 Describing a Hierarchy

Read concepts are placed in a hierarchy, so that a health carer may select a more general (e.g. *Asthma*) or a more specific (e.g. *Intrinsic asthma*, *Extrinsic asthma*) concept according to need. The hierarchical arrangement also allows searches in clinical records systems for groups of codes (e.g. all respiratory diseases, all asthma's), no matter at which level of detail these are specified.

To permit any Read concept to have not only several children, but also several parents, hierarchies are now represented as parent-child links in a database. Take the following excerpts from a hierarchy, which is a fabricated exemplar (for simplicity):

```

Infective disorder (A....)
    Tuberculosis (A1...)
        Pulmonary tuberculosis (A11..)
        Tuberculous meningitis (A13...)
Respiratory disorder (H...)
    Pulmonary tuberculosis (A11..)
  
```

This would be represented in the hierarchy file as follows:

read_code*	parent_read_code	list_order
A1...	A....	00
A11..	A1...	00
A13..	A1...	01
A11	H....	00

* relates to the child Read Code

Note that *Pulmonary tuberculosis* (A11..) has two parents, *Tuberculosis* (A1...) and *Respiratory disorder* (H....). This illustrates how the Read Code in CTV3 itself no longer represents the hierarchy (that is the code A1234 does not necessarily have a parent A123. as was the case in Version 2).

In addition, to build a second quite different hierarchy we only need to substitute a different database of parent-child links.

The **list order field** (third column in the previous table) shows the order in which children should appear on screen when a hierarchy is viewed. The first item that should appear is labelled 00, the second 01, etc. In CTV3, by using the **order field**, each Read coded concept has an explicit numbering with

respect to one of its parents. Where a concept has more than one parent, it sits in more than one sequence and has an ordering number in each sequence. These ordering numbers can be changed at will in the master database to generate a new ordering.

The above mechanism is an improvement on that used in the earlier versions of the Read Codes where the Read Code determined the sequence in which items appeared on the screen. This mechanism was used to ensure that the children of a category such as smoking consumption (*Never smoked, Smokes 1-9 a day, Smokes 10-19 a day, Smokes 20-39 a day, etc.*) appeared in a reasonable order. However, if a new Read Code was introduced at a later date for an intermediate category, the mechanism failed.

5.5 Accessing Terms and Concepts (Key File)

Users traditionally access clinical terms either by browsing down the hierarchy or by using a key to jump straight to a relevant term, whose children may then be browsed. To improve the thoroughness of keys, these are produced by a computerised algorithm which will key on the first 10 letters of every word in a term (except excluded words such as of, and, any, other).

Keys are linked to terms. This allows users to gain access to the terms they would naturally use in clinical dialogue. But a second reason for mapping terms, rather than to Read Codes, is that disorienting anomalies may appear if this is not done. For example, a user may type the string 'Hear' expecting to recover all Hearing disorders. If keys are mapped to terms, then typing 'Hear' retrieves *Hearing disorders* and *Heart attack*. If keys are only mapped to Read Codes, then the terms retrieved might be *Hearing disorders* and *Myocardial infarct* (which has the synonym *Heart attack*). It may be very unclear to the user why *Myocardial infarct* was retrieved.

Finally, keys are marked as whole words (W), partial words (P) or abbreviations (A) in the key type field. Certain keying behaviours can take advantage of these markings.

term-key	term_id	key-type
TUBERCULOS	Y1040	P
TUBERCULOS	Y71HU	P
INFARCT	Ya0vx	W
CORONARY	Ya0vv	W
MI	Ya0vw	A

5.6 Describing a Speciality's Interest in Clinical Concepts (Specialty Files)

Speciality files offer clinical groups the option of indicating a subset of the Read Codes in which they are particularly interested. These declarations open

up the possibility of restricting picking lists of terms to those of immediate interest, to make term selection easier. This mechanism is explained in more detail in the document “Clinical Terms Version 3 – General Practice Flagged Subset”.

