SDTM-ETL 5.0 User Manual and Tutorial

Author: Jozef Aerts, XML4Pharma

Last update: 2025-08-25

SDTM/SEND generation for selected domains / dataset definitions only

Table of Contents

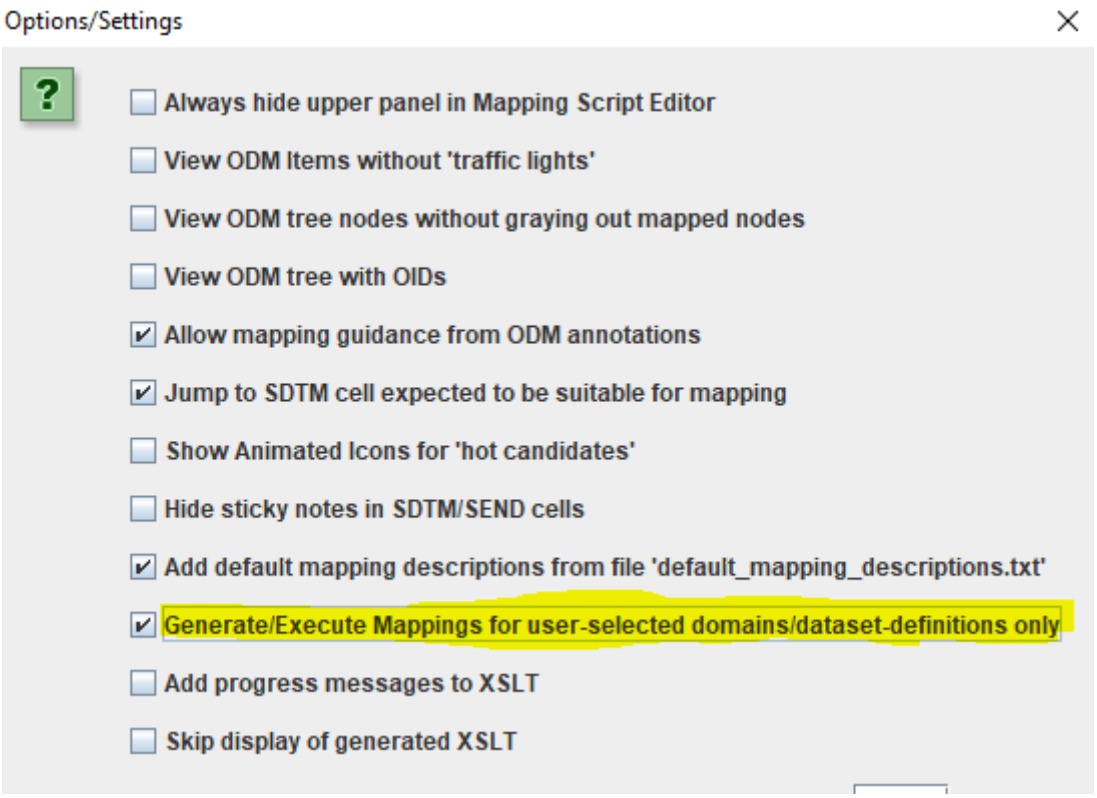
Introduction

Some of our users have a define.xml with lots of dataset definitions (sometimes 30 or more), but want to be able to generate the SDTM/SEND files for a limited set of domains or dataset definitions only, especially during testing, or just because new source data for a single domain have become available.

As of version 5.0, it is now possible to select a limited set of domain / dataset definitions, even when many more are loaded.

Setting the option for limited domain/dataset execution

In order to set the option to only generate the SDTM/SEND datasets for a specific set of domain / dataset definitions, instead of all of the loaded study-specific ones (which is the default), use the menu "Options - Settings", and check the checkbox "Generate/Execute Mappings for user-selected domains/dataset-definitions only":



