SDTM-ETL 4.5: Summary of New Features

Author: Jozef Aerts, XML4Pharma

Last update: **2024-05-24**

Summary

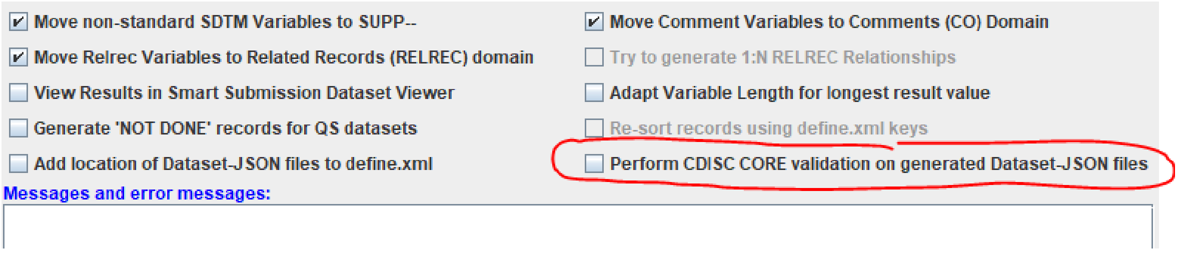
This document contains a summary of the most important new features of SDTM-ETL 4.5 and bug fixes.
There are many minor improvements and new features that are not described in this document, but that can be found in other manuals / tutorials of SDTM-ETL 4.5.

Table of Contents

New CDISC-CORE Validation Engine

SDTM-ETL v.4.5 now comes with the newest CDISC CORE Engine v.0.7.1 which also supports Dataset-JSON as submission format. The implementation is however in such a way that when a new CORE version becomes available, it can just be replaced by the new one, without an update of the SDTM-ETL software. Exception is when the CORE command parameters to start CORE have been changed.
One of such changes in near future will be that users (or companies as a whole) can add their own additional validation rules, for example as "quality measures" to the engine.
If this happens, we will make a new version of SDTM-ETL available, and publish a special manual on how to write company-internal validation rules and add them to the engine.

This new CORE engine also means that CORE can be executed not only for the outdated SAS-XPT format, but also for the modern CDISC Dataset-JSON format.

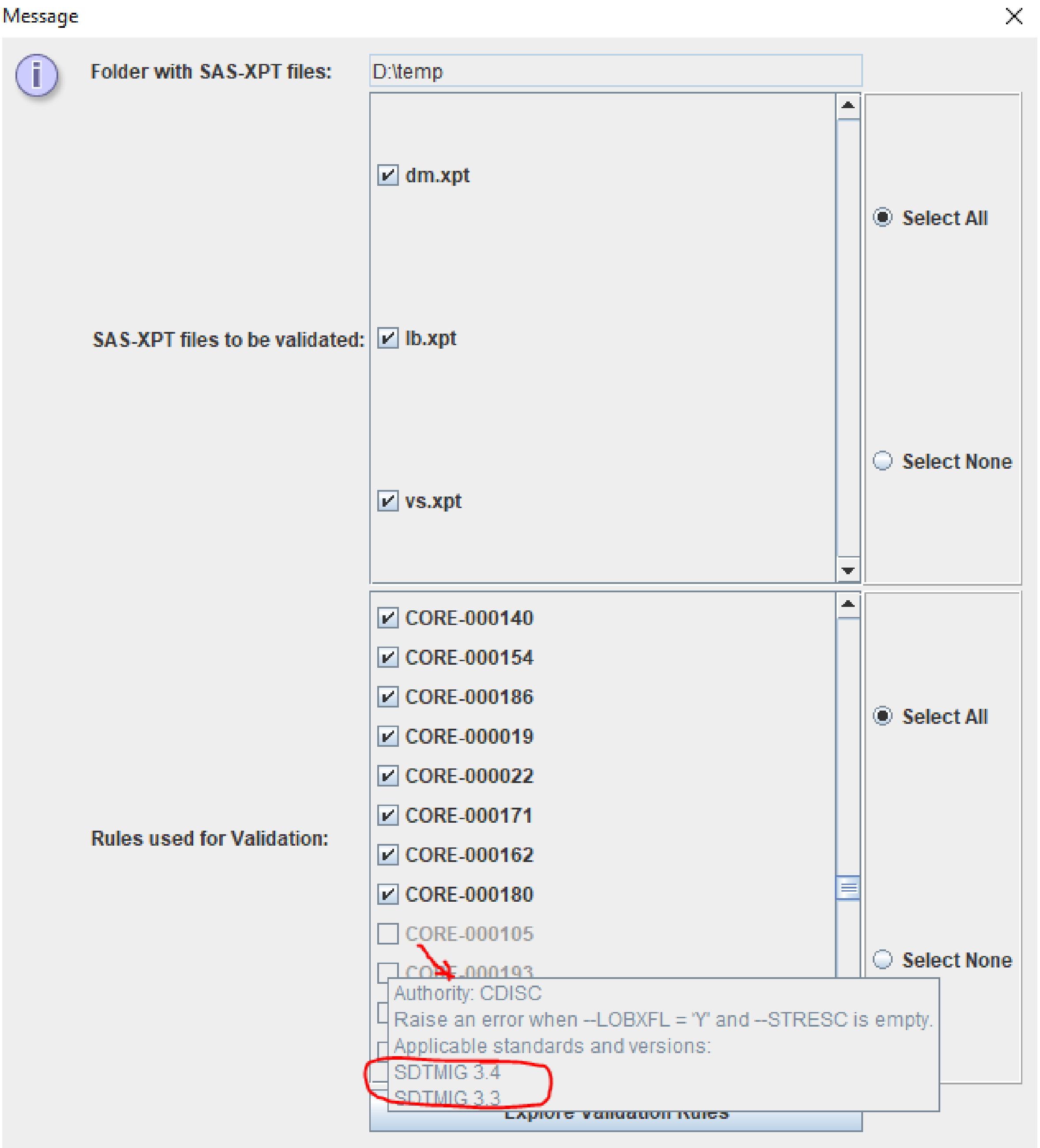


☑ Move non-standard SDTM Variables to SUPP--
☑ Move Relrec Variables to Related Records (RELREC) domain
☐ View Results in Smart Submission Dataset Viewer
☐ Generate 'NOT DONE' records for QS datasets
☐ Add location of Dataset-JSON files to define.xml
☑ Move Comment Variables to Comments (CO) Domain
☐ Try to generate 1:N RELREC Relationships
☐ Adapt Variable Length for longest result value
☐ Re-sort records using define.xml keys
☐ Perform CDISC CORE validation on generated Dataset-JSON files

Messages and error messages:

We must however remark that CDISC, in cooperation with the FDA, is working on a new version of Dataset-JSON, which is expected to become official later this year. We then also expect that the FDA will first start accepting submissions in the new Dataset-JSON and later completely move to Dataset-JSON and abandon SAS Transport 5 ("XPT") format.

We also improved the way the user can select individual rules for the generated files. For example, when the SDTMIG-3.2 is still used, all rules that do not apply to v.3.2, but only to 3.3 and 3.4, will be blended out:



Message

Folder with SAS-XPT files: D:\temp

SAS-XPT files to be validated:
☑ dm.xpt
☑ lb.xpt
☑ vs.xpt

● Select All
○ Select None

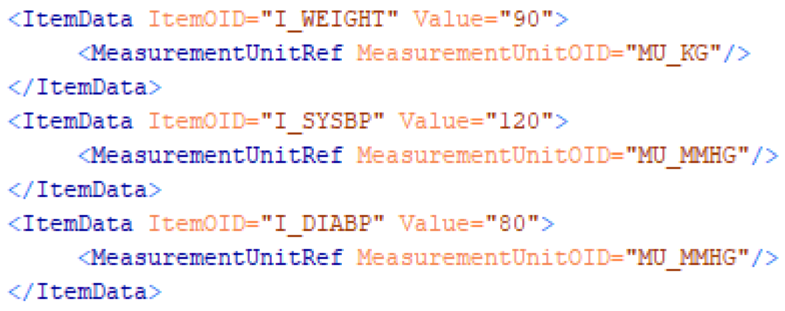
Rules used for Validation:
☑ CORE-000140
☑ CORE-000154
☑ CORE-000186
☑ CORE-000019
☑ CORE-000022
☑ CORE-000171
☑ CORE-000162
☑ CORE-000180
□ CORE-000105
□ CORE-000193

