SDTM-ETL 4.0 User Manual and Tutorial

Author: Jozef Aerts, XML4Pharma

Last update: 2019-01-28



Subsetting and Extending Codelists

In SDTM and SEND, you will often want and need to make subset of codelists published by CDISC. Typical examples are the "UNIT" codelist and the "LBTESTCD" and "LBTEST" codelists. When doing so, you will need to create a list of all terms that either were planned (e.g. taken from the CRF) or that appeared in the results (e.g. units for lab tests). Also, you might have items that are not represented in the corresponding CDISC codelist, and then you will want to add (codelist extensions).

Creating such "subset-extended" codelists is normally not an easy task, but with the new features in SDTM-ETL 4.0, it becomes "piece of cake".

In this tutorial, we will create a subset of the "LBTESTCD" and "LBTEST" codelist for a few blood tests. We will only retain "leukocytes", "erythrocytes" and "platelets", "mean platelet volume" and "platelet distribution width" (all already present in the "LBTESTCD" and "LBTEST" codelists) and add a new non-CDISC term "<u>plateletcrit</u>" which is not covered by CDISC controlled terminology. The latter is covered by LOINC (e.g. LOINC code <u>51637-7</u>) but CDISC still refuses to allow to use LOINC codes as unique identifiers for lab tests.

In order to generate a subset/extended codelist, use the menu "Insert – Create new SDTM/SEND CodeList from existing CodeList". The following dialog appears:

Create new SDIM CodeList from	n existing CodeList					
?	Find CodeList					
	CL.C66742.Y - No Yes Respon	ise (Yes or	ıly)			-
	New OID:					
	New Name:					
	New DataType: text					-
	New SASFormatName:					
	Remove rows until only those Several rows can be removed	remain tha I simultane	at you want to appear in the CodeL eously.	ist.		
	You can also add new rows.					
		Insert	row		Remove row	
	CodedValue		Rank	NCI Code	ExtendedValue	
	Y			C49488		-
						· ·
			Search in CodeList	Sort CodeList Items on exit		
			ОК	Cancel		

You can then search for the codelist. In our case, we use the field and button "Find CodeList" using the search term "laboratory":

Find CodeList	laboratory
CL.C66742.Y - No Yes	Response (Yes only)
New OID:	
New Name:	
New Dete Trees	44

When clicking "Find CodeList", we first find:

Find CodeList	laboratory					
CL.C102580.LBSTRESC	- Laboratory Test §	standard Character Resu	lt			
New OID:	CL.C102580.LBST	RESC.SUBSET				
New Name:	Laboratory Test Sta	ndard Character Result s	ubset			
New DataType:	text	text				
New SASFormatName:						
Remove rows until only Several rows can be rea	those remain that y moved simultaneou	ou want to appear in the sly.	CodeList.			
You can also add new re	0 ws.					
	Insert rov	v			Remove	
CodedValu	ie	Rank		NCI Code		
BORDERLINE			C141	157		
INDETERMINATE			C486	658		
INVALID			C509)13		
NEGATIVE			C387	/57		
POSITIVE			C387	758		

But this is not the one we want. Again clicking "Find CodeList" leads to the codelist that we want to subset:

Find CodeList	laboratory					
CL.C65047.LBTE STCD	Laboratory Te	st Code				
New OID:	CL.C65047.LE	BTESTCD.SUBSET				
New Name:	Laboratory Tes	st Code subset				
New DataType:	text	text				
New SASFormatName:						
Remove rows until only Several rows can be re	those remain t moved simultar	hat you want to appear in the Code neously.	List.			
You can also add new r	ows.					
	Inse	rt row				
CodedValu	le	Rank	NCI Code			
A1AGLP C100429						
A1ANTRYP C80167						
A1MCREAT			C100462			
A1MICG			C100461			
A0MACC C00460						

As you however see, most of the codes are not very well understandable, so it will be very hard to find the ones we need. This is due to the fact that CDISC publishes most of the codelists just as "enumerated" lists, without the human-readable "decode". In such cases, the "code" and "decode" are usually connected by the "NCI code".

