

Document filename:	UK_SNOMED_CT_DMD_How_To_Re-Assemble_Parsed_Descriptors_YYYYMMDD.docx		
Project / Programme	Terminology and Classifications Delivery Service	Project	Pharmacy Terminology Development Service
Document Reference			
Project Manager	Stuart Abbott	Status	Approved
Owner	Emma Melhuish	Version	10.0
Author	Rik Smithies	Version issue date	01/04/2018

How to Re-Assemble Parsed dm+d Descriptor Files

Document management

Revision History

Version	Date	Summary of Changes
1.0	November 2008	Unknown
2.0	April 2009	Unknown
3.0	July 2009	Unknown
4.0	October 2010	Unknown
5.0	July 2012	Unknown
6.0	22-Mar-2013	Rebranded to HSCIC Updates to author and owner
7.0	05-Feb-2014	Minor updates
8.0	17-Sep-2014	Updated contact details
9.0	14-Sep-2016	Rebranded to NHS Digital
10.0	01-Apr-2018	Change UKTC to UK NRC (UK National Release Centre)

Reviewers

This document must be reviewed by the following people:

Reviewer name	Title / Responsibility	Date	Version
Ian Spiers	Pharmacy Terminology Specialist	26-Oct-2010	4.0
Peter Trigg	Technical Consultant	08-Jan-2014	7.0

Approved by

This document must be approved by the following people:

Name	Signature	Title	Date	Version
Stuart Abbott		Senior Terminology Specialist	01-Sep-2016	9.0

Glossary of Terms

Term / Abbreviation	What it stands for
AMP	Actual Medicinal Product
dm+d	NHS Dictionary of Medicines and Devices
VMP	Virtual Medicinal Product
XML	Extensible Markup Language

XSL	Extensible Stylesheet Language
-----	--------------------------------

Contact Information

This document is produced by:

Address:	NHS Digital The UK National Release Centre (UK NRC) 1 Trevelyan Square Leeds LS1 6AE
Telephone:	0300 30 34 777
Email:	information.standards@nhs.net
Internet:	www.digital.nhs.uk

Document Control:

The controlled copy of this document is maintained in the NHS Digital corporate network. Any copies of this document held outside of that area, in whatever format (e.g. paper, email attachment), are considered to have passed out of control and should be checked for currency and validity.

Contents

1	Background	5
2	dm+d Parsed Descriptor Files - VMP Data	6
2.1	Guiding Principles	6
2.2	Naming convention of XML elements in the Parsed Descriptor files	6
2.3	To Re-Assemble the VMP names	7
3	dm+d Parsed Descriptor Files - AMP Data	9
3.1	To Re-Assemble the AMP names	9

1 Background

These files have been produced by parsing the dm+d VMP name and AMP name fields. (Please note the dm+d XML AMP files also contain a file called AMP description which is created for medicinal products by concatenating the AMP name and supplier fields).

Appliance VMPs, VMPs for foodstuffs, VMPs termed Generic (e.g. “Generic Gaviscon”) and Combination packs along with the AMPs linked to these VMP Combination pack concepts have been excluded from these files.

It is important to note that only original dm+d descriptions should be presented for selection in pick lists. However this extra parsed data may be used to structure pick lists and to aid conformance to certain CUI requirements (guidance on these requirements will be produced and distributed separately).

dm+d parsed descriptor files contain terms broken down into identified words and phrases, allowing more flexible use of the data. VMP names and AMP names can be re-built from their constituent parts in the parsed files. For example:

- file f_vmp2_3270907_parsed.xml

Please note - These files are produced in addition to the official dm+d XML files in order to provide extra information with respect to dm+d textual descriptors.

It must be noted that these files are not an officially endorsed dm+d product, but files produced by the Health and Social Care Information Centre (HSCIC) to aid implementers in utilising dm+d data.

2 dm+d Parsed Descriptor Files - VMP Data

2.1 Guiding Principles

With a single exception (identified below) all fields in the VMP parsed files are in the same order as the text appears in the original term.

Hence the basic method¹ is to join all the fields for a given XML record together in "document order". Most fields must be concatenated with a single space inserted between to separate the words, some fields use other delimiters. Empty XML fields need not be concatenated.

2.2 Naming convention of XML elements in the Parsed Descriptor files

A given term can have several "moieties" in the parsed record and each "moiety phrase" has several fields describing it. It should be noted that the first moiety has the possibility of being described by a second strength (where the moiety relates to multiple ingredients) or a dual expression of strength (where a single strength is expressed in two ways).

The different moieties in the term are separated with a forward slash surrounded by two spaces i.e. " / ".

An example with two moiety phrases is:

- Coal tar 2.5% / Lecithin 0.3% scalp lotion

The moiety phrases here are "Coal tar 2.5%" and "Lecithin 0.3%". The words "scalp lotion" apply to the term as a whole rather than to either moiety.

