

SDTM-ETL 3.1 New Features



SDTM-ETL 3.1 has a lot of new features. The most important ones are listed and explained in this document which lists all major new features of versions 3.0 and 3.1.

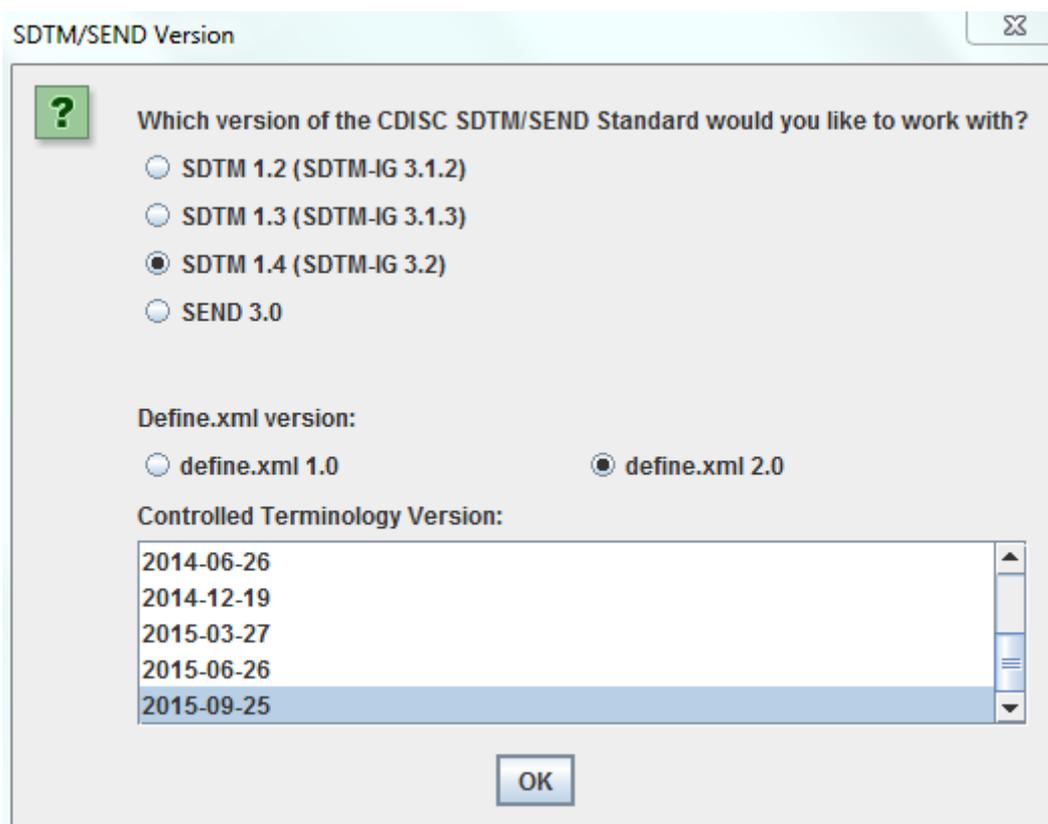
Support for the newest CDISC standards

SDTM-ETL 3.1 has full support for the following new CDISC standards:

- [define.xml 2.0](#)
- [Dataset-XML 1.0](#)
- [newest CDISC Controlled Terminology](#) (june 2014) + all older versions
- [SDTM 1.4 / SDTM-IG 3.2](#)

The user can now choose between define.xml 1.0 and define.xml 2.0 for keeping the mappings and exporting a „near-submission-ready“ define.xml. For new projects, the use of define.xml 2.0 is strongly recommended.

When starting a new project using define.xml 2.0, the user is asked which version of CDISC controlled terminology (either for SDTM or SEND) need to be used.



New versions of CDISC-CT can simply be added later – the software will recognize them automatically¹. Also other codelists can be loaded additionally.