Each of the entries in the table has a tooltip showing the decode (retrieved by using the NCI code):

	CodedValue	Rank
A1AGLP		
A1ANTRYP		
A1MCREAT		
A1MICG	A1ANTRYP: Alpha-1 Ant	titrypsin
A2MACG		
AAMAPAC		
AAMBTAC		

But even then, it would be very difficult to quickly find a code like for "erythrocytes".

Therefore, when such an "enumerated" codelist is found, and there is a corresponding codelist with "decodes" (this will usually be the case for pairs of "TESTCD"-"TEST" codelists), the system will ask you:

Find CodeList laboratory					
CL.C65047.LB	TESTCD - Laboratory Test Code				
New OID:	Use decodes X				
New Name: New DataType: New SASForma	A corresponding codelist named Laboratory Test Name contains the 'decoded' values for the current codelist.				
Remove rows u Several rows c You can also ac	Do you want to use the 'decodes' from this corresponding codelist to be used in the generation of the current (subset) codelist? (this has major advantages when searching and selecting items)				
	<u>Y</u> es <u>N</u> o				

When clicking "OK", the "decodes" will be retrieved and also be presented and made "searchable". In our case this leads to:

Find CodeList laboratory									
CL.C65047.LBTE STO	CL.C65047.LBTESTCD - Laboratory Test Code								
New OID:	CL.C65047.LE	TESTCD.SUBSET							
New Name:	Laboratory Tes	at Code subset							
New DataType:	text					-			
New SASFormatNam	ne:								
Remove rows until o Several rows can be You can also add new	nly those remain t removed simultar w rows.	hat you want to appear in the CodeList. neously.							
	Inser	t row		Remove row					
CodedValue	Language	Decode	Rank	NCI Code	ExtendedValue				
A1AGLP		Alpha-1 Acid Glycoprotein		C100429					
A1ANTRYP		Alpha-1 Antitrypsin		C80167					
A1MCREAT		Alpha-1 Microglobulin/Creatinine		C100462					
A1MICG		Alpha-1 Microglobulin		C100461					
A2MACG		Alpha-2 Macroglobulin		C80168					
AAMAPAC		Alpha-Aminoadipic Acid		C154761					
AAMBTAC		Alpha-Aminobutyric Acid		C154759					
AAP		Alanine Aminopeptidase		C100430					
ABNCE		Abnormal Cells		C111124					
ABNCECE		Abnormal Cells/Total Cells		C150835					
ABNCELE		Abnormal Cells/Leukocytes C150834							
ABO		ABO Blood Group C125939							
ABOA1		ABO A1 Subtype		C135397					
ACANT		Acanthocytes		C74699					
ACANTRBC		Acanthocytes/Erythrocytes		C74633					

Remark that an additional column "Decode" appears containing the corresponding "test name" which comes from the "LBTEST" codelist.

Depending on your system configuration and the way the codelists have been installed, it may be that the "LBTESTCD" codelist DOES already contain the "decodes", so that this step is not necessary.

It now becomes easy to find "leukocytes", "erythrocytes" etc.. In order to do so, use the button "Search in CodeList" near the bottom:

Search in CodeList	Sort CodeList Items on exit
ОК	Cancel

This will add a new dialog:

Find CodeList	laboratory				
CL.C65047.LBTE STCD -	Laboratory Test Code	Search			×
New OID:	CL.C65047.LBTESTCD.SUB	Jearch			
New Name:	Laboratory Test Code subse	?	Search for:	Search	Find Next
New DataType:	text	_	Match case Whole words only	Add to list	Display list
New SASFormatName:					
Remove rows until only Several rows can be ren	those remain that you want t moved simultaneously.			Recreate ta	ble from list
You can also add new ro	ows.		Close		
	Insert row				

In which we can start searching, e.g. for "leukocytes":

Search			×
?	Search for: leukocytes	Search	Find Next
	Match case Whole words only	Add to list	Display list
		Recreate ta	ble from list
	Close		

Clicking the "Search" button will start a search from the top of the list, clicking "Find Next" will start searching from the current position in the list.