The XML records have repeating groups of fields, one per moiety phrase in the original term. These groups of fields are numbered with a suffix of an underscore plus the number. In fact the first group is itself un-numbered but has the base name, so the result is:

```
MOIETY
SVN
SVU
...
MOIETY_2
SVN_2
SVU_2
...
MOIETY_3
```

and so on.

Some fields have an intrinsic number e.g.

¹ The basic method could potentially be varied to comply with CUI requirements, such as having two spaces between words.

- Strength2, for the secondary, bracketed, strength in a moiety phrase
e.g. Adrenaline 1mg/ml (1 in 1,000) topical solution
- A second strength2, in a second Moiety phrase would use an underscore and be Strength2_2
e.g. Lidocaine 400mg/20ml (2%) / Adrenaline 100micrograms/20ml (1 in 200,000) solution for injection 20ml vials

The expressed numeric strength in a moiety can be a ratio, and the numerator, possible secondary numerator and denominator are also separated by forward slashes, this time however without a space either side. For example:

- "3mg" as numerator and "0.3ml" as denominator in the term:
Ranibizumab 3mg/0.3ml solution for injection vials

2.3 To Re-Assemble the VMP names

All non-empty fields for a given term (XML VMP element) are concatenated in XML document order. Between moiety phrase groups, a space-slash-space sequence should be inserted (" / ").

Preceding a second strength numerator (e.g. SVN2) or a denominator (e.g. SVD or SVD_2) a single un-spaced slash should be inserted ("/").

Other fields are concatenated with a single space in between (" "), with the exception that units fields (e.g. SVU, SVU_2, SDU etc) are not prefixed with spaces. They join directly onto the preceding numeric field (e.g. "25mg"), with just three exceptions. The three units that expect a space prefix to aid readability of terms are:

- Kallikrein inactivator units
- SQ-T
- tuberculin units

eg Aprotinin 3,000 Kallikrein inactivator units/1ml solution for irrigation vials

An additional requirement is that the secondary strength field (e.g. Strength2, Strength2_2) must be surrounded by round brackets, in addition to the space in front of it

e.g. Lidocaine 100mg/10ml (1%) solution for injection pre-filled syringes

There are a small number of identified exceptions to the above rules:

Some terms do not use " / " to separate the moieties, but instead use the word " in ". This applies to "half strength Hartmann's", as in "Glucose 5% *in* half strength Hartmann's solution for injection 500ml bags", and to "aqueous cream" as in "Menthol 0.5% *in* aqueous cream" and other similar percent strengths.²

² The full list of terms is:

- 5121111000001106 Glucose 5% in half strength Hartmann's solution for injection 1litre bags
- 5121411000001101 Glucose 5% in half strength Hartmann's solution for injection 500ml bags

One term doesn't fit the rules above relating to word order:

- Muromonab-CD3 solution for injection 5mg/5ml ampoules

This term has been authored with the dose form before the strength rather than after, so as to separate the numeric part of the name from the numerics denoting the strength of the product. To illustrate, by the standard naming this term would be "Muromonab-CD3 5mg/5ml *solution for injection* ampoules". This must be considered when reassembling.

Note that the full expression of these rules has been coded into the parsed file displaying stylesheet, dm+dVMPParsedViewer.xsl, in the template identified: `<xsl:template match="TERM" mode="reassemble">`.

3 dm+d Parsed Descriptor Files - AMP Data

The AMP parsed file is very similar to the VMP one. The same criteria for record inclusion apply. XML element names are as for the VMP with a few exceptions.

The AMP file has additional fields:

DELIM
FLAVOUR
SIZE
OTHER__AMP

DELIM is used to represent the joining characters that connect the first and second moiety phrases. In a VMP this is almost always the text “ / “, as described above (rarely, it is “in”). Since there is more variety in this part of AMPs, specifically many having the word “with” instead of a slash, a special field is used to show what is present. This can be re-inserted where the “ / “ would be in a VMP, and will then correctly show “ / “, “ with “, “ or “, “ + “ etc. as appropriate. There is only one DELIM field as the issue only arises between first and second moiety phrases. Second and third phrases, which are rare with AMPs, are always joined with “ / “ as with VMPs.

Where an AMP includes flavour wording or an explicit size, or gauge, FLAVOUR and SIZE will be populated accordingly.

There is a greater variety of text in AMPs compared to VMPs. Consequently it is necessary to have un-parsed text at the end of some AMPs, that doesn't fit into any other category. Some examples are “for children”, “film coated”, and “P” (pharmacy).

3.1 To Re-Assemble the AMP names

AMP names are re-assembled using the same method as the VMP names with the exception of using the DELIM field, documented above.

The method to re-assemble both VMP names and AMP names is also laid out in the XSL stylesheet supplied with the files (see VMP section of this document).