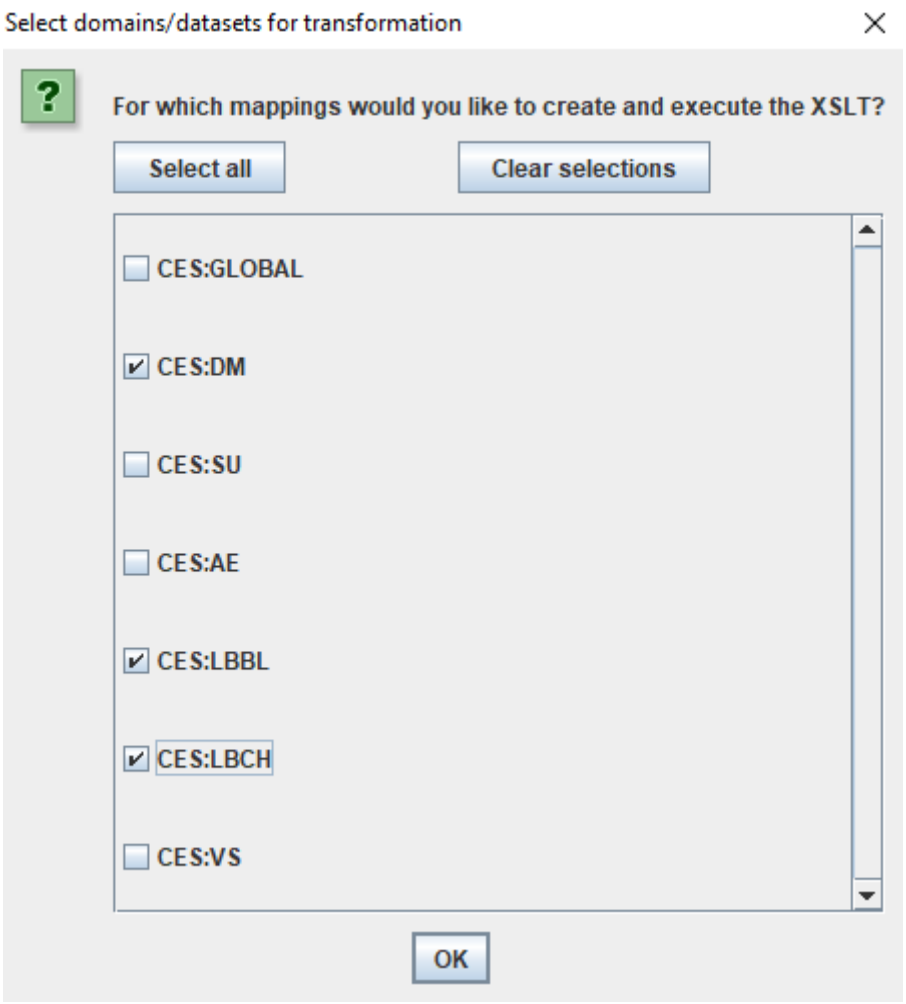
This will cause that when generation of the transformation is started, the user will be asked for which of the loaded dataset definitions the SDTM/SEND datasets need to be generated.

Execution of selected domain / dataset definitions

For example, when we have following study-specific dataset definitions loaded:

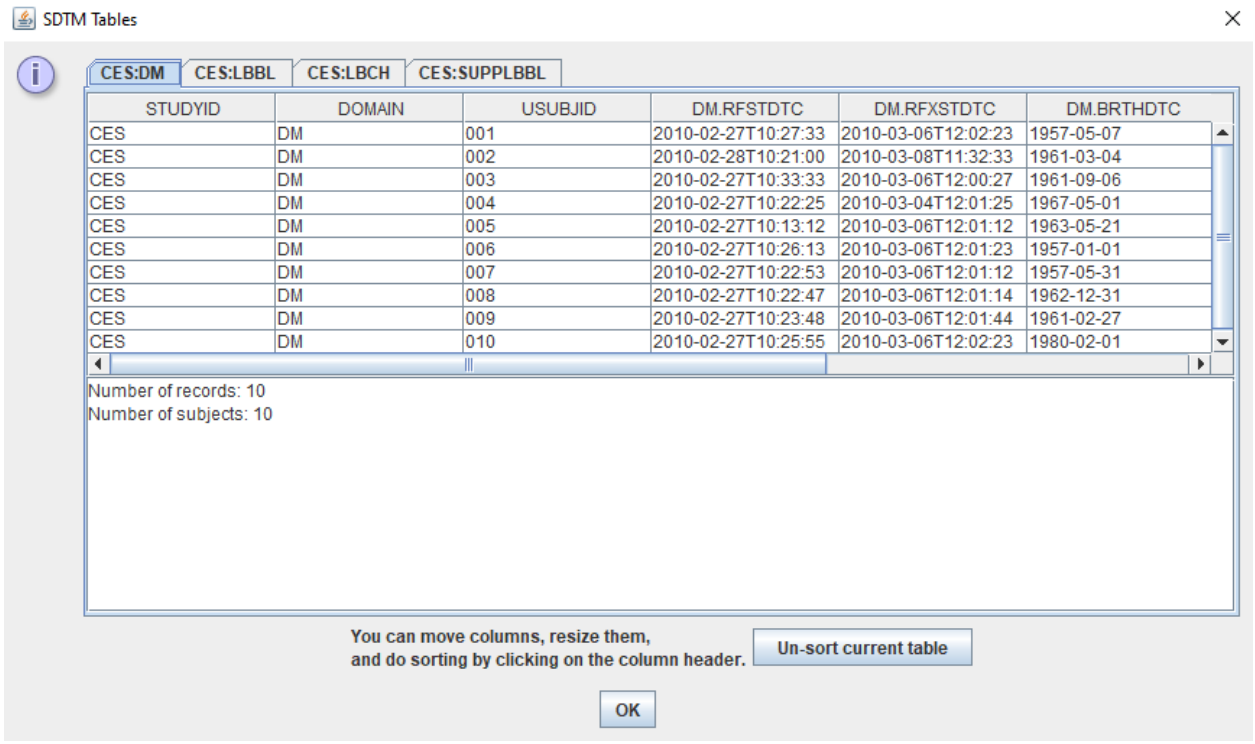
TU	STUDYID	DOMAIN	USUBJID	TU.TUSEQ	TU.TUGRPID	TU.TUREFID	TU.TUSPID
TR	STUDYID	DOMAIN	USUBJID	TR.TRSEQ	TR.TRGRPID	TR.TRREFID	TR.TRSPID
RS	STUDYID	DOMAIN	USUBJID	RS.RSSEQ	RS.RSGRPID	RS.RSREFID	RS.RSSPID
VS	STUDYID	DOMAIN	USUBJID	VS.VSSEQ	VS.VSGRPID	VS.VSSPID	VS.VSTESTCD
FA	STUDYID	DOMAIN	USUBJID	FA.FASEQ	FA.FAGRPID	FA.FASPID	FA.FATESTCD
SR	STUDYID	DOMAIN	USUBJID	SR.SRSEQ	SR.SRGRPID	SR.SRREFID	SR.SRSPID
RELREC	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
SUPPQUAL	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL
CES:GLOBAL	REFSTARTDATE						
CES:DM	STUDYID	DOMAIN	USUBJID	SUBJID	DM.RFSTDTC	DM.RFENDTC	DM.RFXSTDTC
CES:SU	STUDYID	DOMAIN	USUBJID	SU.SUSEQ	SU.SUGRPID	SU.SUSPID	SU.SUTRT
CES:AE	STUDYID	DOMAIN	USUBJID	AE.AESEQ	AE.AEGRPID	AE.AEREFID	AE.AESPID
CES:LBBL	STUDYID	DOMAIN	USUBJID	LB.LBSEQ	LB.LBGRPID	LB.LBREFID	LB.LBSPID
CES:LBCH	STUDYID	DOMAIN	USUBJID	LB.LBSEQ	LB.LBGRPID	LB.LBREFID	LB.LBSPID
CES:VS	STUDYID	DOMAIN	USUBJID	VS.VSSEQ	VS.VSGRPID	VS.VSSPID	VS.VSTESTCD

and we use the menu "Transform - Generate Transformation (XSLT) Code" and the option is switched on, the following dialog is displayed, in which we can select the dataset definitions for which datasets need to be generated:



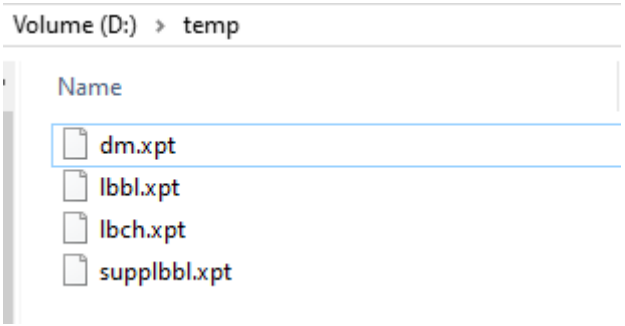
In this example case, we will only generate datasets for DM, LBBL and LBCH.
No datasets will then be generated for SU, AE and VS.

The result of the execution will then be:



Remark that also a SUPPLBBL dataset is automatically generated, as the LBBL dataset definition had mappings for a "non-standard variable" (NSV), and the option to "move" NSVs to a "Supplemental Qualifier dataset" was on.

In the case of SAS-XPT as the dataset format, the files then generated are:



Limitations

One must take into account that when using this feature, one should be careful about dependencies between datasets. For example, for --LOBXFL generation in "Findings" datasets, is it necessary that the DM dataset is loaded and selected for execution.
Similarly, when one e.g. want to have RELREC automatically generated for the relation between AE and CM, both dataset definitions need to be loaded and selected.

Similarly for "comment variables". For example, when one has a "comment variable" for both DM as LB, and one does not select LB to be included, and using the option "Move Comment Variables to Comments (CO) domain", then the generated dataset will only contain the comments for DM, and not for LB.

Conclusions

This new feature will allow users to work considerably faster in the case that a large number of mappings for a larger number of domains / dataset definitions are present, but the user wants to only generate (or update) datasets for one or a limited number domains / dataset definitions.
This is essentially beneficial especially in the case of a large number of subjects.