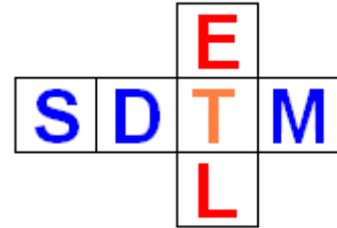


SDTM-ETL 4.0 User Manual and Tutorial

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 "[plateletcrit](#)" which is not covered by CDISC controlled terminology. The latter is covered by LOINC (e.g. LOINC code [51637-7](#)) 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 SDTM CodeList from existing CodeList

?

Find CodeList

CL.C66742.Y - No Yes Response (Yes only)

New OID:

New Name:

New DataType:

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 rows.

Insert row		Remove row	
CodedValue	Rank	NCI Code	ExtendedValue
Y		C49488	<input type="checkbox"/>

Search in CodeList Sort CodeList Items on exit

OK 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: [highlighted]

New Name: [highlighted]

New DataType: text

When clicking "Find CodeList", we first find:

Find CodeList | laboratory

CL.C102580.LBSTRESC - Laboratory Test Standard Character Result

New OID: CL.C102580.LBSTRESC.SUBSET

New Name: Laboratory Test Standard Character Result subset

New DataType: text

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 rows.

CodedValue	Rank	NCI Code
BORDERLINE		C14157
INDETERMINATE		C48658
INVALID		C50913
NEGATIVE		C38757
POSITIVE		C38758

Insert row | Remove

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.LBTESTCD - Laboratory Test Code

New OID: CL.C65047.LBTESTCD.SUBSET

New Name: Laboratory Test Code subset

New DataType: text

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 rows.

Insert row

CodedValue	Rank	NCI Code
A1AGLP		C100429
A1ANTRYP		C80167
A1MCREAT		C100462
A1MICG		C100461
A2MACG		C80168

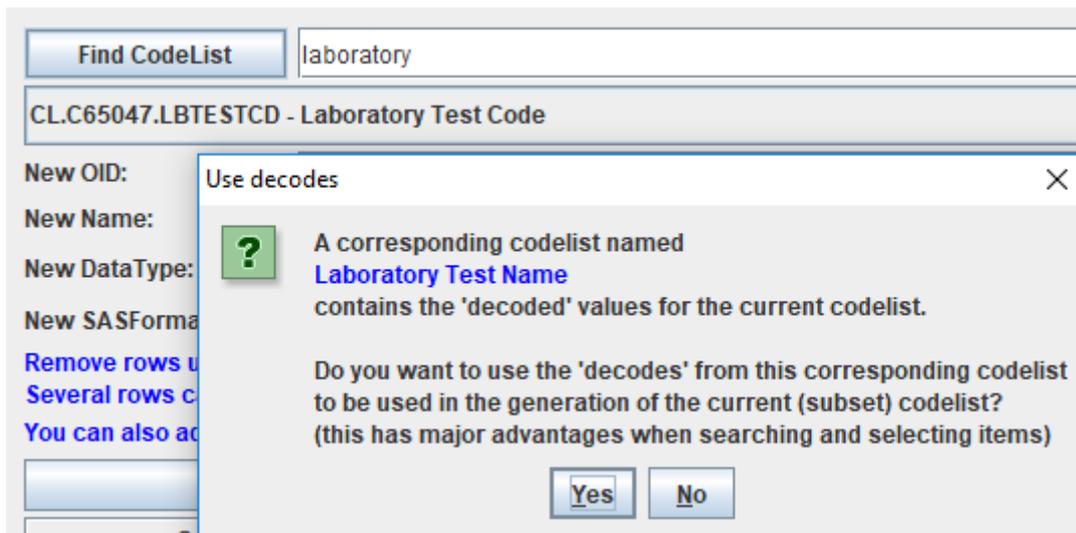
As you however see, most of the codes are not very well understandable, so it will be very hard to find the ones we need. This is due to the fact that CDISC publishes most of the codelists just as "enumerated" lists, without the human-readable "decode". In such cases, the "code" and "decode" are usually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

