INTERNATIONAL HEALTH TERMINOLOGY
STANDARDS DEVELOPMENT ORGANISATION
UK Terminology Centre (UKTC)

NHS
Connecting for Health

THE CLINICAL TERMS VERSION 3
(THE READ CODES)

GENERAL PRACTICE FLAGGED SUBSET

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	172v1.0

Published by the UK Terminology Centre

Table of Contents

1. Introduction

The illustrative General Practice (GP) flagged subset provides a list of Clinical Terms Version 3 (The Read Codes) – (CTV3) concepts which have been identified as being of interest to GPs, from the large selection available in the CTV3 Thesaurus. This document aims to provide an understanding of the mechanisms used to generate the subset and guidance as to the content of the GP Subset Explicit and Instruction set files.

2. Background

The CTV3 Thesaurus is a comprehensive source of coded clinical terms which reflect the terming needs of all healthcare professionals. However, different users have different perspectives on the contents of the Thesaurus, prompting the notion of a “subset” of pre-selected codes for a given speciality.

3. What is a Flagged Subset?

A subset is a list of codes which may have been selected by a particular specialty, a group or an individual. The aim of a subset is to provide a mechanism for users to have access to a pre-selected set of terms appropriate to their terming needs in everyday use. The result, when applied to key word searches, is shorter picking lists of terms.

Please note: the use of a subset should not preclude access to the complete set of terms in the thesaurus, but should provide preferential access to pre-selected terms for any one specialty or professional group.

4. Concept Status in a Flagged Subset

A code can be in one of three states in relation to a subset (see also sections 10.3 and 11.3):

1. Specifically included

2. Specifically excluded

3. Included as an ancestor of a specifically included code. These ancestor codes, which may not themselves be selected, provide a path to the

top of the hierarchy from the selected code and (are illustrated as "grey") to allow access to the selected codes from the top down.

5. Types of Flagged Subset

From the hierarchical relationship of the codes in a subset, three theoretical types of subset can exist and are known as 'skirt', 'sliver' or 'speckle' sets, respectively. These three types may exist in a 'pure' state or else in combination with each other.

5.1 Skirt Subset

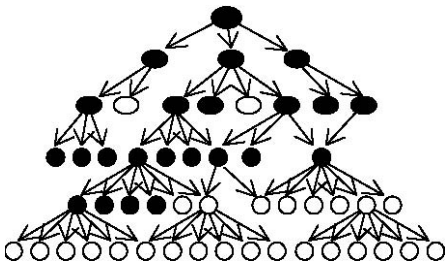
A selected code and all its ancestors to the top level of the hierarchy are included. Its descendants are specifically excluded. Skirt subsets do not, however, preclude wide coverage across the entire CTV3 hierarchy. An example is the general practice set, where excessive detail below a particular level in the hierarchy is not usually required.

Key to diagrams:

Dark = Specifically included code

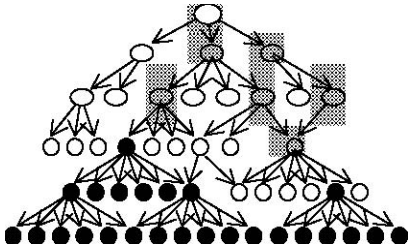
Grey = Included as an ancestor of a specifically included code

White = Specifically excluded code



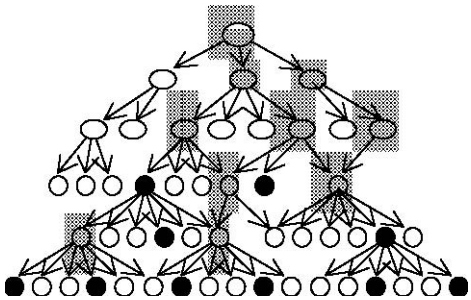
5.2 Sliver Subset

A code and all its descendants are included in the subset. The ancestors to the selected code are included as "grey" codes. Thus a narrower set of codes with greater detail can be generated. An example might be the urological surgery set.



5.3 Speckle Subset

Selected codes in this subset type are typically scattered throughout the CTV3 hierarchy. The ancestor codes are again “grey” to allow access to the selected codes using the hierarchy structure. An example of such a subset might be the pathology messaging set.



The view created from a subset can be applied to searches based on key words in order to create shorter picking lists for the user. In addition, the user may wish to adapt pre-defined subsets to reflect more local healthcare terming needs. An alternative could be the complete self-selection of a set.

