

ICD-10 5th Edition: Codes and Titles and Metadata File Specification

August 2024



Glossary of Terms

Term / Abbreviation	What it stands for
category	A category is a three-character code and rubric/description.
subcategory	<p>A subcategory is a four-character code and rubric/description. A decimal point may be shown between the third and fourth character.</p> <p>Most three-character categories are subdivided into four-character subcategories.</p> <p>Insertion of the character 'X' into the fourth field of codes where only three characters exist is mandatory, so the codes are of a standard length for data processing and validation. This provides a standard 'filler' character.</p> <p>Codes with a format of AnnX are included in references to 4 character codes in this specification.</p> <p>Many systems, but not all systems, automatically insert the filler 'X', therefore all users of coded clinical data should be aware of this requirement.</p>
(sub)category	Refers to both category and subcategory in this document
rubric	The text associated with a (sub)category. Also known as title or code description
Preferred rubric	The title associated with a (sub)category.
Preferred long rubric	The title associated with a (sub)category. Some preferred rubrics are not self-explanatory and need to be read in conjunction with their superordinated category. See Appendix A for demonstration of difference between Preferred rubric and Preferred long rubric.
Field name	The official name for a field within a record
Char	Character (A letter (e.g., a,A, b, and B), number (0, 1, 2, or 3) etc
DSV	<p>Delimiter Separated Values (DSV) format</p> <p>Specifically, the variant is "Tab Separated Values" (the delimiter being a tab character).</p>
WHO	World Health Organisation
HES	Hospital Episode Statistics
TRUD	<p>Technology Reference data Update Distribution</p> <p>https://isd.hscic.gov.uk/trud3/user/guest/group/0/home</p>

Contents

Status of this document	6
Changes to the GENDER_MASK metadata field	6
Introduction	7
Purpose	7
Audience	7
Background	7
ICD-10	7
ICD-10 Codes and Titles and Metadata file	7
WHO Copyright	7
WHO updates to ICD-10	8
Maintenance and release	8
Implementing ICD-10 5th Edition	8
Outputting data for central requirements	8
Viewing the DSV file – advice for non-technical users	8
Derivation, formats and distribution	9
Derivation of codes and titles	9
Access to ClAML	9
Derivation of validation checks	9
ICD-10 Volumes 1, 2 and 3	9
Format	9
Filename and distribution	9
Abbreviated descriptions	9
Availability in other formats.....	9
File content	10
Record format	11
Field metadata.....	11
Field list	12
ICD-10 Codes and Titles and Metadata file	13
Code and usage fields	14
CODE	14
ALT_CODE	14
Description.....	14
Addition of 5 th character codes to support subclassifications.....	15
Type of metadata check	15
Important Note on 5 character codes.....	15
Use	15

USAGE	16
USAGE_UK.....	16
Special features	16
Type of validation check	17
Use	17
Ordering of the dagger/asterisk	17
More resources on daggers and asterisks.....	17
Description fields	18
DESCRIPTION.....	18
Special features	18
Type of validation check*	18
Use*	18
MODIFIER_4.....	18
MODIFIER_5.....	19
Metadata fields	20
QUALIFIERS	20
Special features	20
Type of validation check	20
Use	20
GENDER_MASK.....	21
Type of validation check	21
Use	21
Age restriction fields	21
MIN_AGE	21
Use	21
MAX_AGE	22
Use	22
Type of check	22
TREE_DESCRIPTION	23
Future reissue policy	24
Enquiries	24
Appendix A: Additional information on rubrics	25
Preferred versus Long Preferred Rubric (ClAML).....	25
Appendix B: Regular expressions	26
Appendix C: Unicode	27

Change system to support Unicode	27
Data storage	27
Executable code	27
Presentation issues	27
Degrade data to legacy encoding.....	27
Appendix D: Selection of fonts that support subscript 6	29
Annex 1: Viewing the DSV file using Excel data import	30
Annex 2: USAGE_UK	32
Absolute dagger, with asterisk(s) in title (value 1)	32
Absolute asterisk (value 2)	33
Other (non-dagger/asterisk, inclusion dagger or inclusion asterisk) (value 3)	33
(Absolute) dagger, without asterisk(s) in title (value 4)	34
Not applicable (previously expressed as space) (value 0)	35

Status of this document

This document was reissued in August 2024 and contains a change to the type of validation check for the [GENDER_MASK](#) metadata field.

Changes have also been made to URL's, links and names of publications, where they have changed since the original publication. The format of the document has been updated.

All other information within this document remains the same as originally published in July 2015.

Changes to the [GENDER_MASK](#) metadata field

The [GENDER_MASK](#) metadata field was an absolute check meaning that where this validation is present it will lead to record failure when:

- A code with a [GENDER_MASK](#) value of 1 is assigned on a male patients' record
- A code with a [GENDER_MASK](#) value of 2 is assigned on a female patients' record.

This check has been changed to a scrutiny check so that when it is applied it does not automatically lead to rejection of the record but should generate a warning message to alert the coder to investigate further. This is in order to comply with the Equality Act 2010 to ensure equal treatment of patients regardless of a chosen gender.

This will ensure that a record will not be rejected where a condition is appropriate to the anatomy of the patient but does not match with their chosen gender identity.

This should not impact data quality because the [GENDER_MASK](#) validation applied as a scrutiny check will still provide a warning to check the correct record is being coded for that episode of care, if appropriate.

Introduction

Purpose

This document specifies the structure, content and use of the ICD-10 5th Edition: Codes and Titles and Metadata file to enable implementation in software applications in readiness for “go live” date of 1 April 2016.

It must be read in full in conjunction with the ICD-10 5th Edition Codes and Titles and Metadata file to ensure that systems are updated correctly, and data quality is not compromised. The data file is available at [NHS ICD-10 5th Edition data files - TRUD \(digital.nhs.uk\)](http://digital.nhs.uk)

Audience

Suppliers and other technical professionals that incorporate the ICD-10 data file into patient administration and other systems to support Commissioning Data Sets, Central Returns etc. as specified in the NHS Data Model and Dictionary at <http://www.datadictionary.nhs.uk/>

Informatics specialists, including clinical coders, who are involved in the implementation of the ICD-10 5th Edition and quality assurance of coded clinical data, will find this document useful.