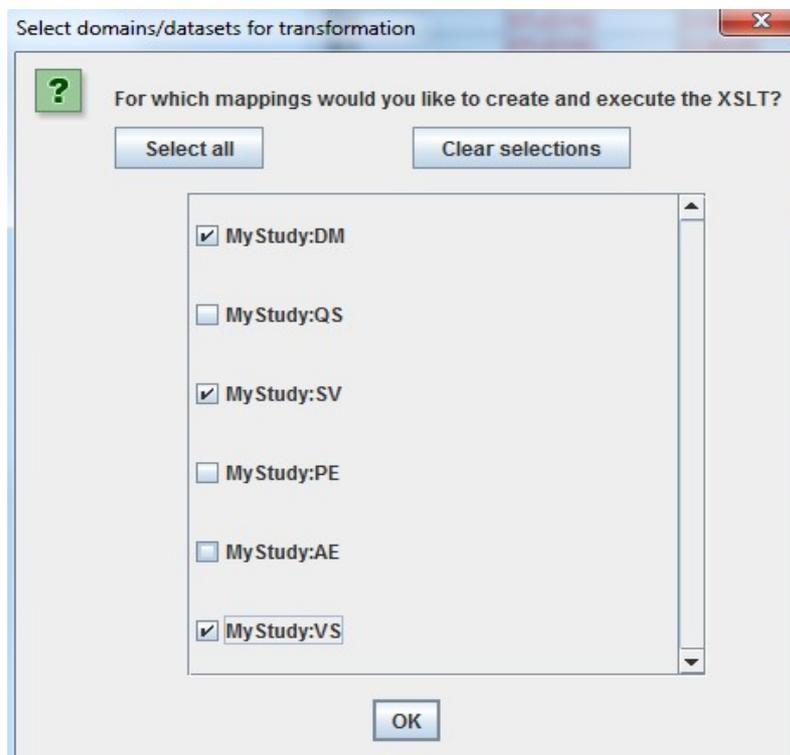
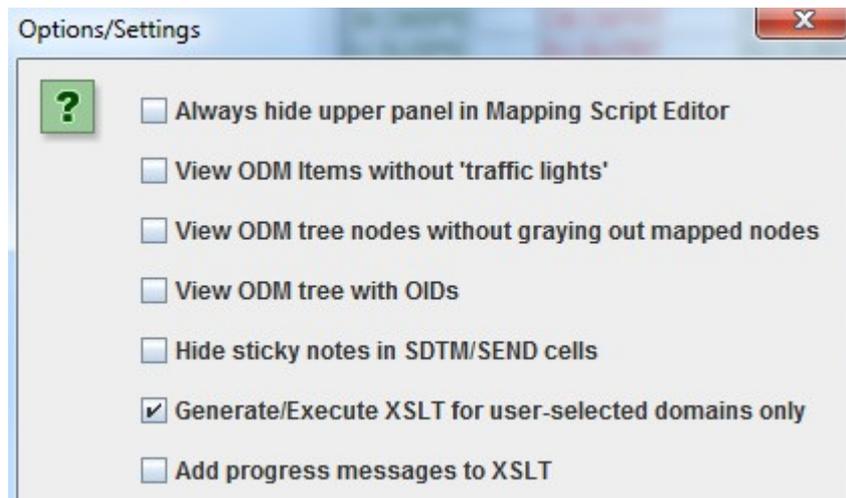
Also additional templates such as for Medical Devices and Associates Persons can be loaded additionally.