● Select All
○ Select None

Authority: CDISC
Raise an error when --LOBXFL = 'Y' and --STRESC is empty.
Applicable standards and versions:
SDTMIG 3.4
SDTMIG 3.3

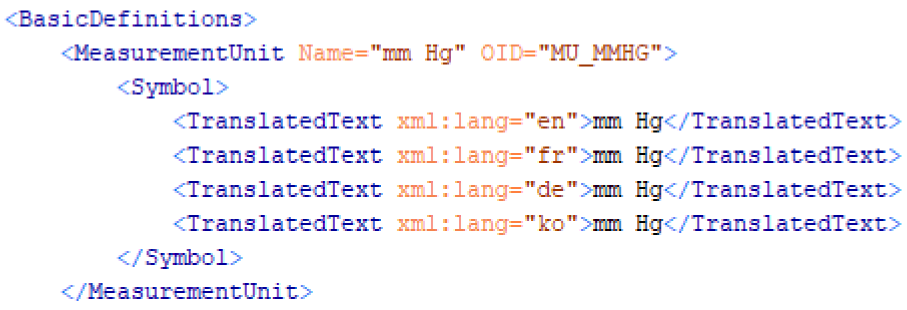
Better handling of ODM "MeasurementUnitRef"

Units of measure of observations are handled in two different ways in CDISC ODM. Either they come as separate data points, or they come as a child "MeasurementUnitRef" element of "ItemData" which represents the measurement itself. For example:



```
<ItemData ItemOID="I_WEIGHT" Value="90">
     <MeasurementUnitRef MeasurementUnitOID="MU_KG"/>
</ItemData>
<ItemData ItemOID="I_SYSBP" Value="120">
     <MeasurementUnitRef MeasurementUnitOID="MU_MMHG"/>
</ItemData>
<ItemData ItemOID="I_DIABP" Value="80">
     <MeasurementUnitRef MeasurementUnitOID="MU_MMHG"/>
</ItemData>
```

where "MeasurementUnitRef" is a reference to the definition in the ODM metadata, e.g.:



```
<BasicDefinitions>
    <MeasurementUnit Name="mm Hg" OID="MU_MMHG">
        <Symbol>
            <TranslatedText xml:lang="en">mm Hg</TranslatedText>
            <TranslatedText xml:lang="fr">mm Hg</TranslatedText>
            <TranslatedText xml:lang="de">mm Hg</TranslatedText>
            <TranslatedText xml:lang="ko">mm Hg</TranslatedText>
        </Symbol>
    </MeasurementUnit>
```

As of version 4.5, one can now automatically copy the set of ODM "MeasurementUnit"s into the define.xml and transform that into a define.xml "CodeList" and it to it. This is especially useful in the case of SEND, as in SDTM, one will usually want to the use the SDTM "UNIT" codelist, which however is not always possible in the case of animal studies.
All this is further described in the separate manual "Mapping Units of Measure".

Also the automated mapping of units from the ODM to "CDISC Units" has been made more easy: when one drag-and-drop an item in the ODM tree that has a "MeasurementUnit" attached to a --ORRESU cell in the SDTM/SEND table, the system will automatically suggest to start a "mapping wizard" for mapping the ODM units to the CDISC units.
Also this is further explained in the separate manual "Mapping Units of Measure".

However, essentially should really move to the UCUM system for units of measure, as this allows to fully automate conversions, as for example between "conventional units" and "SI units" and the other way around. SDTM-ETL contains some functions that work with the NLM RESTful Web Services for automated unit conversions, which are also described in the separate manual.

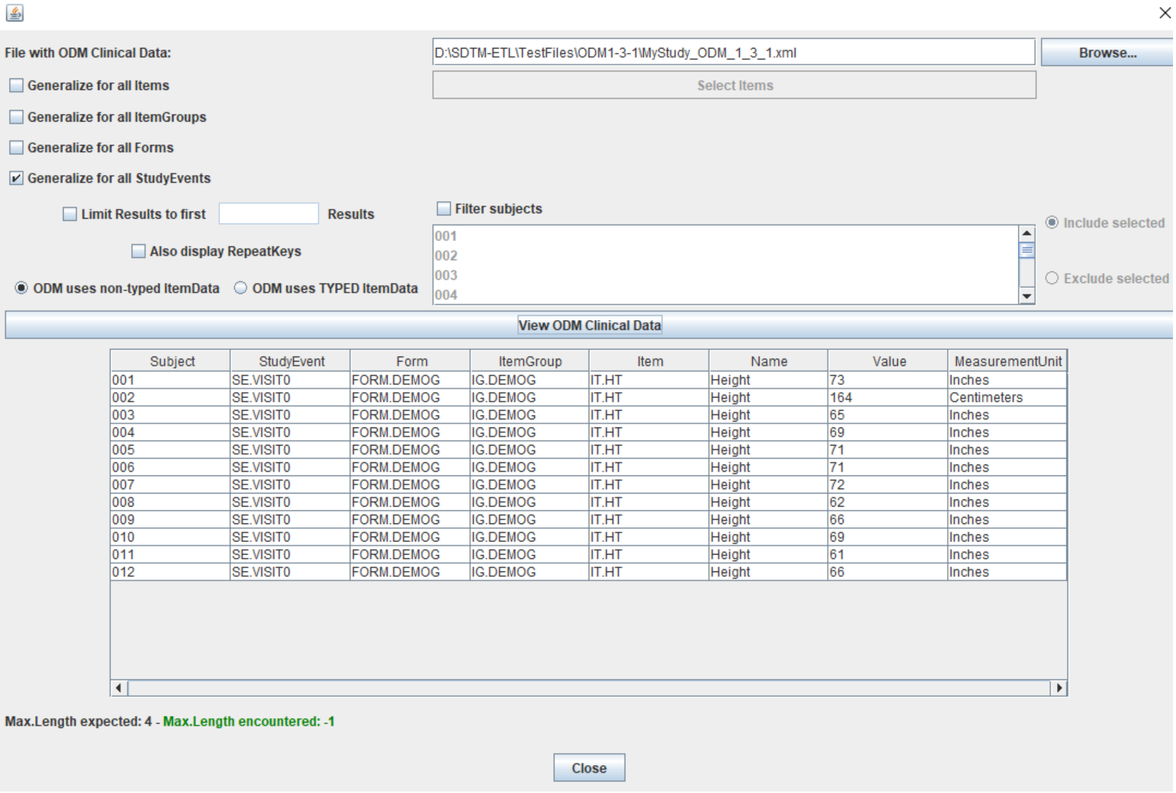
Working with Viedoc-ODM exports

Most of the modern EDC systems allow to export the data and metadata of the studies in CDISC-ODM format. Some of these have "vendor extensions" and/or use the ODM "typed ItemData" mechanism. This is e.g. the case with Viedoc EDC. Viedoc also has a somewhat special mechanism to treat different stages of the study design development, using different "ODM MetaDataVersion" instances.
As we have more and more users using Viedoc EDC, we further improved the way Viedoc-

ODM is handled. These new features are described in the separate manual "Working with Viedoc ODM Files".

