Smart Submission Dataset Viewer User Manual

By XML4Pharma Last update: 2024-12-16

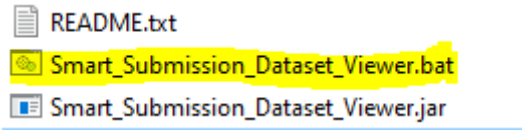
This user manual and also other manuals and tutorials can always be found on our website:
http://www.xml4pharma.com/Smart_Submission_Dataset_Viewer/

Table of contents

Starting the application

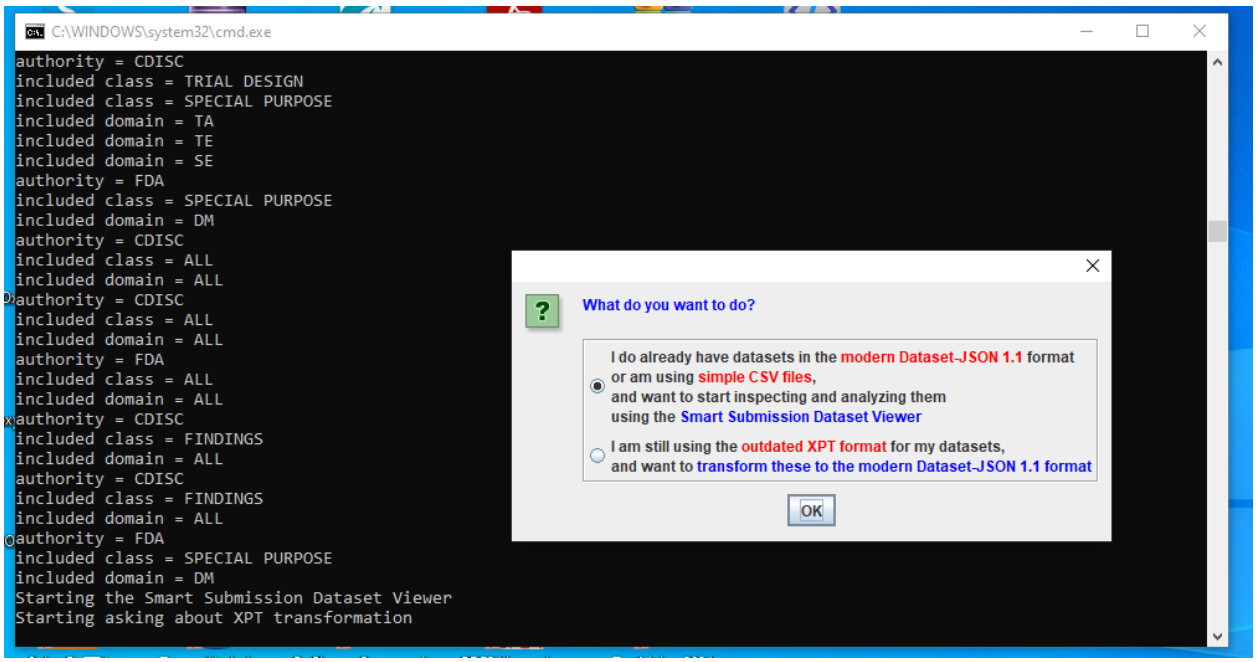
The software comes as a Java application. For installation instructions, please find them here. The document also contains information about setting up a desktop shortcut for the software.

Essentially, one starts the software by a double-click on "Smart_Submission_Dataset_Viewer.bat" (Windows) or "Smart_Submission_Dataset_Viewer.sh" (Linux and Mac) in the folder where you installed the "Smart Submission Dataset Viewer". For example:



or on the corresponding shortcut item on the desktop.

This will start with the opening screen:

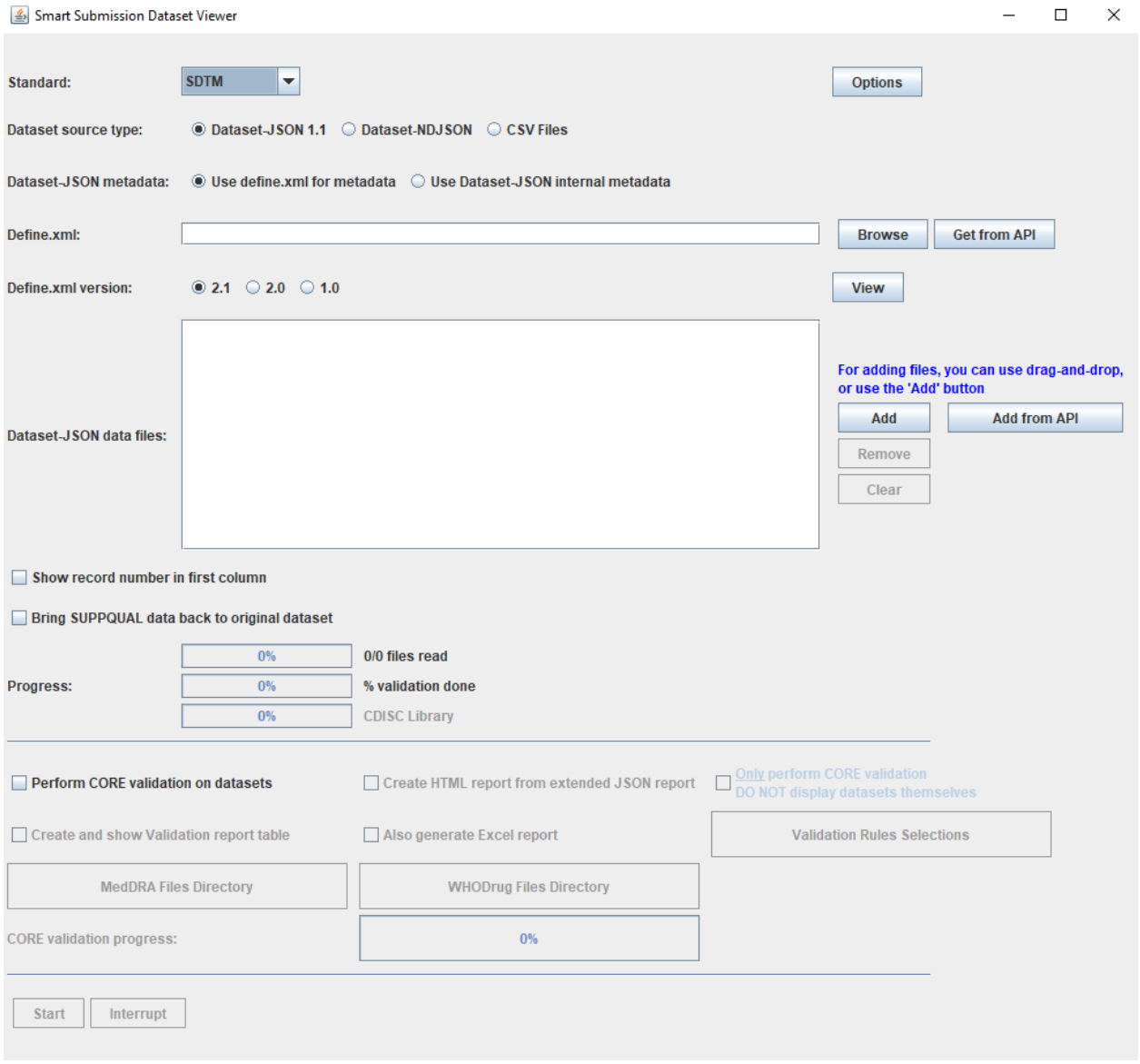


At the same time, a "console window" (the "black" window) is opened, that will display all information messages produced by the software, and which are also stored in a log file, which can be find in the folder "logs".

The opening screen asks whether one already has modern CDISC Dataset-JSON 1.1 files available, or that one wants to convert existing SAS-XPT files to Dataset-JSON. For the latter, see the separate tutorial "Generating Dataset-JSON from SAS-XPT".
If one already uses Dataset-JSON, and one does not want to be asked again about conversion from SAS-XPT, this opening screen can be skipped by setting a parameter in the "properties.dat" file. See the section "The 'properties.dat' file".

Selecting data standard, define.xml and to-be-loaded files

In case no conversion from SAS-XPT file is needed as one already has files in Dataset-JSON format, clicking "OK" leads to the main screen:



First, one will want to select the standard for which the viewer will be used. Current choices are: "SDTM", "SEND", "ADaM" and "Other tabular". In this manual, we will use an SDTM example.
These choices are there in order to optimize display and additional features. For example, CDISC ADaM does not have the concept of "Supplemental Qualifier datasets" (SUPPQUAL), so when ADaM is selected, the checkbox "Bring SUPPQUAL data back to original dataset" will be disabled.
Also, when one selects "ADaM, some new choices appear, such as:

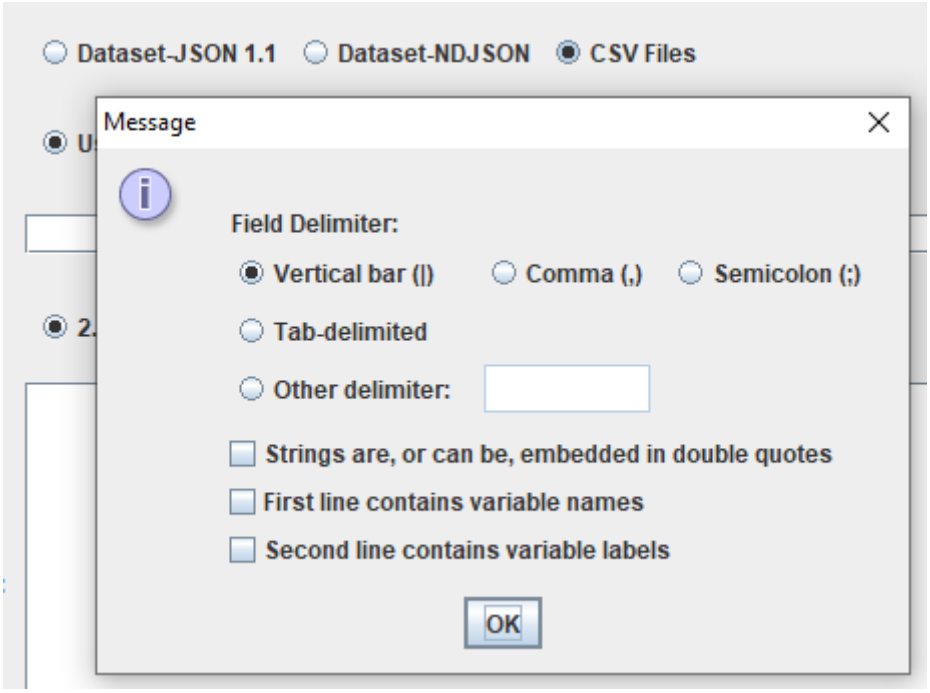
We will explain the "Options" button on the right top later.

After having selected the standard, one must select in which format the data will delivered to the viewer. The choices are:

- Dataset-JSON 1.1: This is the "regular" CDISC Dataset-JSON 1.1 format, for which the specification can be found at: https://cdisc-org.github.io/DataExchange-DatasetJson/doc/dataset-json1-1.html
You will probably want to have your submission data files in this format in 90% of the cases.
Information about software packages and tools to develop the mappings for, and generate submission files in Dataset-JSON can be found at:
https://wiki.cdisc.org/display/DSJSONHACK/Hackathon+I+Projects
Some SDTM/SEND mapping tools, such as the "SDTM-ETL" software, also allow to generate submission datasets in Dataset-JSON format.

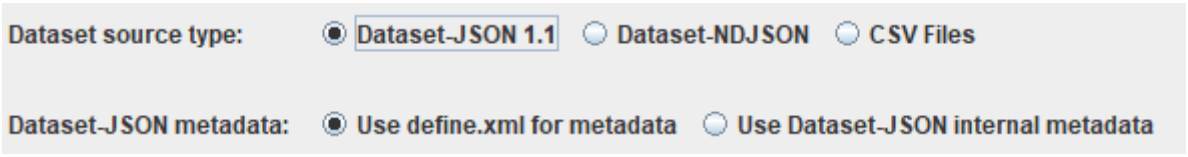
- Dataset-NDJSON
This is a special form of Dataset-JSON 1.1 in which each record is on a separate line in the file, each line containing a separate JSON object (ND stands for "newline delimited"). It can be advantageous to be able to process for very large datasets in software tools that store the JSON in memory.
The specification can also be found on the CDISC Wiki.
The Smart Submission Dataset Viewer however uses "streaming" when reading the contents of the files, so that, at least for our software, there is no difference in performance between Dataset-JSON and Dataset-NDJSON. But as people might want to generate Dataset-NDJSON for other reasons, we of course need to support it.

- CSV (Comma-separated Values).
When selecting "CSV", the following dialog is shown:



allowing the user to select or assign the delimiter between fields to be used and further details how the data is organized.
Using CSV can be useful when e.g. obtaining SDTM/SEND/ADaM files as Excel or any other spreadsheet format.

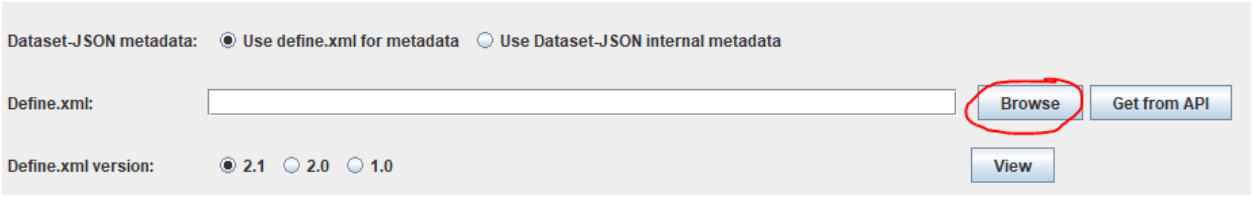
Suppose we selected "SDTM" and "Dataset-JSON", then we need to select whether we want to use a define.xml file for loading the metadata for the files, or that we just want to use the metadata that is already present in the Dataset-JSON files themselves.



As the amount of metadata in the Dataset-JSON files themselves is very limited, it is always best to use the define.xml for loading the metadata. This however requires to have a good quality define.xml.
Generating such a define.xml starting from an Excel file (as a "specification") is not always a good idea, as it often produces low-quality define.xml files. There are however different mapping tools on the market that generate a quality define.xml "synchronously" with the mappings themselves. There is also a good number of software tools for developing, generating and updating/extending define.xml on the market that come with a user-friendly graphical user interface.

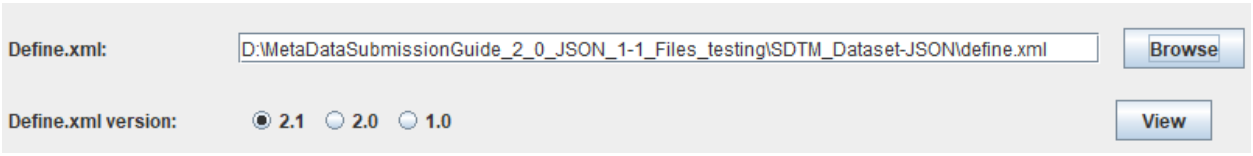
Depending on whether one selects "Use define.xml for metadata" or "Use Dataset-JSON internal metadata", some features will become available or disabled. For example, the feature "Bring SUPPQUAL data back to original dataset" requires information from the define.xml. As such, this checkbox will be disabled when one selected "Use Dataset-JSON internal metadata". Also many data validation features, such as checking values against their variable codelist, are only possible when using the define.xml for loading the metadata (see later).

When having selected "Use define.xml for metadata", one should now first select the define.xml, by using the "Browse" button, and then select the define.xml that is applicable to the datasets to be loaded.



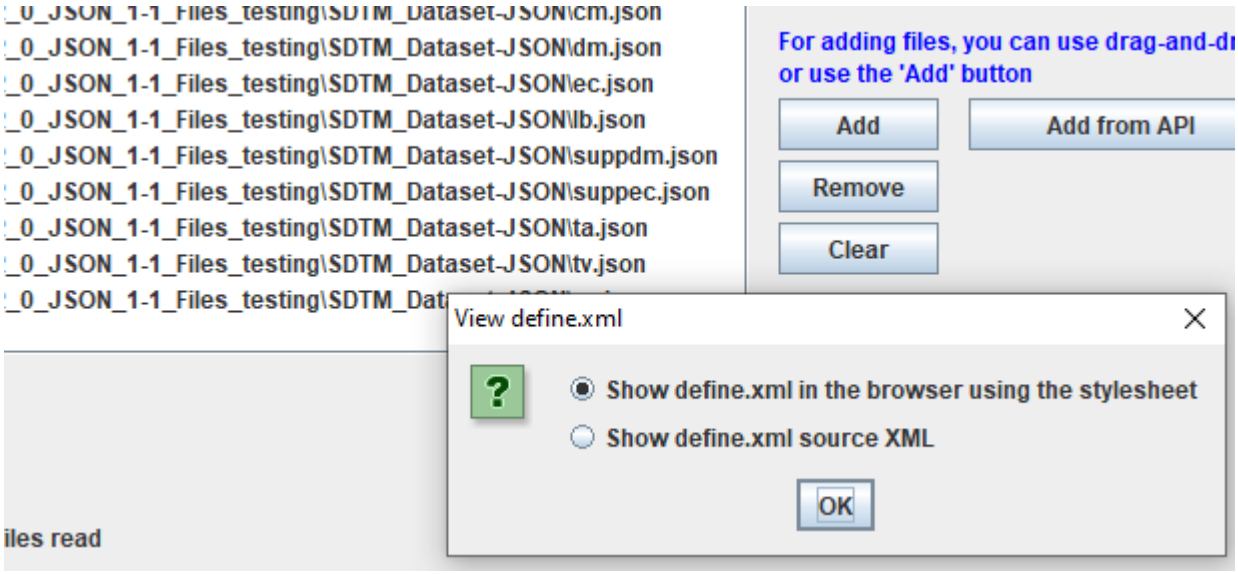
In the current version of the software, one also sees a "Get from API" button. This is currently for demonstration purposes only, and is explained in the section "Using APIs".

In this manual, we will explain many of the features using the Dataset-JSON 1.1 files from the "SDTM Metadata Submission Guidelines v2.0" published by CDISC.
So we select its define.xml e.g.:

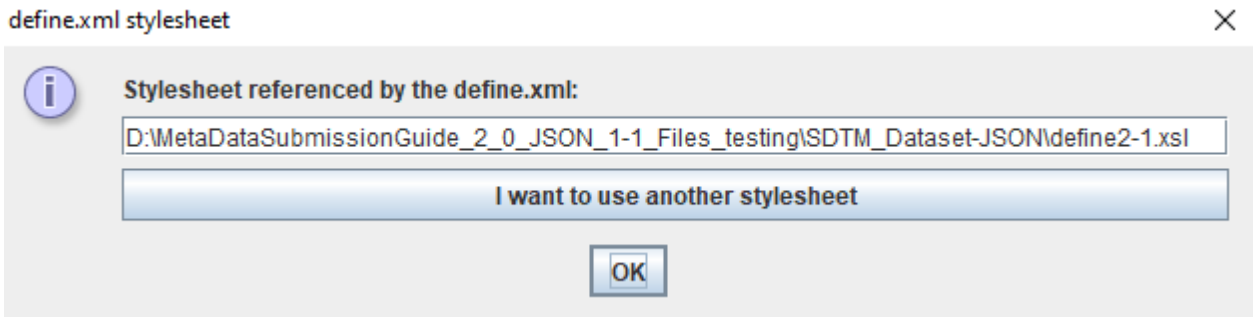


Normally, the software will recognize which version of Define-XML the define.xml file implements, and in this case automatically select "2.1" for "Define.xml" version, but it is always a good idea to check this.