Background

ICD-10

The International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, (ICD-10) is a comprehensive classification of causes of morbidity and mortality. It contains an alphanumeric listing of codes for diseases, injuries, external causes of injury and poisonings, and factors influencing health status and contact with health.

It is maintained and published in three Volumes by the World Health Organisation. See [International Classification of Diseases \(ICD\) \(who.int\)](http://www.who.int)

ICD-10 Codes and Titles and Metadata file

The Delimiter Separated Values (DSV) data file contains the codes and titles listed in ICD-10 5th Edition, Volume 1: Tabular list of Inclusions and four-character subcategories. Additional information, such as sex or age range, is associated with codes where appropriate.

The data file, where integrated into validation software, is the initial source of validation checks performed by systems on ICD-10 codes and related clinical data. The data file will also form part of local quality assurance and audit functions prior to the submission of Hospital Episode Statistics (HES) data.

WHO Copyright

The World Health Organization (WHO) is the copyright holder of ICD-10 which is used under licence for United Kingdom government purposes.

ICD-10 codes, terms and text used by permission of the WHO, from: International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10) Vols 1-3. Geneva, World Health Organization, 1992-2016.

WHO updates to ICD-10

ICD-10 5th Edition includes the WHO cumulative changes up to and including those with a WHO implementation date of January 2016.

Maintenance and release

The Terminology and Classifications Delivery Service at NHS England is responsible for maintaining and issuing the ICD-10 Codes and Titles and Metadata file.

The file is updated and released to support new versions of ICD-10.

Implementing ICD-10 5th Edition

All Finished Consultant Episodes (FCEs) that end on or after 1 April 2016 must be coded to the above edition. Trusts need to implement procedures to manage the transition from the ICD-10 4th Edition to the 5th Edition to prevent contamination of FCEs with invalid ICD-10 codes.

Ideally systems should be able to support multiple versions of ICD-10 (and OPCS-4) in line with the HSCIC's commitment to update classifications regularly.

Outputting data for central requirements

Trusts must continue to abide by the interoperability guidance that regulates the submission of Commissioning Data Sets (CDS) etc to Secondary Uses Service (SUS). More information is available at [Secondary Uses Service \(SUS\) - NHS England Digital](#).

The DSV file specification addresses the generation of ICD-10 data from the data file for use within your application. It does not address formats required for interoperability with other services.

Viewing the DSV file – advice for non-technical users

The simplest way to view the file is to drag the file icon into an open, empty, Excel window. Note that this does not select the correct encoding so some characters may be garbled (see **Appendix D**), but rows affected are in a minority.

See **Annex 1** for instructions on how to create a more formal data import. Selecting the Unicode encoding will rectify the above issue.

It is NOT recommended that you save the file from Excel because it will break the file structure. Setting the read-only flag on the file is advised.

Derivation, formats and distribution

Derivation of codes and titles

The data within the file is derived from ICD-10 Classification Markup Language (ClAML). This is an XML document schema - provided by the WHO - used for the purpose of conveying the ICD-10 classification. The list of members, their codes, usage, and descriptive rubrics, are primarily derived from this source.

The ICD-10 5th Edition contains no description longer than 207 characters. However, future updates may contain longer descriptions.

Access to ClAML

If used instead of the ICD-10 data files from TRUD, implementations must conform to the NHS Clinical Coding Standards.

Derivation of validation checks

WHO ICD-10 Volume 2 Section 3.1.5 provides a source of advice on metadata for example, gender checks, associated with specific codes in the file. For ICD-10 5th Edition, metadata has been researched and added for new and changed codes, as necessary.

ICD-10 Volumes 1, 2 and 3

The printed reference materials remain the canonical source of the ICD-10 classification.

Format

The data is presented in Delimiter Separated Values (DSV) format. The variant is “Tab Separated Values” (the delimiter being a tab character). Field width is unrestricted, thus enabling the full code descriptions presented in ICD-10 Volume 1 to be denoted.

Filename and distribution

The ICD-10 5th Edition Codes and Titles and Metadata file (ICD10_Edition5_CodesAndTitlesAndMetadata_GB_20160401.txt) is distributed at [NHS ICD-10 5th Edition data files - TRUD \(digital.nhs.uk\)](#).

Abbreviated descriptions

As stated above, field width is unrestricted in the DSV file. A separate file of abbreviated descriptions is available – on receipt of a request via the Information Standards Helpdesk - for systems that cannot handle full descriptions at this time.

Email: information.standards@nhs.net and ask for ICD10_Edition5_AbbreviatedDescriptions_GB_20160401.txt. Please provide a brief business case describing your particular requirement for the file. This will help to inform future releases.

Availability in other formats

The ICD-10 5th Edition Codes and Titles and Metadata file is also available in Extensible Markup Language (XML), which is the required standard for the NHS. It is called ICD10_Edition5_CodesAndTitlesAndMetadata_GB_20160401.xml and is distributed at [NHS ICD-10 5th Edition data files - TRUD \(digital.nhs.uk\)](#)

File content

The file has the following entries:

	ICD-10 ALT_CODE	Count	Required for outputs: Commissioning Data Sets (CDSs) Hospital Episode Statistics (HES).
1	3 char categories	2049 See row 3	No – see below
2	4 char subcategories	12309	Yes
3	Un-split 3 char categories with X*	250** Also included at row 1	Yes
4	5 char subclassifications	3576 Includes five char codes with 'X' filler	Yes
5	Total distinct codes	17934	

*Three-character categories that do not have 4-character subdivisions are designated with an 'X filler', e.g. **A46X Erysipelas**. Semantically, these are three-character categories.

** Includes un-split 3 char categories with X, that have 5th digit subclassifications at **T08**, **T10**, **T12** and **M45**.

The above table indicates valid codes for use in Commissioning Data Sets (CDS) and Hospital Episode Statistics (HES).

Record format

The format will follow the Delimiter Separated Values¹ (DSV) format as described below. This is a plain-text based format that uses characters to separate fields.

