Technical Report
UK Edition: NHS Realm Description Refsets
Published May 2016
## Contents

### Introduction 3
- Purpose: 3
- Audience: 3

### Background 4

### LIMITATIONS 5
1. One product, two refsets, two files 5
2. Omissions 5

### SAMPLE IMPLEMENTATION (MYSQL) 6

### Additional Refinement 7
Introduction

Please note this document is a draft. We would be grateful if any suggestions for improvement, for clearer guidance or matters of accuracy were sent to denise.downs@nhs.net.

Purpose:

This document illustrates how to use the UK Realm Description Refsets to identify the appropriate descriptions from the full set of descriptions provided in the UK Edition of SNOMED CT.

Audience:

This document is written for system implementers who need to configure the SNOMED CT RF2 release data for use in UK systems.
Background

The SNOMED CT RF2 distribution of the UK Edition includes three files that when loaded into a single database table contains all the descriptions within the UK Edition:

- `sct2_Description_Snapshot-en_INT_yyyymmdd` from the IHTSDO International Edition
- `sct2_Description_Snapshot-en-GB_GB1000000_yyyymmdd` from the UK Clinical Extension
- `xsct2_Description_Snapshot-en-GB_GB1000001_yyyymmdd` from the UK Pharmacy Extension

As loaded, these descriptions are initially one of two types (indicated by the data item: Typeld): Fully Specified Name (FSN) and Synonym. (Note. As per RF2 format this metadata is also a SNOMED CT concept, FSN has a conceptId of 900000000000003001 and synonym a conceptId of 900000000000013009). At this stage of the data load, therefore, there is no notion of which synonym is the preferred term. Additionally, some concepts may be associated with more than one Fully Specified Name.

The NHS realm description Language Refset (NHS-RLS) resolves this issue and also provides UK English acceptable descriptions for the preferred term. Its intended behaviour is that, when used alongside the raw data according to the algorithm in this briefing, all descriptions will be further categorised so that:

- Every concept has exactly one current, active Fully Specified Name (FSN)
- Every concept has exactly one current, active Preferred Term (PT)
- Every concept has between zero and many synonyms

The NHS realm description refset is provided via two files; one in the Clinical Extension and one in the Drug Extension. Equivalents are also provided in the RF1 release as subsets. The files referred to are:

<table>
<thead>
<tr>
<th>Subset Name</th>
<th>Subset Original Id</th>
<th>Refset FSN</th>
<th>Refset Id</th>
<th>Refset Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHS realm description subset (pharmacy part)</td>
<td>238301000001135</td>
<td>National Health Service realm language reference set (pharmacy part) (foundation metadata concept)</td>
<td>999000691000001104</td>
<td>Component</td>
</tr>
<tr>
<td>NHS realm description subset (clinical part)</td>
<td>39141000000139</td>
<td>National Health Service realm language reference set (clinical part) (foundation metadata concept)</td>
<td>999001261000000100</td>
<td>Component</td>
</tr>
</tbody>
</table>
LIMITATIONS

The April 2016 release of the NHS Realm Language Refset has the following known limitations:

1. One product, two refsets, two files

Although conceived as a single data artefact, it is currently actually published in two parts; ie. as two different SNOMED CT Reference Sets and thus two different files. One is in the Clinical Extension and one in the Drug Extension. Some concepts provided in the UK Clinical Extension have the preferred term and/or synonyms located within the Drug Extension.

2. Omissions

The April 2016 release of the NHS-RLS has some known omissions and does not quite achieve the ‘one FSN, one PT’ goal for the full Edition due to the inclusion of metadata in the RF2 release. So

- 65 concepts have no Fully Specified Name
- 110 concepts have no Preferred Term

All affected concepts are within the Metadata hierarchy of SNOMED CT (they’re almost exclusively component identifiers for new reference sets). Since clinicians would not normally be expected to have (or need) access to the metadata content, this is anticipated to have little or no impact. However, a suggested solution should these be required is included here.
SAMPLE IMPLEMENTATION (MYSQL)

The following code will result in a descriptions table for the full UK Edition of SNOMED CT that has two additional columns:

- One which indicates if this description is acceptable within the UK or not;
- One which indicates for those acceptable whether the description is an FSN, a PT or a synonym. Note. Currently there is no Welsh translation but this may be developed in the future in which case it is likely that there may be a Welsh PT provided by a Welsh Extension to the UK Edition.