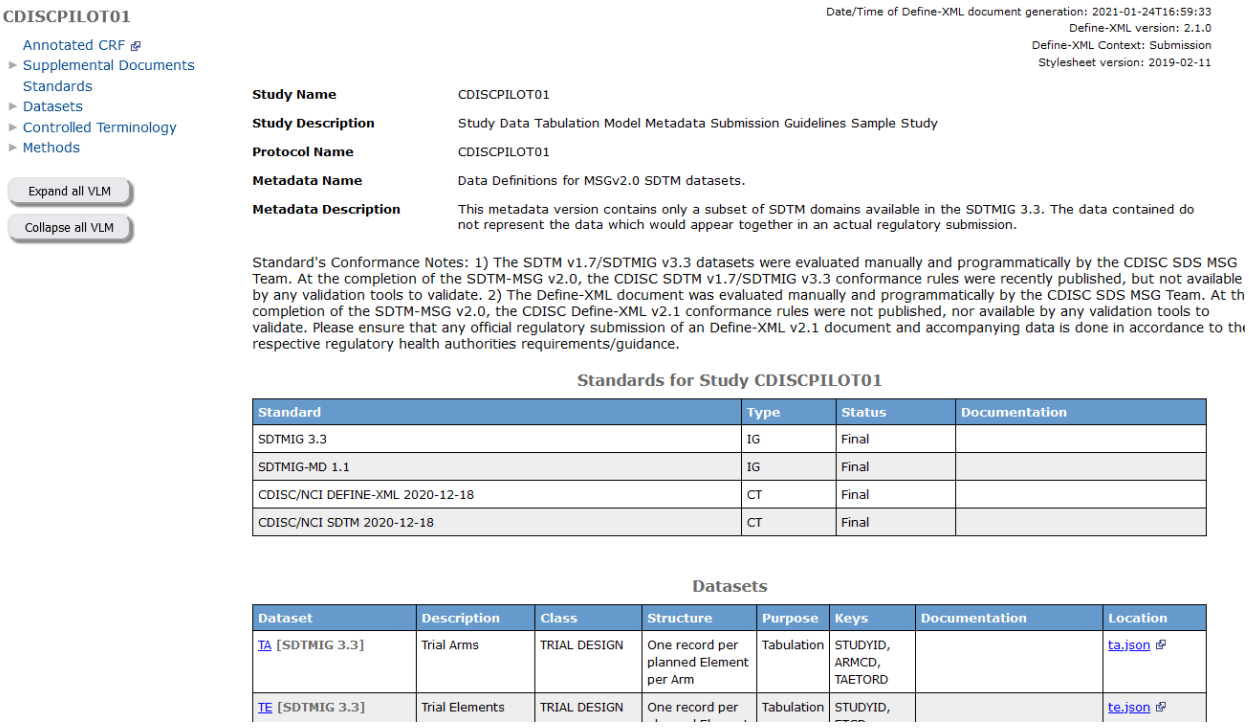
One can now also inspect the contents of the define.xml by clicking the "View" button. The following dialog is then displayed:



If one chooses "Show define.xml in the browser …", the system proposes to use a stylesheet it found in the same folder as the define.xml itself, but also allows the user to select another stylesheet:



Remark that the "Smart Submission Dataset Viewer" does not come with Define-XML stylesheets, as the stylesheet is the responsibility of the user's organization.
The stylesheet will then be applied and then display the define.xml as HTML in your default browser, e.g.



CDISCPILOT01

Annotated CRF
Supplemental Documents
Standards
Datasets
Controlled Terminology
Methods

Expand all VLM
Collapse all VLM

Date/Time of Define-XML document generation: 2021-01-24T16:59:33
Define-XML version: 2.1.0
Define-XML Context: Submission
Stylesheet version: 2019-02-11

Study Name	CDISCPILOT01
Study Description	Study Data Tabulation Model Metadata Submission Guidelines Sample Study
Protocol Name	CDISCPILOT01
Metadata Name	Data Definitions for MSGv2.0 SDTM datasets.
Metadata Description	This metadata version contains only a subset of SDTM domains available in the SDTMIG 3.3. The data contained do not represent the data which would appear together in an actual regulatory submission.

Standard's Conformance Notes: 1) The SDTM v1.7/SDTMIG v3.3 datasets were evaluated manually and programmatically by the CDISC SDS MSG Team. At the completion of the SDTM-MSG v2.0, the CDISC SDTM v1.7/SDTMIG v3.3 conformance rules were recently published, but not available by any validation tools to validate. 2) The Define-XML document was evaluated manually and programmatically by the CDISC SDS MSG Team. At th completion of the SDTM-MSG v2.0, the CDISC Define-XML v2.1 conformance rules were not published, nor available by any validation tools to validate. Please ensure that any official regulatory submission of an Define-XML v2.1 document and accompanying data is done in accordance to the respective regulatory health authorities requirements/guidance.

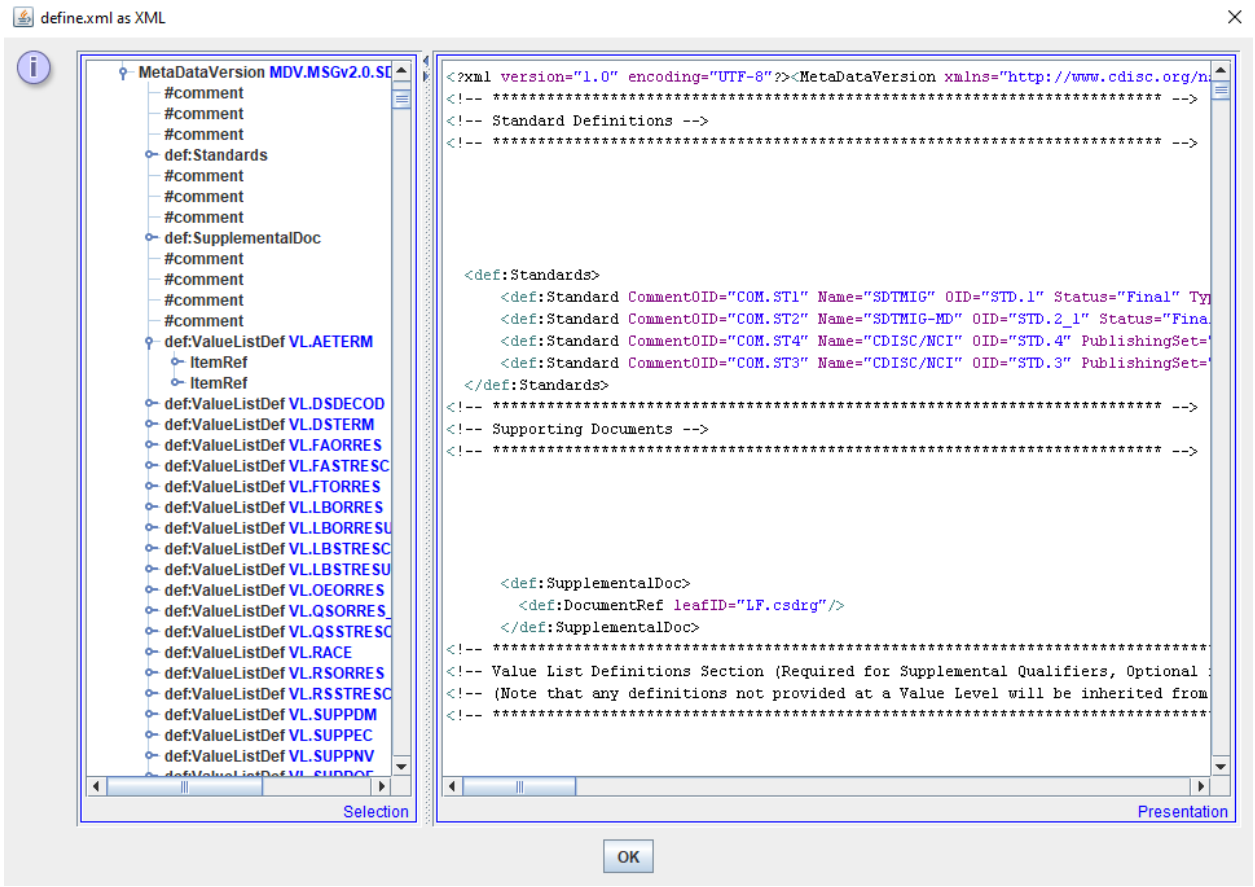
Standards for Study CDISCPILOT01

Standard	Type	Status	Documentation
SDTMIG 3.3	IG	Final	
SDTMIG-MD 1.1	IG	Final	
CDISC/NCI DEFINE-XML 2020-12-18	CT	Final	
CDISC/NCI SDTM 2020-12-18	CT	Final	

Datasets

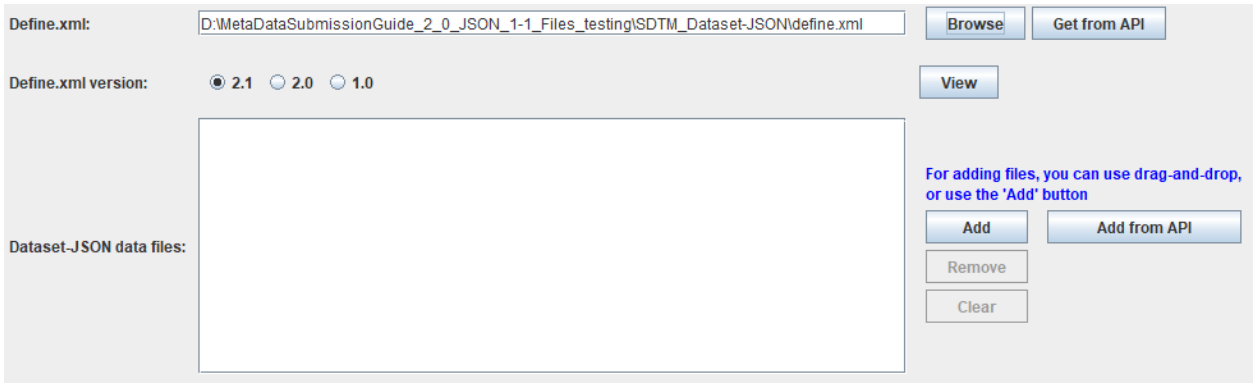
Dataset	Description	Class	Structure	Purpose	Keys	Documentation	Location
TA [SDTMIG 3.3]	Trial Arms	TRIAL DESIGN	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD		ta.json
TE [SDTMIG 3.3]	Trial Elements	TRIAL DESIGN	One record per	Tabulation	STUDYID,		te.json

The second choice "show define.xml source XML" allows to display the define.xml as XML itself in a separate window, e.g.:



where one can quickly navigate to specific sections by clicking on one of the tree nodes in the left part of the window.
This feature is of course only of interest to people who understand the XML structure of Define-XML.

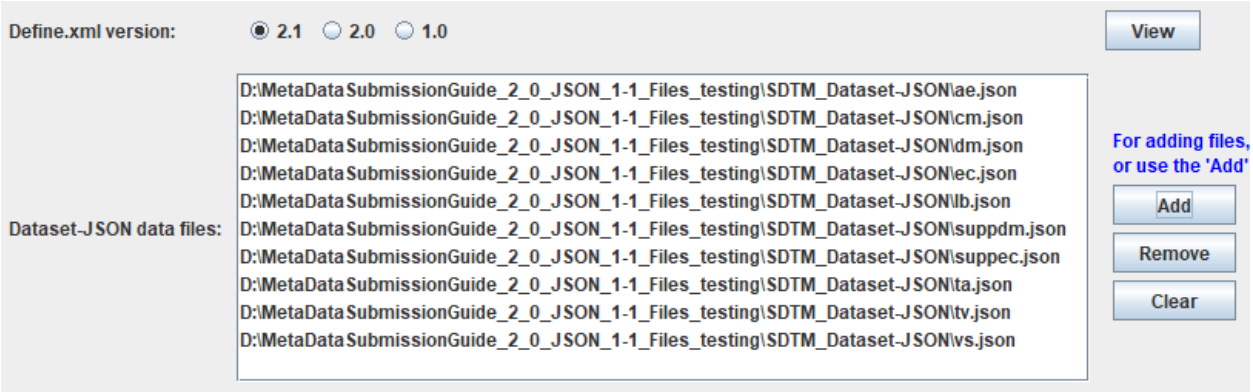
We will then want to load our Dataset-JSON files that we want to inspect.



One can either use drag-and-drop, or use the "Add" button, and then use the file chooser that is displayed.
Once one or more files have been selected, also the "Remove" and "Clear" buttons become available.
The "Remove" button is used to remove some files from the list after having selected them, and the "Clear" button to remove all files from the list.

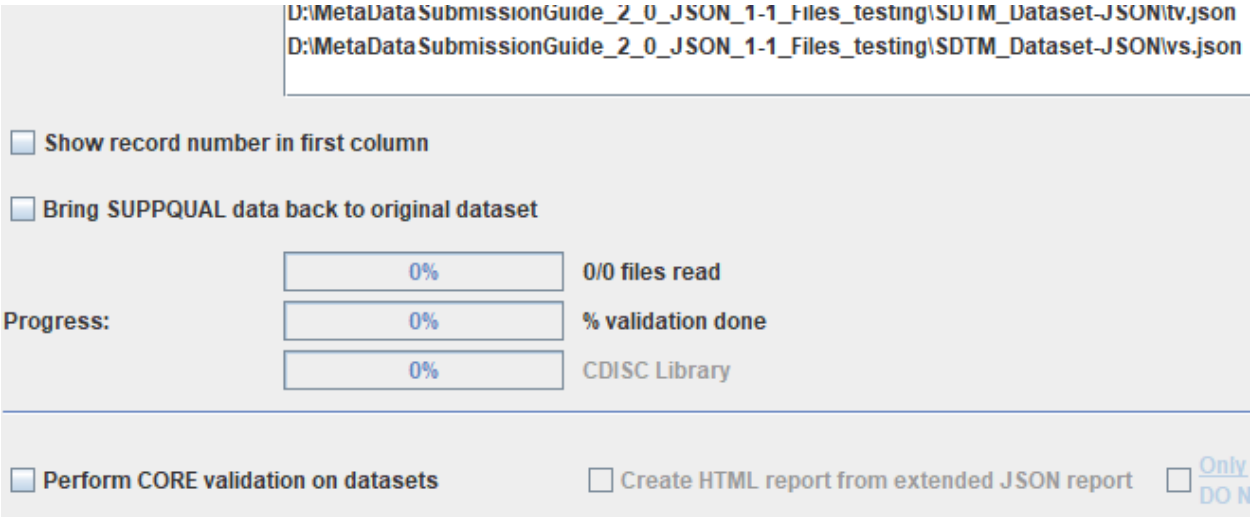
Also here, the "Add from API" button is for demonstration purposes only and will be explained in the section "Using APIs".

When loading submission files, it is always a good idea to load the DM (Demographics) dataset in the case of SDTM and SEND, or the ADSL (ADaM Subject-level Analysis) in the case of ADaM, as this will not only to validate some of the data against DM/ADSL, but also to "toggle" between a record in any dataset and the "basic" subject data in the DM/ADSL dataset.
Let us first add a few datasets already, e.g.:



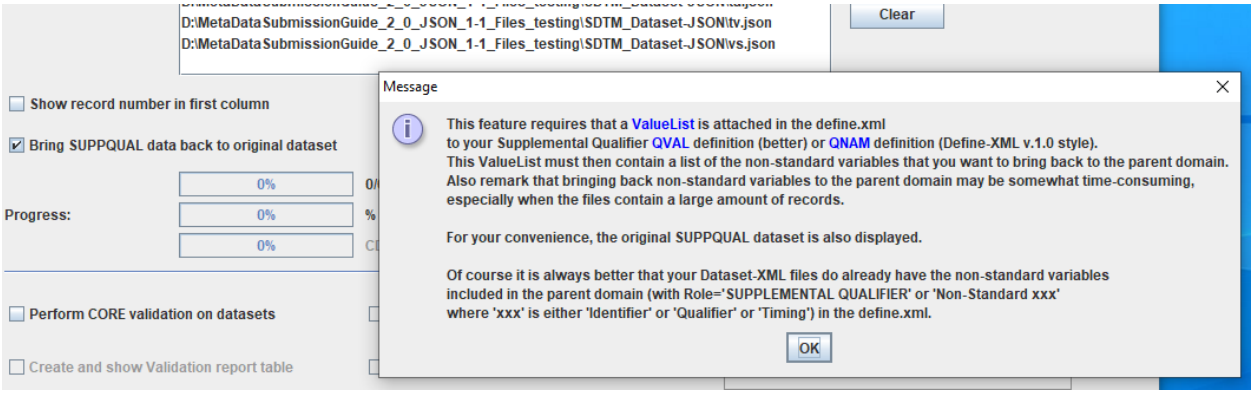
Don't mind about the order of the datasets in the list, the system will itself set up a more natural order.

Depending on which standard was selected, and whether one will use the define.xml for the metadata or not, a number of features become available. We explain them here for the case that one wants to use the define.xml for the metadata, as this should always be the preferred way.



First of all, when the checkbox "Show record number in first column" is checked, the system will add an extra column as the first column, and assign a record number to each record, which is fixed to that record. This also means that when later applying sorting, filtering, or any action that rearranges the order of the records, the record number will remain the same. Such a "pinned" record number can then enormously help when referring to a specific record later. Also see the section "sorting and filtering".

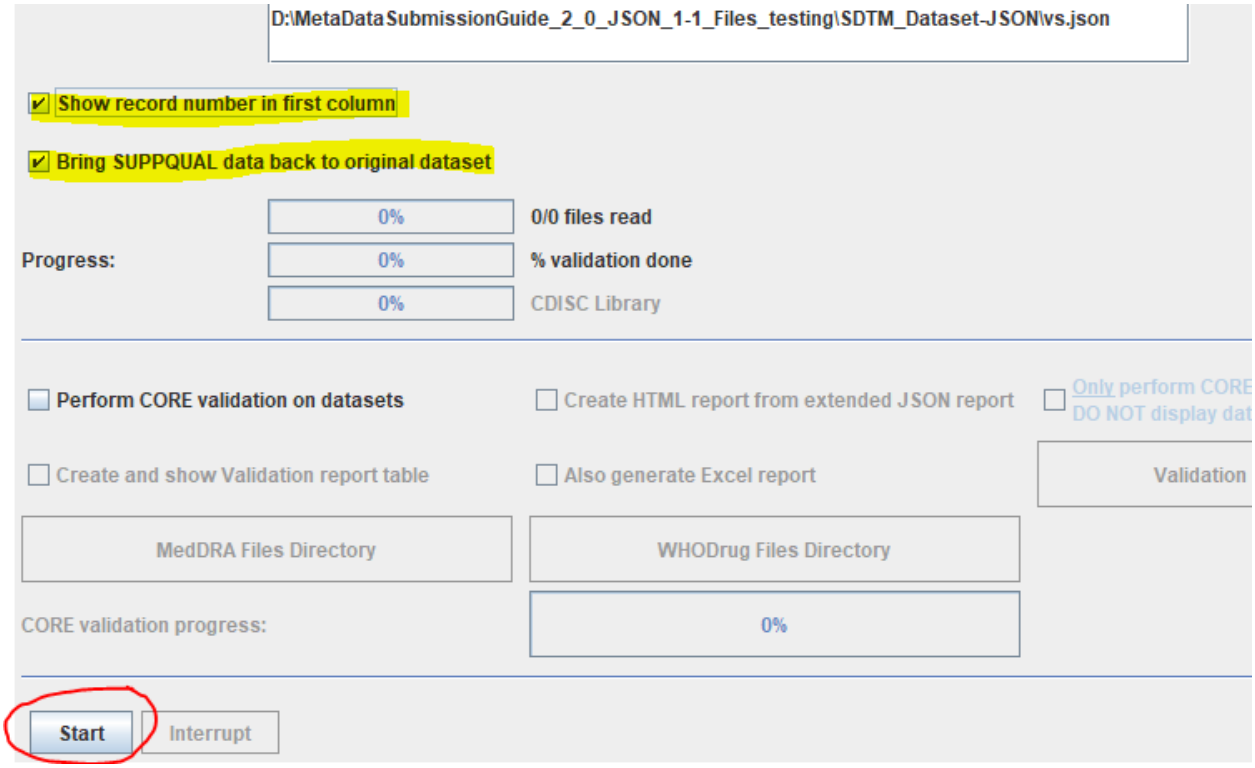
The second checkbox "Bring SUPPQUAL data back to original dataset" will only be enabled when the define.xml is to be used for the metadata. When one selects, some additional information is shown:



Essentially, it states that this feature will only work when the define.xml has "ValueLists" attached to QVAL for the variable definitions in the SUPPxx datasets.
As we have DM and SUPPDM, and EC and SUPPEC in our list, we check the checkbox. We will discuss the results a bit later.

