

Concept set SOP:

This document outlines the standard operating procedure that should be followed to create a concept set. The goal of this SOP is to create concept sets that are understandable and reproducible.

There are three general steps:

1. Identify standard concept_id and descendants for all given codes
2. Review by a clinician
3. Upload to N3C enclave

For demonstration purposes Diabetes will be used in this example. The starting codes were given for T1D as part of the Q4 autoimmune query are used as an example.

1. Create a list of concepts from given codes

1.1 A Standard format is required for clinicians to submit list of conditions

A standard format to be submitted by clinicians to request conversion of identified codes defining a condition to Standard codes.

Table 1:

M	M	M	O	M= Mandatory, O = Optional		
Concept Set Name	Vocabulary	Code	Comment			
Endocrine: T1D	SNOMED	31321000119102	Diabetes mellitus type 1 without retinopathy			
Endocrine: T1D	SNOMED	23045005	Insulin dependent diabetes mellitus type 1A			
Endocrine: T1D	SNOMED	28032008	Insulin dependent diabetes mellitus type 1B			
Endocrine: T1D	SNOMED	426875007	autoimmune diabetes in adult			
Endocrine: T1D	SNOMED	60956600	Pregnancy - type 1 diabetes mellitus			
Endocrine: T1D	SNOMED	46635009				
Endocrine: T1D	ICD	E10	Type 1 diabetes			
Endocrine: T1D	SNOMED	190372001	Maturity Onset			
Endocrine: T1D	SNOMED	314893005	Arthropathy			
Endocrine: T1D	SNOMED	190369008	w/gangrene			
Endocrine: T1D	ICD	E10.65	w/hyperglycemia			
Endocrine: T1D	SNOMED	190330002				

Columns: Concept Set Name(M), Vocabulary(M), Code(M), Comment(O)

1.2 Fusion Sheet for Processing ICD10/SNOMED Codes Template:

Table 1 will be used for processing. A Condition Group Number can be added to differentiate between multiple conditions if given.

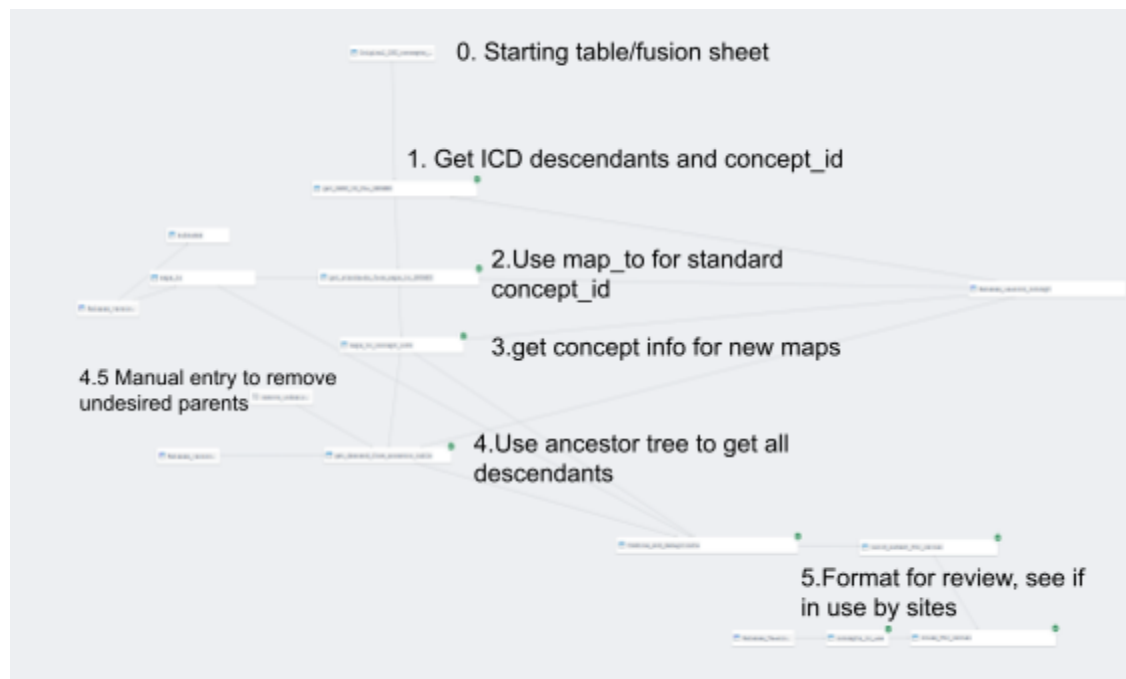
In the example below 15 was assigned to the Endocrine T1D condition.