The general features are:

UTF-8 Encoding		Suitable for multi-language text with extended character sets and the usual modern standard
Escape Character	\	All characters following the escape character MUST be taken as literal input rather than command characters, excepting the character following an escaped escape character (i.e. \\)
Record Separator	LF	Each record MUST be separated by this character
Field Separator	TAB	Each field MUST be separated by this character
Header Row		The first row of the file will contain the field names

Appendix B: Regular Expressions contains more information on notation

Appendix C: Unicode contains more information on UTF-8

Field metadata

Name	The name of the field (as per header row)
Null?	Whether this field can be empty or not
Description	A short description of the field
Form	A description of the form the field should take
Regex	A regular expression describing the content of the field more rigorously

¹ Eric S Raymond : Data File Metaformats
<http://www.catb.org/~esr/writings/taoup/html/ch05s02.html>

Field list

Number	Name	Null?	Description
1	CODE	No	ICD-10 Code
2	ALT_CODE	No	ICD-10 Code, alternate form
3	USAGE	No	Dagger / Asterisk indication
4	USAGE_UK	No	Dagger / Asterisk indication
5	DESCRIPTION	No	Longest preferred rubric
6	MODIFIER_4	Yes	4 th character modifier suffix
7	MODIFIER_5	Yes	5 th character modifier suffix
8	QUALIFIERS	Yes	Dual classification(asterisk codes)
9	GENDER_MASK	Yes	Gender mask
10	MIN_AGE	Yes	Minimum age
11	MAX_AGE	Yes	Maximum age
12	TREE_DESCRIPTION	Yes	See Section 18

ICD-10 Codes and Titles and Metadata file

The codes and titles assist the process of validation by providing a comprehensive listing of acceptable ICD-10 codes for inclusion in Commissioning Data Sets (CDS) and Hospital Episode Statistics (HES). The metadata provide additional validation checks on gender, age etc.

The metadata are used in conjunction with patient and other data to identify invalid or inconsistent use of a code, for example, male pregnancy.

There are two types of metadata checks:

- I. absolute checks which are intended to lead to rejection of a record that fails the check, for example invalid code; and
- II. scrutiny checks that generate a warning message to alert the coder to investigate further, for example apparent violation of “normal” age. This check will not automatically lead to rejection of a record.

Scrutiny checks are mostly restricted to those for age, which are considered reliable and aspects of 5th characters/digits

The DSV file does not include validation software. It only contains the data, which would be suitable for inclusion in such software and used for validation purposes.

Code and usage fields

CODE

(ICD-10 Code)

Form	See description below
Null?	No
Regex	<code>^[A-Z]\d{2}(\.(X\d \d{1,2})?)?\$</code>

This field will carry the ICD-10 code identifying the (sub)category. This is the primary key of the file.

The code is required to consist of a single uppercase character followed by exactly two digits, followed (optionally), by a period and between one and two digits or an X followed by a single digit.

ALT_CODE

(ICD-10 Code, alternate form)

Form	See description
Null?	No
Regex	<code>^[A-Z]\d{2}(X \d{1,2} X\d)?\$</code>

Description

This form strips the decimal point from the code and appends the filler X where the 3 character category is undivided.

The 5 char codes are optional where the information is unavailable, therefore the 4 char code is present (for use) in the file as well as the 5 char grandchild. This includes any 3 char codes with X, for example M45X.

The ALT_CODE is composed as follows:

1st char A-Z

2nd & 3rd chars Numeric = represent 1st two numeric characters

4th char alpha numeric = represents 4th character

4th char X = filler for 3-character Undivided code

5th char = represents 5th character where it exists

This is the format of the code perpetuated from the pre April 2012 Codes and Titles and Metadata files. The filler “X” for undivided 3 character codes is mandatory.

No decimal point is included e.g.

CODE	ALT_CODE
A00	A00
A00.1	A001
A46	A46X
A15.0	A150
I70.00	I7000

Addition of 5th character codes to support subclassifications

The 4 character code has been extended to include a 5th character for specified codes starting with I, J, M, S and T. The use of 5th characters to support the collection of ICD-10 subclassifications is mandatory where the information is available.

Type of metadata check

Absolute check to be applied to the first 4 characters.

Where the first 4 characters of a code are valid and where a 5th character exists for that code, but is absent, the code should be presented to the coder for scrutiny to establish if the omission of the 5th character is deliberate or accidental. If the coder asserts that the 4 character code is correct for that medical record, the code should be accepted by the system. It should NOT be re-presented for scrutiny.

Absolute check must be applied to the 5th character where it has been used by the coder.

Important Note on 5 character codes

For codes starting with M the 4 character code is valid if the 5th character information is not available for coding i.e. **M10.0 Idiopathic gout**, is valid, if none of the 5 character variants, for example **M10.00 Idiopathic gout Multiple sites**, are present for coding.

Where a 5th character exists for codes starting with I, J, S and T the 5th character must always be assigned.

The National Clinical Coding Standards ICD-10 Reference Book, available from [TRUD](#), also covers 5th character codes.

Use

The ALT_CODE is the basic parameter against which to test the validity of a code used and/or manually input.

The ALT_CODE field contains the list of 3, 4 and 5 character codes. **Coding is completed to 4 character level or greater granularity.** The 3 character codes are included for completeness and to give essential context to 4/5 character codes where this is absent in the lower level code. See **16.1.1** for more details.

Undivided three-character codes have a filler “X” and therefore qualify as four-character codes in terms of length but retain the classification characteristics of a three-character category.

USAGE

Form	Enumeration
Null?	No
Regex	^DEFAULT DAGGER ASTERISK\$

This field contains the usage of the (sub) category as declared in the ClAML data.

USAGE_UK

(dagger/ asterisk indication)

Form	Enumeration
Null?	No
Regex	^[0-4]\$

This field contains the descriptions as expressed in ICD-10 Metadata² Specification v3.1 of the data. See below:

Value	Description
1	absolute dagger, with asterisk(s) in title
2	absolute asterisk
3	other (non-dagger/asterisk, inclusion dagger or inclusion asterisk)
4	(absolute) dagger, without asterisk(s) in title See 15.1.2
0 (zero)	not applicable (previously expressed as space in ICD-10 Metadata Specification v3.1)

Special features

² ICD-10 Metadata Specification v3.1 was issued with icd10met.txt

In cases where value = 1, it should be combined with field QUALIFIERS (Dual classification (asterisk codes)) in order to make full use of its potential.