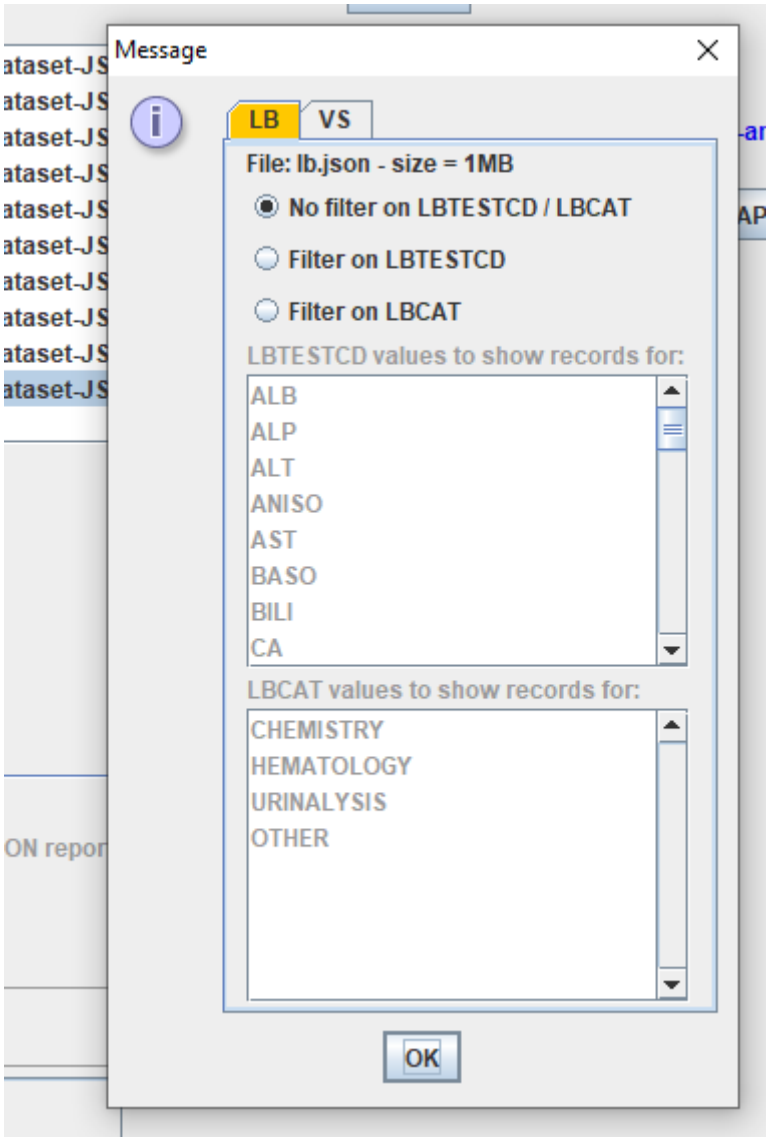
At this moment, we do not set any special options, using the "Options" button yet, so that only essential validation will be performed. We will later explain some of these in a separate section.

With the current settings:

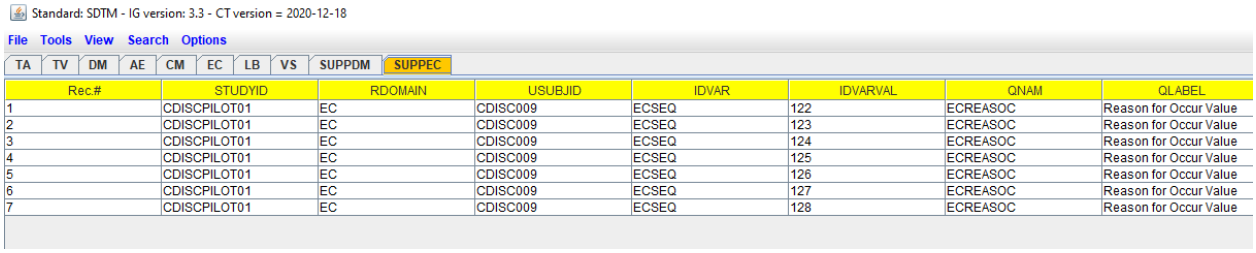


we now click the "Start" button to start the viewing process.

First, the system loads the define.xml , which can take some time, depending on the complexity of the contents of the define.xml (progress bar), in most cases followed by the following dialog:



This dialog is only displayed when some of the to-be-loaded files have a file size that exceeds a certain threshold (see later when treating all the options) and the define.xml has codelists attached to variables such as "--TESTCD", "--CAT" allowing to load subsets of data, such as only "hematology" data for LB. This feature is expected to facilitate the reviewing process considerably[1].
As this dialog is based on information about codelists in the define.xml, it is skipped when the user has decided to use the metadata from the Dataset-JSON files themselves.
We will first not do any "pre-loading filtering".
After clicking "OK", the loading process proceeds and after a few seconds (or more, depending on the number of records to be loaded), the window with all the tables is displayed:



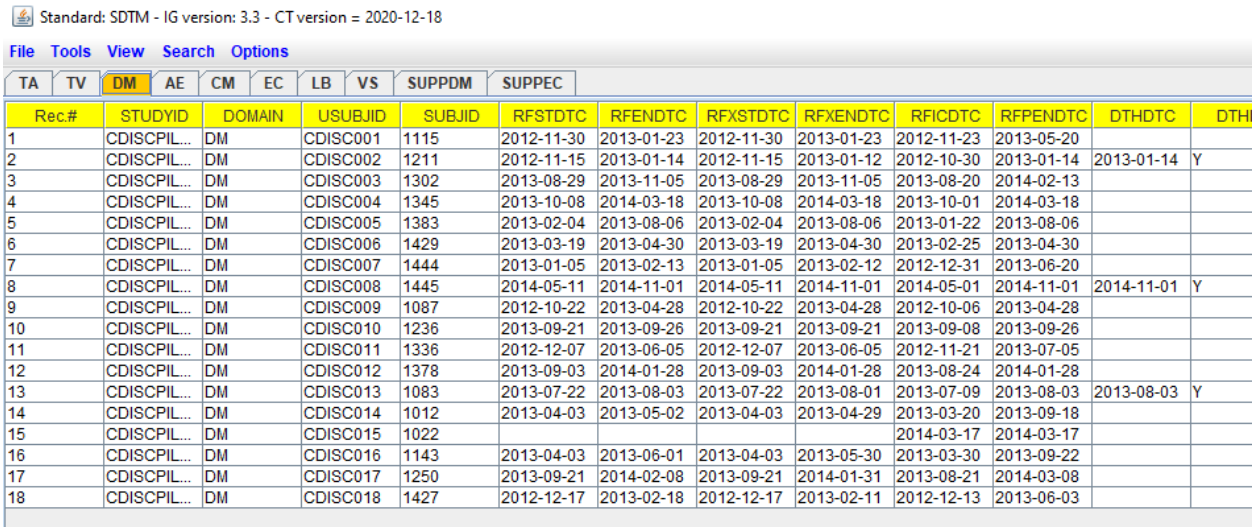
Rec.#	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL
1	CDISCPILOT01	EC	CDISC009	ECSEQ	122	ECREASOC	Reason for Occur Value
2	CDISCPILOT01	EC	CDISC009	ECSEQ	123	ECREASOC	Reason for Occur Value
3	CDISCPILOT01	EC	CDISC009	ECSEQ	124	ECREASOC	Reason for Occur Value
4	CDISCPILOT01	EC	CDISC009	ECSEQ	125	ECREASOC	Reason for Occur Value
5	CDISCPILOT01	EC	CDISC009	ECSEQ	126	ECREASOC	Reason for Occur Value
6	CDISCPILOT01	EC	CDISC009	ECSEQ	127	ECREASOC	Reason for Occur Value
7	CDISCPILOT01	EC	CDISC009	ECSEQ	128	ECREASOC	Reason for Occur Value

[1] When using the classic "SAS Universal Viewer", the user has no other choice than loading all the records, often leading to "pagination". When then doing additional filtering, the filter only applies to the current "page". As such, using the "SAS Universal Viewer" can easily lead to review problems.

with one tab for each dataset.
As one sees, the system has reordered the datasets into a more natural order, with the "Trial Design" datasets coming first, followed by DM (Demographics) and with the "Supplemental Qualifier" datasets coming at the end

Viewing the results - basic features

Let us first inspect the results for the DM dataset, by clicking on the DM tab:

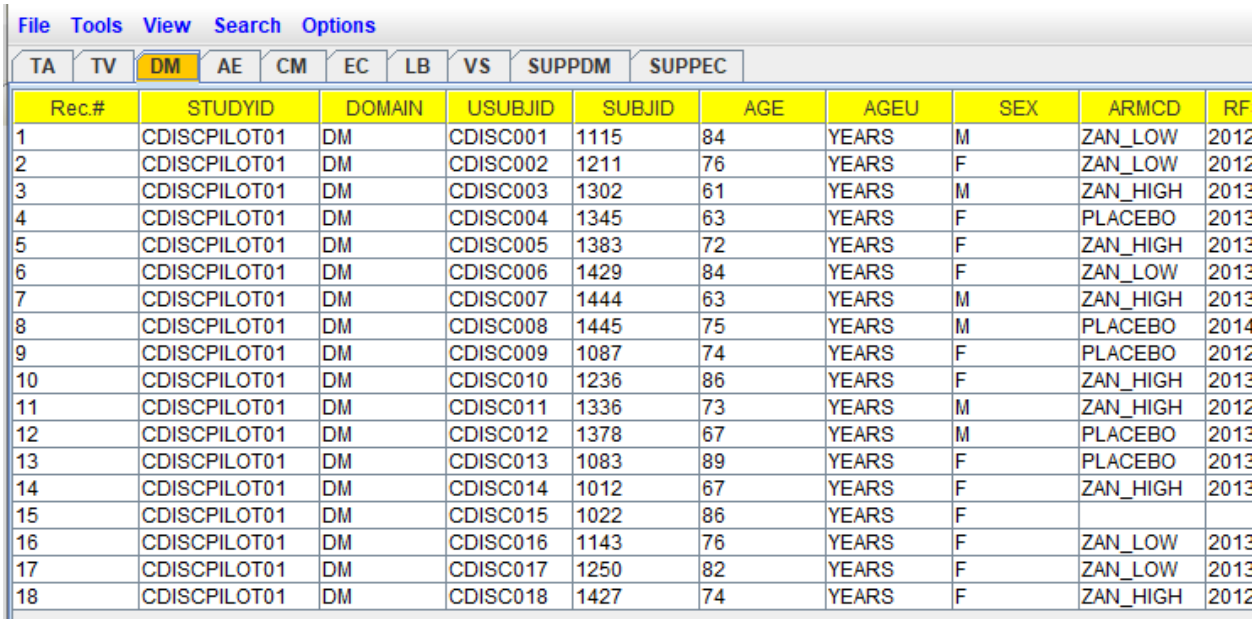

Standard: SDTM - IG version: 3.3 - CT version = 2020-12-18

File Tools View Search Options

TA TV DM AE CM EC LB VS SUPPDM SUPPEC

Rec.#	STUDYID	DOMAIN	USUBJID	SUBJID	RFSTDTC	RFENDTC	RFXSTDTC	RFXENDTC	RFICDTC	RFPENDTC	DTHDTC	DTHI
1	CDISCPIL...	DM	CDISC001	1115	2012-11-30	2013-01-23	2012-11-30	2013-01-23	2012-11-23	2013-05-20		
2	CDISCPIL...	DM	CDISC002	1211	2012-11-15	2013-01-14	2012-11-15	2013-01-12	2012-10-30	2013-01-14	2013-01-14	Y
3	CDISCPIL...	DM	CDISC003	1302	2013-08-29	2013-11-05	2013-08-29	2013-11-05	2013-08-20	2014-02-13		
4	CDISCPIL...	DM	CDISC004	1345	2013-10-08	2014-03-18	2013-10-08	2014-03-18	2013-10-01	2014-03-18		
5	CDISCPIL...	DM	CDISC005	1383	2013-02-04	2013-08-06	2013-02-04	2013-08-06	2013-01-22	2013-08-06		
6	CDISCPIL...	DM	CDISC006	1429	2013-03-19	2013-04-30	2013-03-19	2013-04-30	2013-02-25	2013-04-30		
7	CDISCPIL...	DM	CDISC007	1444	2013-01-05	2013-02-13	2013-01-05	2013-02-12	2012-12-31	2013-06-20		
8	CDISCPIL...	DM	CDISC008	1445	2014-05-11	2014-11-01	2014-05-11	2014-11-01	2014-05-01	2014-11-01	2014-11-01	Y
9	CDISCPIL...	DM	CDISC009	1087	2012-10-22	2013-04-28	2012-10-22	2013-04-28	2012-10-06	2013-04-28		
10	CDISCPIL...	DM	CDISC010	1236	2013-09-21	2013-09-26	2013-09-21	2013-09-21	2013-09-08	2013-09-26		
11	CDISCPIL...	DM	CDISC011	1336	2012-12-07	2013-06-05	2012-12-07	2013-06-05	2012-11-21	2013-07-05		
12	CDISCPIL...	DM	CDISC012	1378	2013-09-03	2014-01-28	2013-09-03	2014-01-28	2013-08-24	2014-01-28		
13	CDISCPIL...	DM	CDISC013	1083	2013-07-22	2013-08-03	2013-07-22	2013-08-01	2013-07-09	2013-08-03	2013-08-03	Y
14	CDISCPIL...	DM	CDISC014	1012	2013-04-03	2013-05-02	2013-04-03	2013-04-29	2013-03-20	2013-09-18		
15	CDISCPIL...	DM	CDISC015	1022					2014-03-17	2014-03-17		
16	CDISCPIL...	DM	CDISC016	1143	2013-04-03	2013-06-01	2013-04-03	2013-05-30	2013-03-30	2013-09-22		
17	CDISCPIL...	DM	CDISC017	1250	2013-09-21	2014-02-08	2013-09-21	2014-01-31	2013-08-21	2014-03-08		
18	CDISCPIL...	DM	CDISC018	1427	2012-12-17	2013-02-18	2012-12-17	2013-02-11	2012-12-13	2013-06-03		

As one sees, and extra column with the (generated) record number has been added
One can now resize columns with the mouse, or even move columns from one place to another (i.e. changing the order of the columns) which is always very useful, as one can then put the columns of interest together. For example, when one is mainly interested in age, sex, and assigned arm of the subjects, one can drag the columns in the following order:


File Tools View Search Options

TA TV DM AE CM EC LB VS SUPPDM SUPPEC

Rec.#	STUDYID	DOMAIN	USUBJID	SUBJID	AGE	AGEU	SEX	ARMCD	RF
1	CDISCPILOT01	DM	CDISC001	1115	84	YEARS	M	ZAN_LOW	2012
2	CDISCPILOT01	DM	CDISC002	1211	76	YEARS	F	ZAN_LOW	2012
3	CDISCPILOT01	DM	CDISC003	1302	61	YEARS	M	ZAN_HIGH	2013
4	CDISCPILOT01	DM	CDISC004	1345	63	YEARS	F	PLACEBO	2013
5	CDISCPILOT01	DM	CDISC005	1383	72	YEARS	F	ZAN_HIGH	2013
6	CDISCPILOT01	DM	CDISC006	1429	84	YEARS	F	ZAN_LOW	2013
7	CDISCPILOT01	DM	CDISC007	1444	63	YEARS	M	ZAN_HIGH	2013
8	CDISCPILOT01	DM	CDISC008	1445	75	YEARS	M	PLACEBO	2014
9	CDISCPILOT01	DM	CDISC009	1087	74	YEARS	F	PLACEBO	2012
10	CDISCPILOT01	DM	CDISC010	1236	86	YEARS	F	ZAN_HIGH	2013
11	CDISCPILOT01	DM	CDISC011	1336	73	YEARS	M	ZAN_HIGH	2012
12	CDISCPILOT01	DM	CDISC012	1378	67	YEARS	M	PLACEBO	2013
13	CDISCPILOT01	DM	CDISC013	1083	89	YEARS	F	PLACEBO	2013
14	CDISCPILOT01	DM	CDISC014	1012	67	YEARS	F	ZAN_HIGH	2013
15	CDISCPILOT01	DM	CDISC015	1022	86	YEARS	F		
16	CDISCPILOT01	DM	CDISC016	1143	76	YEARS	F	ZAN_LOW	2013
17	CDISCPILOT01	DM	CDISC017	1250	82	YEARS	F	ZAN_LOW	2013
18	CDISCPILOT01	DM	CDISC018	1427	74	YEARS	F	ZAN_HIGH	2012

Viewing values for Supplemental Qualifiers

As we checked the checkbox "Bring SUPPQUAL data back to original dataset", we should now also see the values of the "non-standard variables", which were "banned" to SUPPDM, in the DM dataset itself. Remark again that this will only work when a "ValueList" was assigned to the QVAL variable in the define.xml.

When scrolling to the right, we find:

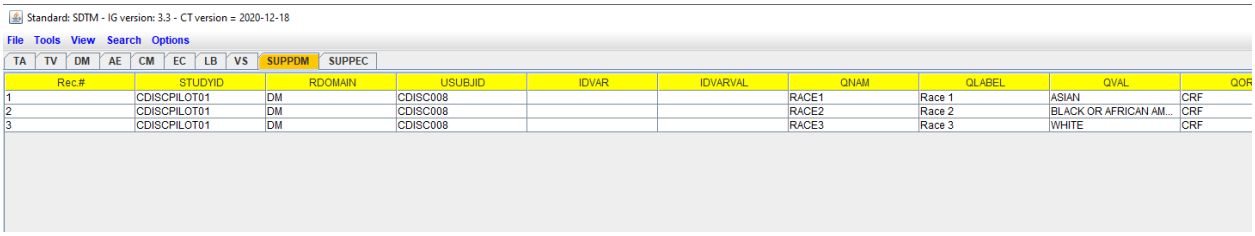
BRTHDTC	RACE	ETHNIC	ARM	ACTARMCD	ACTARM	ARMNRS	ACTARMUD	COUNTRY	RACE1	RACE2	RACE3	RA
1928	WHITE	NOT HISPA...	Zanomalin...	ZAN_LOW	Zanomalin...			USA				
1936	WHITE	NOT HISPA...	Zanomalin...	ZAN_LOW	Zanomalin...			USA				
1951	WHITE	NOT HISPA...	Zanomalin...	ZAN_HIGH	Zanomalin...			USA				
1950	WHITE	NOT HISPA...	Placebo	PLACEBO	Placebo			USA				
1941	WHITE	NOT HISPA...	Zanomalin...	ZAN_HIGH	Zanomalin...			USA				
1929	WHITE	NOT HISPA...	Zanomalin...	ZAN_LOW	Zanomalin...			USA				
1949	WHITE	HISPANIC ...	Zanomalin...	ZAN_HIGH	Zanomalin...			USA				
1939	MULTIPLE	NOT HISPA...	Placebo	PLACEBO	Placebo			USA	ASIAN	BLACK OR AFRICAN AMERICAN	WHITE	
1938	WHITE	NOT HISPA...	Placebo	PLACEBO	Placebo			USA				
1927	WHITE	NOT HISPA...	Zanomalin...	ZAN_HIGH	Zanomalin...			USA				
1939	WHITE	NOT HISPA...	Zanomalin...	ZAN_HIGH	Zanomalin...			USA				
1946	BLACK OR ...	NOT HISPA...	Placebo	PLACEBO	Placebo			USA				
1924	WHITE	NOT HISPA...	Placebo	PLACEBO	Placebo			USA				
1945	WHITE	NOT HISPA...	Zanomalin...	ZAN_HIGH	Zanomalin...			USA				
1928	WHITE	NOT HISPA...				SCREEN F...		USA				
1936	WHITE	NOT HISPA...	Zanomalin...	ZAN_LOW	Zanomalin...			USA				
1931	WHITE	HISPANIC ...	Zanomalin...	ZAN_LOW	Zanomalin...			USA				
1938	BLACK OR ...	NOT HISPA...	Zanomalin...	ZAN_HIGH	Zanomalin...			USA				

that there is one record having "MULTIPLE" for "RACE", and the individual races being depicted in "RACE1", "RACE2", and "RACE3". Also here, we can drag columns so that we get a better oversight of what we are really interested in:

SUBJID	AGE	AGEU	SEX	ARMCD	RACE	RACE1	RACE2	RACE3
1115	84	YEARS	M	ZAN_LOW	WHITE			
1211	76	YEARS	F	ZAN_LOW	WHITE			
1302	61	YEARS	M	ZAN_HIGH	WHITE			
1345	63	YEARS	F	PLACEBO	WHITE			
1383	72	YEARS	F	ZAN_HIGH	WHITE			
1429	84	YEARS	F	ZAN_LOW	WHITE			
1444	63	YEARS	M	ZAN_HIGH	WHITE			
1445	75	YEARS	M	PLACEBO	MULTIPLE	ASIAN	BLACK OR AFRICAN AMERICAN	WHITE
1087	74	YEARS	F	PLACEBO	WHITE			
1236	86	YEARS	F	ZAN_HIGH	WHITE			
1336	73	YEARS	M	ZAN_HIGH	WHITE			
1378	67	YEARS	M	PLACEBO	BLACK OR ...			
1083	89	YEARS	F	PLACEBO	WHITE			
1012	67	YEARS	F	ZAN_HIGH	WHITE			
1022	86	YEARS	F		WHITE			
1143	76	YEARS	F	ZAN_LOW	WHITE			
1250	82	YEARS	F	ZAN_LOW	WHITE			

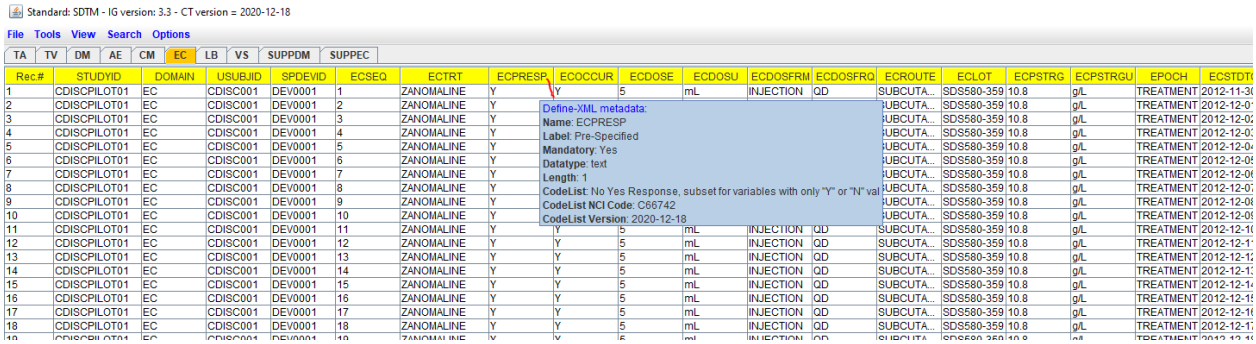
so that we now see that this subject is male, 75 years old, and received placebo.

