



THE CLINICAL TERMS VERSION 3 (THE READ CODES)

INTRODUCTION AND OVERVIEW

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	166v1.0

© Crown Copyright 2008
Published by the UK Terminology Centre

Table of Contents

Table of Contents.....	3
1. The Purpose of this Document.....	4
2. The Read Codes.....	5
3. History of Read Code Versions.....	6
3.1 4-Byte Set.....	6
3.2 5-Byte Version 1.....	6
3.3 5-Byte Version 2.....	6
4. Clinical Terms Version 3 (The Read Codes) – Background.....	7
4.1 The Need for the Clinical Terms	7
4.1.1 Redundancy.....	7
4.1.2 Hierarchy Impurity.....	8
4.1.3 Synonym Impurity	8
4.2 The Clinical Terms Projects	9
4.3 Features of Clinical Terms Version 3.....	9
4.4 The Structure of CTV3.....	13

1. The Purpose of this Document

This document is intended to provide an introduction to the development, structure and function of Clinical Terms Version 3 (The Read Codes). Along with the accompanying documents listed below it can be utilised as a manual for potential users and developers of the Clinical Terms Version 3 (The Read Codes), referred to as CTV3 for brevity in these documents.

This introduction and overview contains brief notes on the salient features of CTV3 that are explained in more depth in the complementary documents. This account includes a brief discussion of previous versions of the Read Codes, the reasons for introducing CTV3 and its main advantages and features. A brief description of the contents of CTV3 is also provided.

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

Clinical Terms Version 3 – Contents

Clinical Terms Version 3 – Main File Structure: Overview and Technical Description

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

2. The Read Codes

The Read Codes are a set of coded terms for use in clinical practice. The codes are arranged in a hierarchical structure, so that an extensive clinical terminology can be readily accessed and used by computer software.

The Read Codes were initially developed in the mid-1980s, as the 4-Byte Set for use in General Practice. This version provides a terminology for describing relevant clinical summary and administrative data. It is known as the 4-Byte Set since each code is four characters long.

The codes were subsequently adapted for use in hospitals and were extended to allow more detail, leading to Version 2 of the Read Codes which was released in 1990. To accommodate extra detail, an additional alphanumeric character was added to the Read Code (5-Byte Sets – Version 1 and Version 2).

CTV3 was developed as a result of the Clinical Terms Projects, which ran from 1992 to 1995 and was first released in 1994. This version introduced a new, more complex structure, allowing greater coverage and flexibility. In particular, CTV3:

- Permits future changes in terminology to be evolutionary, by using a structure which can cope with new perspectives, details and functions demanded of the codes;
- Allows multiple perspectives of the same concepts (to be equally useful for nursing, general practitioners (GPs), central returns, specialists, etc.); and
- Produces more accurate and complete cross-mappings to other classifications

The features of CTV3 are described in greater detail in section 4 of this introduction.

The primary use of the Read Codes has remained that of underpinning a computerised medical record which can be used for viewing a patient's record from different perspectives (e.g. clinical audit, producing reports, meeting central returns, research, etc.). As medical science and practice advance, the structure of the Read Codes allows new terms and concepts to be incorporated. All versions of the clinical Read Codes are currently released biannually to allow updating; for example the inclusion of medical developments; and the incorporation of appropriate terminology to reflect new diseases, procedures and organisational changes.

Read drugs Version 2 is updated and released monthly. CTV3 drugs ceased to be maintained or updated in July 2006. A transform of Version 2 data into CTV3 format is carried out to satisfy the "Superset" principle.

3. History of Read Code Versions

3.1 4-Byte Set

The original version of the Read Codes was developed for GPs in the mid-1980s. In this version, Read Codes consist of only 4 alphanumeric characters. The files are much smaller than those in later versions and contain fewer codes. This version contains approximately 40,000 codes. Text descriptions consist of up to 30 characters. “Keys”¹, which may be entered by users to select a group of related clinical terms, are 4 characters long.

3.2 5-Byte Version 1

This version was developed to include specific functions for cross-references to central returns for hospitals, as well as providing functionality for GPs. Read Codes are extended to 5 alphanumeric characters, allowing a 5-level hierarchy. Text descriptions still consist of up to 30 characters, with keys of 4 characters long.