Extended features for "View - Clinical Data"

Often, it makes sense to inspect the collected clinical data before starting using a selected data point in a mapping, this in order to better understand what the data is about, check on the data type, etc.. One then selects an Item of the ODM tree and uses the menu "View - ODM Clinical Data", for example, when one has selected "Height" in the ODM tree:



File with ODM Clinical Data: D:\SDTM-ETL\TestFiles\ODM1-3-1\MyStudy_ODM_1_3_1.xml

☐ Generalize for all Items

☐ Generalize for all ItemGroups

☐ Generalize for all Forms

☑ Generalize for all StudyEvents

☐ Limit Results to first Results

☐ Also display RepeatKeys

● ODM uses non-typed ItemData ○ ODM uses TYPED ItemData

☐ Filter subjects

001
002
003
004

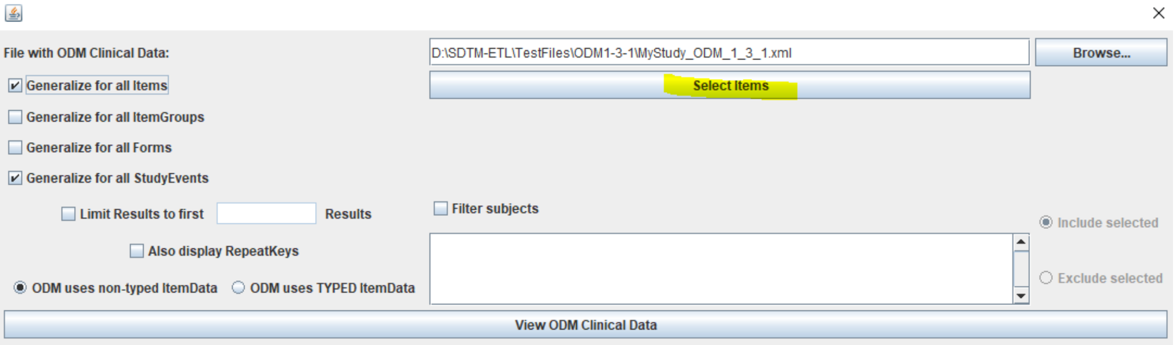
● Include selected ○ Exclude selected

Subject	StudyEvent	Form	ItemGroup	Item	Name	Value	MeasurementUnit
001	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	73	Inches
002	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	164	Centimeters
003	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	65	Inches
004	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	69	Inches
005	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	71	Inches
006	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	71	Inches
007	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	72	Inches
008	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	62	Inches
009	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	66	Inches
010	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	69	Inches
011	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	61	Inches
012	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	66	Inches

Max.Length expected: 4 - Max.Length encountered: -1

New in v.4.5 of the software is that also the "decode" for the units of measure "MeasurementUnits" is displayed (last column in the screenshot above).

Some of our users also asked to extend this feature to enable to inspect data for more than one selected item at the same time, for example both "Height" and "Weight".
This is now possible using the button "Select Items" when one has checked the checkbox "Generalize for all Items":



File with ODM Clinical Data: D:\SDTM-ETL\TestFiles\ODM1-3-1\MyStudy_ODM_1_3_1.xml

☑ Generalize for all Items

☐ Generalize for all ItemGroups

☐ Generalize for all Forms

☑ Generalize for all StudyEvents

☐ Limit Results to first Results

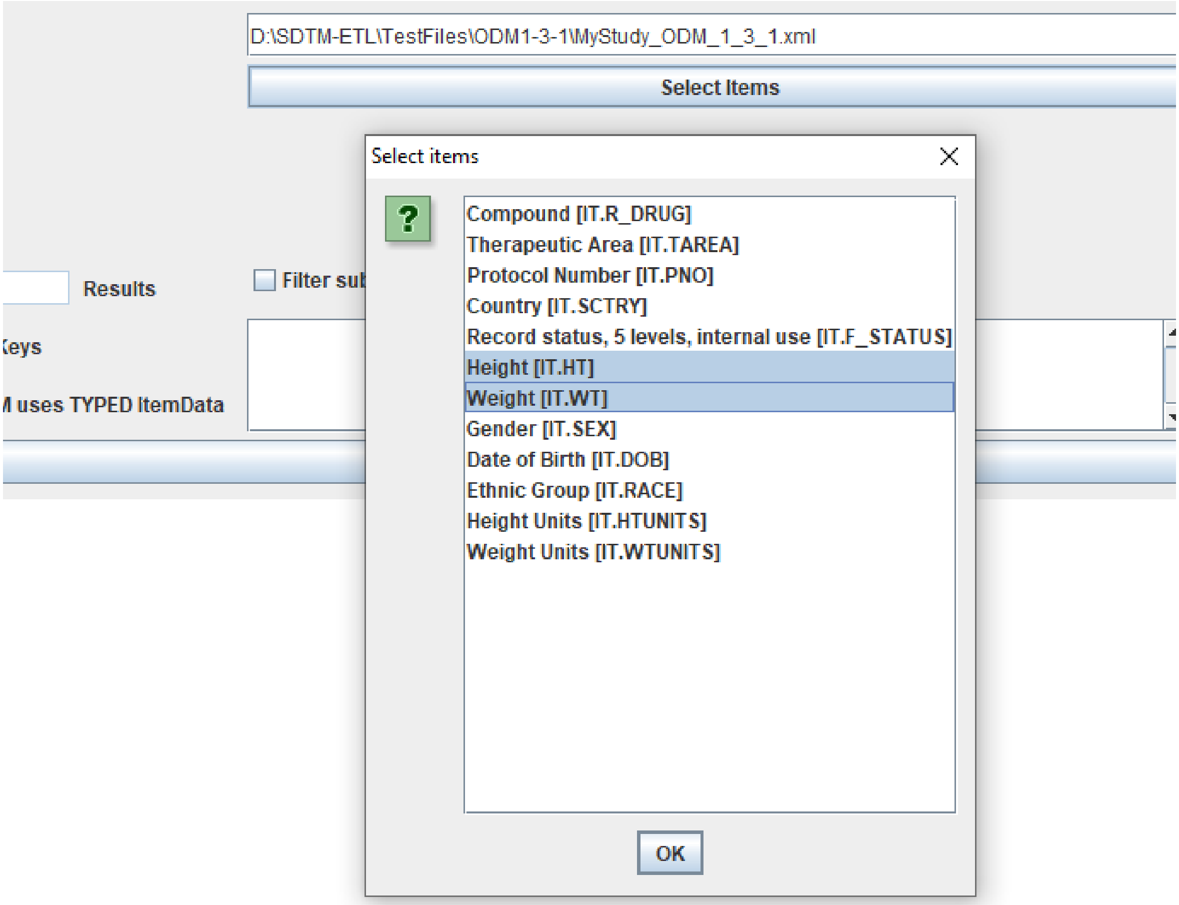
☐ Also display RepeatKeys

● ODM uses non-typed ItemData ○ ODM uses TYPED ItemData

☐ Filter subjects

● Include selected ○ Exclude selected

When one then clicks "Select Items", a dialog is displayed where we can do a selection, e.g.:

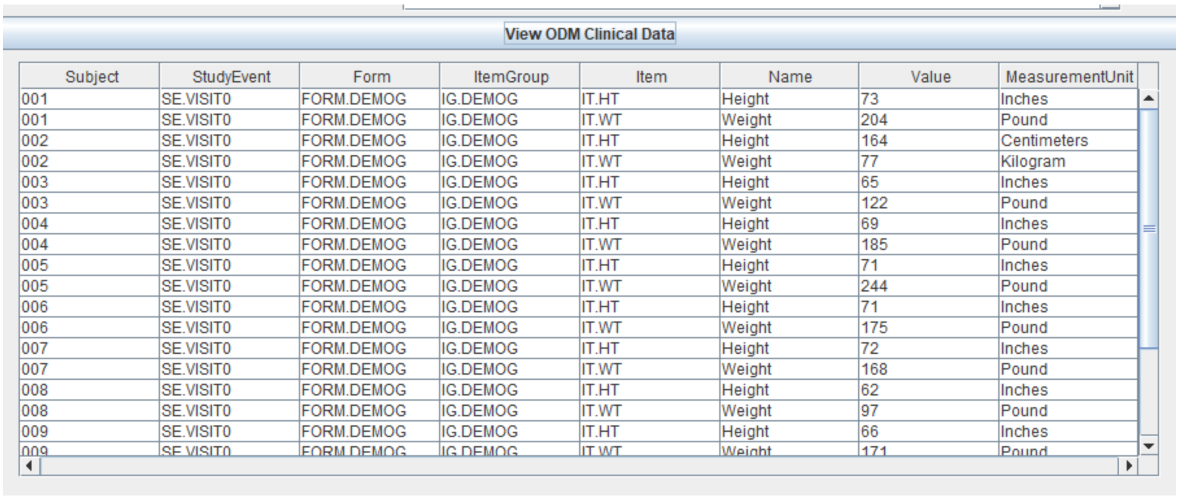


D:\SDTM-ETL\TestFiles\ODM1-3-1\MyStudy_ODM_1_3_1.xml

Select items

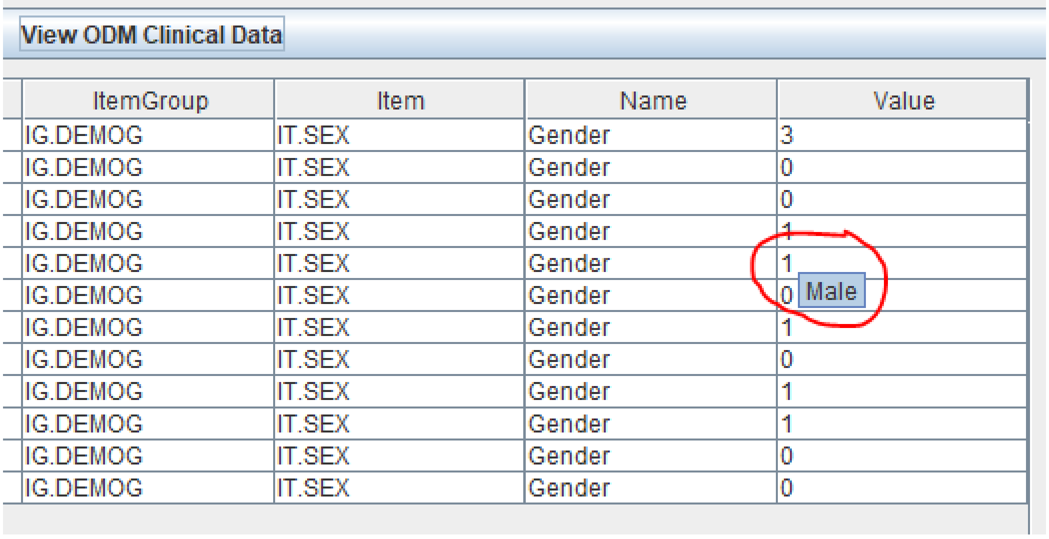
- Compound [IT.R_DRUG]
- Therapeutic Area [IT.TAREA]
- Protocol Number [IT.PNO]
- Country [IT.SCTRY]
- Record status, 5 levels, internal use [IT.F_STATUS]
- Height [IT.HT]
- Weight [IT.WT]
- Gender [IT.SEX]
- Date of Birth [IT.DOB]
- Ethnic Group [IT.RACE]
- Height Units [IT.HTUNITS]
- Weight Units [IT.WTUNITS]

then leading to, after clicking "View ODM Clinical Data agains" to:



Subject	StudyEvent	Form	ItemGroup	Item	Name	Value	MeasurementUnit
001	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	73	Inches
001	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.WT	Weight	204	Pound
002	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	164	Centimeters
002	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.WT	Weight	77	Kilogram
003	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	65	Inches
003	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.WT	Weight	122	Pound
004	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	69	Inches
004	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.WT	Weight	185	Pound
005	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	71	Inches
005	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.WT	Weight	244	Pound
006	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	71	Inches
006	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.WT	Weight	175	Pound
007	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	72	Inches
007	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.WT	Weight	168	Pound
008	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	62	Inches
008	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.WT	Weight	97	Pound
009	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.HT	Height	66	Inches
009	SE.VISIT0	FORM.DEMOG	IG.DEMOG	IT.WT	Weight	171	Pound

Another feature is that when a value is coded, holding the mouse over the value also provides the "decode", e.g.:

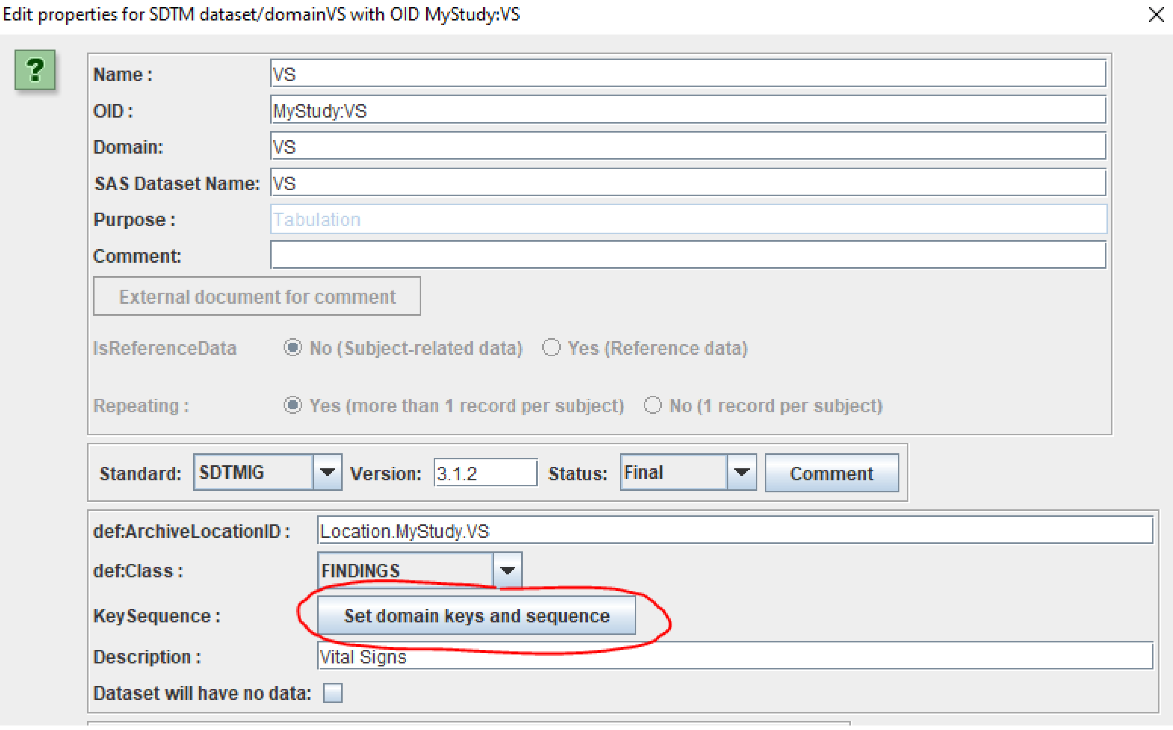


View ODM Clinical Data

ItemGroup	Item	Name	Value
IG.DEMOG	IT.SEX	Gender	3
IG.DEMOG	IT.SEX	Gender	0
IG.DEMOG	IT.SEX	Gender	0
IG.DEMOG	IT.SEX	Gender	1
IG.DEMOG	IT.SEX	Gender	1
IG.DEMOG	IT.SEX	Gender	0 Male
IG.DEMOG	IT.SEX	Gender	1
IG.DEMOG	IT.SEX	Gender	0
IG.DEMOG	IT.SEX	Gender	1
IG.DEMOG	IT.SEX	Gender	1
IG.DEMOG	IT.SEX	Gender	0
IG.DEMOG	IT.SEX	Gender	0

Sorting using the define.xml "keys"

One of the features already longer present is the ability to sort the generated data according to the keys in the define.xml ("KeySequence"). Adding these keys is done by a double-click on the first cell of the row for the study-specific domain, or using the menu "Edit - SDTM/SEND Domain Properties" or using Shift-E on the keyboard. For example:



Edit properties for SDTM dataset/domainVS with OID MyStudy:VS

Name : VS

OID : MyStudy:VS

Domain: VS

SAS Dataset Name: VS

Purpose : Tabulation

Comment:

IsReferenceData ● No (Subject-related data) ○ Yes (Reference data)

Repeating : ● Yes (more than 1 record per subject) ○ No (1 record per subject)