Of course, we can still inspect the SUPPDM dataset itself by clicking the SUPPDM tab:



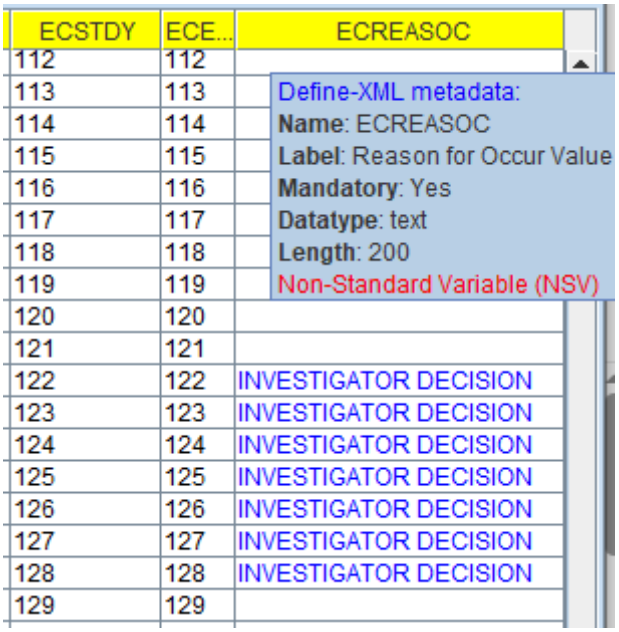
Showing the reviewer the same information, but in a much less user-friendly way.

Let us know at EC (Exposure as Collected) and SUPPEC. For EC we find:



One sees that when hovering the mouse over a column header, one gets the metadata information as a tooltip.

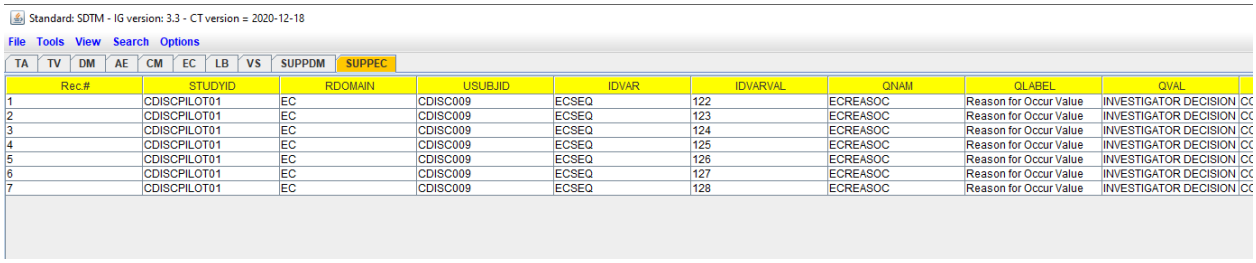
As we have a SUPPEC dataset, and checked the checkbox "Bring SUPPQUAL data back to original dataset", we also expect one or more additional columns for the NSVs (Non-Standard Variables). When we scroll to the right we indeed find:



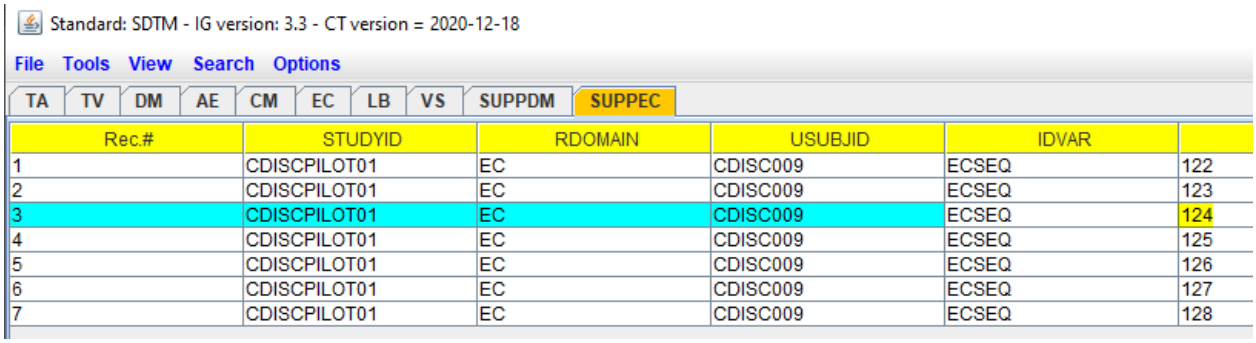
As this is related to non-treated subjects, we can better move that column, e.g.:

CTRT	ECPRESP	ECOCCUR	ECREASOC	ECDOSE	ECDOSU	ECDOS
BO	Y	Y		5	mL	INJECT
BO	Y	Y		5	mL	INJECT
BO	Y	Y		5	mL	INJECT
BO	Y	Y		5	mL	INJECT
BO	Y	Y		5	mL	INJECT
BO	Y	Y		5	mL	INJECT
BO	Y	Y		5	mL	INJECT
BO	Y	Y		5	mL	INJECT
BO	Y	Y		5	mL	INJECT
BO	Y	Y		5	mL	INJECT
BO	Y	N	INVESTIGATOR DECISION			INJECT
BO	Y	N	INVESTIGATOR DECISION			INJECT
BO	Y	N	INVESTIGATOR DECISION			INJECT
BO	Y	N	INVESTIGATOR DECISION			INJECT
BO	Y	N	INVESTIGATOR DECISION			INJECT
BO	Y	N	INVESTIGATOR DECISION			INJECT
BO	Y	N	INVESTIGATOR DECISION			INJECT
BO	Y	Y		5	mL	INJECT
BO	Y	Y		5	mL	INJECT
BO	Y	Y		5	mL	INJECT

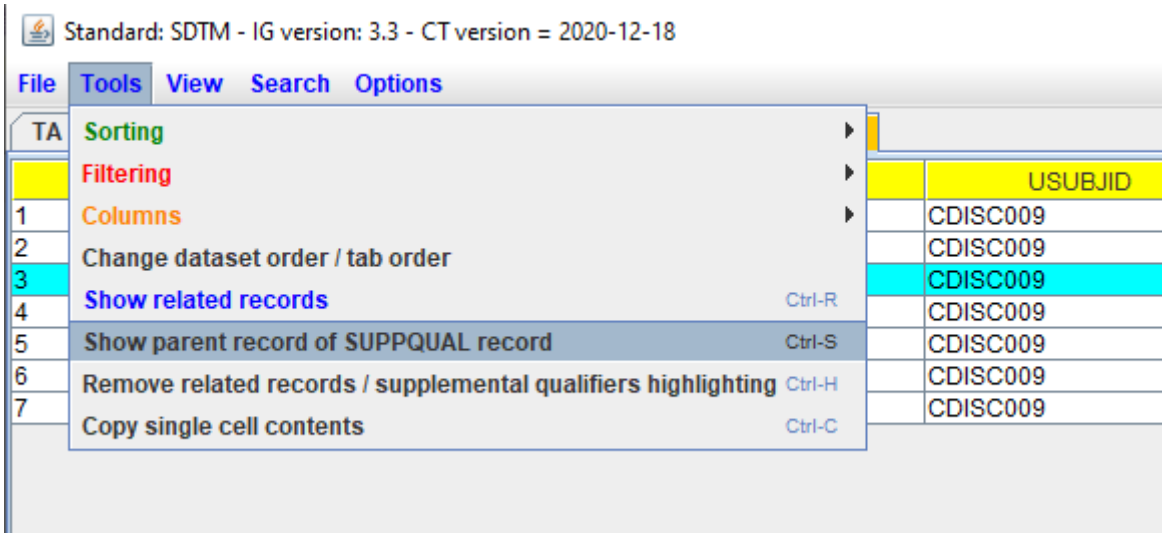
And the SUPPEC dataset:



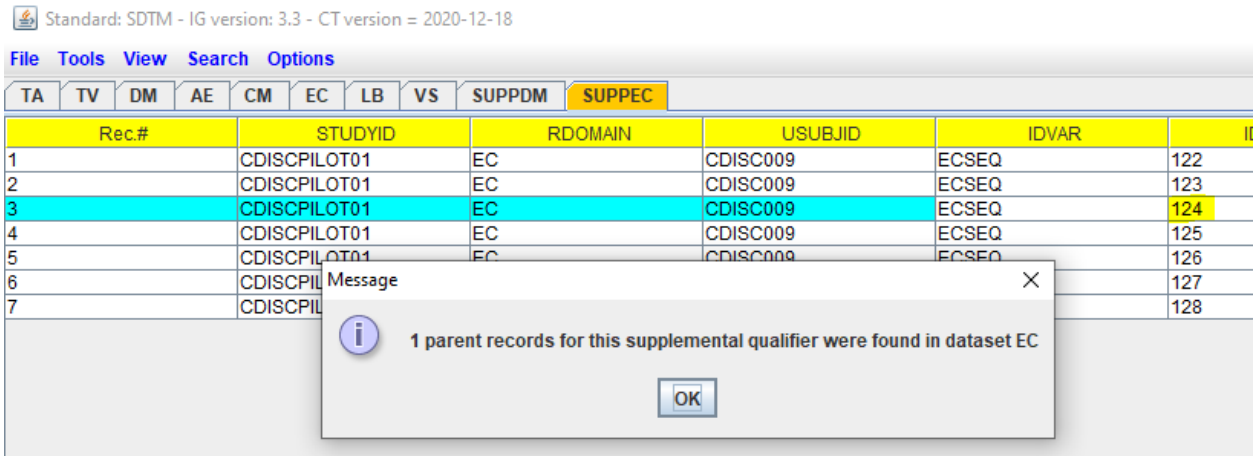
Even when one has <u>not</u> checked the checkbox "Bring SUPPQUAL data back to original dataset", one can easily find the parent record of a SUPPQUAL record. This can be done by selecting a row in the SUPPQUAL table (here SUPPEC):



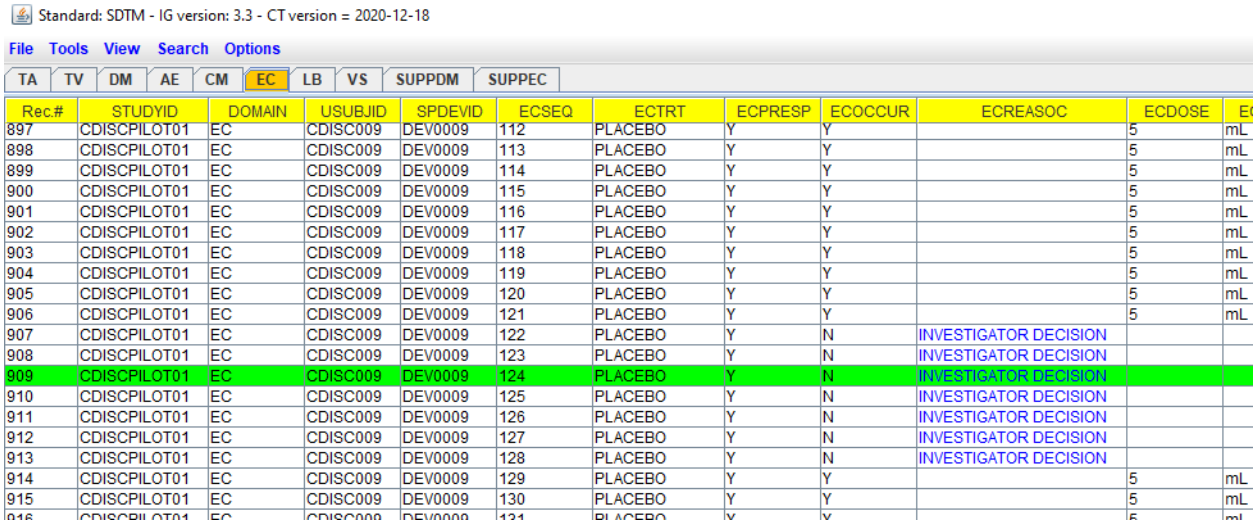
and then using the menu "Tools - Show parent record of SUPPQUAL record" (or just using Ctrl-S on the keyboard)



leading to:



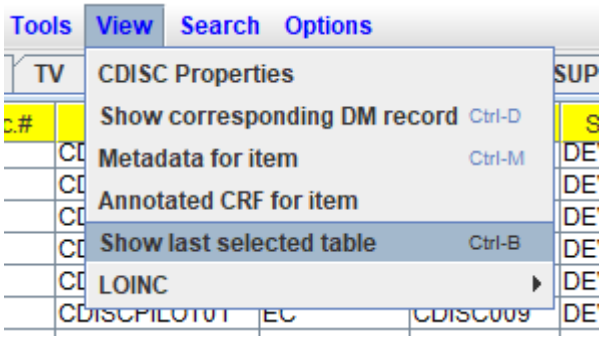
and after clicking "OK", the row for subject CDISC009 and ECSEQ=124 is selected (also the tab for EC is automatically selected) and highlighted:



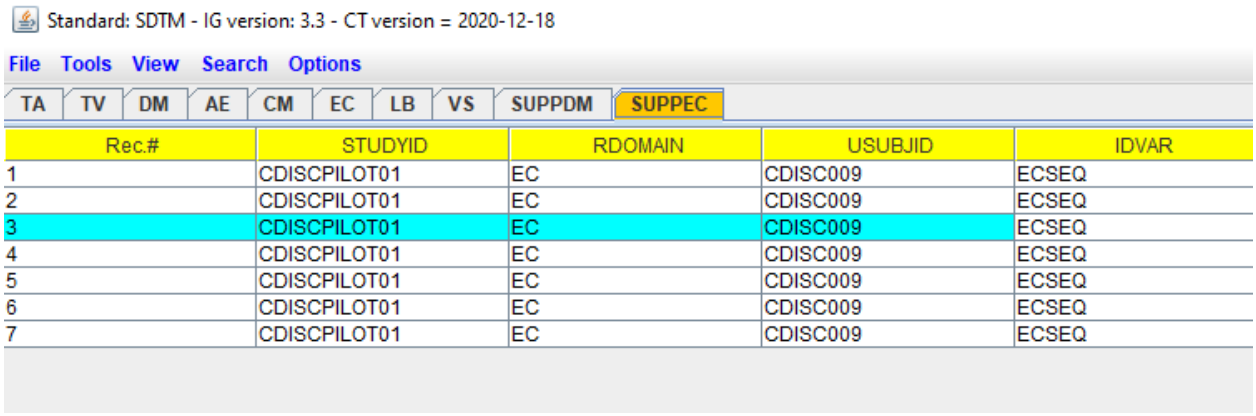
Remark that when the "identifying variable" (IDVAR) is e.g. a --CAT variable, this will usually lead to more than 1 parent record, and all these will then be highlighted.

One can now "toggle" between the record in EC and the record in SUPPEC (also when the checkbox "Bring

SUPPQUAL data back to original dataset" was not checked) by using the menu "View - Show last selected table", or easier and much faster by just using Ctrl-B on the keyboard:



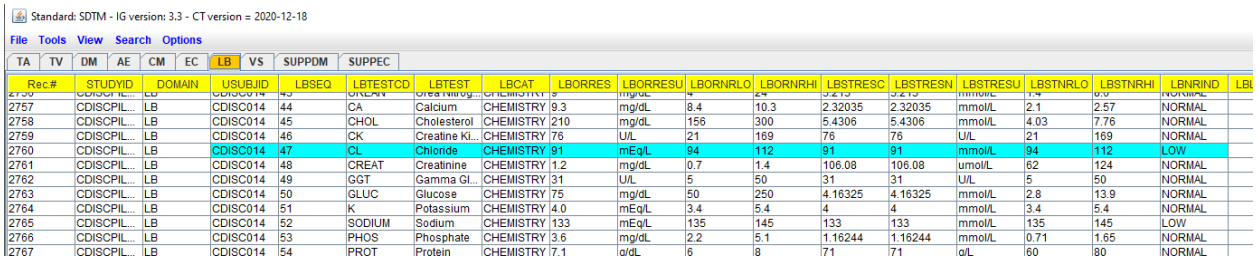
showing SUPPEC again with the selected record:



Rec.#	STUDYID	RDOMAIN	USUBJID	IDVAR
1	CDISCPILOT01	EC	CDISC009	ECSEQ
2	CDISCPILOT01	EC	CDISC009	ECSEQ
3	CDISCPILOT01	EC	CDISC009	ECSEQ
4	CDISCPILOT01	EC	CDISC009	ECSEQ
5	CDISCPILOT01	EC	CDISC009	ECSEQ
6	CDISCPILOT01	EC	CDISC009	ECSEQ
7	CDISCPILOT01	EC	CDISC009	ECSEQ

Looking up basic subject information in DM or ADSL

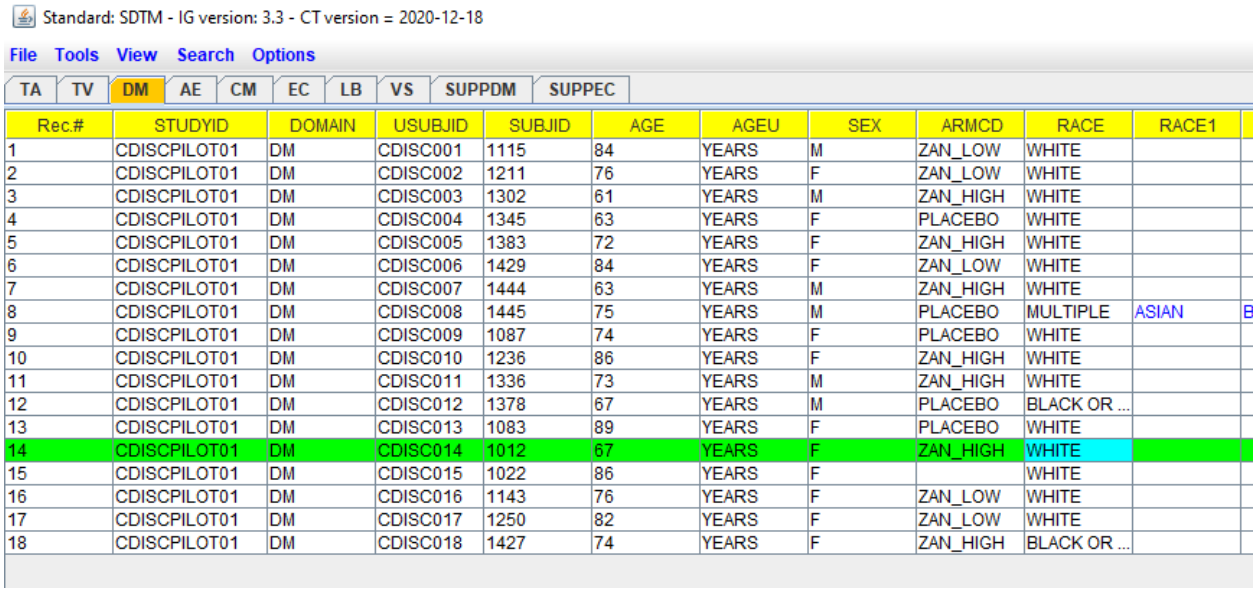
When inspecting a record in any of the datasets, one can always quickly "jump" to the corresponding record in DM (for SDTM and SEND) or ADSL (for ADaM). For example, when having selected a record in LB:



Rec.#	STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	LBORRES	LBORRESU	LBORNRLO	LBORNRHI	LBSTRESC	LBSTRESN	LBSTRESU	LBSTNRLO	LBSTNRHI	LBNRIND
2757	CDISCPIL...	LB	CDISC014	44	CA	Calcium	CHEMISTRY	9.3	mg/dL	8.4	10.3	2.32035	2.32035	mmol/L	2.1	2.57	NORMAL
2758	CDISCPIL...	LB	CDISC014	45	CHOL	Cholesterol	CHEMISTRY	210	mg/dL	156	300	5.4306	5.4306	mmol/L	4.03	7.76	NORMAL
2759	CDISCPIL...	LB	CDISC014	46	CK	Creatine Ki...	CHEMISTRY	76	U/L	21	169	76	76	U/L	21	169	NORMAL
2760	CDISCPIL...	LB	CDISC014	47	CL	Chloride	CHEMISTRY	91	mEq/L	94	112	91	91	mmol/L	94	112	LOW
2761	CDISCPIL...	LB	CDISC014	48	CREAT	Creatinine	CHEMISTRY	1.2	mg/dL	0.7	1.4	106.08	106.08	umol/L	62	124	NORMAL
2762	CDISCPIL...	LB	CDISC014	49	GGT	Gamma Gl...	CHEMISTRY	31	U/L	5	50	31	31	U/L	5	50	NORMAL
2763	CDISCPIL...	LB	CDISC014	50	GLUC	Glucose	CHEMISTRY	75	mg/dL	50	250	4.16325	4.16325	mmol/L	2.8	13.9	NORMAL
2764	CDISCPIL...	LB	CDISC014	51	K	Potassium	CHEMISTRY	4.0	mEq/L	3.4	5.4	4	4	mmol/L	3.4	5.4	NORMAL
2765	CDISCPIL...	LB	CDISC014	52	SODIUM	Sodium	CHEMISTRY	133	mEq/L	135	145	133	133	mmol/L	135	145	LOW
2766	CDISCPIL...	LB	CDISC014	53	PHOS	Phosphate	CHEMISTRY	3.6	mg/dL	2.2	5.1	1.16244	1.16244	mmol/L	0.71	1.65	NORMAL
2767	CDISCPIL...	LB	CDISC014	54	PROT	Protein	CHEMISTRY	7.1	g/dL	6	8	71	71	g/L	60	80	NORMAL

for subject CDISCPILOT14 with a low Chloride value, we can "jump" to the DM record for that subject by using the menu "View - Show corresponding DM record"[2], or much faster, just using Ctrl-D on the keyboard, leading to:

[2] In the case of ADaM, the menu item will show "Show corresponding ADSL record"



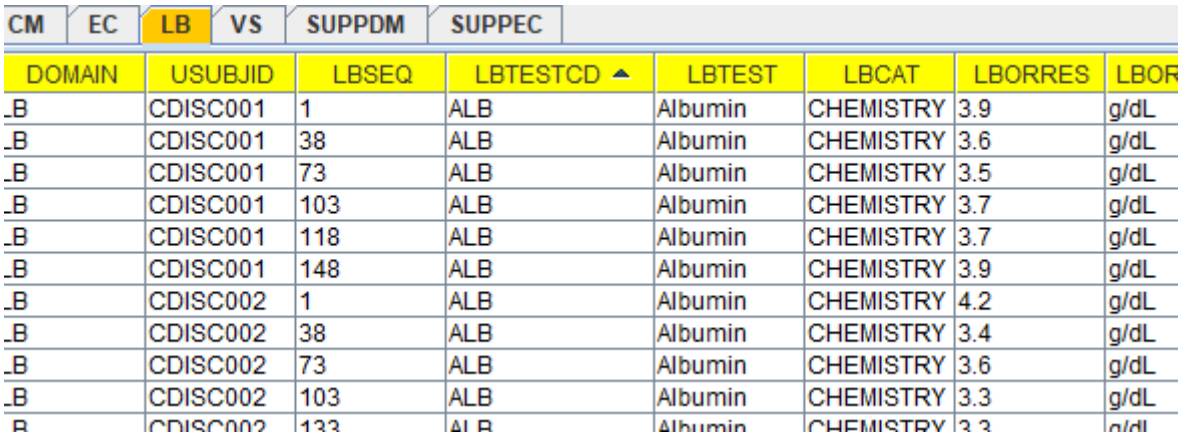
Rec.#	STUDYID	DOMAIN	USUBJID	SUBJID	AGE	AGEU	SEX	ARMCD	RACE	RACE1	
1	CDISCPILOT01	DM	CDISC001	1115	84	YEARS	M	ZAN_LOW	WHITE		
2	CDISCPILOT01	DM	CDISC002	1211	76	YEARS	F	ZAN_LOW	WHITE		
3	CDISCPILOT01	DM	CDISC003	1302	61	YEARS	M	ZAN_HIGH	WHITE		
4	CDISCPILOT01	DM	CDISC004	1345	63	YEARS	F	PLACEBO	WHITE		
5	CDISCPILOT01	DM	CDISC005	1383	72	YEARS	F	ZAN_HIGH	WHITE		
6	CDISCPILOT01	DM	CDISC006	1429	84	YEARS	F	ZAN_LOW	WHITE		
7	CDISCPILOT01	DM	CDISC007	1444	63	YEARS	M	ZAN_HIGH	WHITE		
8	CDISCPILOT01	DM	CDISC008	1445	75	YEARS	M	PLACEBO	MULTIPLE	ASIAN	B
9	CDISCPILOT01	DM	CDISC009	1087	74	YEARS	F	PLACEBO	WHITE		
10	CDISCPILOT01	DM	CDISC010	1236	86	YEARS	F	ZAN_HIGH	WHITE		
11	CDISCPILOT01	DM	CDISC011	1336	73	YEARS	M	ZAN_HIGH	WHITE		
12	CDISCPILOT01	DM	CDISC012	1378	67	YEARS	M	PLACEBO	BLACK OR ...		
13	CDISCPILOT01	DM	CDISC013	1083	89	YEARS	F	PLACEBO	WHITE		
14	CDISCPILOT01	DM	CDISC014	1012	67	YEARS	F	ZAN_HIGH	WHITE		
15	CDISCPILOT01	DM	CDISC015	1022	86	YEARS	F		WHITE		
16	CDISCPILOT01	DM	CDISC016	1143	76	YEARS	F	ZAN_LOW	WHITE		
17	CDISCPILOT01	DM	CDISC017	1250	82	YEARS	F	ZAN_LOW	WHITE		
18	CDISCPILOT01	DM	CDISC018	1427	74	YEARS	F	ZAN_HIGH	BLACK OR ...		

showing us that the subject was in the "high xanomeline" arm.

Also here, one can then use Ctrl-B to "toggle" between the two datasets and records.

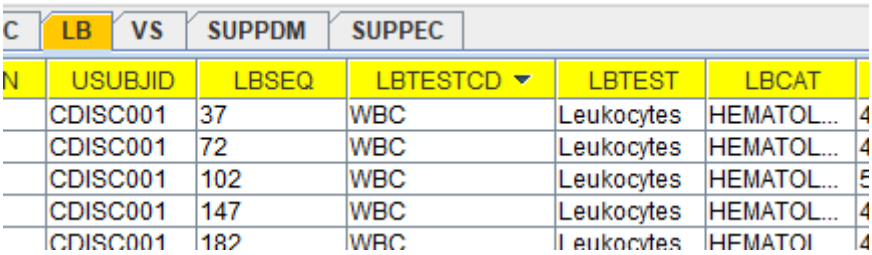
Sorting and filtering

The result tables can easily be sorted and filtered,
One can sort within a column by clicking on the column header. For example, in LB by clicking on the LBTESTCD column header, the rows will be ordered in alphabetic order for the value of LBTESTCD:



DOMAIN	USUBJID	LBSEQ	LBTESTCD ▲	LBTEST	LBCAT	LBORRES	LBOR
_B	CDISC001	1	ALB	Albumin	CHEMISTRY	3.9	g/dL
_B	CDISC001	38	ALB	Albumin	CHEMISTRY	3.6	g/dL
_B	CDISC001	73	ALB	Albumin	CHEMISTRY	3.5	g/dL
_B	CDISC001	103	ALB	Albumin	CHEMISTRY	3.7	g/dL
_B	CDISC001	118	ALB	Albumin	CHEMISTRY	3.7	g/dL
_B	CDISC001	148	ALB	Albumin	CHEMISTRY	3.9	g/dL
_B	CDISC002	1	ALB	Albumin	CHEMISTRY	4.2	g/dL
_B	CDISC002	38	ALB	Albumin	CHEMISTRY	3.4	g/dL
_B	CDISC002	73	ALB	Albumin	CHEMISTRY	3.6	g/dL
_B	CDISC002	103	ALB	Albumin	CHEMISTRY	3.3	g/dL
_B	CDISC002	133	ALB	Albumin	CHEMISTRY	3.3	g/dL

The next click on the column header then sorts the rows in the reverse order:

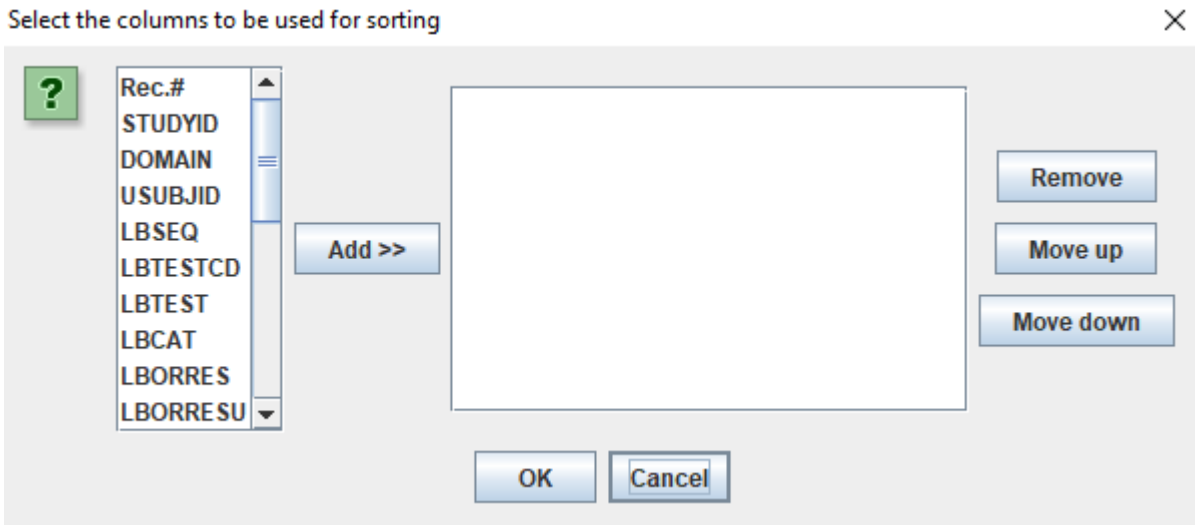


N	USUBJID	LBSEQ	LBTESTCD ▼	LBTEST	LBCAT	
	CDISC001	37	WBC	Leukocytes	HEMATOL...	4
	CDISC001	72	WBC	Leukocytes	HEMATOL...	4
	CDISC001	102	WBC	Leukocytes	HEMATOL...	5
	CDISC001	147	WBC	Leukocytes	HEMATOL...	4
	CDISC001	182	WBC	Leukocytes	HEMATOL	4

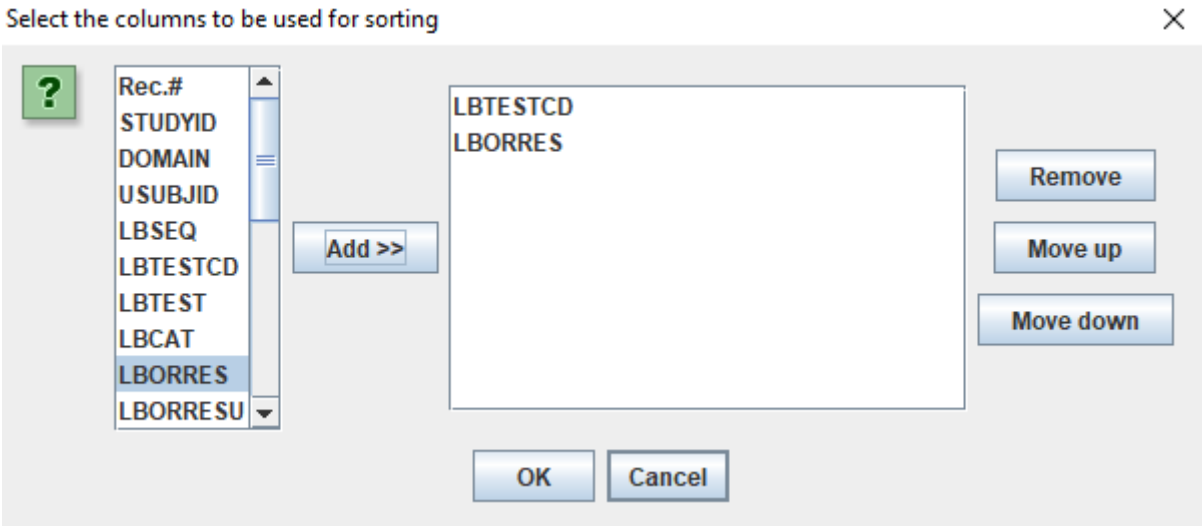
Remark that for numeric variables ("integer" or "float" in the metadata) the rows will be sorted numerically for the variable value.

Sorting can also be accomplished using the menu "Tools - Sorting":

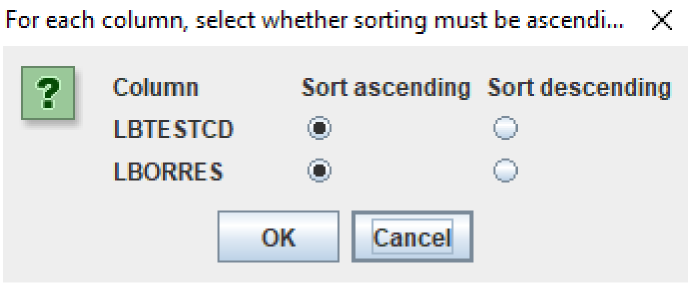
with keyboard shortcuts Ctrl-T (for sorting) Ctrl-U (for unsorting). When using "Sorting", this leads to:



where one can add and remove columns to be sorted on. For example, for sorting on LBTESTCD and LBORRES:



This leads to a new dialog:

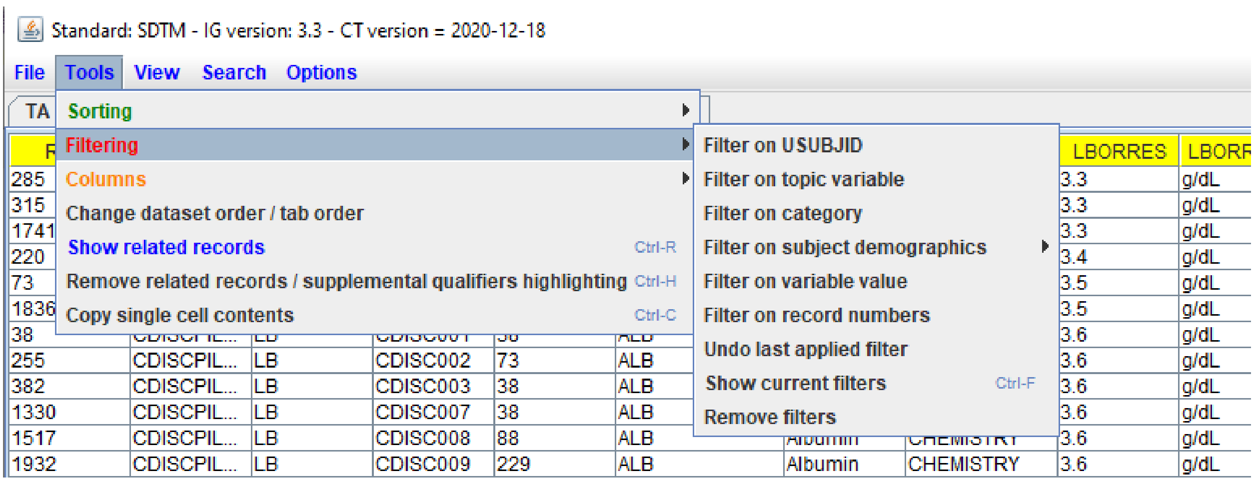


where one can decide to sort ascending or descending for each of the variables, e.g. leading to:

LBTESTCD ▲	LBTEST	LBCAT	LBORRES	LBORRESU	LBORNRLO	LBORNRHI	LBSTRI
ALB	Albumin	CHEMISTRY	3.3	g/dL	3.5	4.6	33
ALB	Albumin	CHEMISTRY	3.3	g/dL	3.5	4.6	33
ALB	Albumin	CHEMISTRY	3.3	g/dL	3.5	4.6	33
ALB	Albumin	CHEMISTRY	3.4	g/dL	3.5	4.6	34
ALB	Albumin	CHEMISTRY	3.5	g/dL	3.5	4.6	35
ALB	Albumin	CHEMISTRY	3.5	g/dL	3.5	4.6	35
ALB	Albumin	CHEMISTRY	3.6	g/dL	3.5	4.6	36
ALB	Albumin	CHEMISTRY	3.6	g/dL	3.5	4.6	36
ALB	Albumin	CHEMISTRY	3.6	g/dL	3.3	4.9	36
ALB	Albumin	CHEMISTRY	3.6	g/dL	3.3	4.9	36
ALB	Albumin	CHEMISTRY	3.6	g/dL	3.5	4.6	36
ALB	Albumin	CHEMISTRY	3.6	g/dL	3.5	4.6	36
ALB	Albumin	CHEMISTRY	3.7	g/dL	3.5	4.6	37
ALB	Albumin	CHEMISTRY	3.7	g/dL	3.5	4.6	37
ALB	Albumin	CHEMISTRY	3.7	g/dL	3.3	4.9	37
ALB	Albumin	CHEMISTRY	3.7	g/dL	3.5	4.6	37
ALB	Albumin	CHEMISTRY	3.7	g/dL	3.5	4.6	37
ALB	Albumin	CHEMISTRY	3.7	g/dL	3.5	4.6	37
ALB	Albumin	CHEMISTRY	3.7	g/dL	3.5	4.6	37
ALB	Albumin	CHEMISTRY	3.7	g/dL	3.5	4.6	37
ALB	Albumin	CHEMISTRY	3.7	g/dL	3.5	4.6	37
ALB	Albumin	CHEMISTRY	3.7	g/dL	3.5	4.6	37
ALB	Albumin	CHEMISTRY	3.7	g/dL	3.5	4.6	37
ALB	Albumin	CHEMISTRY	3.7	g/dL	3.5	4.6	37
ALB	Albumin	CHEMISTRY	3.8	g/dL	3.5	4.6	38

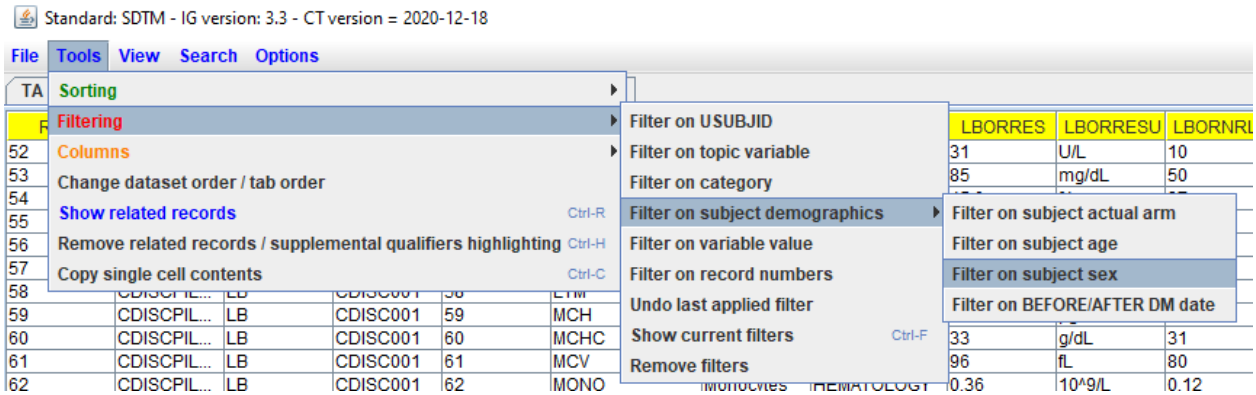
Bringing the table back to the original order can then be done using Ctrl-U.