¹ In future we hope to do this by connecting to CDISC SHARE

Graphical User Interface

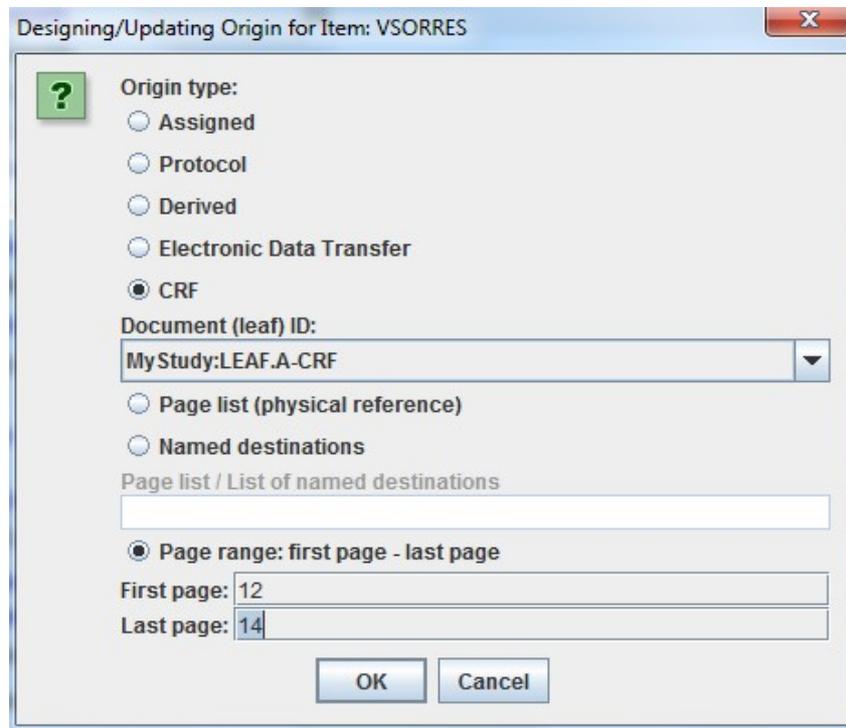
As for each version of SDTM new domains are added, the user easily loses the overview. Therefore, it is now possible in SDTM-ETL 3.1 to hide domains in the GUI. This applies as well to template domains as to study-specific domains. Domains (table rows in the GUI) can be hidden or (re)displayed with a few mouse clicks only.

In SDTM-ETL 3.1, the user can even decide and select for which of the loaded domains the mappings are executed. Executing the mappings for a selection (or only one) of the loaded domains especially saves a lot of time during the testing of the mappings.

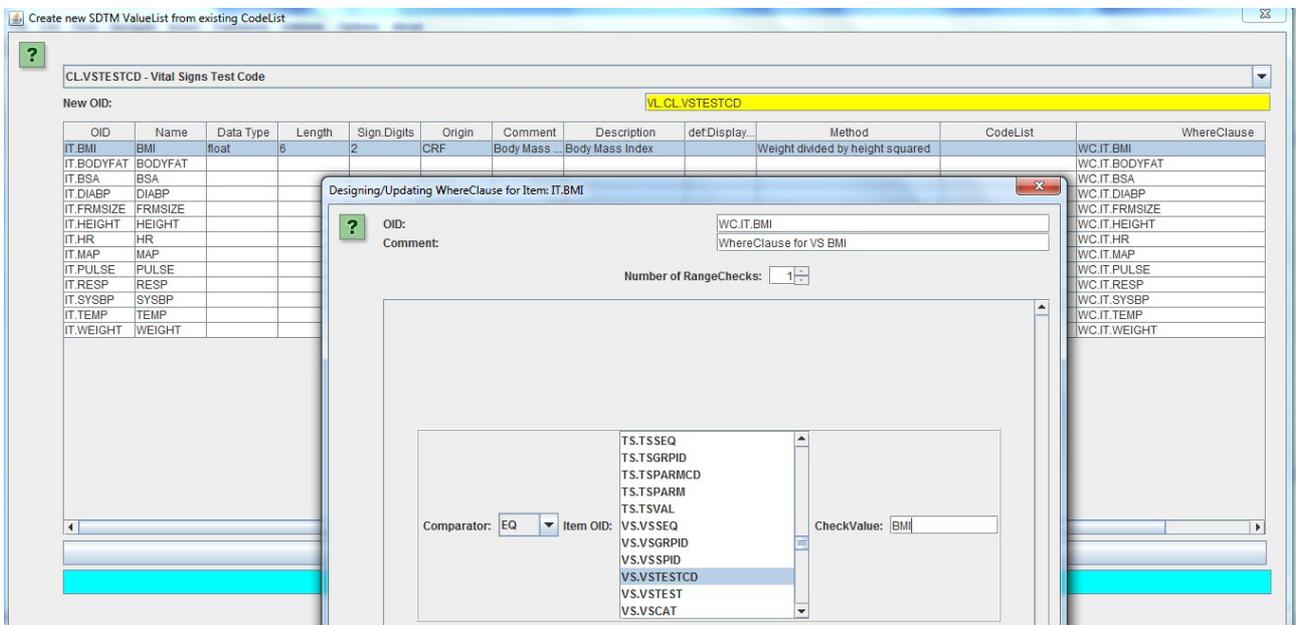


Define.xml 2.0 support

SDTM-ETL 3.1 has full support for define.xml 2.0. A large number of new wizards and dialogs have been added to support the new features of define.xml 2.0 like the def:Origin element.



and the def:WhereClause:



Even very complicated „Where Clauses“ can be generated in a very user-friendly GUI-only way, without any XML editing.