Table 2:

M	M	M	M	O	M= Mandatory, O = Optional
Group Number	Concept Set Name	Vocabulary	Code	Comment	
15	Endocrine: T1D	SNOMED	31321000119102	Diabetes mellitus type 1 without retinopathy	
15	Endocrine: T1D	SNOMED	23045005	Insulin dependent diabetes mellitus type 1A	
15	Endocrine: T1D	SNOMED	28032008	Insulin dependent diabetes mellitus type 1B	
15	Endocrine: T1D	SNOMED	426875007	autoimmune diabetes in adult	
15	Endocrine: T1D	SNOMED	60956600	Pregnancy - type 1 diabetes mellitus	
15	Endocrine: T1D	SNOMED	46635009		
15	Endocrine: T1D	ICD	E10	Type 1 diabetes	
15	Endocrine: T1D	SNOMED	190372001	Maturity Onset	
15	Endocrine: T1D	SNOMED	314893005	Arthropathy	
15	Endocrine: T1D	SNOMED	190369008	w/gangrene	
15	Endocrine: T1D	ICD	E10.65	w/hyperglycemia	
15	Endocrine: T1D	SNOMED	190330002		

1.3 Workbook Processing:

- Copy Table 2 to a fusion sheet within the enclave and import into a workbook.
- Import template [insert name, finish making], which works as described:



0. The starting code table described in the previous step is imported into the workbook
1. The given codes are joined on the concept table to convert from the given vocabulary to an OMOP id.
 - a. At this step, any ICD codes are also mapped to descendants using regular expression. ex: E10 is given, so E10.1, E10.11, E10.2, etc. are selected
 - b. If for any code a valid concept_id cannot be found, it is removed from the list of concepts

2. Using map_to from the relationship table, the standard concept is mapped
3. Concept name, standard, and invalid reason are retrieved from the concept table for the newly mapped concepts.
4. The now standard concepts are joined to the ancestor table to get all descendants (children, grandchildren, etc.)
 - a. Due to the way ICD codes map to standard, this may retrieve a large amount of unrelated descendants. To combat this, a manual entry is included where a parent_concept_name can be provided to exclude all concepts which are retrieved due to that parent. This table can be updated after running “get_descend_from_ancestor_table” (step 4), but will require rerunning step 4 for the exclusions to take effect.
5. The final table of concepts is formatted and checked against current site usage.
 - a. The table produced is **not** deduplicated and may include the same concept_id multiple times if it is mapped by multiple starting codes, or the descendant of multiple starting codes. The column instance number is provided if a single instance of each concept is more important than knowing the inclusion source.

Final table will have columns:

- A. **Instance:** a numeric value indicating how many times the concept_id has appeared.
- B. **Concept_Group_Number:** numerical indicator of which group a concept is in.
- C. **Concept_Set_Name:** name given to a concept set from the original table. Each Concept_Set_Name is assigned a different Concept_Group_Number.
- D. **Starting_code:** code given in the original table that is the parent or mapping code.
 - a. Some codes may have multiple starting_codes. The “minimum” code is reported.
- E. **Concept_ID:** OMOP concept_id
- F. **Concept_Name:** Name for concept_id
- G. **Standard:** S if concept is standard, *null* if concept is not standard.
 - a. Note: All concepts given at this point will be standard. Non-standard concepts have much lower use, and could be picked up by including “mapped”.
- H. **Site_count:** The number of sites that have at least one use of a concept_id

2. Review by a clinician

The table created in the code workbook [can be copied *approval?]* into a spreadsheet for clinical review. This spreadsheet will serve as a record of which concepts are removed and why. **Do not delete concepts from the spreadsheet**, instead, indicate in the “final inclusion” column whether the concept should be included (Y) or excluded (N).