Type of validation check

Values 1 and 2 are normally considered absolute checks. Occasionally category 4 may not be a dagger code, so codes with this value should be considered as a scrutiny check.

Use

Identifies many, but not all, occasions where dagger and asterisk combinations should be used.

Ordering of the dagger/asterisk

By default dagger codes are sequenced before asterisk codes, however this order can be reversed, and an asterisk code may precede a dagger code in the primary or secondary code positions. You may have A/D or D/A in the primary field and secondary fields.

More resources on daggers and asterisks

A full explanation of the dagger and asterisk system is provided in Volume 2 of ICD-10, and Annex 2 of this document.

The National Clinical Coding Standards ICD-10, available from [Delen](#), also covers daggers and asterisks.

Description fields

These fields are used to contain the description or rubric from the ICD-10 (sub)category and subclassifications.

Where **MODIFIER_4** or **MODIFIER_5** contains data, applications **MUST** present it in conjunction with the main text in **DESCRIPTION**.

DESCRIPTION

Form	Alphanumeric text
Null?	No
Regex	[^\n\r\t]* - no carriage returns, line feeds, or tabs

Applications may trim whitespace from either end of the description.

This field is populated with the description of the code.

See **Appendix A** for additional information about these descriptions.

Special features

The full code description, as presented in ICD-10 Tabular list, is provided. The number of characters is unrestricted.

There are a few occasions where the description, standing on its own, is insufficient to supply the full meaning of the subcategory code. These subcategories should employ their parent category (3 char) description for completeness. The subcategories of the following codes are affected:

- Y06, Y07, Y60, Y61, Y62, Y83, Y84,
- and subcategories of codes between V19 and V79 where the MODIFIER_4 field is empty e.g. including V39.- but not V38.-

Type of validation check*

Not normally used as a validation check.

Use*

Normally used for display purposes and as a response to the entry of an ICD-10 code. The latter provides assurance to the coder that the input has been accurate.

*Special features, type of validation check and use also apply to MODIFIER_4 or MODIFIER_5.

MODIFIER_4

Where a subcategory is the result of applying an additional fourth significant character modifier, the modifier text will appear in this field.

Form	Alphanumeric text
Null?	Yes
Regex	[^\n\r\t]* - no carriage returns, line feeds, or tabs

Where **MODIFIER_4** contains data, applications **MUST** present it in conjunction with the main text in **DESCRIPTION**, [in addition to **MODIFIER_5** if present]. See below:

CODE	DESCRIPTION	MODIFIER_4
E10.1	Type 1 diabetes mellitus	With ketoacidosis

MODIFIER_5

Where a subcategory is the result of applying an additional fifth significant character modifier, the modifier text will appear in this field

Form	Alphanumeric text
Null?	Yes
Regex	[^\n\r\t]* - no carriage returns, line feeds, or tabs

Where **MODIFIER_5** contains data, applications **MUST** present it in conjunction with the main text in **DESCRIPTION**, in addition to **MODIFIER_4** if present. See below:

CODE	DESCRIPTION	MODIFIER_5 (SUBCLASSIFICATION_5 TH DIGIT)
M00.10	Pneumococcal arthritis and polyarthritis	Multiple sites

Metadata fields

QUALIFIERS

(Dual classification (asterisk codes))

QUALIFIER Dual classification (asterisk codes) Dagger/asterisk Form	List of ICD-10 code
Null?	Yes
Regex	$^((^ (?<=\d)\^)([A-Z]\d{2}(\.\d{1,2})?))+\$$

Each code will be separated by the sub-field separator (^)

e.g.

`M07.0^M07.1^M07.2^M07.3^M09.0`

Note that the use of the primary key, CODE, in this field serves solely to identify the data row required, and that presentation of this data should be consistent with that required for its context.

Special features

This field is applicable only to those absolute dagger codes with asterisks in their title³: (Dagger/asterisk indication = 1 in UK USAGE). For any code with an indicator of 1 there is an asterisk code(s) with which this dagger code must be associated.

Some asterisk codes have 5 characters.

Type of validation check

Absolute check. One of the specified (asterisk) codes must be used in combination with a specified dagger code.

Use

If the diagnosis has been coded to a dagger and asterisk combination (dual classification) and the Dagger/asterisk indication = 1, the associated asterisk code must be one of the specified codes; otherwise the record should be rejected.

³ Reference to ICD-10, Volume 1, code description/title. The links to asterisk codes are stripped from the titles used in the file description fields.

GENDER_MASK

Form	Enumeration
Null?	Yes
Regex	<code>^1 2\$</code>

Value	Description
1	Males not accepted
2	Females not accepted

Type of validation check

Scrutiny check

Use

It identifies conditions that are usually, but not always, associated with a single sex. It should be noted that record failure associated with the use of this field might indicate an incorrect patient record or an incorrect diagnostic code.

Age restriction fields

Form	Integer
Null?	Yes
Regex	<code>^\d{1,2}\$</code>

These fields indicate the “normal” age range for the code, expressed as a positive integer between 1 and 99, denoting age in years. Zero padding is not required.

MIN_AGE

This field indicates the minimum age that applies to this code

Empty field = no check required

1-99 = reject if age (in years) is less than this value

Use

It identifies conditions occurring below their ‘normal’ age range. It should be noted that record failure associated with the use of this field might indicate an incorrect age rather than an incorrect diagnostic code.

MAX_AGE

This field indicates the maximum age that applies to this code

Empty field = no check required.