5.7 Cross-Mapping Files

These are discussed in a separate document, titled “Clinical Terms Version 3 – Cross Mapping File”.

Technical Annex

In this section, a simplified version of the conceptual model underlying the CTV3 file structure is described and then related to the file structure for distribution of the codes. The file structure is described field-by-field after which the proposed changes are analysed.

T.1 Conceptual Model of the Read Codes

The CTV3 file layouts reflect much more closely the entities and relationships in the conceptual model of the Read Codes than does Version 2. The CTV3 structure attempts to reduce redundant data to a minimum in order to clarify the meanings of the data elements. System developers may optimise the file structure to give the required functionality and performance.

T.1.1 Entities

The major entities are described below.

Concept

A concept is some kind of action, process, state, event or object in the world that is relevant to medicine. Concepts (other than dates and numbers) are represented by Read Codes.

Term

Terms are the rubrics or linguistic labels (e.g. *Myocardial infarct*, *Heart attack*) used in clinical dialogue when referring to concepts.

Description

The combination of one concept and one term identifier is a description. It is equivalent to one ‘sense’ in a typical (e.g. English) dictionary entry. Preferred descriptions include the preferred term for representing a concept, while synonymous descriptions include other (synonym) terms for a concept.

Hierarchical relationship

Hierarchies represent the relationships between concepts and in CTV3 this is based upon the principle of “is_a” or sub-type. As each concept in a hierarchy may have several children and each child several parents, these should more strictly be called directed acyclic graphs – see the later discussion on the hierarchy files in Section T.2.5 for clarification of this. However, in this document we will persist in calling such graphs – hierarchies.

Key

A string that relates to a clinical term. It is usually all or part of a word in the term to which it relates.

Speciality flag

The interest a group of health carers declares in a concept. Initially the speciality flag was used to distinguish the interests of different hospital specialities, though other care groups such as health visitors, nurses and general practitioners may also be regarded as specialties.

Cross reference

An entry in another coding scheme such as ICD-10 or OPCS-4.

T.1.2. Entity Model

The diagram below describes the relationship between these concepts. This is in simplified form.

Entities are formally described in the text above. But the relationships (arcs or interconnecting lines) in the entity diagram are not labelled or explained formally.

Boxes are entities. Boxes with a second line at the left hand edge are subtypes of entities. There are therefore two types of description, preferred descriptions and synonymous descriptions.

The cardinality of the relationships (one-to-one, many-to-one, etc.) is given by interconnecting lines and ‘crows feet’. Thus one **term** can take part in just one **preferred description** (one-to-one), but in several **synonymous descriptions** (one-to-many). The entities ‘term’ and ‘concept’ are highlighted simply for clarity.