6. GP Flagged Subset Structure

The GP subset is a combination of a ‘skirt’ and a ‘sliver’ set. In most parts of the Thesaurus a ‘skirt’ was generated by excluding more detailed lower level codes in the hierarchy. However, there are two notable exceptions where a ‘sliver’ was included. These are the hierarchies of GP Administration terms and the Drug dictionary.

7. GP Flagged Subset Production

The GP Flagged Subset was produced in two stages. A member of the GP Speciality Working Group (SWG) scanned the Thesaurus for concepts thought to be required for the subset. A second pass involved all members of the SWG reviewing concepts both included and excluded from the first process.

The GP Flagged Subset was included as test data in the March 1997 release, after which additional codes were reviewed and added to the subset before its release as a Value Added File in October 1997, with further updates in March 1998. The October 1997 GP Flagged Subset contained 28,352 codes, excluding the Drug Dictionary and 60,124 including the Drug Dictionary. This total has increased to 41,573 codes, excluding the Drug Dictionary for the updated March 1998 GP Flagged Subset.

8. GP Flagged Subset Maintenance

No commitment to maintaining the Subset in the long term can be made at present. Amendments to content will be made in response to user feedback.

9. GP Flagged Subset File Structure

The GP Flagged Subset is being issued in the same file structure as earlier release data.

10. GP Flagged Subset Explicit File

It is intended that the file be used as an adjunct to the core CTV3 release of the same date. Limiting user access to codes based only on the contents of this file will not be desirable in the majority of situations and we recommend that such filtering be user-configurable.

10.1 Warning

This subset will continue to undergo an iterative process. Amendments arising from operational testing of the subset by GP users may result in significant changes to its content. Changes are therefore possible to both the size and scope of the subset and also to the file description.

10.2 File Description

The full product description for the GP Flagged Subset Explicit file is attached in Appendix A.

File name:	**gpeset.v3**
File description:	A file containing a list of Read Codes and associated Term IDs to enable filtering of codes or

terms presented to users during keyword searching and hierarchy navigation.

Field number	Title	Size	Unique
1	READ_CODE	5 characters	#
2	TERM_ID	5 characters	#
3	TONE	1 character	
4	RELEASE	8 characters	

10.3 Field Descriptions

The following field descriptions apply to the GP Flagged Subset Explicit file shown above:

READ_CODE
This field contains the five character concept code and is either current or optional concept status. This is a primary key.

TERM_ID
The term identifier of the natural term of this code for this subset. Initially, this field will be mainly populated with the preferred term identified in the descriptions table. This is a primary key.

TONE
This field has two possible values:

1 = A code ancestral to one specifically included in the subset and acting to join the included codes together in the hierarchy. A “grey” code.

2 = A code specifically included in the subset.

The vast majority of codes will be flagged with a tone of 2.

RELEASE
This field denotes the release date using the International Standard Organisation (ISO) 8601 format YYYY-MM-DD. For example, the first day of September 1997 is 1997-09-01.

11. GP Flagged Subset Instruction Set File

It is intended that the file be used as an adjunct to the core CTV3 release of the same date. Limiting user access to codes based only on the contents of this file will not be desirable in the majority of situations and we therefore recommend that such filtering be user-configurable.

11.1 Warning

This subset may continue to undergo an iterative process. Amendments arising from operational testing of the subset by GP users may result in significant changes to its content. Changes are therefore possible to both the size and scope of the subset and also to the file description.

11.2 File Description

The full product description for the GP Flagged Subset Instruction set file is attached in Appendix B.

File name: **gpiset.v3**

File description: A file containing a list of Read Codes and associated instructions to enable filtering of codes or terms presented to users during keyword searching and hierarchy navigation.

Field number	Title	Size	Unique
1	SUBSET_ID	8 characters (variable)	
2	READ_CODE	5 characters	#
3	INSTRUCTION	7 character (variable)	
4	RELEASE	8 characters	

11.3 Field Descriptions

The following field descriptions apply to the GP Flagged Subset Instruction set file shown above:

SUBSET_ID
This field contains the identifier for a subset. It is a variable character field with a maximum of 8 characters and for example, will contain 'GP' for the GP instruction subset.

READ_CODE
This field contains the five character concept code and is either current or optional concept status. This is the primary key.