Also filtering is easily possible, using the menu "Tools - Filtering":

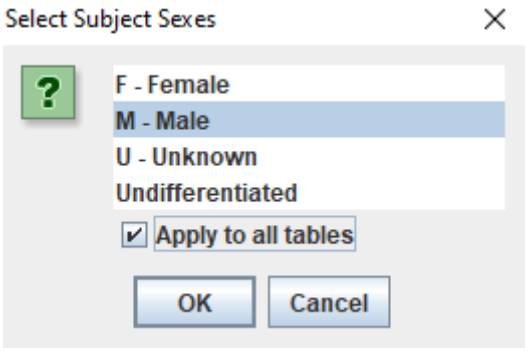


One can filter on the subject, one the topic variable (--TESTCD for "Findings", --TERM for "Events", --TRT for "Interventions", on category (--CAT variable), on subject demographics (e.g. SEX, RACE, AGE, …), but also on variable values (ranges) and of course on record numbers. One can apply "filters on filters", so making combination of filters, display them, and remove all the filters.

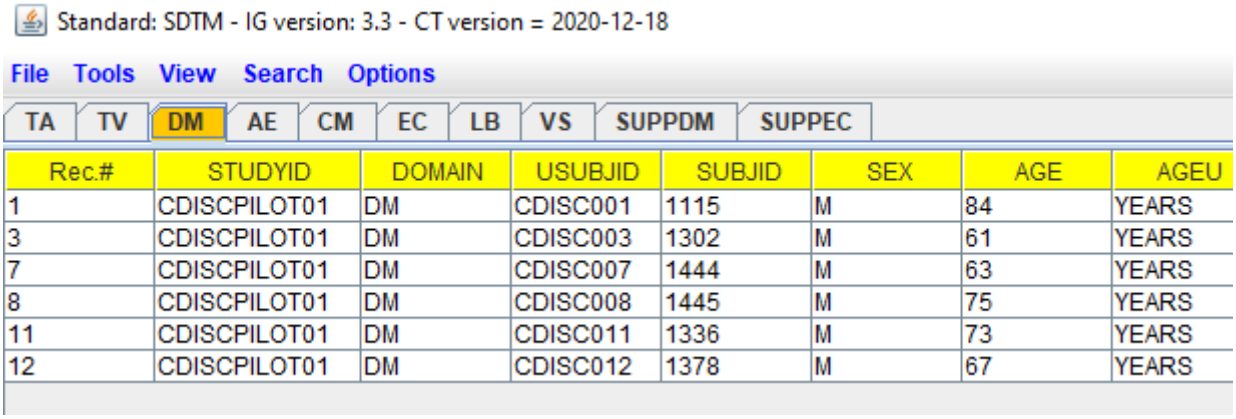
For example, if we want to filter on the "ketone" lab values for the male subjects, we start by applying a filter on "subject demographics", using the corresponding dialog:



followed by:



and ensure that the checkbox "Apply to all tables" is checked. For DM, we will then see:



Rec.#	STUDYID	DOMAIN	USUBJID	SUBJID	SEX	AGE	AGEU
1	CDISCPILOT01	DM	CDISC001	1115	M	84	YEARS
3	CDISCPILOT01	DM	CDISC003	1302	M	61	YEARS
7	CDISCPILOT01	DM	CDISC007	1444	M	63	YEARS
8	CDISCPILOT01	DM	CDISC008	1445	M	75	YEARS
11	CDISCPILOT01	DM	CDISC011	1336	M	73	YEARS
12	CDISCPILOT01	DM	CDISC012	1378	M	67	YEARS

Only showing the male subjects: CDISC001, CDISC003, CDISC007, CDISC008, CDISC011 and CDISC012.
In the LB table we then find:

Standard: SDTM - IG version: 3.3 - CT version = 2020-12-18

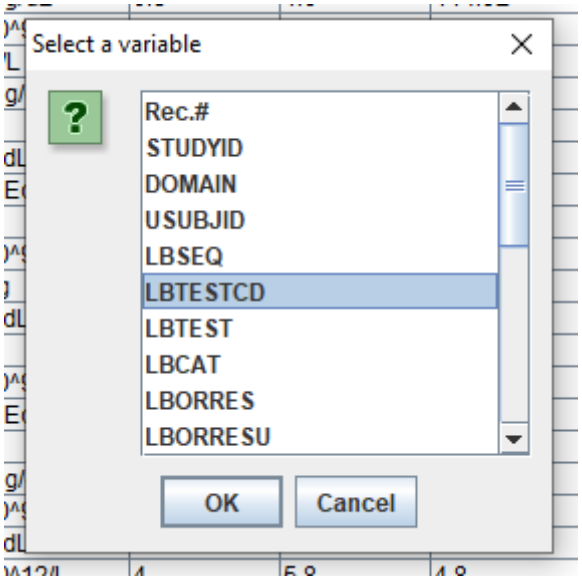
File Tools View Search Options

TA TV DM AE CM EC LB VS SUPPDM SUPPEC

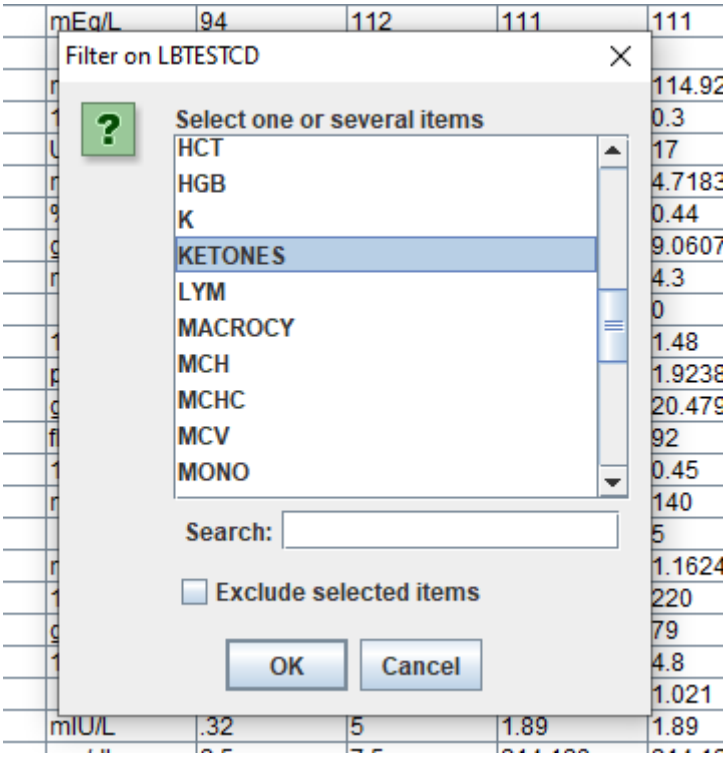
Rec.#	STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	LBORRES	LBORRESU	LBORNRLO	LBORNRHI
178	CDISCPIL...	LB	CDISC001	178	RBC	Erythrocytes	HEMATOL...	4.80	10^12/L	4	5.8
179	CDISCPIL...	LB	CDISC001	179	SPGRAV	Specific Gr...	URINALYSIS	1.023		1.006	1.03
180	CDISCPIL...	LB	CDISC001	180	URATE	Urate	CHEMISTRY	4.5	mg/dL	2.5	7.5
181	CDISCPIL...	LB	CDISC001	181	UROBIL	Urobilinogen	URINALYSIS	0			
182	CDISCPIL...	LB	CDISC001	182	WBC	Leukocytes	HEMATOL...	4.98	10^9/L	3.8	10.7
345	CDISCPIL...	LB	CDISC003	1	ALB	Albumin	CHEMISTRY	4.3	g/dL	3.3	4.9
346	CDISCPIL...	LB	CDISC003	2	ALP	Alkaline Ph...	CHEMISTRY	67	U/L	35	115
347	CDISCPIL...	LB	CDISC003	3	ALT	Alanine Am...	CHEMISTRY	62	U/L	6	43
348	CDISCPIL...	LB	CDISC003	4	AST	Aspartate A...	CHEMISTRY	52	U/L	11	36
349	CDISCPIL...	LB	CDISC003	5	BASO	Basophils	HEMATOL...	0.07	10^9/L	0	0.2
350	CDISCPIL...	LB	CDISC003	6	BILI	Bilirubin	CHEMISTRY	0.6	mg/dL	0.2	1.2
351	CDISCPIL...	LB	CDISC003	7	UREAN	Urea Nitrog...	CHEMISTRY	13	mg/dL	4	24
352	CDISCPIL...	LB	CDISC003	8	CA	Calcium	CHEMISTRY	9.6	mg/dL	8.4	10.3
353	CDISCPIL...	LB	CDISC003	9	CHOL	Cholesterol	CHEMISTRY	255	mg/dL	149	286
354	CDISCPIL...	LB	CDISC003	10	CK	Creatine Ki...	CHEMISTRY	444	U/L	22	198

only containing records for male subjects.

On top of this, we want to apply a filter for LBTESTCD = KETONES.
We can do this by filtering on the "topic variable", as for LB, LBTESTCD is the "topic variable", or we can use "Filter on variable value", select LBTESTCD:



and then select "KETONES":



Remark that If one wants to select all values except for a few ones, one can also use the checkbox "Exclude the selected items".
The result after clicking "OK" then is:

Standard: SDTM - IG version: 3.3 - CT version = 2020-12-18

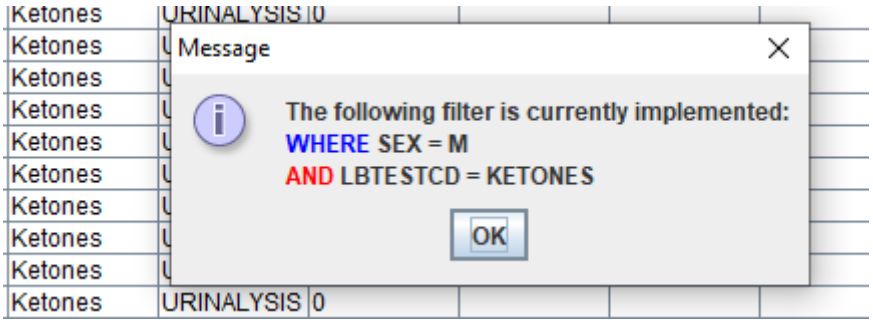
File Tools View Search Options

TA TV DM AE CM EC LB VS SUPPDM SUPPEC

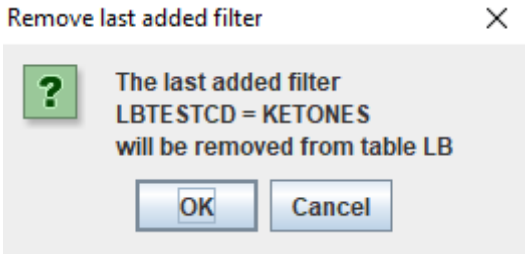
Rec.#	STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	LBORRES	LBOR
20	CDISCPILOT01	LB	CDISC001	20	KETONES	Ketones	URINALYSIS	0	
57	CDISCPILOT01	LB	CDISC001	57	KETONES	Ketones	URINALYSIS	0	
167	CDISCPILOT01	LB	CDISC001	167	KETONES	Ketones	URINALYSIS	0	
364	CDISCPILOT01	LB	CDISC003	20	KETONES	Ketones	URINALYSIS	0	
401	CDISCPILOT01	LB	CDISC003	57	KETONES	Ketones	URINALYSIS	0	
526	CDISCPILOT01	LB	CDISC003	182	KETONES	Ketones	URINALYSIS	0	
1312	CDISCPILOT01	LB	CDISC007	20	KETONES	Ketones	URINALYSIS	0	
1349	CDISCPILOT01	LB	CDISC007	57	KETONES	Ketones	URINALYSIS	0	
1414	CDISCPILOT01	LB	CDISC007	122	KETONES	Ketones	URINALYSIS	0	
1449	CDISCPILOT01	LB	CDISC008	20	KETONES	Ketones	URINALYSIS	0	
1500	CDISCPILOT01	LB	CDISC008	71	KETONES	Ketones	URINALYSIS	0	
1626	CDISCPILOT01	LB	CDISC008	197	KETONES	Ketones	URINALYSIS	0	
1689	CDISCPILOT01	LB	CDISC008	260	KETONES	Ketones	URINALYSIS	0	
2111	CDISCPILOT01	LB	CDISC011	20	KETONES	Ketones	URINALYSIS	0	
2148	CDISCPILOT01	LB	CDISC011	57	KETONES	Ketones	URINALYSIS	0	
2272	CDISCPILOT01	LB	CDISC011	181	KETONES	Ketones	URINALYSIS	0	
2368	CDISCPILOT01	LB	CDISC011	277	KETONES	Ketones	URINALYSIS	0	
2434	CDISCPILOT01	LB	CDISC012	20	KETONES	Ketones	URINALYSIS	0	
2471	CDISCPILOT01	LB	CDISC012	57	KETONES	Ketones	URINALYSIS	0	
2596	CDISCPILOT01	LB	CDISC012	182	KETONES	Ketones	URINALYSIS	0	
2661	CDISCPILOT01	LB	CDISC012	247	KETONES	Ketones	URINALYSIS	0	

only showing the lab records for "ketones" for male subjects CDISC001, CDISC003, CDISC007 etc..

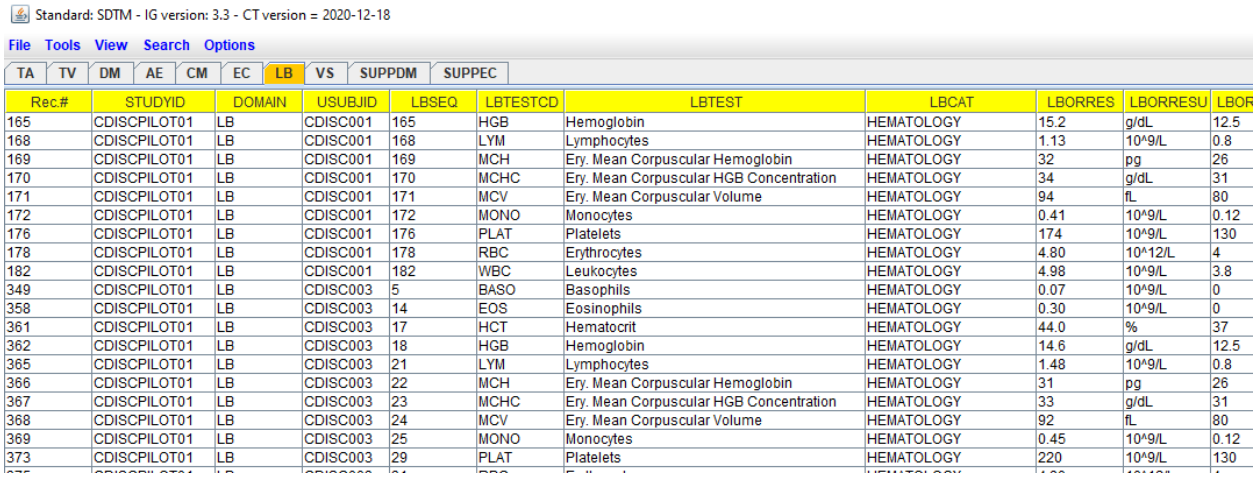
One can always check which filters are on, by using the menu "Tools - Filtering - Show current filters", in our case leading to:



One can also remove the last applied filter using the menu "Tools - Filtering - Undo last applied filter". A message is then displayed:

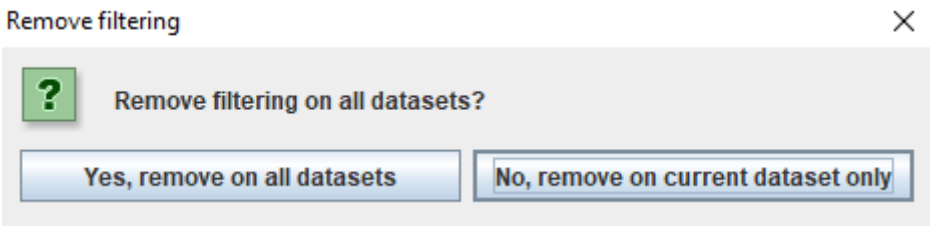


If we do so, and then add a new filter on LBCAT = HEMATOLOGY, the result is:



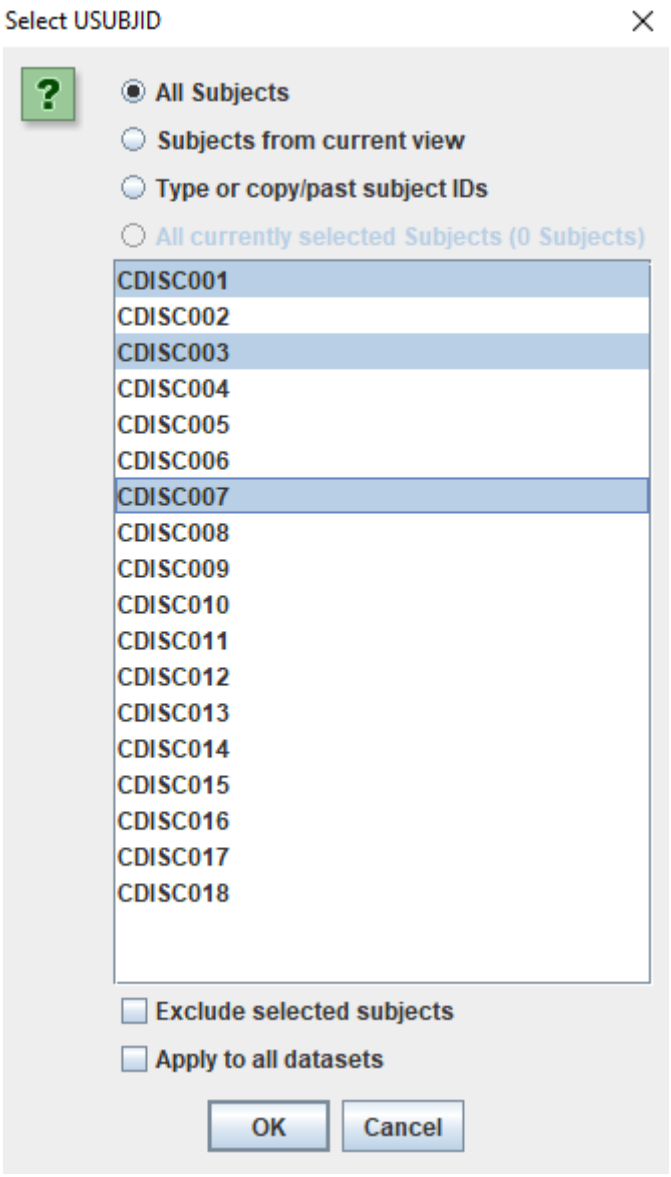
Rec.#	STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	LBORRES	LBORRESU	LBOR
165	CDISCPILOT01	LB	CDISC001	165	HGB	Hemoglobin	HEMATOLOGY	15.2	g/dL	12.5
168	CDISCPILOT01	LB	CDISC001	168	LYM	Lymphocytes	HEMATOLOGY	1.13	10^9/L	0.8
169	CDISCPILOT01	LB	CDISC001	169	MCH	Ery. Mean Corpuscular Hemoglobin	HEMATOLOGY	32	pg	26
170	CDISCPILOT01	LB	CDISC001	170	MCHC	Ery. Mean Corpuscular HGB Concentration	HEMATOLOGY	34	g/dL	31
171	CDISCPILOT01	LB	CDISC001	171	MCV	Ery. Mean Corpuscular Volume	HEMATOLOGY	94	fL	80
172	CDISCPILOT01	LB	CDISC001	172	MONO	Monocytes	HEMATOLOGY	0.41	10^9/L	0.12
176	CDISCPILOT01	LB	CDISC001	176	PLAT	Platelets	HEMATOLOGY	174	10^9/L	130
178	CDISCPILOT01	LB	CDISC001	178	RBC	Erythrocytes	HEMATOLOGY	4.80	10^12/L	4
182	CDISCPILOT01	LB	CDISC001	182	WBC	Leukocytes	HEMATOLOGY	4.98	10^9/L	3.8
349	CDISCPILOT01	LB	CDISC003	5	BASO	Basophils	HEMATOLOGY	0.07	10^9/L	0
358	CDISCPILOT01	LB	CDISC003	14	EOS	Eosinophils	HEMATOLOGY	0.30	10^9/L	0
361	CDISCPILOT01	LB	CDISC003	17	HCT	Hematocrit	HEMATOLOGY	44.0	%	37
362	CDISCPILOT01	LB	CDISC003	18	HGB	Hemoglobin	HEMATOLOGY	14.6	g/dL	12.5
365	CDISCPILOT01	LB	CDISC003	21	LYM	Lymphocytes	HEMATOLOGY	1.48	10^9/L	0.8
366	CDISCPILOT01	LB	CDISC003	22	MCH	Ery. Mean Corpuscular Hemoglobin	HEMATOLOGY	31	pg	26
367	CDISCPILOT01	LB	CDISC003	23	MCHC	Ery. Mean Corpuscular HGB Concentration	HEMATOLOGY	33	g/dL	31
368	CDISCPILOT01	LB	CDISC003	24	MCV	Ery. Mean Corpuscular Volume	HEMATOLOGY	92	fL	80
369	CDISCPILOT01	LB	CDISC003	25	MONO	Monocytes	HEMATOLOGY	0.45	10^9/L	0.12
373	CDISCPILOT01	LB	CDISC003	29	PLAT	Platelets	HEMATOLOGY	220	10^9/L	130

To remove all filters, use the menu "Tools - Filtering - Remove filters", leading to:



Usually, one will want to remove the filters on all the loaded datasets, otherwise things can get a bit messy. When then using "Yes, remove on all datasets", all the tables will be shown in their original state.