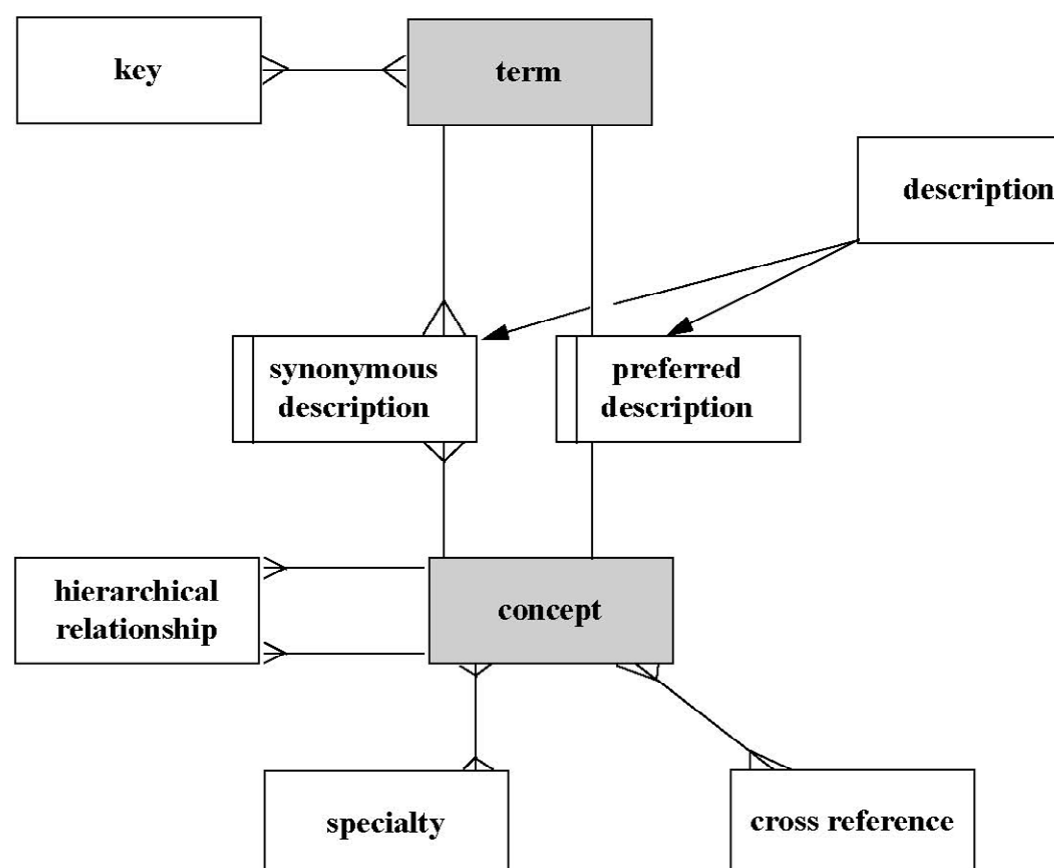


Figure 1: Entity model

T.1.3 Rules

The rules constraining relationships between entities are listed here for completeness, but are discussed in the commentary on each file layout.

Each concept must be in the hierarchy (unless it is redundant), either as a child or as a parent, or both. No concept may be its own parent or child.

A parent concept may have zero, one or several children; and a child concept must have one or more parents (unless it is the root node 'Read Thesaurus' denoted by the Read Code '.....' which has no parents).

All concepts have (are related to) one and only one preferred description (unless it is extinct where it may have both multiple preferred terms and synonyms attached). A concept may have zero, one or more alternative (synonymous) descriptions.

A term may only relate to one preferred description (except where it is attached to extinct concepts).

All current terms have (are related to) at least one description (preferred or alternative) and may relate to several. Obsolete terms do not relate to any Read coded concepts (and thus do not appear in the description file).

Extinct concepts may have synonymous descriptions in which the term also relates to a preferred description of a current, optional or rarely another extinct concept.

In all other cases, a term belonging to a preferred description may not also belong to a synonymous description involving a different concept (Read Code).

A term may have zero or more keys.

Each key must relate to one term and may relate to several.

Each concept may be flagged only once by any one specialty (to denote its relevance to that specialty). A concept does not necessarily have a specialty flagging.

Each concept may relate to several codes in a target coding scheme. But the great majority of cross-mappings (ca 70%) have a default one-to-one mapping. Any one code in another coding scheme may point to several Read coded concepts. A Read coded concept does not necessarily have to have a cross reference.

Redundant codes so not have descriptions, but all have an entry in the Redundant codes mapping file indicating its persistent code of same meaning.

T.2 Description of the File Layouts in CTV3

There is a close relationship between each entity in the conceptual model of the terminology and files (or tables) in the database to be distributed. There are single concept, term and key files.

However, all descriptions (preferred and synonymous) are included in a single descriptions file. A flag in the descriptions file indicates which descriptions are preferred and which are synonymous.

Finally, there is one file for the hierarchy, one file for each terminology cross-referenced by the Read Codes and potentially, one file for each specialty. The file layout is drawn in Figure 2, which is deliberately drawn to reflect the relationships in the entity-relationship model.