INSTRUCTION
This field will have one of a possible 6 values, of variable character length (3-7 characters), to be applied in the following order:

INCUPDN = Include this node, all its ancestors and all its descendants

INCUP = Include this node and all its ancestors

INCDN = Include this node and all its descendants

EXDN = Exclude this node and all its descendants

INC = Include this node alone

EXC = Exclude this node alone

RELEASE
This field denotes the release date using the International Standard Organisation (ISO) 8601 format YYYY-MM-DD. For example, the first day of September 1997 is 1997-09-01.

12. File Release Format

The files will be fixed length fields, except for the Subset ID and Instruction fields in the Instruction Set File.

All fields will be delimited with the vertical bar character "|" (ASCII value 7C hex) which itself will not be an allowable character in any of the data fields.

The file will contain one record per line.

Appendix A

GP Flagged Subset Release Product Description (Explicit File)

Product Name	GP Flagged Subset Explicit File
Contents and purpose of file	A file containing an explicit list of Read Codes and an associated Term ID from which the subset can be generated thus enabling the filtering of codes or terms presented to users during keyword searching and hierarchy navigation.
How should the file be used?	This file should be used as an adjunct to the core CTV3 release. Limiting user access to codes based only on the contents of these files will not be desirable in the majority of situations. We therefore recommend that such filtering be user-configurable.
Warnings	**This subset may continue to undergo iterative review and refinement. Changes are possible in the size and scope of the subset**

	and the file description. No guarantee can be given that this Value Added File will continue to be released or updated in future quarterly releases.
Additional detail	Field descriptions for Explicit format: **Read Code** A Read Code. This is a primary key for this table. **Term ID** The Term ID of the natural term[1] of this code for this specialty. Initially, this field will be largely populated with the preferred term ID identified in the Descriptions table. This is a primary key for this table. **Tone** This field will initially have 2 possible values: 1 = Included only to join all the codes in the subset together into a single tree. 2 = Specifically included The vast majority of codes will be flagged with a tone of 2. **Release** This field contains the date of the release for the file. It is in ISO format of YYYY-MM-DD. For example, a file released on 1st September 1997 would be 19970901.
File names	**gpeset.V3**
File format	ASCII (DOS) Text: Read Code (char5) | Term ID (char5) |Tone (char1) | Release (char8)
Date of first release	October 1997
Release history	Value Added File (VAF) in March 1998 release.
Maintenance routine	No maintenance will be guaranteed.
Change report	None.

[1] For an explanation of natural synonyms, refer to the document "Clinical Terms Version 3 – Main File Structure: Overview and Technical Description".

Notification of forthcoming changes	None.
Produced by	UK Terminology Centre 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
Available 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
Enquiries	UK Terminology Centre

Appendix B

GP Flagged Subset Release Product Descriptions (Instruction Set File)

Product Name	GP Flagged Subset Instruction File
Contents and purpose of file	A file containing a list of Read Codes each with an instruction from which the subset can be generated thus enabling the filtering of codes or terms presented to users during keyword searching and hierarchy navigation.
How should the file be used?	This file should be used as an adjunct to the core CTV3 release. Limiting user access to codes based only on the contents of these files will not be desirable in the majority of situations. We therefore recommend that such

	filtering be user-configurable.
Warnings	**This subset may continue to undergo iterative review and refinement. Changes are possible in the size and scope of the subset and the file description. No guarantee can be given that this Value Added File will continue to be released or updated in future quarterly releases.**
Additional detail	Field descriptions for Instruction set format: **Subset ID** This field contains the identifier for a subset. It is 'GP' for this file. **Read Code** A Read Code. This is the Primary key for this table. **Instruction** This field will have one of a possible 6 values to be applied in the following order: INCUPDN = include this node, all its ancestors and all its descendants INCUP = include this node and all its ancestors INCDN = include this node and all its descendants EXDN = exclude this node and all its descendants INC = include this node alone EXC = exclude this node alone **Release** This field contains the date of the release for the file. It is in ISO format of YYYY-MM-DD. For example, a file released on 1st September 1997 would be 19970901.
File names	**gpiset.V3**
File format	ASCII (DOS) Text: Subset ID (varchar8) | Read Code (char5) | Instruction (varchar7) | Release (char8)
Date of first release	October 1997

Release history	Value Added File (VAF) in March 1998 release.
Maintenance routine	No maintenance will be guaranteed.
Change report	None.
Notification of forthcoming changes	None.
Produced by	UK Terminology Centre 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
Available 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
Enquiries	UK Terminology Centre