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

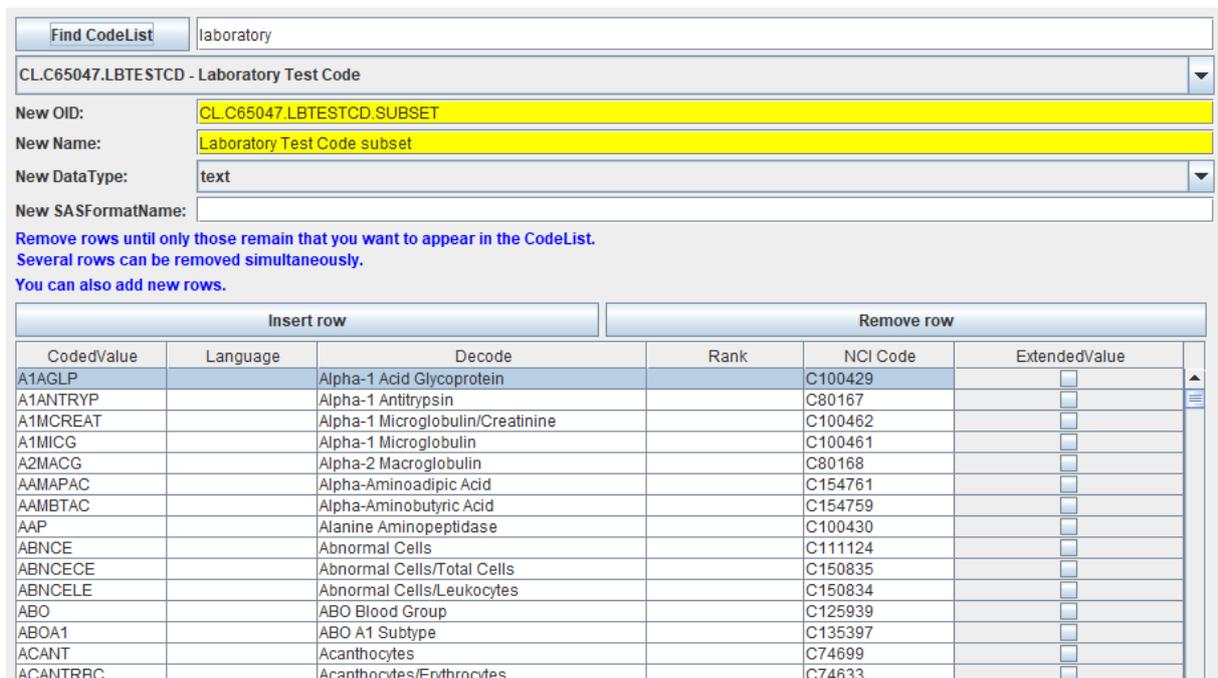
CodedValue	Rank
A1AGLP	
A1ANTRYP	
A1MCREAT	
A1MICG	A1ANTRYP: Alpha-1 Antitrypsin
A2MACG	
AAMAPAC	
AAMBTAC	

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:



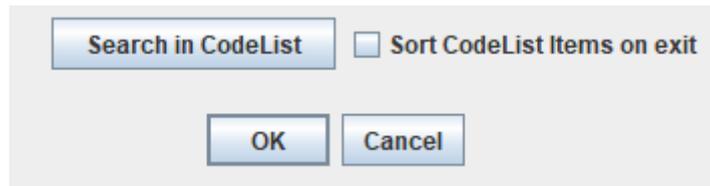
When clicking "OK", the "decodes" will be retrieved and also be presented and made "searchable". In our case this leads to:



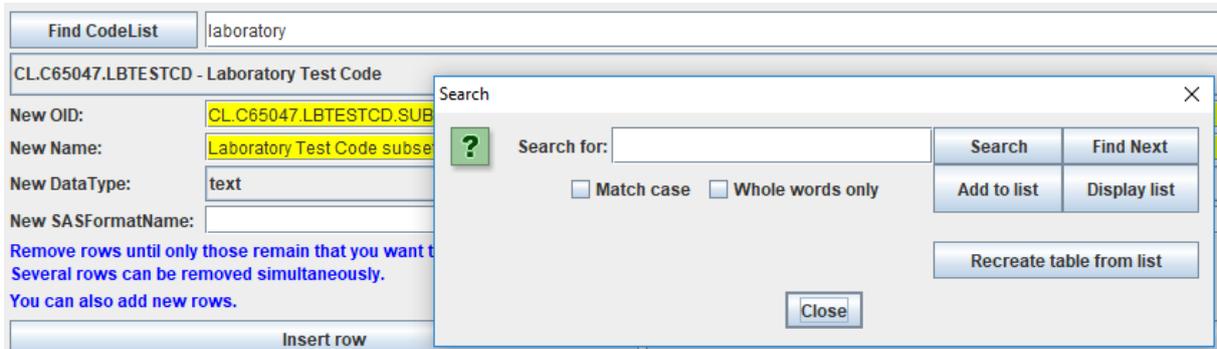
Remark that an additional column "Decode" appears containing the corresponding "test name" which comes from the "LBTEST" codelist.

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.

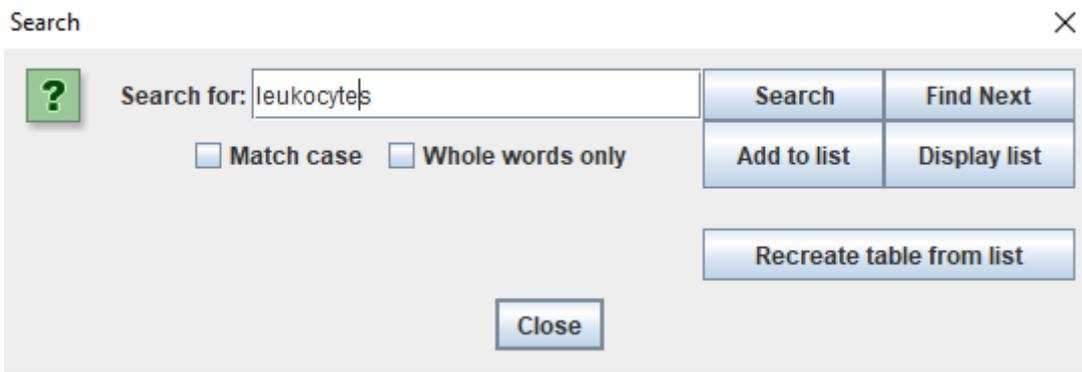
It now becomes easy to find "leukocytes", "erythrocytes" etc.. In order to do so, use the button "Search in CodeList" near the bottom:



This will add a new dialog:

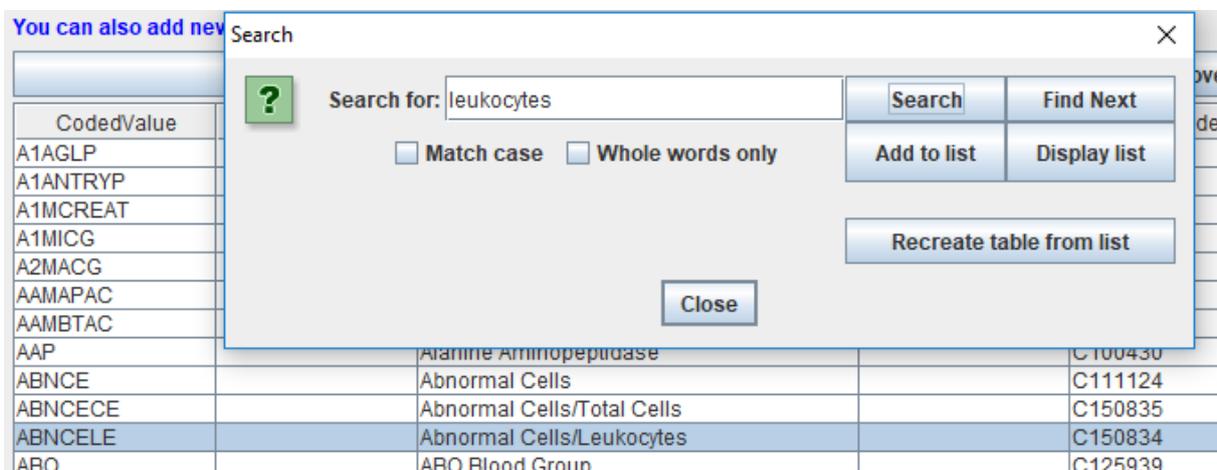


In which we can start searching, e.g. for "leukocytes":



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:



"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		
VITK1		
VLDL		
VLDL1		
VLDL2		
VLDL3		
VLDLPSZ		
VMA		
VOLUME		
VTD2D3IT		
VWFAAC		
VZVAB	Varicella Zoster Virus Antibody	C98797
VZVIGAAB	Varicella Zoster Virus IgA Antibody	C98795
VZVIGGAB	Varicella Zoster Virus IgG Antibody	C98796
VZVIGMAB	Varicella Zoster Virus IgM Antibody	C98797
WBC	Leukocytes	C51948

Search

Search for: leukocytes

Match case Whole words only

Search Find Next

Add to list Display list

Recreate table from list

Close

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:

Language	Decode	Rate	NC
Vitamin E/Cholesterol			C103448

Items to keep

WBC - Leukocytes

Remove item from list

OK

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

Items to keep

WBC - Leukocytes

G6PDRBC - G6PD-Deficient Erythrocytes

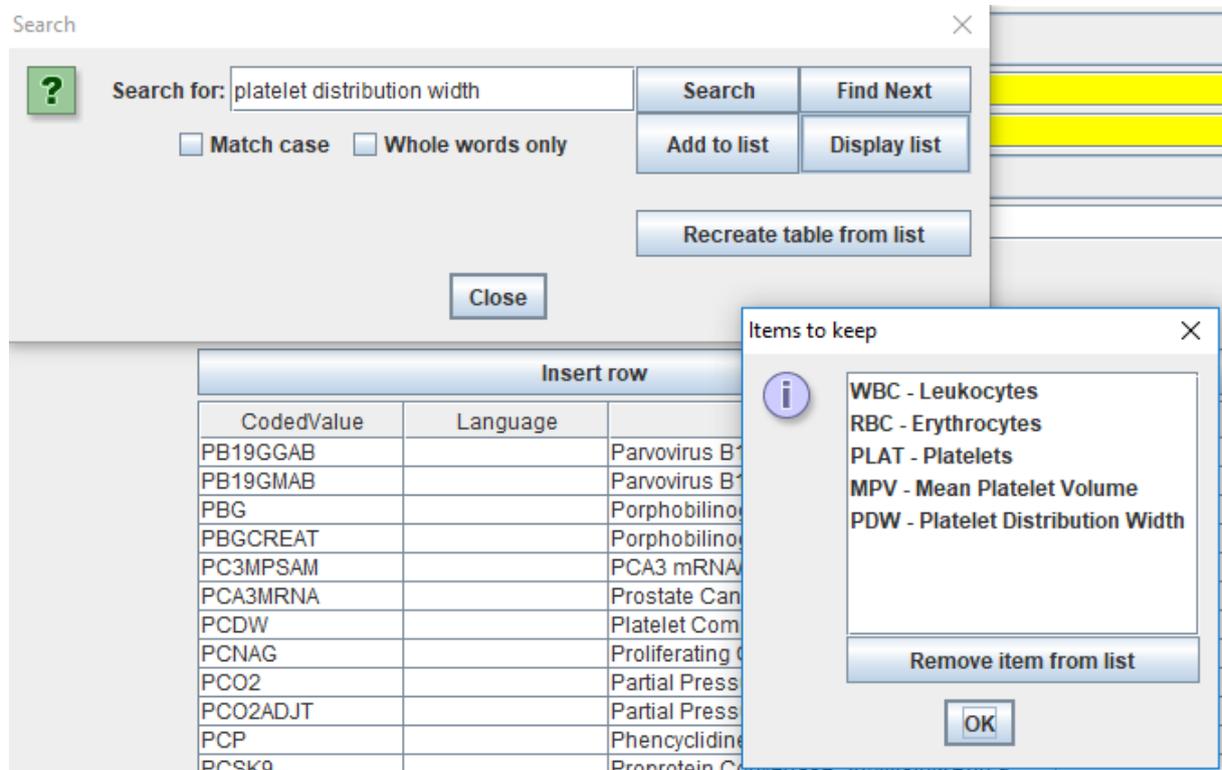
RBC - Erythrocytes

Remove item from list

OK

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

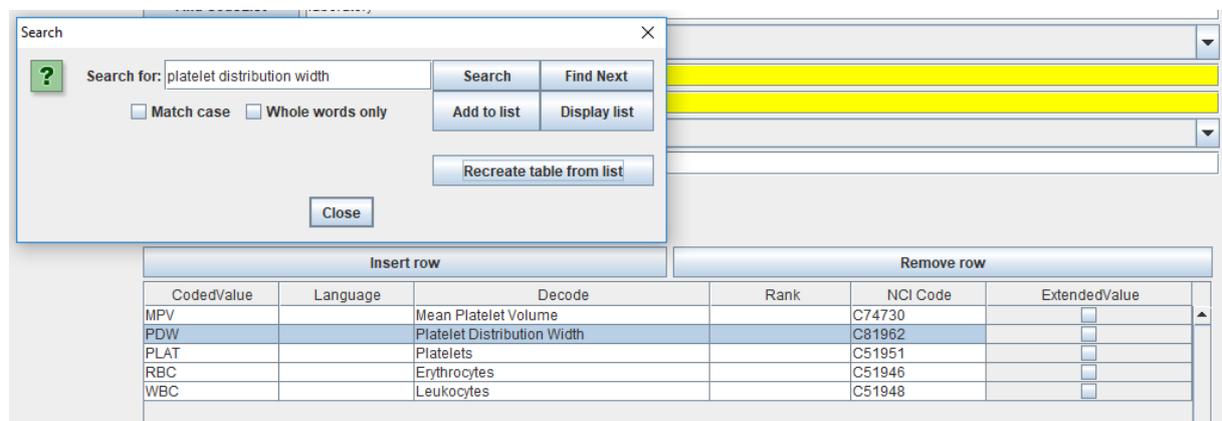
We now also add "platelets", "mean platelet volume" and "platelet distribution width" in the same way as before. Using "Display list" this leads to:



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 need 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:



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":

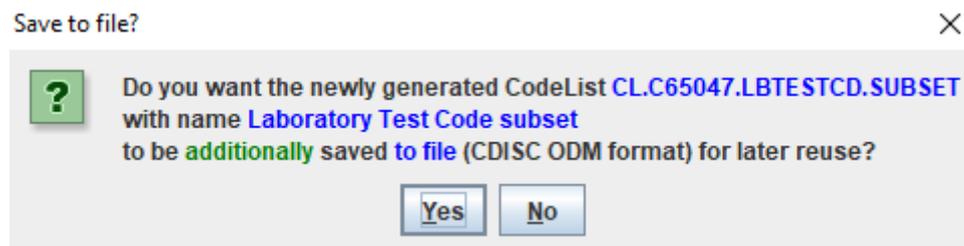
Insert row			Remove row		
CodedValue	Language	Decode	Rank	NCI Code	ExtendedValue
MPV		Mean Platelet Volume		C74730	<input type="checkbox"/>
PDW		Platelet Distribution Width		C81962	<input type="checkbox"/>
PLAT		Platelets		C51951	<input type="checkbox"/>
RBC		Erythrocytes		C51946	<input type="checkbox"/>
WBC		Leukocytes		C51948	<input type="checkbox"/>
PLATCRIT		Plateletcrit			<input checked="" type="checkbox"/>

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!