1-99 = reject if age (in years) is greater than this value

Use

It identifies conditions occurring above their 'normal' age range. It should be noted that record failure associated with the use of this field might indicate an incorrect age rather than an incorrect diagnostic code

Type of check

Apply Scrutiny check to MIN_AGE and MAX_AGE

TREE_DESCRIPTION

Form	Alphanumeric text
Null?	Yes
Regex	[^\n\r\t]* - no carriage returns, line feeds, or tabs

This data field also contains descriptions that are functional in the context of the 3 character descriptions of their parent category, and are thus suitable for presentation where that context is available e.g. in the tree list below:

C00 Malignant neoplasm of lip

C00.0 External upper lip

C00.1 External lower lip

C00.2 External lip, unspecified

However, this represents incomplete information outside of this context. Where such a description is available, it will be placed in the TREE_DESCRIPTION field. See **Appendix A** for discussion of these descriptions.

Implementers must use the description fields at Section 16 for all instances where the data will be presented out of the context of its parent row but may choose to use this field where that context is present, e.g. in a tree view of the data.

Future reissue policy

It is anticipated that new releases of ICD-10 will be undertaken to support future ICD-10 updates issued from the WHO and approved by the Standardisation Committee for Care Information (SCCI).

This data specification will be updated as necessary to reflect changes.

Enquiries

Information Standards Helpdesk:

Tel: 0300 30 34 777 weekdays from 9am to 5pm

Email: information.standards@nhs.net

For information about the Terminology and Classifications Delivery Service visit website:

[Delen: Home - NHS England \(kahootz.com\)](#)

For information about ICD-10 5th Edition visit website:

[ICD-10 - Delen: Home - NHS England \(kahootz.com\)](#)

For information about TRUD visit website:

[Home - TRUD \(digital.nhs.uk\)](#)

Appendix A: Additional information on rubrics

Preferred versus Long Preferred Rubric (ClaML)

or *TREE_DESCRIPTION* versus *DESCRIPTION*

The ClaML standard states “*The attribute kind="preferred" defines a specific unique term that identifies the meaning of a class.*”

However, many of the rubrics with rubric kind "preferred" do not identify the meaning of a class without viewing them in context with the preferred rubric of their parent. Where this is the case, there is typically a rubric of type preferredLong to compensate.

Class	Preferred Rubric	Preferred Long Rubric
C00	Malignant neoplasm of lip	<i>none</i>
C00.0	External upper lip	Malignant neoplasm: External upper lip
O24	Diabetes mellitus in pregnancy	<i>none</i>
O24.0	Pre-existing type 1 diabetes mellitus	Diabetes mellitus in pregnancy: Pre-existing type 1 diabetes mellitus

In the example above, the shorter *preferred* rubrics of **C00.0** and **C24.0** do not clearly identify the full meaning of the class without reference to their parents.

Appendix B: Regular expressions

Regular expressions are a commonly used means of matching patterns in text. A full discussion of the syntax of regular expressions⁴ is beyond the scope of this document, but a short list of the character classes and metacharacters used here is presented for reference.

\d	A digit
^	The beginning of the text
\$	The end of the text
{n, m}	The previous expression should repeat between n and m times.
	The expression to the left, or the expression to the right of this character.
[A-Zq]	Any of the characters in this range (example, A to Z, plus q)
()	Group – anything in a group is considered one expression for the next metacharacter
?	The preceding expression is optional
+	The preceding expression occurs one or more times
.	Any character
\.	A literal dot
*	The preceding expression occurs zero or more times
\n	New line
\r	Carriage return
\t	Tab

⁴ Wikipedia page: Regular expression
http://en.wikipedia.org/wiki/Regular_expression

Appendix C: Unicode

The text encoding of the file has changed from ANSI to UTF-8 to permit the full gamut of international characters in the original material to be used.

Existing systems that do not currently accommodate double-byte characters may exhibit some issues.

Change system to support Unicode

Data storage

Most modern database systems support multi-byte characters and the UTF-8 encoding. Existing schemas may need changing where they have previously used single-byte encodings e.g.

MySQL change the column collation

MS SQL Server change datatype to nvarchar

Executable code

Most modern toolkits support UTF-8; Java and .NET represent strings as UTF-16 internally and load UTF-8 streams easily.

Presentation issues

Systems may need to use a font that supports the full Unicode 2.0 specification to correctly present all characters. The only known problematic character is currently the subscript digit 6 (₆), which is not present in many commonly used fonts. See **Appendix D** for a list of known fonts that support this character.

Tools used to view the data should also support UTF-8. Many tools also guess the encoding of the file by reading the first part, which does not work here as the first double-byte character occurs some distance into the file, so you may need to manually reconfigure which encoding you are using.

The Developer prefers Notepad2 (basic) by Florian Balmer and GVim (powerful but with a learning curve) by Bram Moolenaar – both are freely available open source programs at: <http://code.kliu.org/misc/notepad2/>

Degrade data to legacy encoding

Legacy systems with no capability for double-byte characters may wish to degrade these characters to normalized equivalents in their local encoding, see example script below. Implementers wishing to do this may wish to review [1]

[1] <http://stackoverflow.com/questions/3707977/convert-java-string-to-ascii>

Example script:

```
#!/bin/bash
```

```
#
# Degrade ICD-10 UK release to ASCII
#
# Pass release file name as parameter, output goes to STDOUT

# Notes on conversion ;
# iconv copes with all current characters apart from subscript
# digits, so we translate those with sed first

# Recommend you compare outputs after processing a new release
# baseline release has 18 changed descriptions and all characters
# map to sensible equivalents

# Note : Cygwin iconv mangles accented chars to the accent, then the char
# so use GNU / Linux iconv

RELEASE=$1

sed 'y/0123456789/0123456789/' $RELEASE | \
  iconv -f UTF8 -t ASCII//TRANSLIT
```