Review can be aided by filter and sorting on several columns.

- **parent_concept_name** reflects the concept that was used as the parent/ancestor when looking for descendants. In some cases, it may be appropriate to exclude all descendants of a parent concept. This can be easily accomplished by sorting or filtering **parent_concept_name**, then quickly filling out the inclusion column with “N”.
 - Ex: disease of mouth, periodontal disease, etc. can be excluded from a type 1 diabetes concept set.
- Several instances of a **concept_id** may be included if multiple ancestors of that concept were included. To avoid reviewing the same concept multiple times, the **instance** column can be filtered or sorted.

Table 3:

	A	B	C	D	E	F	G	H
1	instance	Concept_Gr	Concept_Set_Name	starting code	parent_concept_name	concept ID	Concept Name	standard
2	1 15	Endocrine: T1D	E10.0	Coma due to diabetes mellitus	201530	Hyperosmolar coma due to type 2 diabetes mellitus	S	
3	1 15	Endocrine: T1D	E10.0	Coma due to diabetes mellitus	201531	Hyperosmolar coma due to type 1 diabetes mellitus	S	
4	1 15	Endocrine: T1D	E10.0	Coma due to diabetes mellitus	443735	Coma due to diabetes mellitus	S	
5	1 15	Endocrine: T1D	E10.0	Coma due to diabetes mellitus	601160	Hyperosmolar coma due to drug induced diabetes mellitus	S	
6	1 15	Endocrine: T1D	E10.0	Coma due to diabetes mellitus	4095288	Ketoacidotic coma due to diabetes mellitus	S	
7	1 15	Endocrine: T1D	E10.0	Coma due to diabetes mellitus	4221933	Coma due to malnutrition-related diabetes mellitus	S	
8	1 15	Endocrine: T1D	E10.0	Coma due to diabetes mellitus	4223734	Non-ketotic non-hyperosmolar coma due to diabetes mellitus	S	
9	1 15	Endocrine: T1D	E10.0	Coma due to diabetes mellitus	4224254	Ketoacidotic coma due to type 1 diabetes mellitus	S	
15	1 15	Endocrine: T1D	E10.0	Coma due to diabetes mellitus	4226238	Hyperosmolar coma due to diabetes mellitus	S	
28	1 15	Endocrine: T1D	E10.0	Coma due to diabetes mellitus	4226798	Hypoglycemic coma due to diabetes mellitus	S	
44	1 15	Endocrine: T1D	E10.0	Coma due to diabetes mellitus	4228112	Hypoglycemic coma due to type 1 diabetes mellitus	S	
51	1 15	Endocrine: T1D	E10.0	Coma due to diabetes mellitus	4228443	Ketoacidotic coma due to type 2 diabetes mellitus	S	
64	1 15	Endocrine: T1D	E10.0	Coma due to diabetes mellitus	36714116	Hypoglycemic coma due to type 2 diabetes mellitus	S	

3. Upload to enclave

Table 4:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ID	NAME	concept_set_id	concept_set_name	project_id	assigned_informatician	assigned_sme	status	stage	intention	rtc_reviewer	alias	archived	created_by	created_at	INCLUDED
2		333333 another concept														Y
7		444444 something														Y
8		123456789 Ankylosing spondylarthritis and eye lesions														Y
9																
10																
19																
20																
21																

Table 4 can be generated from the included concepts in Table 3. Additional column information is required.

Table setup required for upload. Describe columns, optional/required.

please include any relevant information