The files were originally cross-referenced to the classifications of the Office of Population Censuses and Surveys for procedures – 4th Revision (OPCS-4) and the International Statistical Classification of Diseases and Related Health Problems for diagnoses (ICD-9, and from 1995, ICD-10). Maps are now maintained to OPCS-4 and ICD-10 only.

3.3 5-Byte Version 2

The codes are identical to 5-byte Version 1, but text descriptions are extended to include 60- and 198-character versions and keys are extended to 10 characters. A new code (the term code) allows more than one textual description of a Read-coded concept to be labelled and a “preferred term” to be indicated.

The Version 2 Read Codes also support cross-mapping to mandated classifications, including ICD-9 (up to 1995), ICD-10 (from 1995 onwards) and OPCS-4. Mappings to ICD-10 and OPCS-4 are currently maintained.

¹ A ‘key’ consists of a string of characters of all or part of a word for searching a database

4. Clinical Terms Version 3 (The Read Codes) – Background

4.1 The Need for the Clinical Terms

The structures of the early Read Code versions, while being simple to implement, do present a number of problems. In particular, the rigid hierarchy structure leads to difficulties which can compromise the accuracy of concept placement because of limitation in depth and display.

The hierarchies in the earlier versions of the Read Codes are code-based, with the Read Code serving a dual function: the Read Code identifying a concept and also defining its hierarchy position.

Example:

- Infectious and parasitic diseases (A....)
 - Tuberculosis (A1..)
 - Tuberculosis of the meninges and CNS (A13..)
 - Tuberculous meningitis (A130.)

4.1.1 Redundancy

The code-based structure of the earlier versions of the Read Codes gives rise to **redundancy**. One concept can be placed in only one position of the hierarchy. For example, a problem arises with the disease **Tuberculous meningitis**, which can be positioned under *Infectious and parasitic diseases*:

- Infectious and parasitic diseases (A....)
 - Tuberculosis (A1...)
 - Tuberculosis of the meninges and CNS (A13..)
 - Tuberculous meningitis (A130.)

However, it should also be positioned under *Nervous system and sense organ diseases* (Read Code F....):

- Nervous system and sense organ diseases (F....)
 - Inflammatory diseases of the central nervous system (F0...)
 - Bacterial meningitis (F00..)
 - Meningitis – tuberculous (F004.)

This means that a single disease in these earlier versions (*Tuberculous meningitis*) has to be given two Read Codes (A130. and F004.)

This code redundancy can cause inaccuracies in hierarchy-based analysis data stored using the codes.

4.1.2 Hierarchy Impurity

A further problem is **hierarchy impurity**. The code-based structure allows only 5 levels of hierarchy (one level for each character of the Read Code). This means that a concept such as *Subcutaneous mastectomy* (Read Code 71304) is at the lowest possible level of the hierarchy.

```

Operations, procedures, sites (7....)
  Endocrine system and breast operations (71...)
    Breast operations (713..)
      Total mastectomy operations (7130.)
        Subcutaneous mastectomy (71304)
  
```

A more detailed variant of this operation, *Subcutaneous mastectomy for gynaecomastia*, cannot be placed in the appropriate position as there is no 6th level of the hierarchy. One solution used in the earlier versions is to give *subcutaneous mastectomy for gynaecomastia* a Read Code of 71307, placing it **alongside** *Subcutaneous mastectomy* in the 5th hierarchy level – as a “sibling”:

```

Operations, procedures, sites (7....)
  Endocrine system and breast operations (71...)
    Breast operations (713..)
      Total mastectomy operations (7130.)
        Subcutaneous mastectomy (71304)
        Subcutaneous mastectomy for gynaecomastia (71307)
  
```

This ploy results in sub-optimal placement of the concept as when retrieving cases of *Subcutaneous mastectomy* those recorded as *Subcutaneous mastectomy for gynaecomastia* will be missed.

4.1.3 Synonym Impurity

The restrictions of the code-based hierarchy of earlier versions of the Read Codes have also led to **synonym impurity**. For example an alternative to placing *Subcutaneous mastectomy for gynaecomastia* on the same level as *Subcutaneous mastectomy*, as a “sibling”, would be to create it as an “impure synonym” of *Subcutaneous mastectomy*.

```

71304 Subcutaneous mastectomy (Preferred)
      Subcutaneous mastectomy for gynaecomastia (Synonym)
  
```

This has the advantage of allowing all *Subcutaneous mastectomy* operations to be retrieved easily with a single code, but creates a complex impure concept as all cases represented by the code 71304 will not be a *Subcutaneous mastectomy for gynaecomastia*.