Appendix D: Selection of fonts that support subscript 6

Arial Unicode MS

Cambria

Cambria Math

Consolas # a monospaced font, useful for terminal applications

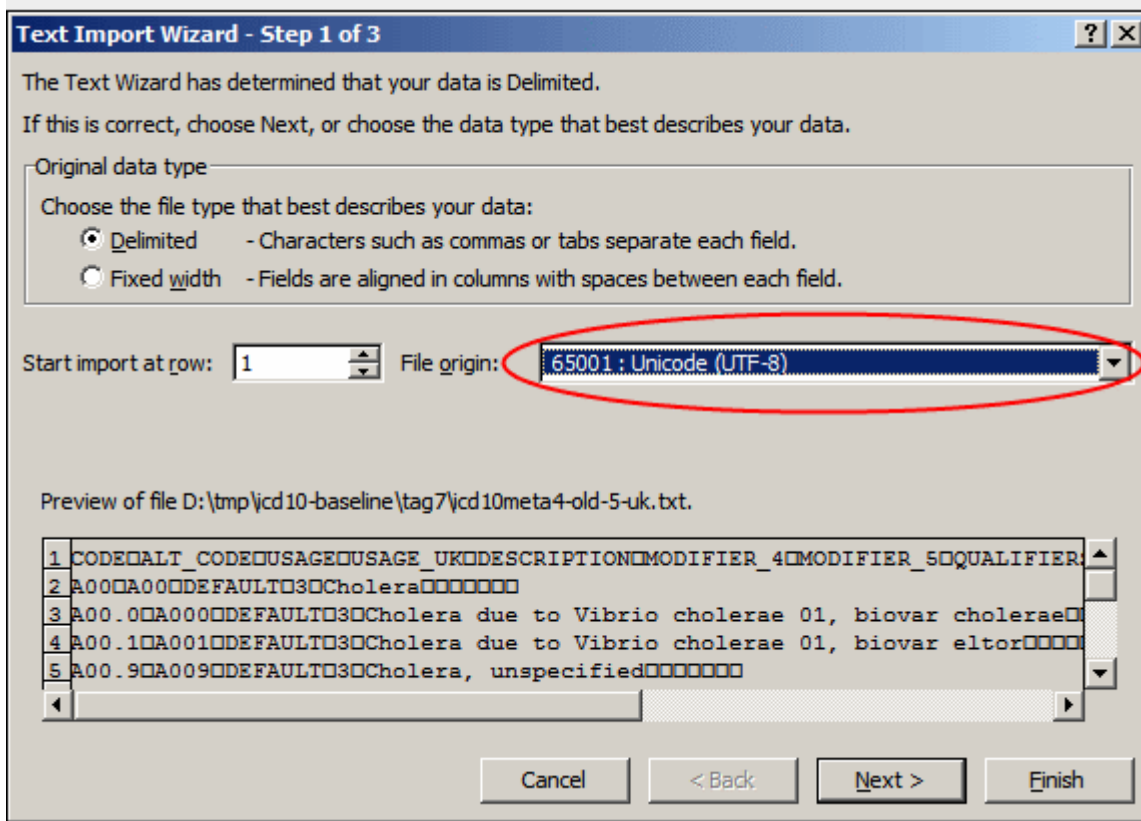
Lucida Sans Unicode

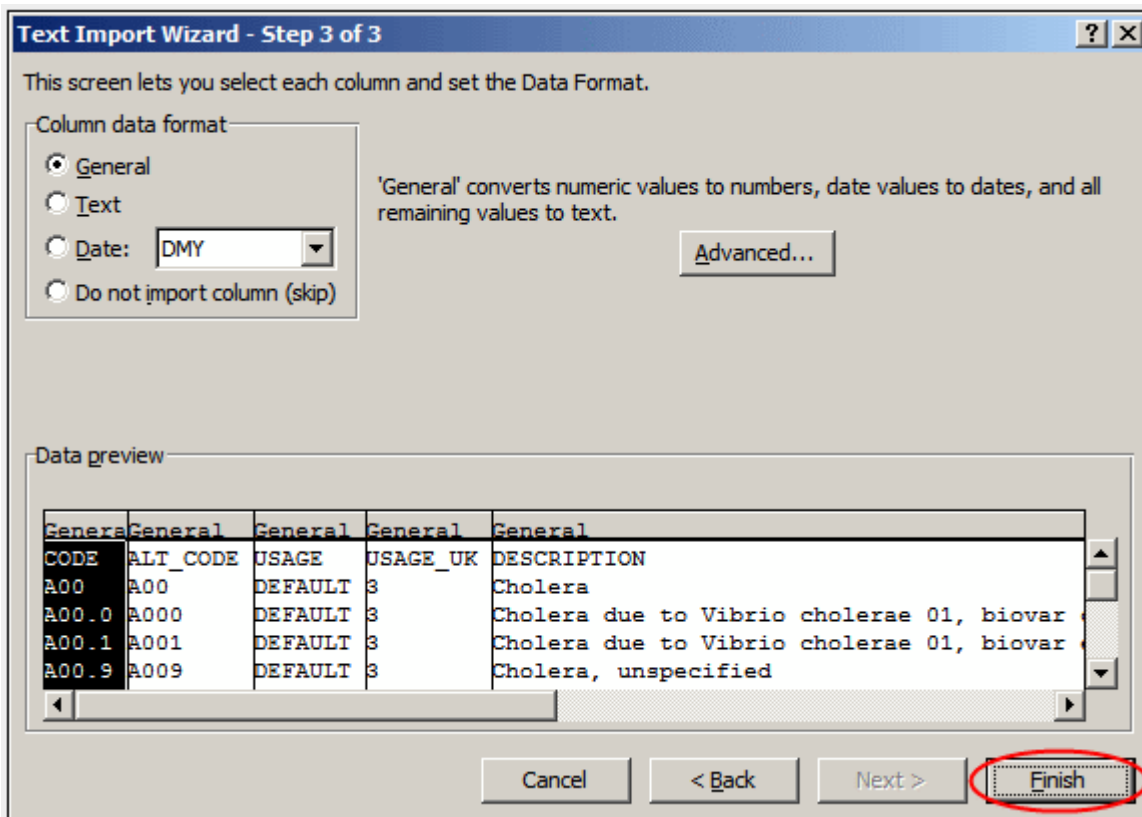
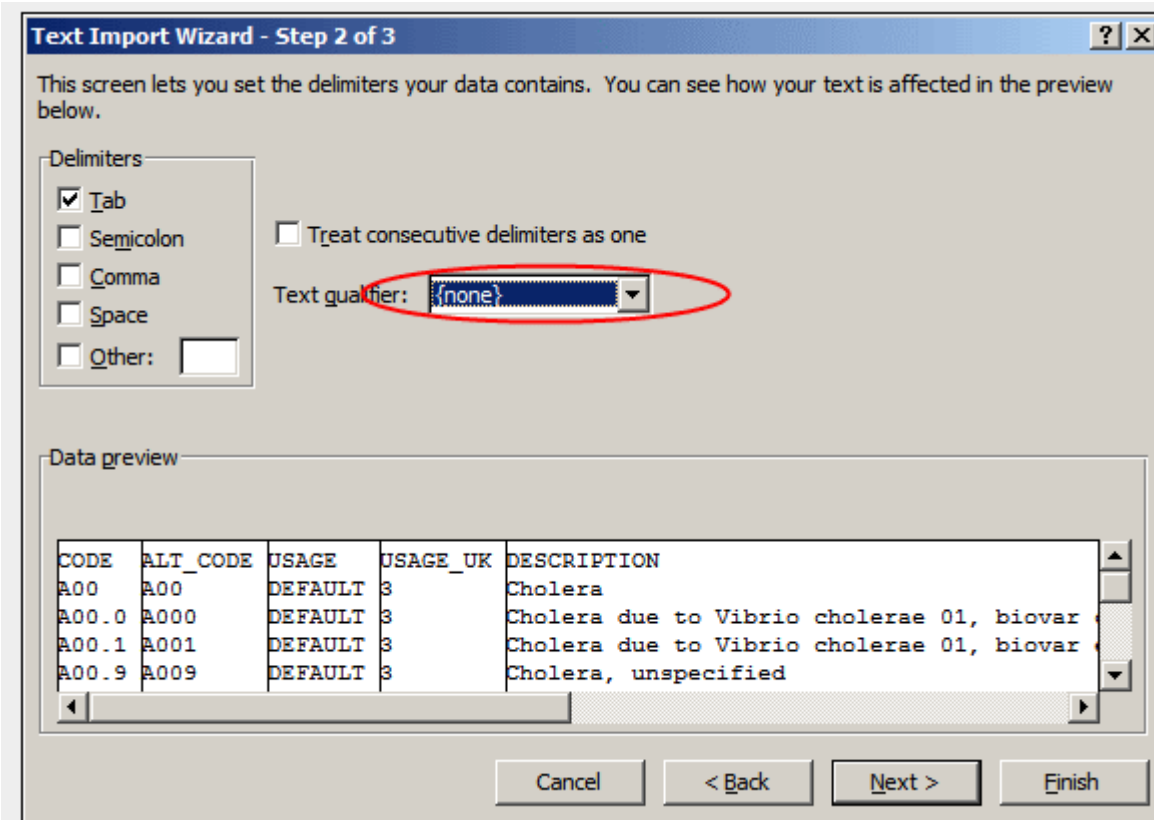
There are many others. You can examine installed fonts by starting the Windows `charmap` tool, grouping by Unicode subrange, choosing the “Super/Subscript” grouping, and verifying that subscript 6 is present.