In the diagram, the link between the files is shown by the label on the arc connecting the files. Fields labelled in small point size next to each box form unique keys (in the normal computer sense of the word 'key' rather than the Read sense) for each record.

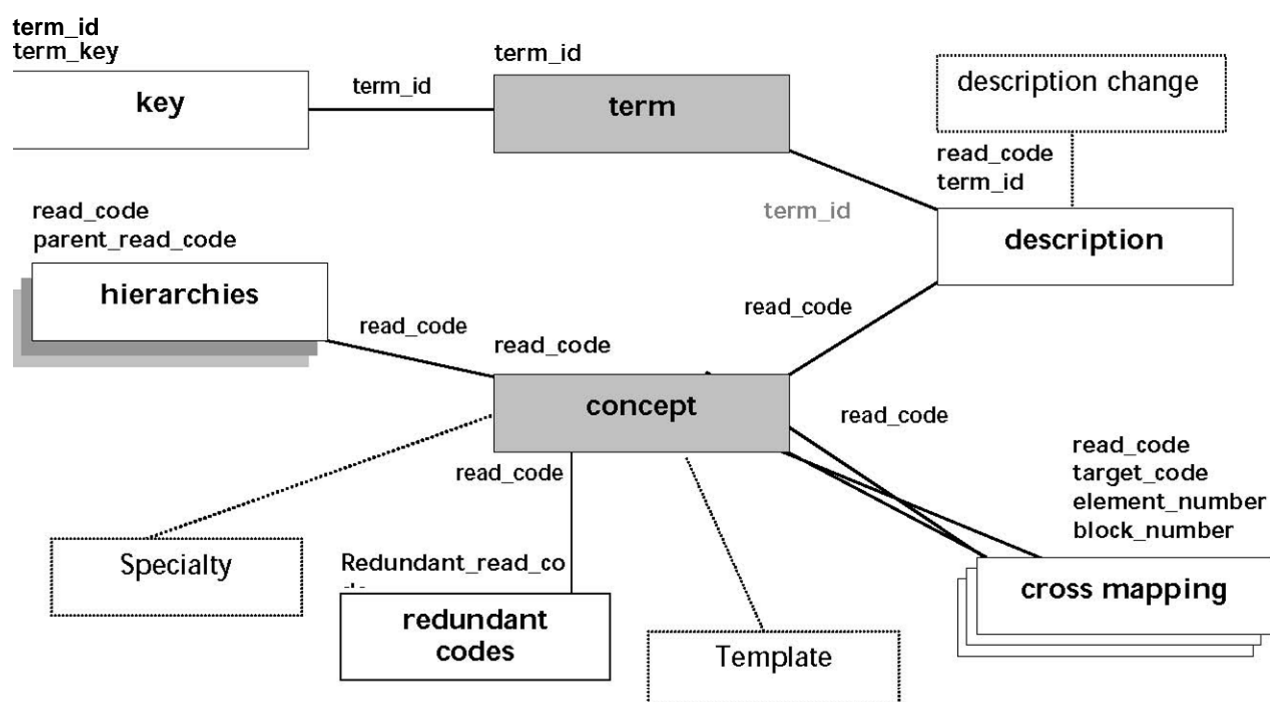


Figure 2: A diagram showing the various files and their relationships

N.B. Those files outlined in broken lines are detailed elsewhere

In the following sections, the files are outlined individually in the following layouts:

char(n)	is a fixed length string n characters long
varchar(n)	is a variable length string with a maximum length of n
M	means that the field is mandatory i.e. will always contain a value (or not null in SQL terminology). It does not mean that you must use the field.
underlined fields	denote the unique key for a table. Some tables have concatenated keys

T.2.1 Concept File

Each clinical concept is uniquely identified by a Read Code. The meaning of each concept is given by the term in the preferred description. A concept's meaning will remain fixed in so far as this is possible and concepts will never be deleted from the terminology, so historical interpretation of a clinical record is not compromised.