4.2 The Clinical Terms Projects

CTV3 was developed as a result of the Clinical Terms Projects. These were a joint undertaking between the clinical professions and the NHS Executive and were carried out using Projects in a Controlled Environment (PRINCE) project management methodology.

Input was obtained from the following speciality working groups:

- Medical specialties, including General Practice and Dentistry
- The nursing professions, including Health Visitors and Midwives
- Professions allied to medicine (PAMs), including Physiotherapists, Speech and Language Therapists, Podiatrists, Occupational Therapists and Dieticians

The needs of those professions were identified, 55 representative working groups set up to select terms which satisfied their requirements. Terms were incorporated from the earlier versions of the Read Codes and from other classifications (including ICD-10). Specialist detail was also gathered from clinical records and the total corpus of these terms were integrated into a common thesaurus of clinical terms – Clinical Terms Version 3.

The emphasis was on natural clinical terms, which could readily be understood by clinicians, rather than classification terms such as those used in, ICD-10 and OPCS-4. Additionally terms including “embedded contextual information” such as *Complaining of, Family History or, Normal/Abnormal* etc. were disallowed. Such information is to be managed using the qualifier mechanism, discussed in detail in the document “Clinical Terms Version 3 – Context and Clinical Records”.

4.3 Features of Clinical Terms Version 3

CTV3 was developed to address the problems of the early versions as described in section 4.1. The main features include:

Natural Clinical Terms

The use of natural clinical terms underlines the philosophy of CTV3. There are a number of terms, derived from classification systems, which have precise and significant meanings in the context of a classification but which are not clinically intuitive. For example, the meaning of a term such as *Asthma NOS* (not otherwise specified) is not clear without knowing what has been specified in other terms. In CTV3, such clinically non-intuitive terms are avoided wherever possible. (However, a number of such classification terms are included for historical reasons due to their inclusion in earlier versions).

Terms and Concepts

A clinical **concept** represents any relevant action, process, state event or object. In CTV3, a concept may be any clinical disorder, procedure, observation, etc. and each is represented by one unique Read Code.

A concept may be described by one or more **terms**. For example, the terms *Osteoarthritis of the hip*, *Osteoarthrosis of the hip* or *OA of the hip* can all be used to describe the same concept.

Conversely, a term can also describe more than one concept. For example, the term *Cord compression* could be used to describe either *Spinal cord compressions* or *Umbilical cord compression*.

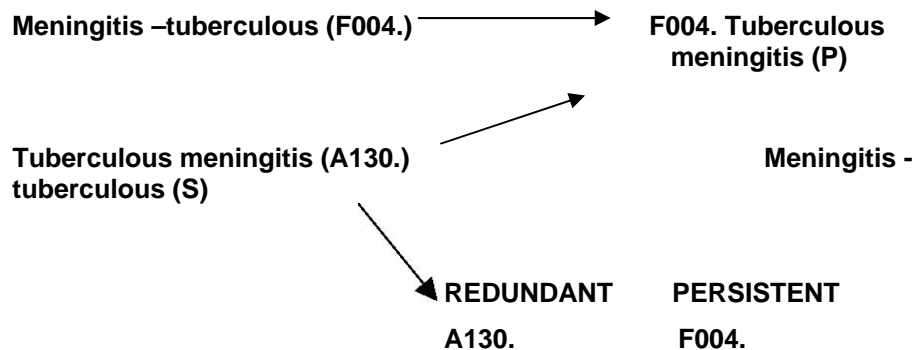
In CTV3, three files manage these many-to-many relationships between concepts and terms: the **Concept file** the **Terms file** and the **Descriptions file**. The way in which these files are used is described in section 4.4 of this document. The files are described in detail in the document “Clinical Terms Version 3 – Main File Structure: Overview and Technical Description”.

Concept Status

In CTV3, each concept is assigned a **concept status**: either current, optional, redundant or extinct.