Annex 1: Viewing the DSV file using Excel data import

Follow the instructions below for a more formal data import. Select the Unicode encoding.

It is NOT recommended that you save the file from Excel because it will break the file structure – even if it is told what it should be. Setting the read-only flag on the file is advised.





Annex 2: USAGE_UK

Value	Description
1	absolute dagger, with asterisk(s) in title
2	absolute asterisk
3	other (non-dagger/asterisk, inclusion dagger or inclusion asterisk)
4	(absolute) dagger, without asterisk(s) in title
0 (zero)	not applicable (previously expressed as space)

Absolute dagger, with asterisk(s) in title (value 1)

B30.0† Keratoconjunctivitis due to adenovirus (H19.2*)

Extract from ICD-10 Codes and Titles and Metadata file (value 1)

COD	ALT_COD	USAGE	USAGE_UK	DESCRIPTION	QUALIFIERS
B30.0	B300	DAGGER	1	Keratoconjunctivitis due to adenovirus	H19.2
B30.1	B301	DAGGER	1	Conjunctivitis due to adenovirus	H13.1
B30.3	B303	DAGGER	1	Acute epidemic haemorrhagic conjunctivitis (enteroviral)	H13.1
B30.8	B308	DAGGER	1	Other viral conjunctivitis	H13.1
B33.4	B334	DAGGER	1	Hantavirus (cardio-)pulmonary syndrome [HPS] [HCPS]	J17.1
B37.3	B373	DAGGER	1	Candidiasis of vulva and vagina	N77.1
B37.5	B375	DAGGER	1	Candidal meningitis	G02.1
B37.6	B376	DAGGER	1	Candidal endocarditis	I39.8
B38.4	B384	DAGGER	1	Coccidioidomycosis meningitis	G02.1
B42.0	B420	DAGGER	1	Pulmonary sporotrichosis	J99.8
B57.0	B570	DAGGER	1	Acute Chagas' disease with heart involvement	I41.2*J98.1
B58.1	B581	DAGGER	1	Toxoplasma hepatitis	K77.0
B58.2	B582	DAGGER	1	Toxoplasma meningoencephalitis	G05.2
B58.3	B583	DAGGER	1	Pulmonary toxoplasmosis	J17.3
B59	B59X	DAGGER	1	Pneumocystosis	J17.3
L40.5	L405	DAGGER	1	Arthropathic psoriasis	M07.0*M07.00*M07
M05.1	M051	DAGGER	1	Rheumatoid lung disease	J99.0

Guidance note (value 1)

In cases where value = 1, it should be combined with field QUALIFIERS (Dual classification (asterisk codes)).

The asterisk codes for VALUE 1 are listed under QUALIFIERS in the ICD10_Edition5_CodesAndTitlesAndMetadata_GB_20160401 because the absolute dagger with asterisk is listed in the code title/description in the ICD-10 Tabular List.