This file contains one and only one record for each clinical concept including those that have been discontinued for some reason. This may be because in 5-Byte Set, several codes were assigned to once concept in some instances (e.g. for *Tuberculous meningitis*). Where this has occurred, all but one of these codes will cease to be used (become **redundant**) and are mapped onto a single persistent code (using the Redundant code mapping file). Some concepts, often coming from earlier versions of the Read Codes, are marked as **extinct**. These are ambiguous and users should not use them when entering new data. A concept may become **optional**, in which case users are advised to avoid them unless no suitable current concept can be found. Note that the Read Codes are concerned only with defining term and concept usage and not in knowledge-based issues such as whether or not drugs are still in use or are to be issued at some future date. Such knowledge is found in other files, such as the drug dictionary.

<u>read_code</u>	char(5)	M	The unique clinical concept identifier
concept-status	char(1)	M	Denotes the current applicability of the concept C= current R = redundant O = optional concept E = extinct
linguistic_role	char(1)	M	The linguistic role a concept may play. Attributes and non-attributes are not interchangeable. Attributes may only be used in the template file (although they are also arranged hierarchically and appear in the hierarchy file) A = attribute N = non-attribute
Subject-type	char(5)	M	Read Code for the category of a code being described. These include: Procedure type Clinical findings type Drug type Attribute type

T.2.2 Term File

Terms are the labels (or strings) used in clinical dialogue to refer to clinical concepts. One term may reference more than one concept, just as some words in natural language have more than one sense or meaning. Each term is uniquely identified by the term_id, and each record in the term file contains one term. Term records will never be deleted, though new term records may be added. The 30, 60 and 198 character rubrics may be amended if the form of the term is mistaken; these are limited to minor spelling corrections and any changes are individually quality assured by two independent authors. It is likely in the near future that no changes of any kind will be allowed. Users are advised that 30 char rubrics may be misleading because of extreme abbreviation.

Developers using 30 character abbreviations for screen display purposes are advised to present users with the full clinical term to confirm the true meaning of the concept. It is recommended that the full clinical term be offered to the user, along with the Read Code and the Term ID as stored in the clinical record.

<u>term_id</u>	char(5)	M	The term_id, to identify a specific term that describes one or more clinical concepts
term_status	char(1)	M	Describes whether the term is: C = Current O = Obsolete
term-30	varchar(30)	M	The 30 character term
term_60	varchar(60)		The 60 character term
term-198	Varchar(198)		The 198 character term

T.2.3 Descriptions File

Each **current** term is associated with at least one description as all labels (terms) must be attached to some concept to become meaningful. As terms may have multiple senses, they may occur in several records in the description file (one occurrence for each sense of the term). Each concept may be associated with several terms, to reflect the various synonyms that describe it. Optional (and extinct) concepts will still be associated with at least one current term through the descriptions file.

Records in this table may be added, deleted, or changed without notice. The effect of this is that the table will not act as a historical record of the association between terms and concepts and so both Read Code and term identifier (as well as the longest term) must be stored in any computerised

document using the terminology. The Description Change File contains the changes that have been made to this file between each release (Described in the document “Clinical Terms Version 3 – Managing Change: Description Change File).

<u>read_code</u>	Char(5)	M	The unique identifier for a clinical concept
<u>term_id</u>	Char(5)	M	The unique term identifier
desc_type	Char(1)	M	The type of description. P = preferred description S= synonymous description

T.2.4 Key File

Keys are used for accessing terms. They are produced by a computerised algorithm that indexes on every word, other than those on a specified exclusion list. However, only the first 10 characters of each term (or less for shorter words) are stored as a key. Hand crafted changes may be made where necessary.

Any one key may access several terms and any one term may have several key words. Each record in the key file holds a relation between one key and one term, the latter represented by the term_id. (There are no “blank keys” as in Read Codes Version 2).