Dataset-XML support

Dataset-XML is a new XML-based format for submission of SDTM, SEND and AdAM data to the regulatory authorities, replacing the legacy SAS-XPORT format. The FDA is currently conducting a pilot in the use of this format, which has a lot of advantages over SAS-XPORT such as:

- no more 8-, 40- and 200-character limitations
- open, vendor-neutral format
- non-standard variables can be kept in the parent domain
- no more COVAL1, COVAL2, ... variables
- perfect fit with define.xml 2.0

This new format is also expected to boost the development of new tools, such as smart viewers, validation tools and analysis tools.

The user can choose between generating SDTM/SEND datasets in either CDISC Dataset-XML format or in classic SAS-XPORT format (or do both).

The open source „[Smart Dataset-XML Viewer](#)“ has been integrated into the software in a modular way, meaning that the user can easily replace the default copy of the viewer by his own one (e.g. an extended one) or a newer version. The viewer comes into action when the user selected to generate the SDTM/SEND files in the new CDISC Dataset-XML format.

Vendor extensions support

SDTM-ETL does already have build-in support for OpenClinica and ClinCapture ODM exports. This support has now been extended with new features. Due to its modular structure, also support for other vendor extensions, such as for Medidata can easily be added.

When working with files with vendor extensions, SDTM-ETL 3.1 now allows to remove the vendor extension elements and attributes as a part of the loading procedure.

Creating and Editing Trial Design Datasets

When starting from an ODM file with metadata including SDM-XML ([Study Design Model in XML](#)), it is pretty straightforward to generate trial design datasets from the SDM-XML using the SDTM-ETL software (see separate manual).

However, not everyone is using SDM-XML yet, and most EDC systems do even not export SDM-XML. So, how can one generate the trial design datasets when the information is not, or only limited, in the ODM file with the study metadata?

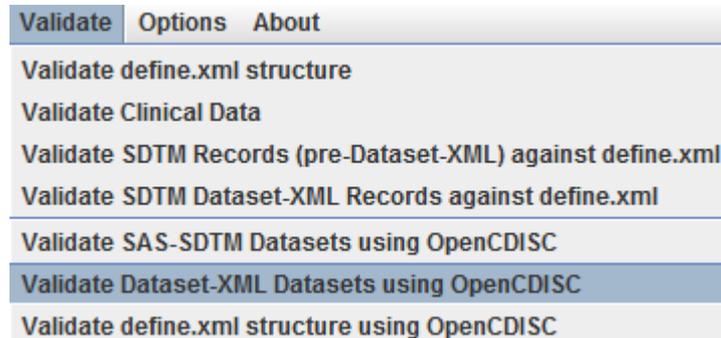
We have recognized this problem, and added a new module to the SDTM-ETL software: a "trial design dataset editor". This editor does not only allow to generate trial design datasets from scratch, but also allows to edit existing trial design in Dataset-XML format².

The editor can also be run in standalone method (so without using the SDTM-ETL software)

² At this moment, using SAS Transport 5 format is not foreseen. One can however transform Dataset-XML files into SAS Transport 5 files using other tools. See <http://wiki.cdisc.org/display/PUB/CDISC+Dataset-XML+Resources> for a list of such tools.