First clicking the "Search" button shows us a first possibility:

You can also add ne	Search			×
CodedValue	?	Search for: leukocytes	Search	Find Next
A1AGLP	-	Match case Whole words only	Add to list	Display list
A1ANTRYP				
A1MCREAT				
A1MICG			Recreate ta	ble from list
A2MACG				
AAMAPAC		Close		
AAMBTAC		01030		
AAP		Alanine Aminopepiidase		C100430
ABNCE		Abnormal Cells		C111124
ABNCECE		Abnormal Cells/Total Cells		C150835
ABNCELE		Abnormal Cells/Leukocytes		C150834
ABO		ABO Blood Group		C125939

"Abnormal Cells/Leukocytes" which however is not what we want. After several times clicking "Find Next" we get to:

VITECHOL		Vitamin E/Cholesterol		C103448
VITK	Search			×
VITK1				
VLDL	2	Search for: leukocytes	Search	Find Next
VLDL1	•			
VLDL2		Match case Whole words only	Add to list	Display list
VLDL3				
VLDLPSZ				
VMA			Recreate ta	able from list
VOLUME			nooroato a	
VTD2D3IT		Close		
VWFAAC		Close		
VZVAB	L	Vancena Zoster Viras / Viras day		0111001
VZVIGAAB		Varicella Zoster Virus IgA Antibody		C98795
VZVIGGAB		Varicella Zoster Virus IgG Antibody		C98796
VZVIGMAB		Varicella Zoster Virus IgM Antibody		C98797
WBC		Leukocytes		C51948

Which is the one we need. Remark that using the checkbox "Whole words only" would have speed up the process and considerably decreased the number of clicks.

As we want to create a "positive" list, we click "Add to list". At any time, the (temporary) list can be displayed using the "Display list" button:

Lanyuaye		Decode	Rdiir	Rdin NO		
_		Vitamin E/Cholesterol		C103448	Items to k	ceep X
-	Search			×	(i)	WBC - Leukocytes
-	? Search fo	r: leukocytes	Search	Find Next	0	
_		Match case 🔲 Whole words only	Add to list	Display list		
-			Recreate ta	ble from list		
		Close				Remove item from list
4		Valicella Zuster Virus Antibudy		0117007		
		Varicella Zoster Virus IgA Antibody		C98795		OK
		Varicella Zoster Virus IgG Antibody		C98796		
		Variable Zester Virus Jaki Antihadu		000707		

We can now repeat the same process for "erythrocytes". When an error is made and an item is accidently added to the list, it can always be removed again using the "Remove from list" button. For example:

ltems to k	eep X
i	WBC - Leukocytes G6PDRBC - G6PD-Deficient Erythrocytes RBC - Erythrocytes
	Remove item from list
	ок

We do not want to have "Deficient Erythrocytes" in the list, so we just select it and then click "Remove item from list".

We now also add "platelets", "mean platelet volume" and "platelet distribution width" in the same way as before. Using "Display list" this leads to:

Search X	
? Search for: platelet distribution width Search Find Next	
Match case Whole words only Add to list Display list	
Close Items to keep	×
Insert row (i) WBC - Leuk	ocytes
CodedValue Language RBC - Erythr	rocytes
PB19GGAB Parvovirus B1 PLAT - Plate	lets
PB19GMAB Parvovirus B1 MPV - Mean	Platelet Volume
PBG Porphobilino PDW Plate	let Distribution Width
PBGCREAT Porphobilino	
PC3MPSAM PCA3 mRNA	
PCA3MRNA Prostate Can	
PCDW Platelet Com	
PCNAG Proliferating Remo	ve item from list
PCO2 Partial Press	
PCO2ADJT Partial Press	DK
PCP Phencyclidine	

We now have all the terms that are represented in CDISC controlled terminology. When searching for "plateletcrit" we don't find anything, but we do nee it, so we need to extend the subset.