The linguistic class of the key is also recorded. Keys may be labelled as abbreviations or acronyms, or may be whole words (10 characters or less), or may be parts of words, or unspecified. (System designers may well not have a use for this field).

<u>term_key</u>	varchar(10)	M	The key word
<u>term_id</u>	char(5)	M	The term that the key refers to
key_type	char(1)	M	The linguistic class of the key: A = acronym or initial letter abbreviation W = whole word P = part word

T.2.5 Hierarchy Files

The Read Codes are not simply a list of concepts and terms, but are arranged according to their meaning (therefore creating a thesaurus) in a sub-type (is_a) hierarchy in which increasingly detailed or specialised concepts are placed at lower levels. The purpose of a hierarchy is to act as an access mechanism to the terms and concepts and as a basis for grouping concepts for searching in clinical records. (Although currently only sub-type hierarchy is released as other forms of relationships could be fashioned in the future).

Technically the hierarchy is really a directed acyclic graph. This is a graph (structure) in which a concept may have both several children and several parents, but cannot have itself as a parent. Other cycles or circular references (e.g. A is a parent of B, B is a parent of C, C is a parent of A) are also forbidden.

The hierarchies are held as lists of simple pairings between parent Read concepts and their children. If a parent concept has 6 children then it will be found in the parent field of 6 records in the hierarchy file. Likewise if a concept has 2 parents, it will be found in the (child) Read Code field of 2 records.

Hierarchy records may be added or deleted as clinical thinking changes over time.

A set of concepts often has a natural ordering within the codes (e.g. never smoked, smokes 1-9 cigs, smokes 10-19 cigs....). This is described in a separate ordering field. Note concepts are ordered with respect to a parent as this allows an ordering to be specified on each occurrence of a concept in the hierarchy.

read_code	Char(5)	M	The concept
parent_read_code	Char(5)	M	A parent concept of the concept
list_order	Char(2)	M	A numeric field stating the order in which concepts with the same parent are to be listed. A value of 99 indicates an unordered value (to be shown after any ordered values)

Current, optional and extinct codes may appear within the hierarchy file (but not redundant). Current codes can be a parent of any concept status; optional concepts may be a parent of only optional or extinct codes; extinct codes may only be parents of extinct codes.

T.2.6 Release Format

Some concepts particularly those derived from the earlier Read Code versions (4-Byte Set and the 5-Byte Set) have more than one Read Code. This occurs partly because of limitations of older file structures and partly because duplicates introduced into CTV3 cannot always be easily spotted. As CTV3 allows only a unique identifying Read Code for each concept, all other codes are flagged as redundant and are not included in other files, such as the hierarchy file. But users may already have stored one of these redundant codes in their clinical records systems and need to find these during a reporting or analysis procedure.

Two resources are available for mapping redundant codes onto current Read Codes. The Description Change file may be the preferred approach for many

uses. This is described in the document “Clinical Terms Version 3 – Mapping Change: Description Change File”.

The second resource is the Redundant codes mapping file, which is a simple list of redundant Read Codes and the persisting Read Codes they are mapped on to. The persisting codes can have any concept status other than redundant (e.g. Current, Optional or Extinct).

<code>persisting_read_code</code>	<code>char(5)</code>	<code>M</code>	The current CTV3 Read Code for a concept
<u><code>redundant_read_code</code></u>	<code>char(5)</code>	<code>M</code>	A Read code made redundant in CTV3

T.2.7 Release Format

In release form, the files are of variable length fields, delimited with the vertical bar character “|” (ASCII value 7C hex). The bar is not surrounded by spaces and is not an allowable character in any of the data fields.

T.3 Analysis of the Differences of CTV3 in comparison to Earlier Read Code Versions

The advantages of CTV3 are presented prior to discussing the ramifications of the file structure in detail.