Standard: SDTMIG Version: 3.1.2 Status: Final

def:ArchiveLocationID : Location.MyStudy.VS

def:Class : FINDINGS

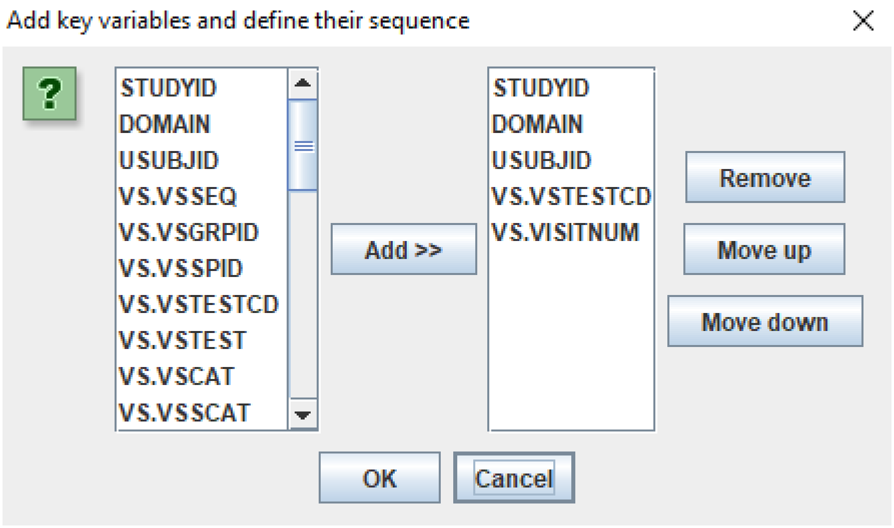
KeySequence : Set domain keys and sequence

Description : Vital Signs

Dataset will have no data: □

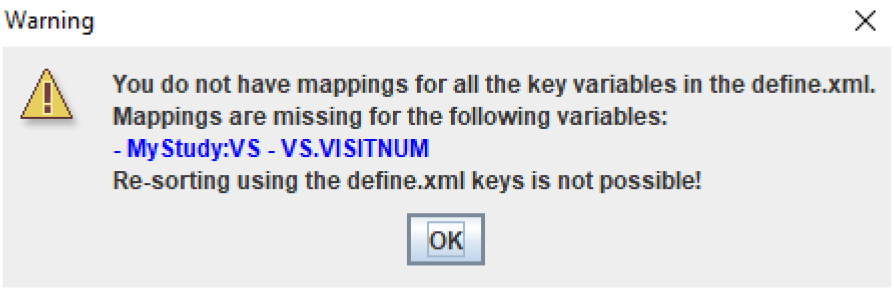
When then clicking "Set domain keys and sequence, one can add the keys, and their order, which one want to use for sorting[1]. For example:

[1] Interestingly, these keys were never meant for sorting, but only for defining record uniqueness. Many people however also use them for (re)sorting as it looks as the tools of the regulatory reviewers are not capable to do sorting themselves.



There are however a few "no-go"s: one should not request to sort according to --SEQ, as this is an "artificial key", which is added after sorting ...

New in v.4.5 is that when adds a variable to the keys for which there is no mapping, the system will protest, issue a warning message, and disallow sorting for that domain/dataset. For example, when there is no mapping (yet) for VS.VISITNUM, and when trying to sort after execution of the mappings:



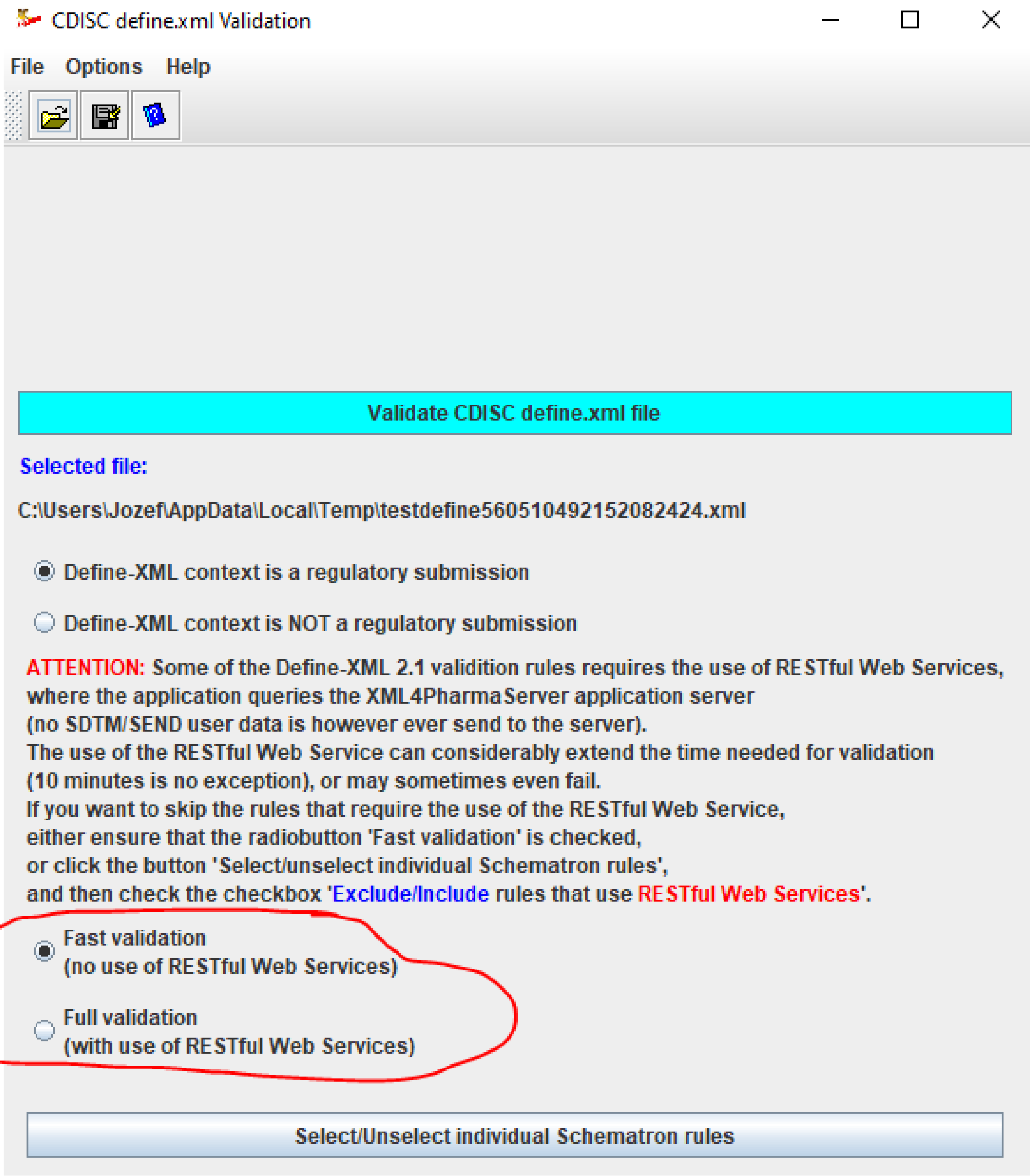
You do not have mappings for all the key variables in the define.xml.
Mappings are missing for the following variables:
- MyStudy:VS - VS.VISITNUM
Re-sorting using the define.xml keys is not possible!

The execution can then performed without sorting, or one may first decide to add a mapping for the key variable VS.VISITNUM first.

Define.xml validation

The dialog for doing define.xml validation has been improved to make it even more user-friendly. The background is that some of the validation uses RESTful web services (RWS), e.g. to check the content of CDISC controlled terminology[2]. This can take considerable time, or even fail when there is no good internet connection or the server is e.g. under maintenance. To make this more clear, the GUI for the validation was slightly changed into:

[2] clinical data / subject data is never submitted to these RESTful Web Services.



Validate CDISC define.xml file

Selected file:

C:\Users\Jozef\AppData\Local\Temp\testdefine560510492152082424.xml

● Define-XML context is a regulatory submission

○ Define-XML context is NOT a regulatory submission

ATTENTION: Some of the Define-XML 2.1 validition rules requires the use of RESTful Web Services, where the application queries the XML4Pharma Server application server (no SDTM/SEND user data is however ever send to the server).
The use of the RESTful Web Service can considerably extend the time needed for validation (10 minutes is no exception), or may sometimes even fail.
If you want to skip the rules that require the use of the RESTful Web Service, either ensure that the radiobutton 'Fast validation' is checked, or click the button 'Select/unselect individual Schematron rules', and then check the checkbox 'Exclude/Include rules that use RESTful Web Services'.