First click "Recreate table from list" in the "Search" panel.

This leads to that all items from the (very large) list are removed, and only our 5 items are retained:

Search				×				
?	Search for: platelet distribution	on width	Search	Find Next				
_	Match case Whole words only		Add to list Display list					
			Recreate ta	able from list				
		Close						
Insert row				Remove row				
	CodedValue	Language		Decode	Rank	NCI Code	ExtendedValue	
	MPV		Mean Platelet Volu	me		C74730		-
	PDW		Platelet Distribution	n Width		C81962		
	PLAT		Platelets			C51951		
	RBC		Erythrocytes			C51946		

We will now add "plateletcrit" to the list as an "extension". This can easily be done by using the "Insert Row" button, and then filling the fields "CodedValue" and "Decode":

	Insert	row	Remove row			
CodedValue Language Decode			Rank	NCI Code	ExtendedValue	
MPV		Mean Platelet Volume		C74730	· · · · · · · · · · · · · · · · · · ·	
PDW		Platelet Distribution Width		C81962		
PLAT		Platelets		C51951		
RBC		Erythrocytes		C51946		
WBC		Leukocytes		C51948		
PLATCRIT		Plateletcrit			¥	

Remark that the "CodedValue" must be no longer than 8 characters¹. Also remark that the checkbox "ExtendedValue" is automatically checked, and that it is not possible to add an NCI code. This is smart, as of course sponsor extensions can never have an NCI code.

Done!

Of course, we need to repeat the whole process for the corresponding "test name" subset. This is tedious and error prone, so it is better to let the program do that for us.

After clicking "OK", the system will come with a new dialog:



Asking us whether we want to (additionally) store this created subset to an external file for later re-use, e.g. in other studies. Sponsors can so create a repository with "usually used subset codelist" and then later, in other mappings, reuse them. The latter can be done using the menu "Insert – xxxxxxxxxxx". The file format used for storage is the CDISC ODM format.

When we click "No", another dialog appears:



It invites us to have the corresponding "Test Name" codelist generated automatically. Using this feature may save hours of work and avoid errors and mismatches between "Test

¹ Another crazy SDTM limitation due to the use of the completely outdated SAS Transport 5 (XPT) format.

Code" and "Test Name" codelists. So we click "OK", leading to:

Message		Х
i	A 'Test Name' CodeList with the Name Laboratory Test Name subset and the OID CL.C67154.LBTE ST.SUBSET has been created.	
	It is not a bad idea at all to check this automatically generated codelis for correctnes using the menu 'Edit' - 'SDTM CodeList'.	s,

When then using the menu "Edit – SDTM/SEND CodeList" and searching for "subset", we also find the automatically generated "Test Name" codelist:

Find CodeList	subset							
CL.C67154.LBTE ST. SUB	SET - Laboratory Tes	t Name subset				-		
Name:	Name: Laboratory Test Name subset							
(New) SASFormatName:								
Remove rows until only those remain that you want to appear in the CodeList. Several rows can be removed simultaneously.								
You can also add new row	NS.							
Insert row Remove row								
CodedValue		Rank		NCI Code	ExtendedValue			
Mean Platelet Volume C		C74	730					
Platelet Distribution Width C			C81	962				
Platelets			C51	C51951				
Erythrocytes			C51	946				
Leukocytes		C51	948					
Plateletcrit					¥			

Which is ... just perfect.

We can still change the Name of the codelist, e.g. in "Laboratory Test Name Hematology-1" and similarly for the codelist "Laboratory Test Code subset"².

We now have created a subset for "test code" and one for "test name". To really use them in the submission, do not forget to assign them to either an SDTM/SEND variable or to a "ValueList" variable. For the latter, see the tutorial "Working with the WhereClause in define.xml 2.0"

 $^{^{2}}$ In order to have another identifier (the OID) for the codelist, this should be done when creating the subset.