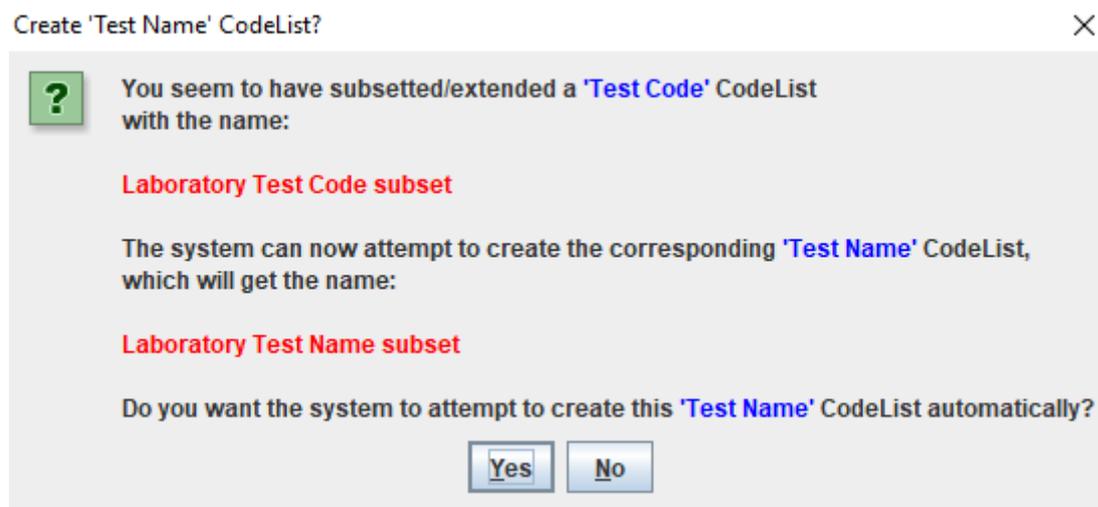
Of course, we need to repeat the whole process for the corresponding "test name" subset. This is tedious and error prone, so it is better to let the program do that for us.

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 – xxxxxxxxxxxx". The file format used for storage is the CDISC ODM format.

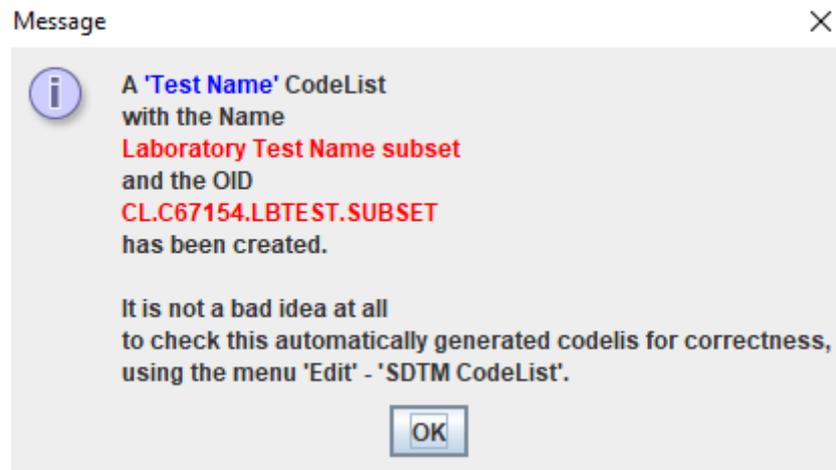
When we click "No", another dialog appears:



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:



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.LBTEST.SUBSET - Laboratory Test Name subset			
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 rows.			
Insert row			
Remove row			
CodedValue	Rank	NCI Code	ExtendedValue
Mean Platelet Volume		C74730	<input type="checkbox"/>
Platelet Distribution Width		C81962	<input type="checkbox"/>
Platelets		C51951	<input type="checkbox"/>
Erythrocytes		C51946	<input type="checkbox"/>
Leukocytes		C51948	<input type="checkbox"/>
Plateletcrit			<input checked="" type="checkbox"/>

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.