Absolute asterisk (value 2)

Inclusion daggers are not listed in the metadata file

H19.2* Keratitis and keratoconjunctivitis in other infectious and parasitic diseases classified elsewhere

Epidemic keratoconjunctivitis (B30.0†)

Keratitis and keratoconjunctivitis (interstitial) in:

- acanthamoebiasis (B60.1†)
- measles (B05.8†)
- syphilis (A50.3†)
- tuberculosis (A18.5†)
- zoster (B02.3†)

Extract from ICD-10 Codes and Titles and Metadata file (value 2)

COD	ALT_COD	USAGE	USAGE_UK	DESCRIPTION	QUALIFIERS
H19.2	H192	ASTERISK	2	Keratitis and keratoconjunctivitis in other infectious and parasitic diseases classified elsewhere	
H19.3	H193	ASTERISK	2	Keratitis and keratoconjunctivitis in other diseases classified elsewhere	
H19.8	H198	ASTERISK	2	Other disorders of sclera and cornea in diseases classified elsewhere	
H22	H22	ASTERISK	2	Disorders of iris and ciliary body in diseases classified elsewhere	
H22.0	H220	ASTERISK	2	Iridocyclitis in infectious and parasitic diseases classified elsewhere	
H22.1	H221	ASTERISK	2	Iridocyclitis in other diseases classified elsewhere	
H22.8	H228	ASTERISK	2	Other disorders of iris and ciliary body in diseases classified elsewhere	

Guidance note (value 2)

Codes defined with USAGE_UK value 2 must add a dagger code. It must be from the list of codes defined as USAGE_UK values 1, 3 or 4.

Other (non-dagger/asterisk, inclusion dagger or inclusion asterisk) (value 3)

Inclusion asterisk (Value 3):

- B60.1 Acanthamoebiasis**
 Conjunctivitis due to *Acanthamoeba*† (H13.1*)
 Keratoconjunctivitis due to *Acanthamoeba*† (H19.2*)

Inclusion dagger (Value 2):

H58.8* Other specified disorders of eye and adnexa in diseases classified elsewhere

Syphilitic ophthalmopathy NEC:

- congenital:
 - early (A50.0†)
 - late (A50.3†)

- early (secondary) (A51.4†)
- late (A52.7†)

Any VALUE 3 code can be combined with an existing asterisk code ie VALUE 2 to give a dagger and asterisk pair.

Dagger and asterisk combinations are not listed in the metadata file.

Extract from ICD-10 Codes and Titles and Metadata file (value 3)

COD	ALT_COD	USAGE	USAGE_UK	DESCRIPTION	QUALIFIERS
B60	B60	DEFAULT	3	Other protozoal diseases, not elsewhere classified	
B60.0	B600	DEFAULT	3	Babesiosis	
B60.1	B601	DEFAULT	3	Acanthamoebiasis	
B60.2	B602	DEFAULT	3	Naegleriasis	
B60.8	B608	DEFAULT	3	Other specified protozoal diseases	
B64	B64X	DEFAULT	3	Unspecified protozoal disease	
B65	B65	DEFAULT	3	Schistosomiasis [bilharziasis]	
B65.0	B650	DEFAULT	3	Schistosomiasis due to Schistosoma haematobium [urinary schistosomiasis]	

CODE	ALT_CODE	USAGE	USAGE_UK	DESCRIPTION	QUALIFIERS
A49.8	A498	DEFAULT	3	Other bacterial infections of unspecified site	
A49.9	A499	DEFAULT	3	Bacterial infection, unspecified	
A50	A50	DEFAULT	3	Congenital syphilis	
A50.0	A500	DEFAULT	3	Early congenital syphilis, symptomatic	
A50.1	A501	DEFAULT	3	Early congenital syphilis, latent	
A50.2	A502	DEFAULT	3	Early congenital syphilis, unspecified	
A50.3	A503	DEFAULT	3	Late congenital syphilitic oculopathy	

Guidance note (value 3)

Codes defined with USAGE_UK value 3 need to be able to be used in combination with an asterisk code to create a dagger and asterisk combination. The asterisk code must be from the list of codes defined as USAGE_UK value 2. The dagger and asterisk combination may or may not be defined in the ICD-10 Tabular List. Therefore the clinical coder must be able to create a dagger code from any VALUE 3 code and combine it with an asterisk code (VALUE 2) to form a dagger and asterisk combination.

(Absolute) dagger, without asterisk(s) in title (value 4)

Dagger and asterisk combinations not listed in the metadata file.

M32.1† Systemic lupus erythematosus with organ or system involvement

- Libman-Sacks disease (I39.-*)
- Lupus pericarditis (I32.8*)
- Systemic lupus erythematosus with:
 - kidney involvement (N08.5*, N16.4*)
 - lung involvement (J99.1*)

Extract from ICD-10 Codes and Titles and Metadata file (value 4)

COD	ALT_COD	USAGE	USAGE_UK	DESCRIPTION	QUALIFIERS
M05.35	M0535	DAGGER	4	Rheumatoid arthritis with involvement of other organs and systems	
M05.36	M0536	DAGGER	4	Rheumatoid arthritis with involvement of other organs and systems	
M05.37	M0537	DAGGER	4	Rheumatoid arthritis with involvement of other organs and systems	
M05.38	M0538	DAGGER	4	Rheumatoid arthritis with involvement of other organs and systems	
M05.39	M0539	DAGGER	4	Rheumatoid arthritis with involvement of other organs and systems	
M32.1	M321	DAGGER	4	Systemic lupus erythematosus with organ or system involvement	

Guidance note (value 4)

Codes defined with USAGE_UK value 4 must add an asterisk code. The asterisk code must be from the list of codes defined as USAGE_UK value 2.

Not applicable (previously expressed as space) (value 0)

External causes of morbidity and mortality (V01-Y98)

Extract from ICD-10 Codes and Titles and Metadata file (value 0)

COD	ALT_COD	USAGE	USAGE_UK	DESCRIPTION	QUALIFIERS
V01	V01	DEFAULT	0	Pedestrian injured in collision with pedal cycle	
V01.0	V010	DEFAULT	0	Pedestrian injured in collision with pedal cycle	
V01.1	V011	DEFAULT	0	Pedestrian injured in collision with pedal cycle	
V01.9	V019	DEFAULT	0	Pedestrian injured in collision with pedal cycle	