- **Current** status is given to all mainstream, clinically useful concepts, suitable for recording clinical data.
- **Optional** status is given to concepts which are mainly derived from incorporation of earlier versions and are not considered clinically intuitive (but which may still be used). These include:
 - classification-type concepts, such as *Asthma NOS*;
 - concepts which contain information better held in qualifiers, such as *Urgent cholecystectomy* (see the discussion of qualifiers in Chapter 4); and
 - combined concepts, such as *Acute gastric ulcer with haemorrhage and perforation*.
- **Redundant** status is assigned in circumstances where more than one code is found to exist for the same concept. This usually arises where a concept is derived from a previous version of the Read Codes which associated with more than one code; in CTV3 it is now represented by only one of those codes (the persistent code). The other code(s) is/are made redundant and mapped to the persistent code via the **Redundant Codes Mapping file** (described in the document “Clinical Terms Version 3 – Main File Structure: Overview and Technical Description”). For example, the duplication of *Tuberculous meningitis*, (discussed in section 4.1 of this document), is resolved by making

F0004. via the Redundant Codes Mapping file. If the terms attached to the redundant code differ from those attached to the persistent code (but are synonymous), then the term(s) from the redundant concept are re-assigned to the persistent concept:



- **Extinct** status is used to deal with ambiguous codes of complex meaning largely derived from earlier versions of the Read Codes, when data is migrated into CTV3. This is described in detail in the document “Clinical Terms Version 3 – Incorporation of Earlier Versions of the Read Codes (The Superset)”.

Synonym Purity

Terms are associated with concepts via descriptions. A concept can have only one preferred description and several synonymous descriptions².

² Extinct concepts due to their complexity may have impure “synonymous terms” that can be preferred terms to other concepts.

Flexible Hierarchy Structure

The code-dependent hierarchy structure of earlier versions has been abandoned in CTV3. Instead, the hierarchy is formed by a set of links between “parent” and “child” concepts. The parent-child links are held in a **Hierarchy file** (described in the document “Clinical Terms Version 2 – Main File Structure: Overview and Technical Description”). This type of hierarchy gives several advantages:

- **Unlimited hierarchy depth** The Read Code in CTV3 serves only as a unique identifier for each concept and does not identify the position of the concept in the hierarchy. This means that the depth of the hierarchy is not limited to 4 or 5 levels as in earlier versions.
- **Pure sub-type hierarchy** Each “child” concept is placed in a position of the hierarchy such that it is a “type of” its parent. This creates a more accurate and coherent hierarchy.
- **Dynamic hierarchy** The contents of hierarchies can be changed as medical knowledge evolves, by adding new concepts, linking “child” concepts with different parents, or linking “child” concepts with additional “parents”, if required.
- **Multiple parents** The hierarchical problem illustrated by *Tuberculous meningitis* (section 4.1) is now resolved. The Hierarchy file can contain two (or more) ‘parent’ links for this Read Code, so that it is positioned in two places in the hierarchy, as a descendant both of *Infectious and parasitic diseases* and *Nervous system and sense organ diseases*.

Combining Concepts (Core Terms and Qualifiers)

A surgical procedure such as *Cholecystectomy* might have a priority of *Emergency*, *Urgent*, *Scheduled* or *Routine*. Similarly, most disorders can have a severity of *Mild*, *Moderate* or *Severe*. To describe all these variations, it might be possible to create a separate Read-coded concept for each combination (i.e. *Emergency cholecystectomy*, *Urgent cholecystectomy*, etc.). However, to do this would require the creation of a vast number of codes.

Clearly, it would be more efficient to give one Read Code to a priority such as *Urgent* and then to combine this with any procedure as required. In CTV3, this is achieved by holding such information in the form of qualifiers. Qualifiers are held in the **Template file**, described in the document “Clinical Terms Version 3 – Template File”.

Semantic Definition (Atoms)

As well as giving additional information about concepts, in the form of qualifiers, it is also desirable to store semantic definitions. For example, in the case of a *Cholecystectomy*, it may be useful to record that the *Site:* of the operation is *Gallbladder* and that the *Method:* of surgical action is *Excision*. The addition of such semantic definitions does not give any additional information, but make searches of the database more efficient (e.g. to ensure that a search for all patients who have had *Gallbladder* operations would recover patients who have been recorded as having had a *Cholecystectomy*).