T.3.1 Summary of the Advantages of CTV3

Two goals have been uppermost in the design process. Firstly to simplify the scheme so that it better reflects the principles and flexibility of a modern terminology and secondly to remove all implementation directed optimisations, normalising the tables to reduce unnecessary space overhead and to facilitate maintenance and quality assessment of the terminology. These changes should make it easier for developers to see the structure of the Read Codes more clearly, but it is assumed that system designers will introduce optimisations in keeping with their requirements and database management systems.

Although there are no changes in the philosophy behind the Read coding scheme itself, advantages accrue from the more flexible representation scheme used.

- Each concept is represented by a unique identifier. The single identifier makes searches easier

- There is no theoretical limit to the number of levels in the hierarchy. Previously the number of levels was linked to the number of characters in the Read Code (4 levels for 4-Byte Set and 5 levels for the 5-Byte Set)
- There is no theoretical limit to the number of parents a concept may have in the display hierarchy (multiple classification is supported)
- Concepts may now be placed more accurately by authors and found more easily by users, as those concepts naturally falling into two or more groups (*Pulmonary tuberculosis* is both an *Infective disorder* and a *Respiratory disorder*) may be sited at the applicable points in the hierarchy
- The order in which items are displayed on the screen may be adjusted as new concepts are inserted into the Read Codes. (In the 4-Byte Set and the 5-Byte Set this was impossible)
- Maintenance of the terminology is easier: changing the hierarchy to encompass the latest clinical thought (for example when introducing changes stemming from the revisions in ICD-10), without having to change Read Codes; concepts can be allocated to more than one category (parent concept); also terms can be reassigned to concepts more easily
- Redundancy is dealt with in a formal way
- Cross mapping to other classifications and coding schemes is more accurate (dealt with in the document "Clinical Terms Version 3 – Cross Mapping File")
- Additional detail can be added using a formal mechanism using qualifiers within a "template". This template is also used as a common mechanism to express the semantic definition of concepts allowing the detection of equivalence between different constructs expressing the same complex concept. It is dealt with in the document "Clinical Terms Version 3 – Template File")
- Changes in the terminology such as alteration of the descriptions and handled using a clear documented mechanism (dealt with in the document "Clinical Terms Version 3 – Managing Change: Description Change File")

T.3.2 on Read Codes

The format of Read Codes (5 alphanumeric characters) remains unchanged from Version 2. CTV3 is a superset of earlier versions. All Version 2 Read Codes and terms are found in CTV3. All 4 byte Read Codes are found in CTV3 (once a preceding dot has been added). All 4 byte terms are included in

CTV3. New Read Codes for CTV3 are unique, but have no hierarchical (meaningless) identifiers.

T.3.3 on Terms and Descriptions

The term code in Version 2 currently performs two functions. It serialises and identifies terms and describes them as preferred, synonymous, etc. In CTV3, these functions are separated to make maintenance, implementation and use easier.

A stand-alone term identifier is used in place of the old Version 2 term code (which defines terms within the compass of a single concept) to highlight those terms which have more than one sense or meaning. The relationship between terms and concepts is handled through the description file.

The term_id field contains an alphanumeric 5 character identifier. All of the read_code/term_code combinations in Version 2 are mapped to new read_code/term_id combinations (see the document “Clinical Terms Version 3 – Incorporation of Earlier Versions of the Read Codes (The Superset)”). This means that conversion from earlier version of the Read Codes to CTV3 is possible.

T.3.4 on Hierarchies

In the 4-Byte Set and the 5-Byte Set the hierarchy is represented as a parentage path, stored in the Read Code itself, whereas in CTV3 the hierarchy is stored as a sequence of parent-child links.

Note that a concept (e.g. *Pulmonary tuberculosis*) has only one set of children (e.g. *Tuberculous pneumonia*, *Tuberculous bronchiectasis*, *Tuberculous pneumothorax*). The members of this set do not vary depending on the parent (e.g. *Infective disorder*, *Respiratory disorder*) of that concept.