Integrated OpenCDISC 1.5 Validator

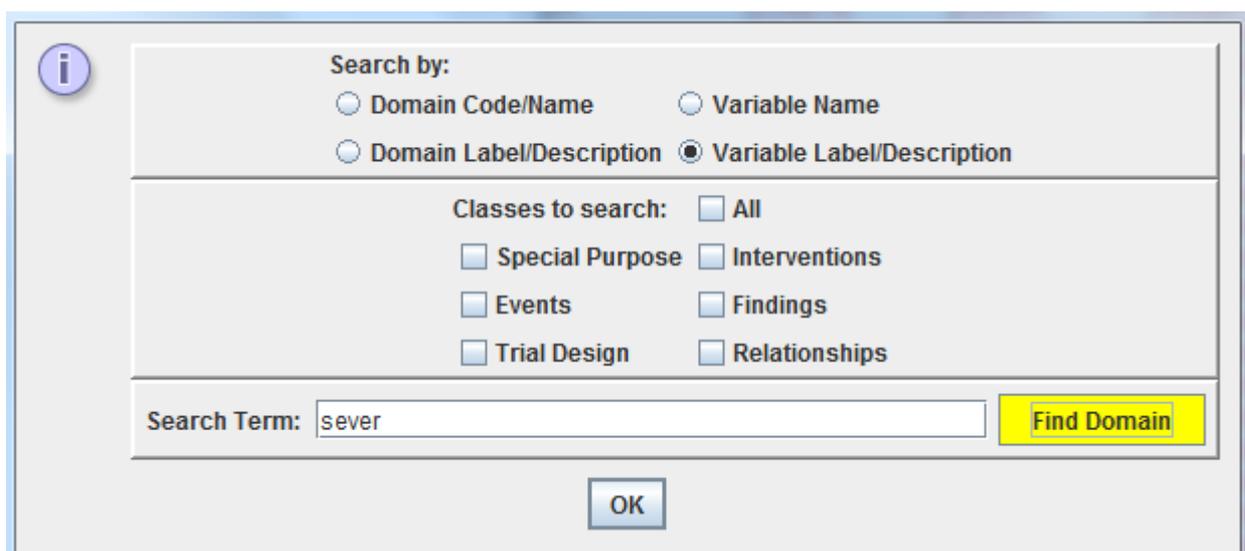
The newest version of the OpenCDISC Validator (v.1.5) has been integrated into the software. It can be used to validate the generated SDTM or SEND files (either in SAS-XPORT or CDISC Dataset-XML format) against the standard. Although SDTM-ETL 3.1 also has its own define.xml validator, the OpenCDISC Validator can also be used to validate the underlying define.xml file („second opinion“).



OpenCDISC versions starting from v.2.0 cannot be started from within the software, as OpenCDISC changed its license terms (i.e. it is not open-source software anymore). Alternatives using formal, by CDISC developed rules are currently in development.

"Find Domain" wizard

With so many new domains, and new ones coming up with the "Therapeutic Area User Guides" (TAUGs), it becomes more and more difficult and complex for the user to decide which SDTM/SEND domain to use. Besides the ability to hide some of the domains as to concentrate on the current important ones, we are now also developing a set of new wizards to help users to decide which SDTM domain needs to be used for a specific use case. As a first step, the new "Find SDTM/SEND Domain" functionality has been added. When it is started (using the menu Navigate → Find SDTM/SEND domain), the following dialog is displayed:



The user can then search by domain code, label and description, or by variable name or label or description. For example, when the user wants to find out which domains have a variable in which the wording "severe" or "severity" appears, one e.g. uses:

Search by:

Domain Code/Name Variable Name

Domain Label/Description Variable Label/Description

Classes to search: All

Special Purpose Interventions

Events Findings

Trial Design Relationships

Search Term: Find Domain

OK

A new window than shows up, displaying the domains in which there is a variable that has a label with the wording "sever":

Domain Name	Label/Description	
AE	Adverse Events	AE Details
CE	Clinical Events	CE Details

OK

Clicking the button "Details" then allows to get more information about the selected domain and its variables, so for example for the CE domain:



Domain Name: CE
 Domain Label/Description: Clinical Events

Name	Label/Description	DataType	CodeList
STUDYID	Study Identifier	text	
DOMAIN	Domain Abbreviation	text	
USUBJID	Unique Subject Identifier	text	
CESEQ	Sequence Number	integer	
CEGRPID	Group ID	text	
CEREFID	Reference ID	text	
CESPID	Sponsor-Defined Identifier	text	
CETERM	Reported Term for the Cl...	text	
CEDECOD	Dictionary-Derived Term	text	
CECAT	Category for Clinical Event	text	
CESCAT	Subcategory for Clinical ...	text	
CEPRESP	Clinical Event Pre-Specifi...	text	CL.C66742.NY
CEOCCUR	Clinical Event Occurrence	text	CL.C66742.NY
CESTAT	Completion Status	text	CL.C66789.ND
CEREASND	Reason Clinical Event N...	text	
CEBODSYS	Body System or Organ Cl...	text	
CESEV	Severity/Intensity	text	
CEBTC	Date/Time of Event Colle...	datetime	
CESTDTC	Start Date/Time of Clinic...	datetime	
CEENDTC	End Date/Time of Clinica...	datetime	
CEDY	Study Day of Event Colle...	integer	
CESTRF	Start Relative to Referen...	text	CL.C66728.STENRF
CEENRF	End Relative to Referenc...	text	CL.C66728.STENRF
CESTRPT	Start Relative to Referen...	text	CL.C66728.STENRF
CESTTPT	Start Reference Time Poi...	text	
CEENRTPT	End Relative to Referenc...	text	CL.C66728.STENRF
CEENTPT	End Reference Time Point	text	

OK

When then going back, the last selected domain will be highlighted in the SDTM/SEND table in the main window. If the domain is hidden, it will be asked whether to make it visible in the table:

Domains (ItemGroups)				
Domain	Variable	Variable	Variable	Variable
DM	STUDYID	DOMAIN	USUBJID	SUBJID
SE	STUDYID	DOMAIN	USUBJID	SE.SESEQ
SV	STUDYID	DOMAIN	USUBJID	SV.VISITNUM
CM	STUDYID	DOMAIN	USUBJID	CM.CMSEQ
EC	STUDYID	DOMAIN	USUBJID	EC.ECSEQ
EX	STUDYID	DOMAIN	USUBJID	EX.EXSEQ

The CE is currently hidden
Do you want to make it visible?

If the user than answers with "Yes", the CE domain will be made visible again, and highlighted so that it can be instantiated for starting generating the mappings:

Domains (ItemGroups)				
Domain	Variable	Variable	Variable	Variable
DM	STUDYID	DOMAIN	USUBJID	SUBJID
SE	STUDYID	DOMAIN	USUBJID	SE.SESEQ
SV	STUDYID	DOMAIN	USUBJID	SV.VISITNUM
CM	STUDYID	DOMAIN	USUBJID	CM.CMSEQ
EC	STUDYID	DOMAIN	USUBJID	EC.ECSEQ
EX	STUDYID	DOMAIN	USUBJID	EX.EXSEQ
CE	STUDYID	DOMAIN	USUBJID	CE.CESEQ

As already stated, this wizard is just a first step in providing features helping the user to decide which domain is suitable for a specific use case: more wizards will follow, e.g. also searching the SDTM and SEND and TAUG-IGs.

Other new features

When inspecting clinical data from an ODM file, also the repeat keys can now be displayed (optional).

The underlying define.xml file can now be displayed using the stylesheet (users can easily add their own one) either within the software (as before) or in the users favorite browser (IE, Firefox, Chrome, ...).

An „autosave“ feature has been added: the system automatically saves a copy of the mappings (as a define.xml file) every 5 minutes. Using the „Options“ menu, the user can define the autosaving time interval.

Features for generating valuelists from codelists have been further improved and made even more user friendly.

When generating SAS XPORT files, numeric values are now generated in double precision (was single precision before).

Generated Dataset-XML files can be validated against the currently loaded define.xml or against a define.xml file on disk.

Last but not least, the memory footprint has further been optimized, so that also extremely large ODM files with clinical data can now be processed with smaller memory requirements.

Minor changes

- links to external documents are now always stored as absolute path links. When exporting to a „cleaned“ define.xml, the absolute links can be automatically converted to relative links, relative to where the define.xml is stored.

Special features

It is also possible to run the SDTM-ETL software in **non-graphical batch mode**. This is very useful for companies that regularly (e.g. daily) receive ODM files with new clinical data and desire to generate new SDTM datasets immediately after having received the new clinical data.

One of our customers has a process in place where new SDTM datasets are generated each night automatically without human intervention, and feed these new datasets into a data warehouse.

The new version also supports the generation of SDTM or SEND datasets in the new CDISC Dataset-XML format.