● Fast validation
(no use of RESTful Web Services)

○ Full validation
(with use of RESTful Web Services)

Select/Unselect individual Schematron rules

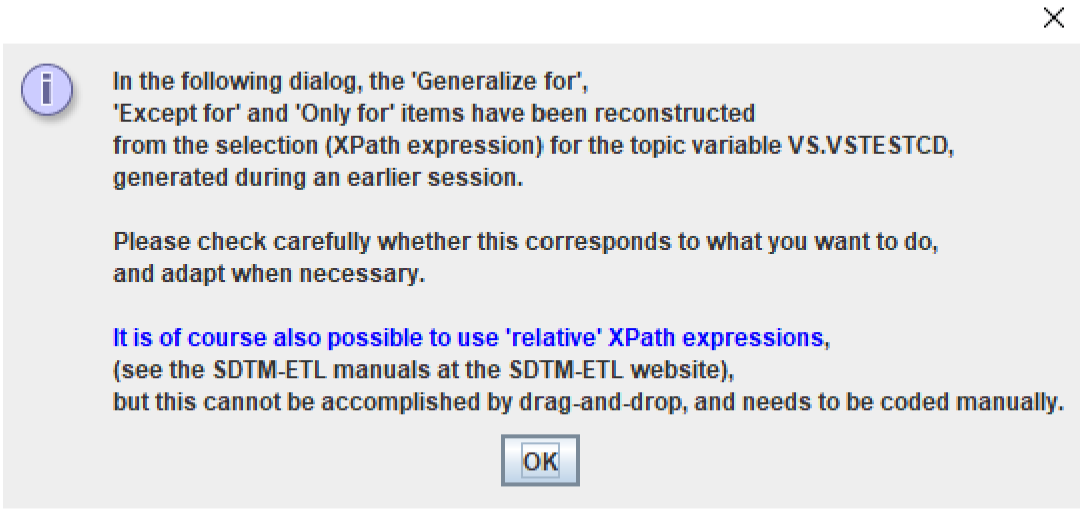
allowing the user to select out those validation rules that use RWS without needing to go into the sub-dialog to select/unselect individual schematron rules.

Further minor improvements and fixes

* Upon request of a customer, the logging system has been extended so that the log files can also be used an "audit trail" of the activities performed by the user such as generation of mappings (with or without the "mapping wizard", codelist subsetting and editing, valuelist generation, etc..
* Some EDC systems have the bad habit to export ODM with identical values for "OID" and "Name" for the items (ItemDef elements), which may make mapping more difficult. In such a case, more information is then in the text of the "Question" child element.
When doing the mappings from the ODM to the SDTM, and the "Mapping Wizard" is used, the "ODM side" of the mapping wizard will usually as well the value of the "OID" as of the

"Name". This doesn't bring much additional information though when both are identical. In such a case, the system will try to retrieve the value of the "Question" and display that instead of the value of "Name".

* At the end of a working day, one will usually want to save the define.xml containing all the mappings developed so far, and continue the next morning.
In the new version, the software will try to reconstruct the "Only for" and "Except for" information after a drag-and-drop when this is not readily available, which is usually the cae when resuming work of the previous day. In such a case, the following dialog is displayed:



In the following dialog, the 'Generalize for',
'Except for' and 'Only for' items have been reconstructed
from the selection (XPath expression) for the topic variable VS.VSTESTCD,
generated during an earlier session.

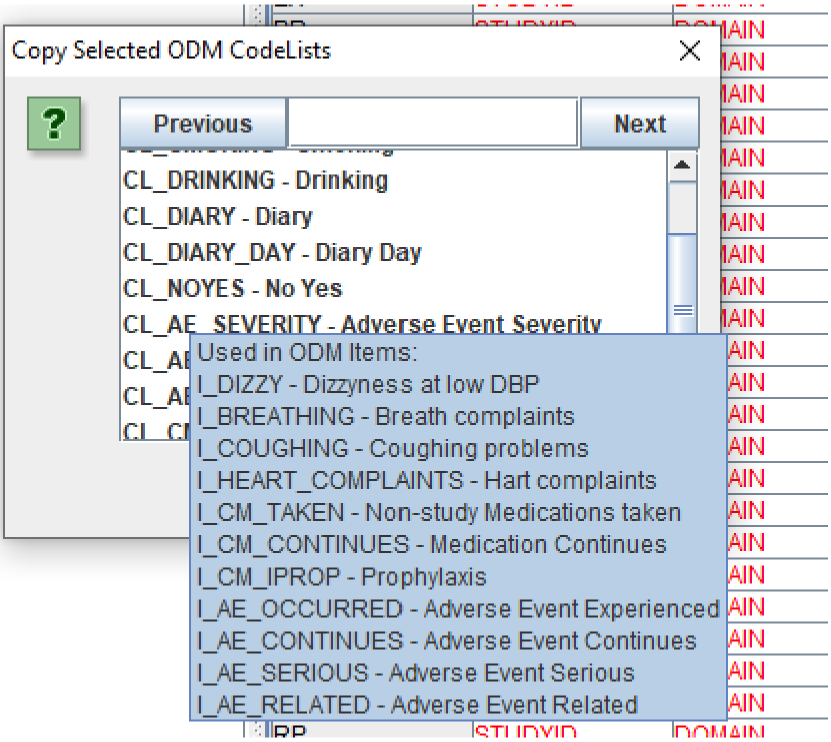
Please check carefully whether this corresponds to what you want to do,
and adapt when necessary.

It is of course also possible to use 'relative' XPath expressions,
(see the SDTM-ETL manuals at the SDTM-ETL website),
but this cannot be accomplished by drag-and-drop, and needs to be coded manually.

* We also made improvements in the code for the case that datasets for the same domain must be merged[3]. This is often the case for datasets for the LB domain, where one wants to have one dataset for e.g. "hematology" (LBHM), one for "urinalysis" (LBUR), etc.. This strategy not only makes it considerably easier to generate the mappings, but also makes it easier for the regulatory authorities to review the data. Often however, the regulatory authority also requires a "merged" LB dataset, although they cannot do anything with it, as their tools cannot handle such large files.
For example, the software now better handles the case that several SUPPLB datasets such as SUPPLBHM, SUPPLBUR must be merged into a single SUPPLB dataset.

* In a number of cases, one will need to copy an ODM codelist to SDTM, for example when the value in the ODM was coded, but there is no SDTM codelist for the variable the item is mapped to. When doing so, the user is asked which of the codelists must be copied to SDTM. In order to facilitate the selection, a new feature has been added so that when holding the mouse over a codelist, it is shown for which items in the ODM, the codelist is used. For example:

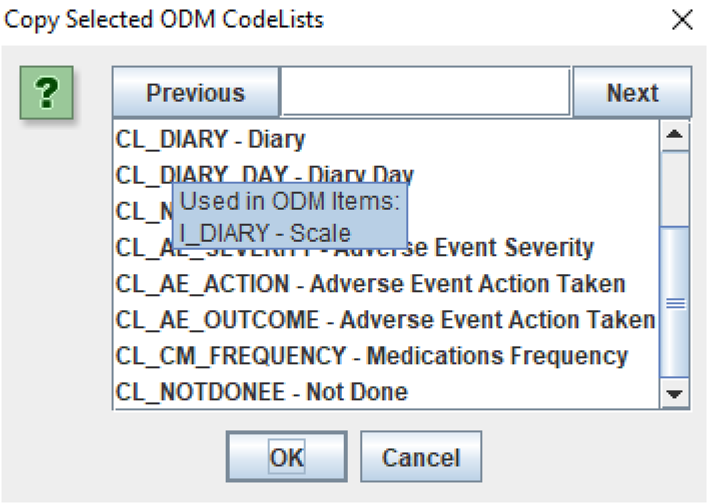
[3] Interestingly, FDA and CDISC name this "split" datasets, although they were never split, as they were essentially generated separately.