Semantic definitions (also called “atoms” of information) are also held in the **Template file** and are described in the document “Clinical Terms Version 3 – Template File”.

Accurate Cross-Mapping

When cross-mapping to target classifications such as ICD-10 or OPCS-4, there is normally a default mapping to a single target code as in Read Codes 5-byte Version2. However, in CTV3, this system is now augmented by alternative mappings – where applicable - which incorporate some of the classification rules in the target terminology (about 70% of the mappings are one-to-one, the other 30% offer alternatives).

Cross-mapping is described in more detail in the document “Clinical Terms Version 3 – Cross Mapping File”.

Keys

Users can find terms by browsing the hierarchy. Alternatively, they can enter a ‘key’, consisting of all or part of a word and searching the database for related terms which will then be displayed on the screen. These keys consist of up to 10 characters and are stored in a separate file.

Updates

CTV3 is updated biannually (monthly for Version 2 drugs), so that new concepts can be added in response to user feedback or as a result of on-going development and consolidation.

4.4 The Structure of CTV3

The functions of CTV3 files are summarised in this section. The files listed below are all described in detail in the document “Clinical Terms Version 3 – Template File” unless otherwise indicated. A summary diagram of the relationships between the various files is also included overleaf.

Concept File

This file allocates a 5-character alphanumeric **Read Code** to each **concept**. The status of each concept is recorded as being current, redundant or extinct. (N.B. the Read Code is case sensitive).

Term File

Terms in CTV3 are coded separately from concepts. In this file, each **Term id**, which is also a 5-character alphanumeric code (this is a separate code from the Read Code). The file can store up to three different lengths of term: 30 characters, 60 characters and 198 characters. It is recommended that the longest available term is used. (N.B. the Term_id is case sensitive).

Description File

The descriptions file links terms to concepts. If a concept can be described by more than one term, the descriptions file designates one term as the “preferred term” and all other terms as “synonymous terms”.

The preferred term must be unambiguous and related to only one concept (with the exception of Extinct concepts, which are discussed in the document “Clinical Terms Version 3 – Incorporation of Earlier Versions of the Read Codes (The Superset)”). Synonymous terms may be shared with other concepts (homonyms). For example, the term *Cord compression* has two potentially different meanings. In an adult, the term can be used to mean *Umbilical cord compression*. The descriptions file will therefore list *Spinal cord compression* and *Umbilical cord compression* as the preferred terms for these two different concepts (i.e. they will be linked to different Read Codes as description type ‘P’ – preferred). The term *Cord compression* will be listed as a synonymous description for both concepts (i.e. will be linked to both Read Codes as description type ‘S’ – synonym). *Cord compression* being ambiguous thus is not used as a preferred description for any concept.

Read Code	Term	P/S	Term_id
X40Cc	Umbilical cord compression	(P)	YaaGm
X40Cc	Cord compression	(S)	Y40xj
Xa0Nk	Spinal cord compression	(P)	Ya1XS
Xa0Nk	Cord compression	(S)	Y40xj

Hierarchy File

In CTV3, the hierarchy is formed by a set of “parent-child” links held in this file, rather than being code-dependent. Any “parent” can have several children and a “child” can have several parents³.

Redundant Codes Mapping File

This file maps redundant concepts to the appropriate persisting current concepts.

Key File

This file links alphanumeric keys to the Term id codes in the Term file. These keys, which are generated automatically, can be entered by users to search for terms.

Cross-Mapping Files

These files map the Read Code to classifications (ICD-10 and OPCS-4). Where the mapping is not one-to-one, default and alternative mappings are provided and identified by a set of flags.

These files are described in detail in the document “Clinical Terms Version 3 – Cross Mapping File”.

Template File

This file allows Read-coded concepts to be combined so that qualifiers can be used to define additional detail. It also links concepts to “atoms” providing semantic definition. This file is described in the document “Clinical Terms Version 3 – Template File”.

Relationships Between Files

A diagram showing the relationship between files is given on the following page.

Further Information

Product descriptions of all CTV3 files, including Value Added Files and Bonus Files, are available on our website on the following link:

<http://www.connectingforhealth.nhs.uk/systemsandservices/data/readcodes>

³ Except for the root node Read thesaurus (Read Code.....) which does not have any parent code.

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