USE rf2_snap_20161001;

# Identifiers for both parts of the NHS Realm Language Reference Set
SET @RLSUKClin = '999001261000000100'; /* National Health Service realm language reference set (clinical part) */
SET @RLSUKPharm = '999000691000001104'; /* National Health Service realm language reference set (pharmacy part) */

# Permitted values for ACCEPTABILITY in Realm Language reference sets
SET @Acceptable = '900000000000549004';
SET @Preferred = '900000000000548007';

# Permitted values for TYPEID in sct2_descriptions table
SET @FSN = '900000000000003001';
SET @SYNONYM = '900000000000013009';

# Add an additional column to the ordinary sct2_description table to hold what kind of description the term is within the NHS
# F = Fully Specified Name
# P = Preferred Term
# S = Synonym
# ? = none of the above

ALTER TABLE sct2_description ADD COLUMN DescriptionType VARCHAR(1);

UPDATE sct2_description t LEFT JOIN sct_refset_c rls ON t.ID = rls.referencedComponentId
SET t.DescriptionType = IF(rls.Attribute1 = @Preferred,IF(t.TYPEID = @FSN,'F','P'),
 IF(rls.Attribute1 = @Acceptable,IF(t.TYPEID = @FSN,'?','S'),'?'))
WHERE rls.REFSETID IN (@RLSUKClin, @RLSUKPharm)
AND rls.active=1
# Additional Refinement

*OPTIONAL* Patches to correct the data errors

**Patch for no FSN**: list all affected concepts in a temporary table and, for each, promote their FSN to being the FSN.

```
USE rf2_snapshot_20161001;
```

- **Permitted values for TYPEID in sct2_descriptions table**
  ```
  SET @FSN = '900000000000003001';
  SET @SYNONYM = '900000000000013009';
  ```

- **This table is required in MySQL as it is not possible to update**
  - a table that is used in a subquery
  - it may be possible in other DBMS environments to directly update the `sct2_description` table.

```
DROP TABLE IF EXISTS DescriptionTypeTemp;
CREATE TABLE DescriptionTypeTemp (Id BIGINT NOT NULL DEFAULT 0, DescriptionType VARCHAR (1));
```

- **Populate DescriptionTypeTemp with descriptions where there is only one FSN per concept and**
  - this does not have a `descriptionType` of `F`

```
INSERT INTO DescriptionTypeTemp
SELECT d.id, 'F'
FROM sct2_description d
INNER JOIN sct2_concept c
ON d.conceptid=c.id
WHERE d.conceptId in (SELECT c.conceptid FROM sct2_description
WHERE typeid=@FSN
GROUP BY conceptid
HAVING count(*)=1)
AND typeid=@FSN
AND d.active=1
AND c.active=1
AND descriptionType is null;
```

- **Add to DescriptionTypeTemp those descriptions where there is only one synonym per concept and**
  - this does not have a `descriptionType` of `P`

```
INSERT INTO DescriptionTypeTemp
SELECT d.id, 'P'
FROM sct2_description d
INNER JOIN sct2_concept c
ON d.conceptid=c.id
WHERE d.conceptid in (SELECT conceptid FROM sct2_description
WHERE typeid=@SYNONYM
GROUP BY conceptid
HAVING count(*)=1)
AND typeid=@SYNONYM
AND d.active=1
AND c.active=1
AND descriptionType is null;
```

- **Update sct2_description table**

```
UPDATE sct2_description d
JOIN DescriptionTypeTemp t
ON d.id=t.id
SET d.descriptionType=t.descriptionType;
```