CL_DRINKING - Drinking
CL_DIARY - Diary
CL_DIARY_DAY - Diary Day
CL_NOYES - No Yes
CL_AE_SEVERITY - Adverse Event Severity

Used in ODM Items:
I_DIZZY - Dizzyness at low DBP
I_BREATHING - Breath complaints
I_COUGHING - Coughing problems
I_HEART_COMPLAINTS - Hart complaints
I_CM_TAKEN - Non-study Medications taken
I_CM_CONTINUES - Medication Continues
I_CM_IPROP - Prophylaxis
I_AE_OCCURRED - Adverse Event Experienced
I_AE_CONTINUES - Adverse Event Continues
I_AE_SERIOUS - Adverse Event Serious
I_AE_RELATED - Adverse Event Related

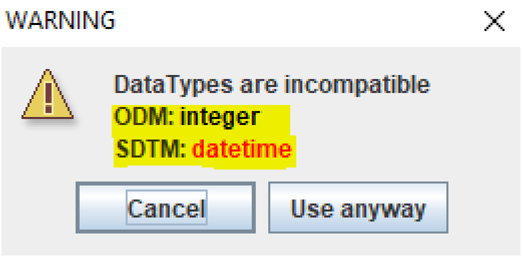
which may not be very useful, as also SDTM has a "No-Yes" codelist, but for e.g. "Diary Scale":



CL_DIARY - Diary
CL_DIARY_DAY - Diary Day
CL_AE_SEVERITY - Adverse Event Severity
CL_AE_ACTION - Adverse Event Action Taken
CL_AE_OUTCOME - Adverse Event Action Taken
CL_CM_FREQUENCY - Medications Frequency
CL_NOTDONEE - Not Done

Used in ODM Items:
I_DIARY - Scale

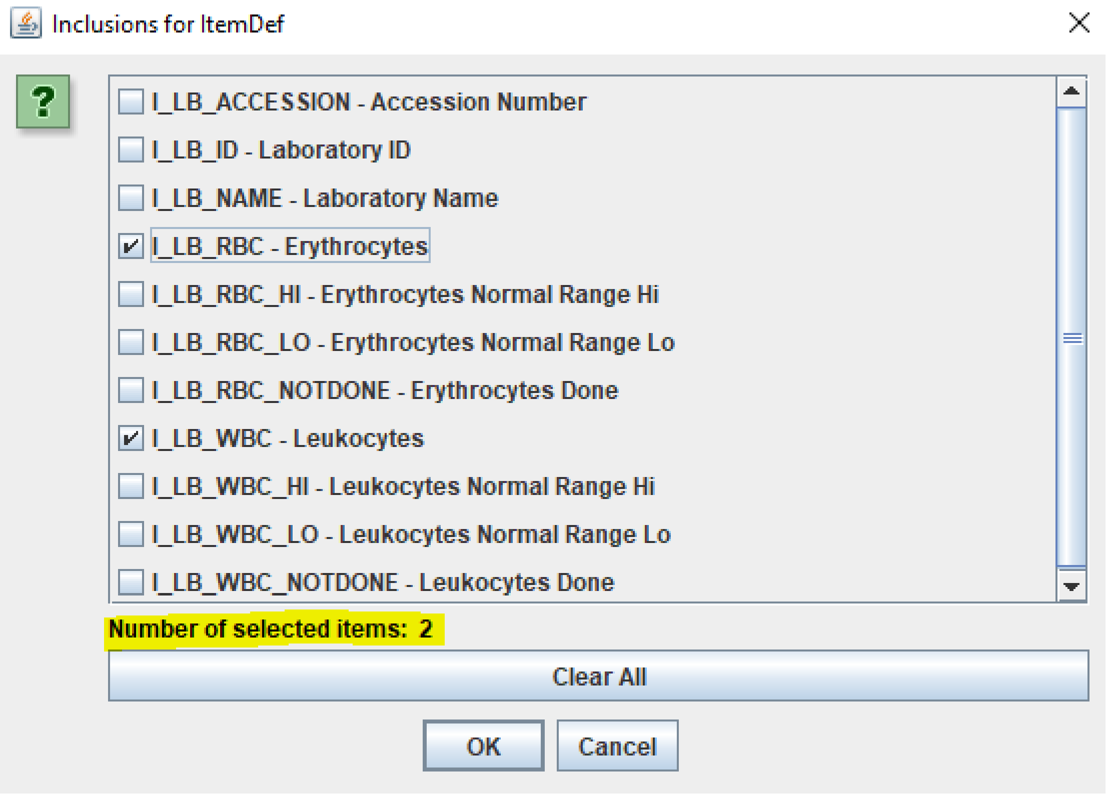
Another minor improvement is that when, upon drag-and-drop, there is a mismatch in data types between ODM item and SDTM/SEND variable, and a warning message is issued, more information is provided about the data types themselves. For example:



DataTypes are incompatible
ODM: integer
SDTM: datetime

* A similar minor improvement is that when using "Only for" or "Except for" when doing a

selection after drag-and-drop, also the number of already selected items is displayed and updated. For example:



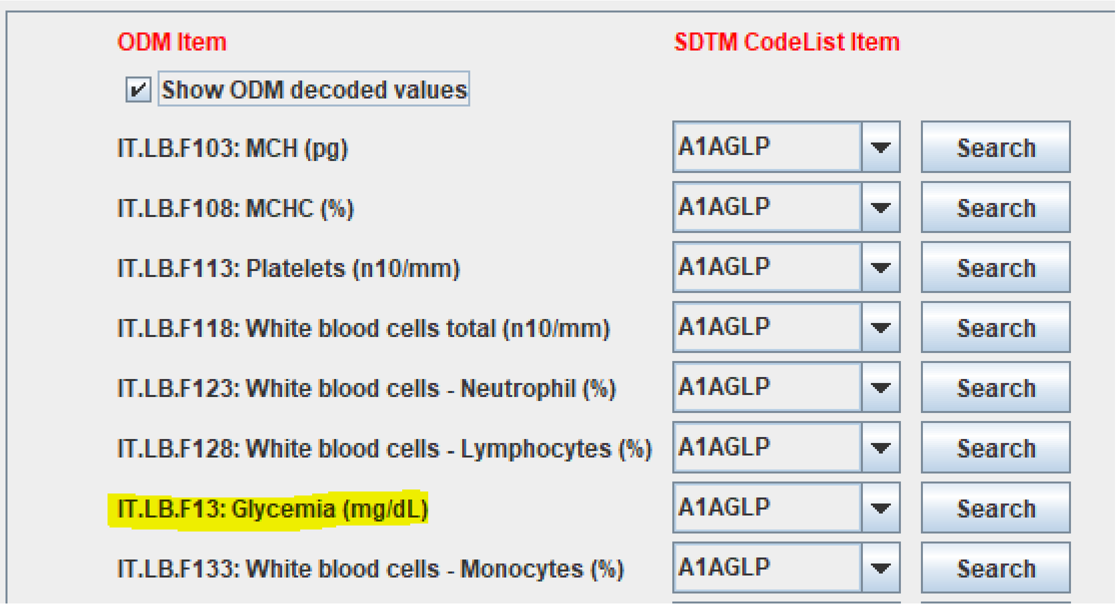
- □ I_LB_ACCESSION - Accession Number
- □ I_LB_ID - Laboratory ID
- □ I_LB_NAME - Laboratory Name
- ☑ I_LB_RBC - Erythrocytes
- □ I_LB_RBC_HI - Erythrocytes Normal Range Hi
- □ I_LB_RBC_LO - Erythrocytes Normal Range Lo
- □ I_LB_RBC_NOTDONE - Erythrocytes Done
- ☑ I_LB_WBC - Leukocytes
- □ I_LB_WBC_HI - Leukocytes Normal Range Hi
- □ I_LB_WBC_LO - Leukocytes Normal Range Lo
- □ I_LB_WBC_NOTDONE - Leukocytes Done

Number of selected items: 2

This can be helpful for the case of very long lists. For example, we had the case that (due to the ODM being generated from an Excel worksheet), there were over 250 fields representing 53 distinct lab tests. If the user know he/she needs to select 53 fields for the 53 tests, than the display of the number of (already) selected items helps in ensuring that nothing was forgotten, or that not too many items were selected.