The rules for displaying terms when stepping up or down a hierarchy remain unchanged; the picking list should contain only preferred terms, then be sorted hierarchically and with that, in the order specified in the order field (also found in the hierarchy table).

T3.5 on the Clinical Record

Any clinical records system using the Read Codes will need to store both the read_code and the term_id, so uniquely identifying both the clinical concept and the term used by the clinician to describe it. This is necessary to capture the precise term used by the clinician and because terms may be reassigned to different Read Codes as their meaning becomes refined. For medico-legal purposes, users are also advised to store the longest term rubric.

T.3.6 Conversion from 4-Byte and Version 2 to CTV3

CTV3 is a superset of older versions of the Read Codes. It contains the terms and Read Codes of the 4-Byte Set and 5-Byte Sets. The 4-Byte set Read Codes are converted to 4-Byte format by adding a lead 'dot'. Information about the conversion files and the conversion process between older versions of the Read Codes and CTV3 is documented in "Clinical Terms Version 3 – Incorporation of Earlier Versions of the Read Codes (The Superset)".

T. 4 Other Available Documents

Several other documents, each covering a different aspect of Clinical Terms Version 3, are available. These include:

Accompanying Suite of Documents making up a Manual for Clinical Terms Version 3

Clinical Terms Version 3 – Introduction

Clinical Terms Version 3 – Contents

Clinical Terms Version 3 – Template File

Clinical Terms Version 3 – Cross Mapping File

Clinical Terms Version 3 – Incorporation of Earlier Versions of the Read Codes (The Superset)

Clinical Terms Version 3 – General Practice Flagged Subset

Clinical Terms Version 3 – Managing Change: Description Change File

Clinical Terms Version 3 – Drug and Appliance Dictionary

Clinical Terms Version 3 – Introduction to Information Retrieval

Clinical Terms Version 3 – Information Retrieval – Experiments

Clinical Terms Version 3 – Context and Clinical Records

Appendix A

A.1 Changes between the Proposed and the Final Structure

The file structure remains unchanged from that described in the official release document (Read Codes File Structure: Overview and Technical Description – November 1993). Many of the examples described in that document have been updated to better reflect the output from the Clinical Terms Projects.

Most of the detail of the file structure presented in this document reflects that described in the original proposal document (December 1992). Some changes were made when producing the final structure in November 1993 and the most important of these are described here. The changes were made after further experience was gained from experiments with the codes and after invited comment from academics, clinicians and developers.

A.2 Changes from File Structure of January 1995

A.3 Concept File

The concept status **obsolete** has been renamed **optional**.

A new concept status '**extinct**' has been introduced.

A.4 New Advice since March 1998

The advice for migrating from earlier version of the Read Codes (e.g. the 4-Byte Set and Versions 1 and 2 of the 5-Byte Set) has been changed. For details, refer to the document "Clinical Terms Version 3 – Incorporation of Earlier Versions of the Read Codes (The Superset)".

A.5 Change Control since October 1998

Detailed documentation has been released describing a mechanism for responding to changes in the descriptions file in Clinical Terms Version 3. This is documented in the document "Clinical Terms Version 3 – Managing Change: Description Change File".

Appendix B

B.1 Medical Types

Users and system developers made great use of the 'chapters' in earlier versions of the Read Codes. These corresponded to the top level of the tree in the hierarchy. The equivalent in CTV3 are nodes found either at the second or third level of hierarchy below the root node (which has the Read Code '.....' and is termed 'Read thesaurus'). Any one concept may fall into several of these chapters and may therefore have several medical types.

Medical types include:

History and observations

Disorders

Operations and procedures

Investigations

.... And others

One chapter contains concepts which all have the medical type 'value only'. It contains concepts that can only be used as qualifiers of core concepts including:

Organisms

Material

Anatomical site

Unit

.... And many others

Attributes are all placed in another chapter. All these have the medical type 'attributes'.

A bonus file available with the CTV3 release (contype.v3) lists the medical types of each concept and the category of each of these (core concept, value only, etc.).