* Mapping lab test names (as exported to ODM from an EDC system) to CDISC-SDTM Lab Controlled Terminology can be extremely tedious. One of the reasons is that the SDTM codelist for LBTESTCD and LBTEST both contain almost 2,400 codes (status May 2024) from which the user needs to choose for each single test as exported into the ODM[4].
When doing such a mapping, one can already use the "Attempt 1:1 mapping" feature, which is based on word similarity, but this will not always select the right code or term. For example:

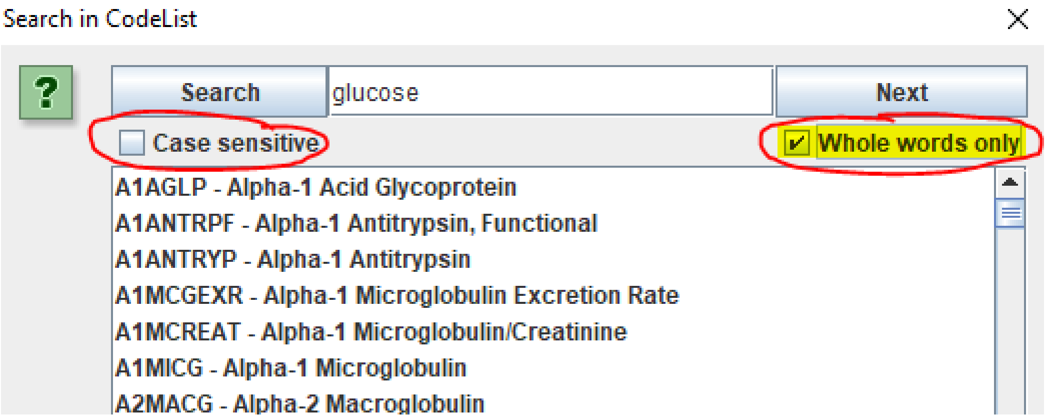
[4] One of the reasons for this long list is that CDISC still considers all controlled terminology as a "set of lists" (one-dimensional) without any possibility for hierarchy, cross-links to other controlled terminology.
CDISC also still refuses to use better systems such as LOINC for lab tests and UCUM for units.



☑ Show ODM decoded values

ODM Item	SDTM CodeList Item	
IT.LB.F103: MCH (pg)	A1AGLP	Search
IT.LB.F108: MCHC (%)	A1AGLP	Search
IT.LB.F113: Platelets (n10/mm)	A1AGLP	Search
IT.LB.F118: White blood cells total (n10/mm)	A1AGLP	Search
IT.LB.F123: White blood cells - Neutrophil (%)	A1AGLP	Search
IT.LB.F128: White blood cells - Lymphocytes (%)	A1AGLP	Search
IT.LB.F13: Glycemia (mg/dL)	A1AGLP	Search
IT.LB.F133: White blood cells - Monocytes (%)	A1AGLP	Search

For "glycemia", the "1:1 attempt" will not find anything useful, but our lab specialist told me it means "Glucose", so we want to search for "Glucose" using the "Search" button:



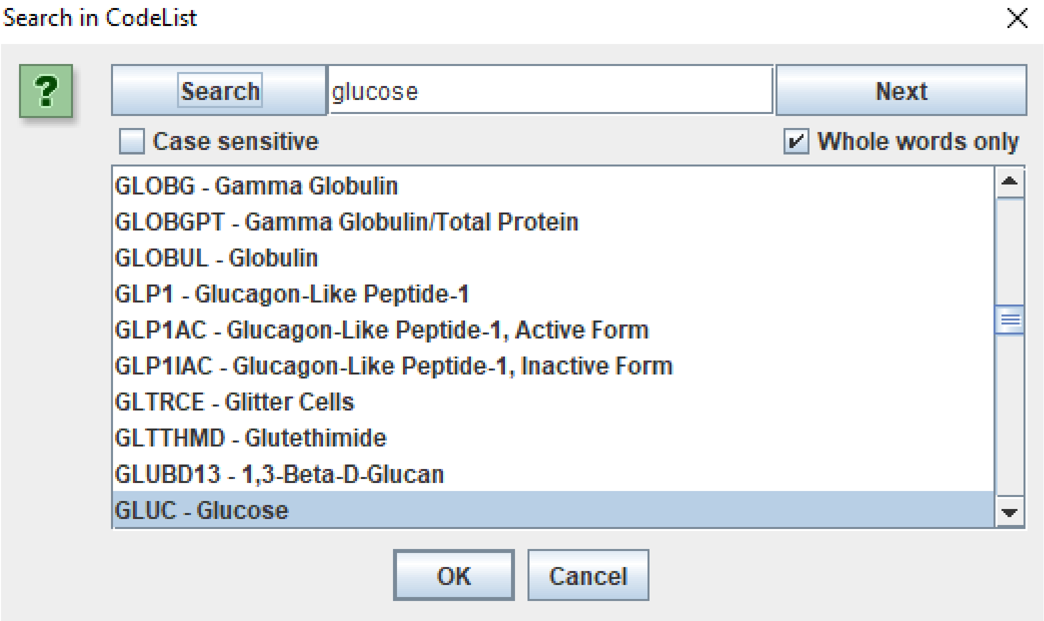
Search in CodeList

Search: glucose

□ Case sensitive ☑ Whole words only

A1AGLP - Alpha-1 Acid Glycoprotein
A1ANTRPF - Alpha-1 Antitrypsin, Functional
A1ANTRYP - Alpha-1 Antitrypsin
A1MCGEXR - Alpha-1 Microglobulin Excretion Rate
A1MCREAT - Alpha-1 Microglobulin/Creatinine
A1MICG - Alpha-1 Microglobulin
A2MACG - Alpha-2 Macroglobulin

New as of v.4.5 of the software are the checkboxes "Case sensitive" and "Whole words only" which enable a considerable faster search. And indeed we immediately find:



Search in CodeList

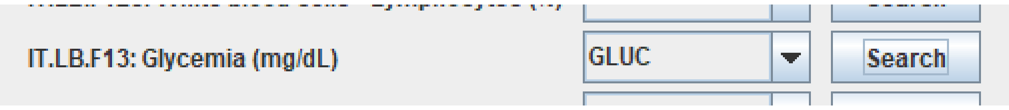
glucose

□ Case sensitive

☑ Whole words only

GLOBG - Gamma Globulin
GLOBGPT - Gamma Globulin/Total Protein
GLOBUL - Globulin
GLP1 - Glucagon-Like Peptide-1
GLP1AC - Glucagon-Like Peptide-1, Active Form
GLP1IAC - Glucagon-Like Peptide-1, Inactive Form
GLTRCE - Glitter Cells
GLTTHMD - Glutethimide
GLUBD13 - 1,3-Beta-D-Glucan
GLUC - Glucose

which leads to "GLUC" being selected as the mapping for "Glycemia":



IT.LB.F13: Glycemia (mg/dL)

GLUC

* One of our customers reported an "Invalid XPath Expression Error" message when working with very long (a large number of selections) XPath expressions. It was however hard to find out what the exact cause of the problem was. The reason was that some versions of Java do not allow to have more than 100 "operators" in an XPath expression, which can easily be solved by adding a parameter "-Djdk.xml.xpathExprOpLimit" to the SDTM-ETL.bat startup file with a parameter value set to a higher value, e.g. "1000".
As the error message dialog did not reveal this level of detail (it just said "Invalid XPath expression), we extended the handling of such errors to provide considerably more details, so that a solution can be found much faster.

Further development of SDTM-ETL

It is now very clear that the FDA is committed to push the replacement of the SAS Transport 5 (XPT) format by the by CDISC developed Dataset-JSON format: the pilot with the FDA was extremely successful, and a CDISC working group (of which we are part) is currently refining the new standard, which will be name Dataset-JSON 1.1.

It is also very clear that FDA fully supports the CDISC CORE project for validation of submission datasets. For example, FDA has asked CDISC to add all "FDA business rules" to CORE. Essentially, this means that for validation, Pinnacle21 is expected to be replaced by CDISC CORE, also at the agency.

Both these have implications for the future versions of the SDTM-ETL software: the next major version (5.0) will have Dataset-JSON 1.1 as the primary format for the generated datasets and SAS Transport 5 as the secondary. Support for the old Dataset-XML format will probably be terminated.

As one of the co-developers of CDISC-CORE we will further support and extend the use of this open-source validation software, so that the world can finally get rid of this buggy and user-unfriendly P21 software.
We are thinking about features such as "validate as you map", providing immediate feedback for the CORE validation engine for each case that a mapping was added or changed.
Also, as soon as it is available, we will add and support the feature that users (or companies) can add their own validation rules to the engine in a secure way[5].

[5] also meaning that any user- or company-specific validation rules are not shared with CDISC nor anyone else outside the company.