One can also always filter on specific subjects, by USUBJID, using the menu "Tools - Filtering - Filter on USUBJID". The system will then present a list of subjects by USUBJID, from which one can select the ones of interest, e.g.:

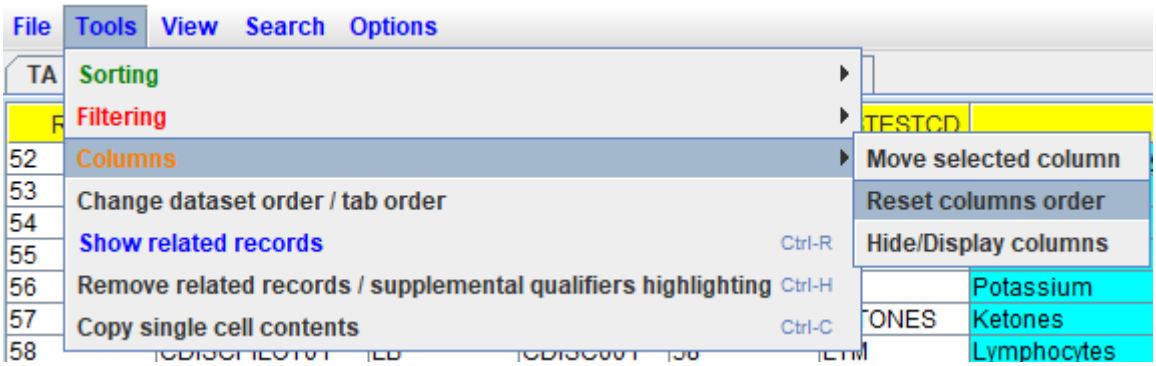


Also here, one can filter by exclusion, and apply the filter on either the current loaded datasets on all loaded datasets.

Remark that it is often to do "pre-loading filtering" as explains in the section "Filtering before loading".

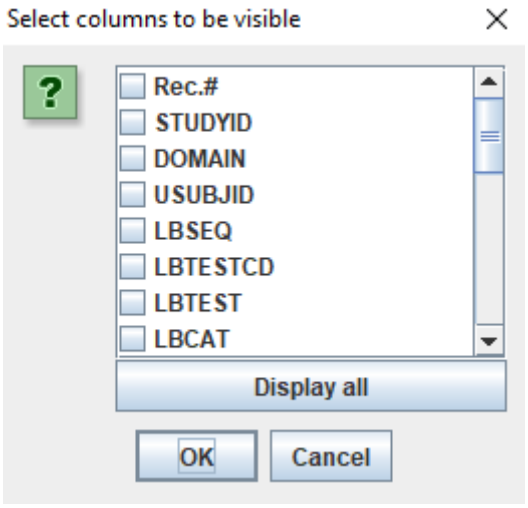
Moving columns around and hiding columns

We already showed how columns can be moved from one place to another, simply by dragging the column header. As in SDTM and SEND, the column order is in principal fixed, one can reset to the original order by using the menu "Tools - Columns - Reset column order":

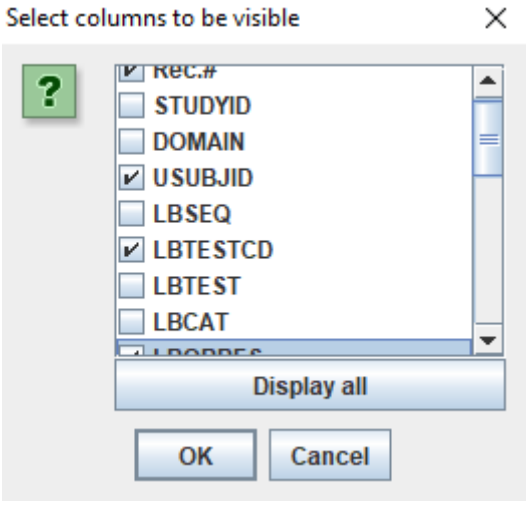


The menu "Tools - Columns - Move selected column" can also be used to move a column to another place (right-click on the column header first to select it), but usually just dragging the column will be the easier way.

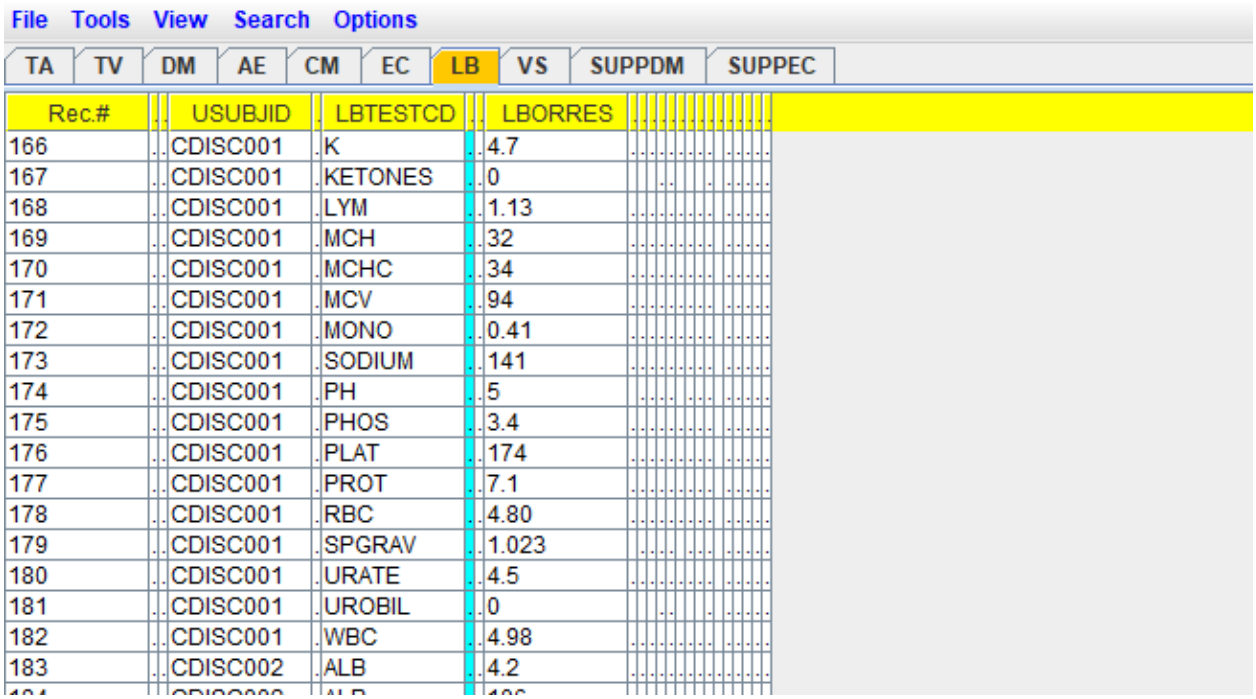
One can also hide (and later re-display) columns by using the menu "Tools - Columns - Hide/Display columns". This brings the following dialog up:



When we e.g. want to have "Rec.#", USUBJID, LBTESTCD, LBORRES and LBORRESU displayed, we select these:



leading to:



Rec.#	USUBJID	LBTESTCD	LBORRES
166	CDISC001	K	4.7
167	CDISC001	KETONES	0
168	CDISC001	LYM	1.13
169	CDISC001	MCH	32
170	CDISC001	MCHC	34
171	CDISC001	MCV	94
172	CDISC001	MONO	0.41
173	CDISC001	SODIUM	141
174	CDISC001	PH	5
175	CDISC001	PHOS	3.4
176	CDISC001	PLAT	174
177	CDISC001	PROT	7.1
178	CDISC001	RBC	4.80
179	CDISC001	SPGRAV	1.023
180	CDISC001	URATE	4.5
181	CDISC001	UROBIL	0
182	CDISC001	WBC	4.98
183	CDISC002	ALB	4.2

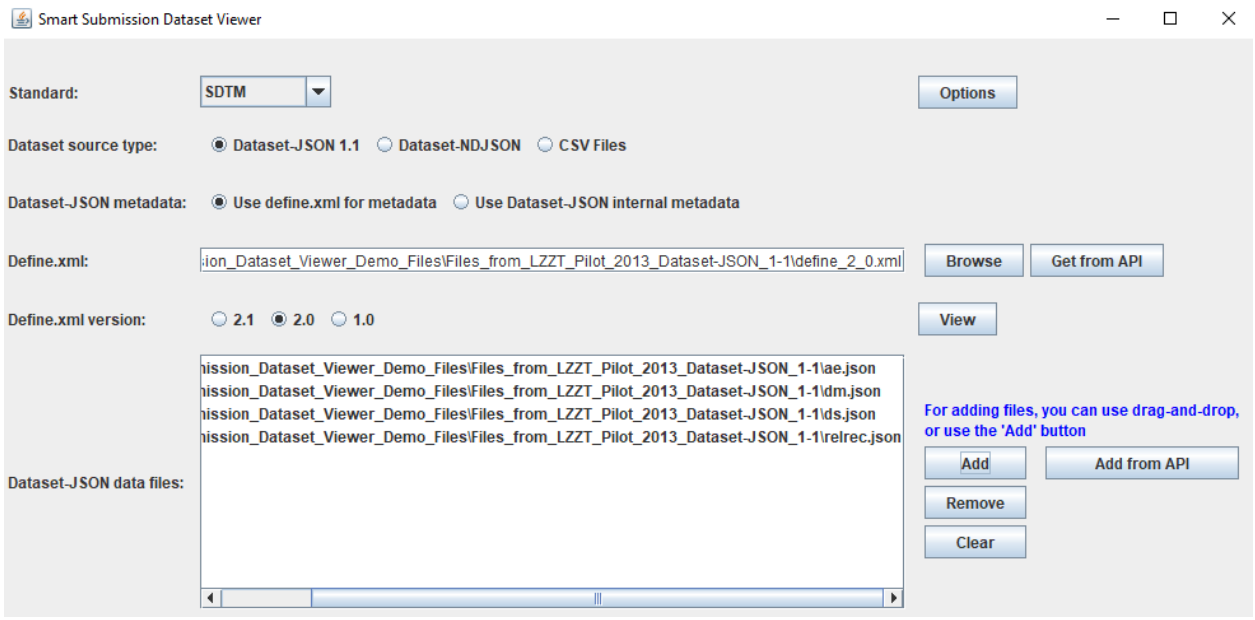
Remark that the columns to be hidden are not completely hidden, but just "minimized", so that dragging on the edge can make individual columns visible again.
We might want to change this behavior in future depending on the user responses.

Making all the columns visible again, i.e. giving them their original width, can then be done using the same menu, and clicking "Display all".

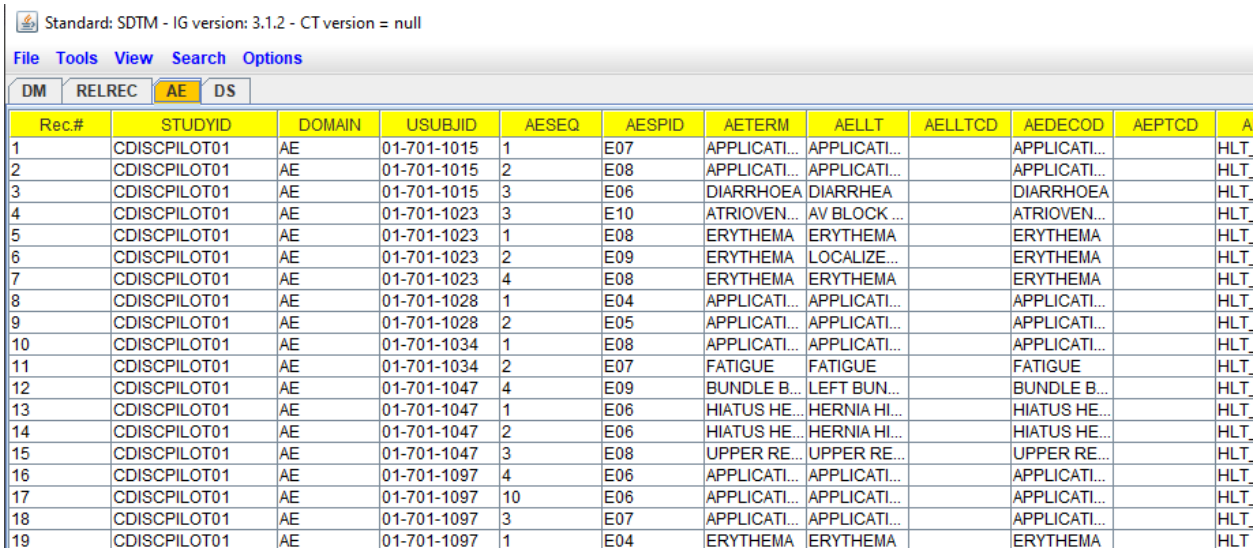
Connecting Related records (RELREC)

We did already show how one can establish the relationship between rows in the SUPPQUAL (supplemental qualifier) datasets and the "parent" records in the "parent" table, either by "bringing back" the "non-standard variables" to the parent dataset as additional columns (which however requires a quality define.xml) or by "toggling" between the related rows in both the datasets.

There is however another importantmechanism of having relationships between records and tables, by the use of the "RELREC" (Related Records) datasets. It is often used to establish the relationship between the AE (adverse events) dataset and the CM (concomitant medications) datasets. Unfortunately, the sample dataset from the "Metadata Submission Guidelines" does not have a good example, so we need to take another example. We find this in the files from the original LZZT pilot 2013 which we converted from XPT to Dataset-JSON 1.1, and which contain records for describing the relation between AE (adverse events) and DS (dispostion). Loading these, together with DM (always a good idea) into the software:

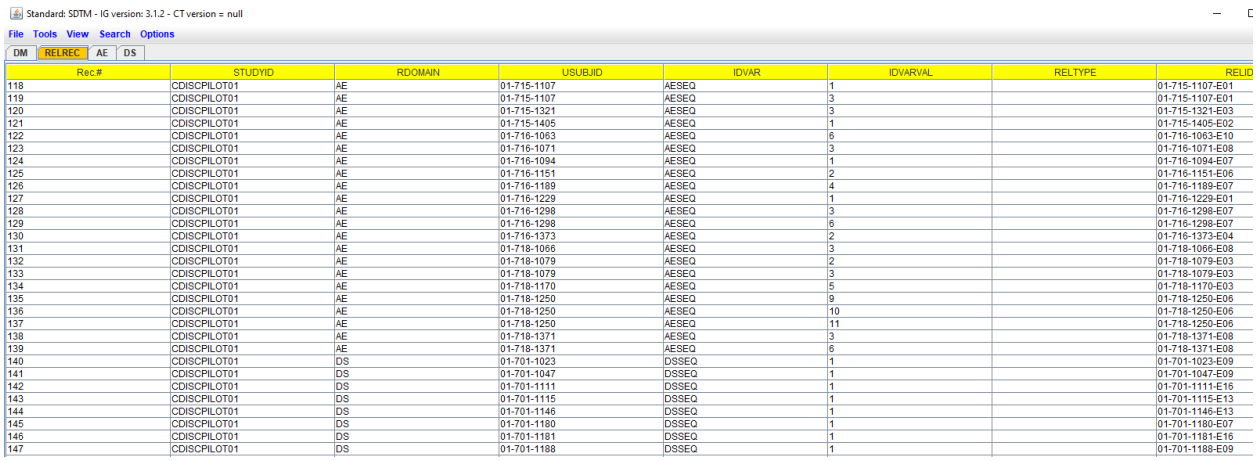


and start loading by clicking the "Start" button, this then leads to:



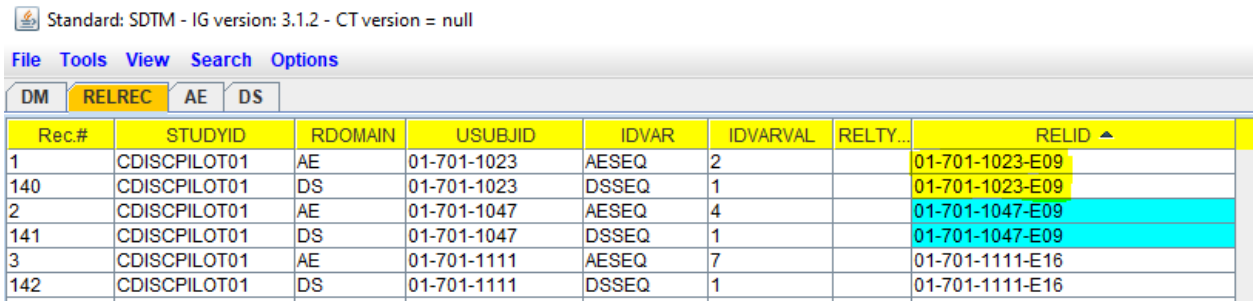
Rec.#	STUDYID	DOMAIN	USUBJID	AESEQ	AESPID	AETERM	AELLT	AELLTCD	AEDECOD	AEPTCD	A
1	CDISCPILOT01	AE	01-701-1015	1	E07	APPLICATI...	APPLICATI...		APPLICATI...		HLT_
2	CDISCPILOT01	AE	01-701-1015	2	E08	APPLICATI...	APPLICATI...		APPLICATI...		HLT_
3	CDISCPILOT01	AE	01-701-1015	3	E06	DIARRHOEA	DIARRHEA		DIARRHOEA		HLT_
4	CDISCPILOT01	AE	01-701-1023	3	E10	ATRIOVEN...	AV BLOCK ...		ATRIOVEN...		HLT_
5	CDISCPILOT01	AE	01-701-1023	1	E08	ERYTHEMA	ERYTHEMA		ERYTHEMA		HLT_
6	CDISCPILOT01	AE	01-701-1023	2	E09	ERYTHEMA	LOCALIZE...		ERYTHEMA		HLT_
7	CDISCPILOT01	AE	01-701-1023	4	E08	ERYTHEMA	ERYTHEMA		ERYTHEMA		HLT_
8	CDISCPILOT01	AE	01-701-1028	1	E04	APPLICATI...	APPLICATI...		APPLICATI...		HLT_
9	CDISCPILOT01	AE	01-701-1028	2	E05	APPLICATI...	APPLICATI...		APPLICATI...		HLT_
10	CDISCPILOT01	AE	01-701-1034	1	E08	APPLICATI...	APPLICATI...		APPLICATI...		HLT_
11	CDISCPILOT01	AE	01-701-1034	2	E07	FATIGUE	FATIGUE		FATIGUE		HLT_
12	CDISCPILOT01	AE	01-701-1047	4	E09	BUNDLE B...	LEFT BUN...		BUNDLE B...		HLT_
13	CDISCPILOT01	AE	01-701-1047	1	E06	HIATUS HE...	HERNIA HI...		HIATUS HE...		HLT_
14	CDISCPILOT01	AE	01-701-1047	2	E06	HIATUS HE...	HERNIA HI...		HIATUS HE...		HLT_
15	CDISCPILOT01	AE	01-701-1047	3	E08	UPPER RE...	UPPER RE...		UPPER RE...		HLT_
16	CDISCPILOT01	AE	01-701-1097	4	E06	APPLICATI...	APPLICATI...		APPLICATI...		HLT_
17	CDISCPILOT01	AE	01-701-1097	10	E06	APPLICATI...	APPLICATI...		APPLICATI...		HLT_
18	CDISCPILOT01	AE	01-701-1097	3	E07	APPLICATI...	APPLICATI...		APPLICATI...		HLT_
19	CDISCPILOT01	AE	01-701-1097	1	E04	ERYTHEMA	ERYTHEMA		ERYTHEMA		HLT

and for the RELREC table:



Rec.#	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
118	CDISCPILOT01	AE	01-715-1107	AESEQ	1		01-715-1107-E01
119	CDISCPILOT01	AE	01-715-1107	AESEQ	3		01-715-1107-E01
120	CDISCPILOT01	AE	01-715-1321	AESEQ	3		01-715-1321-E03
121	CDISCPILOT01	AE	01-715-1405	AESEQ	1		01-715-1405-E02
122	CDISCPILOT01	AE	01-716-1063	AESEQ	6		01-716-1063-E10
123	CDISCPILOT01	AE	01-716-1071	AESEQ	3		01-716-1071-E08
124	CDISCPILOT01	AE	01-716-1094	AESEQ	1		01-716-1094-E07
125	CDISCPILOT01	AE	01-716-1151	AESEQ	2		01-716-1151-E06
126	CDISCPILOT01	AE	01-716-1189	AESEQ	4		01-716-1189-E07
127	CDISCPILOT01	AE	01-716-1229	AESEQ	1		01-716-1229-E01
128	CDISCPILOT01	AE	01-716-1298	AESEQ	3		01-716-1298-E07
129	CDISCPILOT01	AE	01-716-1298	AESEQ	6		01-716-1298-E07
130	CDISCPILOT01	AE	01-716-1373	AESEQ	2		01-716-1373-E04
131	CDISCPILOT01	AE	01-718-1066	AESEQ	3		01-718-1066-E08
132	CDISCPILOT01	AE	01-718-1079	AESEQ	2		01-718-1079-E03
133	CDISCPILOT01	AE	01-718-1079	AESEQ	3		01-718-1079-E03
134	CDISCPILOT01	AE	01-718-1170	AESEQ	5		01-718-1170-E03
135	CDISCPILOT01	AE	01-718-1250	AESEQ	9		01-718-1250-E06
136	CDISCPILOT01	AE	01-718-1250	AESEQ	10		01-718-1250-E06
137	CDISCPILOT01	AE	01-718-1250	AESEQ	11		01-718-1250-E06
138	CDISCPILOT01	AE	01-718-1371	AESEQ	3		01-718-1371-E08
139	CDISCPILOT01	AE	01-718-1371	AESEQ	6		01-718-1371-E08
140	CDISCPILOT01	DS	01-701-1023	DSSEQ	1		01-701-1023-E09
141	CDISCPILOT01	DS	01-701-1047	DSSEQ	1		01-701-1047-E09
142	CDISCPILOT01	DS	01-701-1111	DSSEQ	1		01-701-1111-E16
143	CDISCPILOT01	DS	01-701-1115	DSSEQ	1		01-701-1115-E13
144	CDISCPILOT01	DS	01-701-1146	DSSEQ	1		01-701-1146-E13
145	CDISCPILOT01	DS	01-701-1180	DSSEQ	1		01-701-1180-E07
146	CDISCPILOT01	DS	01-701-1181	DSSEQ	1		01-701-1181-E16
147	CDISCPILOT01	DS	01-701-1188	DSSEQ	1		01-701-1188-E09

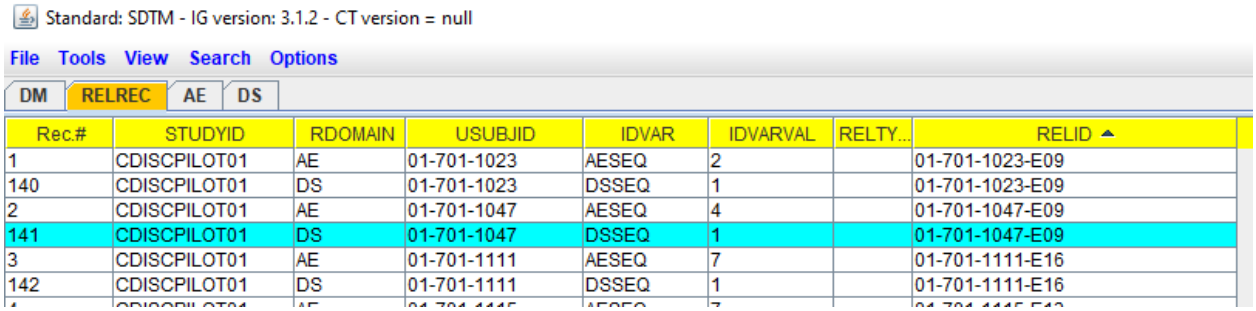
When sorting by RELID, we better see which AE and DS records are related:



Rec.#	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTY...	RELID
1	CDISCPILOT01	AE	01-701-1023	AESEQ	2		01-701-1023-E09
140	CDISCPILOT01	DS	01-701-1023	DSSEQ	1		01-701-1023-E09
2	CDISCPILOT01	AE	01-701-1047	AESEQ	4		01-701-1047-E09
141	CDISCPILOT01	DS	01-701-1047	DSSEQ	1		01-701-1047-E09
3	CDISCPILOT01	AE	01-701-1111	AESEQ	7		01-701-1111-E16
142	CDISCPILOT01	DS	01-701-1111	DSSEQ	1		01-701-1111-E16

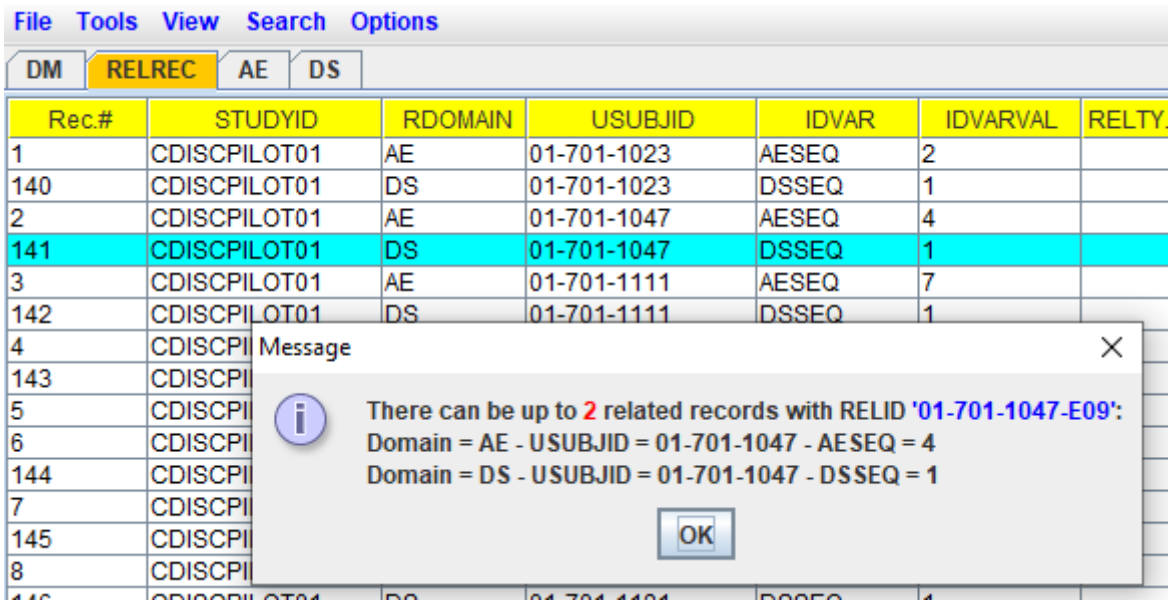
showing that for subject 01-701-1023, the AE record with AESEQ=2 is related to the DS record with DSSEQ=1,
and that for subject 01-701-1047, the AE record with AESEQ=4 is related to the DS record with DSSEQ=1.

If we select the latter (record # 141):

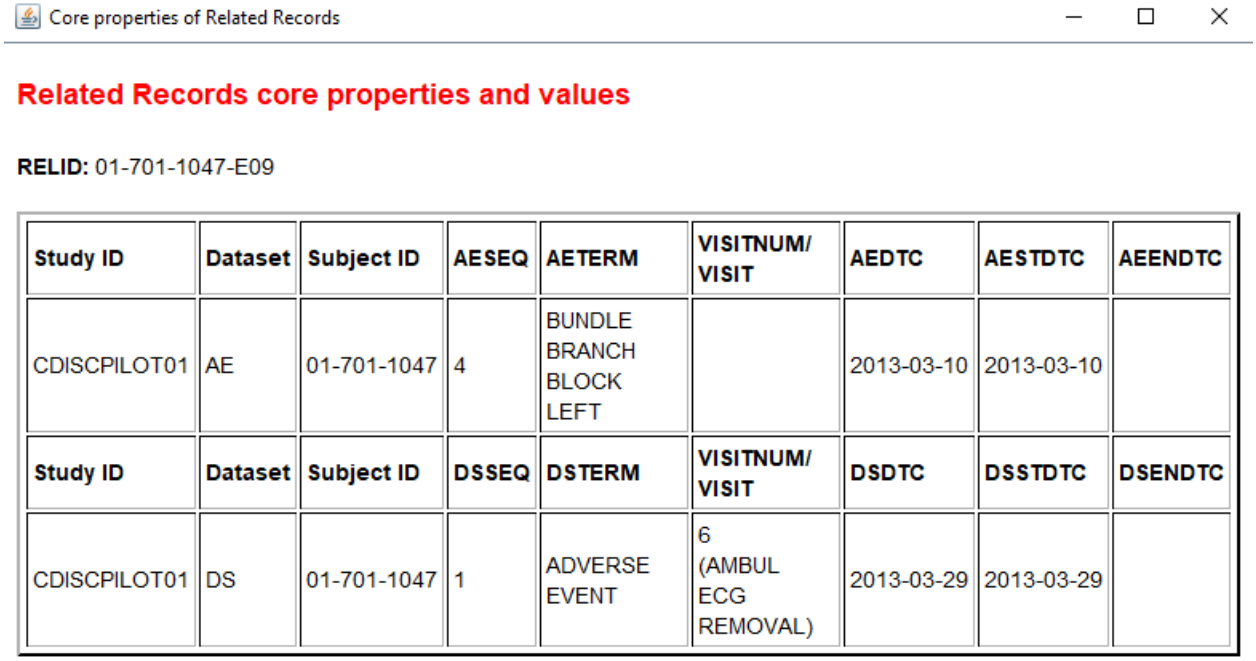


Rec.#	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTY...	RELID
1	CDISCPILOT01	AE	01-701-1023	AESEQ	2		01-701-1023-E09
140	CDISCPILOT01	DS	01-701-1023	DSSEQ	1		01-701-1023-E09
2	CDISCPILOT01	AE	01-701-1047	AESEQ	4		01-701-1047-E09
141	CDISCPILOT01	DS	01-701-1047	DSSEQ	1		01-701-1047-E09
3	CDISCPILOT01	AE	01-701-1111	AESEQ	7		01-701-1111-E16
142	CDISCPILOT01	DS	01-701-1111	DSSEQ	1		01-701-1111-E16

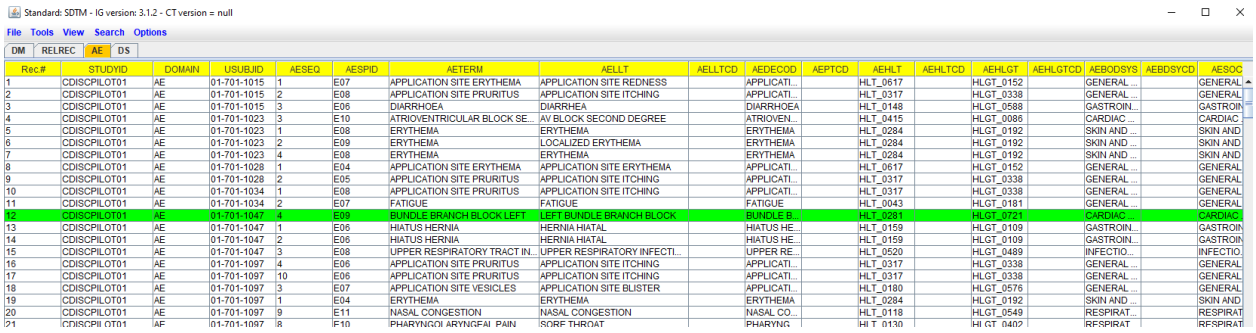
and then use the menu "Tools - Show related records" (or simply use Ctrl-R on the keyboard), this leads to a message:



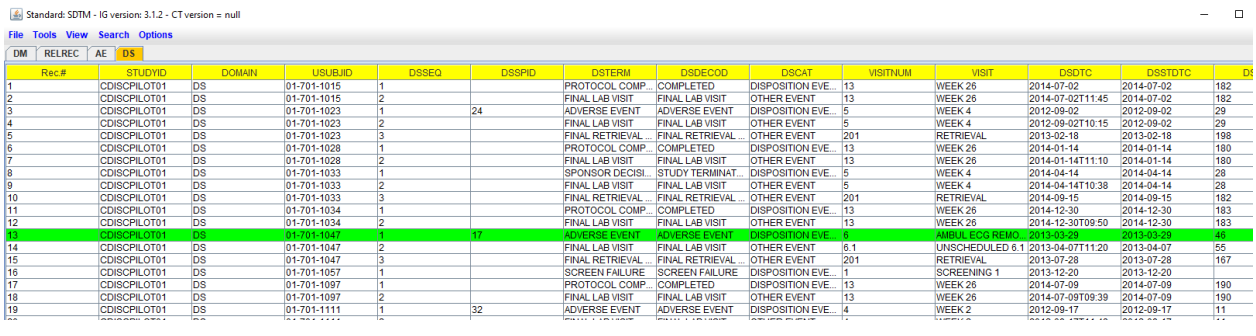
and after clicking "OK", additional information is displayed in a separate window:



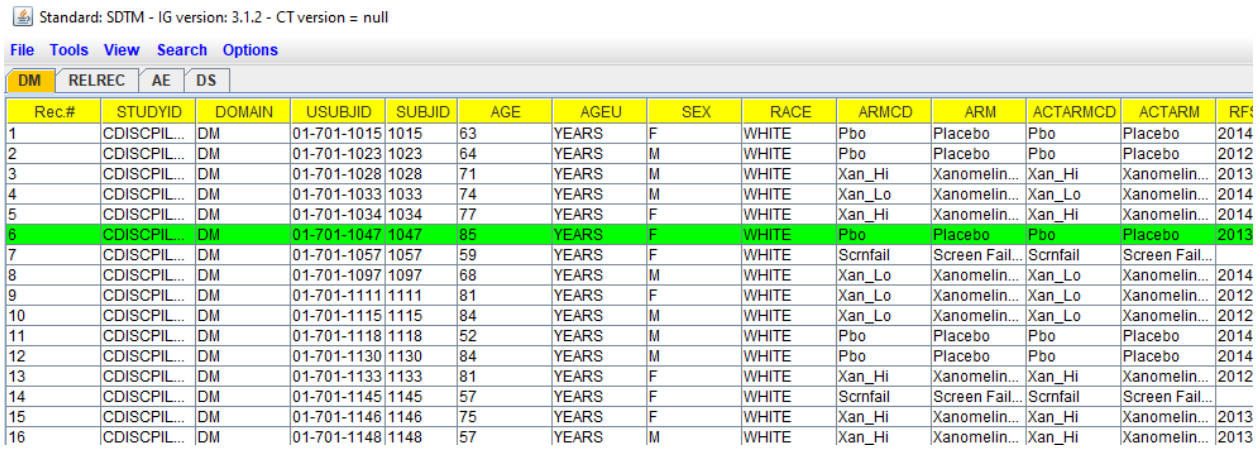
showing the details of the relation.
If we then select the AE tab, the specific record is already highlighted:



and when selecting the DS tab, also there, the specific record is already highlighted:



one can then "toggle" between the DS, AE, and RELREC datasets using Ctrl-B on the keyboard, and using Ctrl-B (or the menu "View - Show corresponding DM record"):



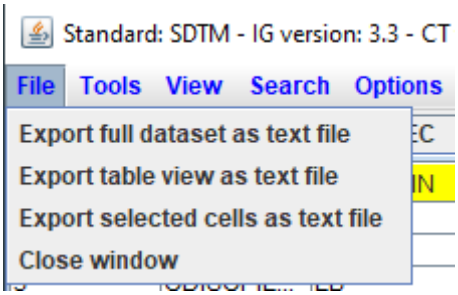
Rec.#	STUDYID	DOMAIN	USUBJID	SUBJID	AGE	AGEU	SEX	RACE	ARMCD	ARM	ACTARMCD	ACTARM	RF
1	CDISCPIL...	DM	01-701-1015	1015	63	YEARS	F	WHITE	Pbo	Placebo	Pbo	Placebo	2014
2	CDISCPIL...	DM	01-701-1023	1023	64	YEARS	M	WHITE	Pbo	Placebo	Pbo	Placebo	2012
3	CDISCPIL...	DM	01-701-1028	1028	71	YEARS	M	WHITE	Xan_Hi	Xanomelin...	Xan_Hi	Xanomelin...	2013
4	CDISCPIL...	DM	01-701-1033	1033	74	YEARS	M	WHITE	Xan_Lo	Xanomelin...	Xan_Lo	Xanomelin...	2014
5	CDISCPIL...	DM	01-701-1034	1034	77	YEARS	F	WHITE	Xan_Hi	Xanomelin...	Xan_Hi	Xanomelin...	2014
6	CDISCPIL...	DM	01-701-1047	1047	85	YEARS	F	WHITE	Pbo	Placebo	Pbo	Placebo	2013
7	CDISCPIL...	DM	01-701-1057	1057	59	YEARS	F	WHITE	Scrnfail	Screen Fail...	Scrnfail	Screen Fail...	
8	CDISCPIL...	DM	01-701-1097	1097	68	YEARS	M	WHITE	Xan_Lo	Xanomelin...	Xan_Lo	Xanomelin...	2014
9	CDISCPIL...	DM	01-701-1111	1111	81	YEARS	F	WHITE	Xan_Lo	Xanomelin...	Xan_Lo	Xanomelin...	2012
10	CDISCPIL...	DM	01-701-1115	1115	84	YEARS	M	WHITE	Xan_Lo	Xanomelin...	Xan_Lo	Xanomelin...	2012
11	CDISCPIL...	DM	01-701-1118	1118	52	YEARS	M	WHITE	Pbo	Placebo	Pbo	Placebo	2014
12	CDISCPIL...	DM	01-701-1130	1130	84	YEARS	M	WHITE	Pbo	Placebo	Pbo	Placebo	2014
13	CDISCPIL...	DM	01-701-1133	1133	81	YEARS	F	WHITE	Xan_Hi	Xanomelin...	Xan_Hi	Xanomelin...	2012
14	CDISCPIL...	DM	01-701-1145	1145	57	YEARS	F	WHITE	Scrnfail	Screen Fail...	Scrnfail	Screen Fail...	
15	CDISCPIL...	DM	01-701-1146	1146	75	YEARS	F	WHITE	Xan_Hi	Xanomelin...	Xan_Hi	Xanomelin...	2013
16	CDISCPIL...	DM	01-701-1148	1148	57	YEARS	M	WHITE	Xan_Hi	Xanomelin...	Xan_Hi	Xanomelin...	2013

easily find out that this relation is about a 85 year old female who was in the placebo arm.

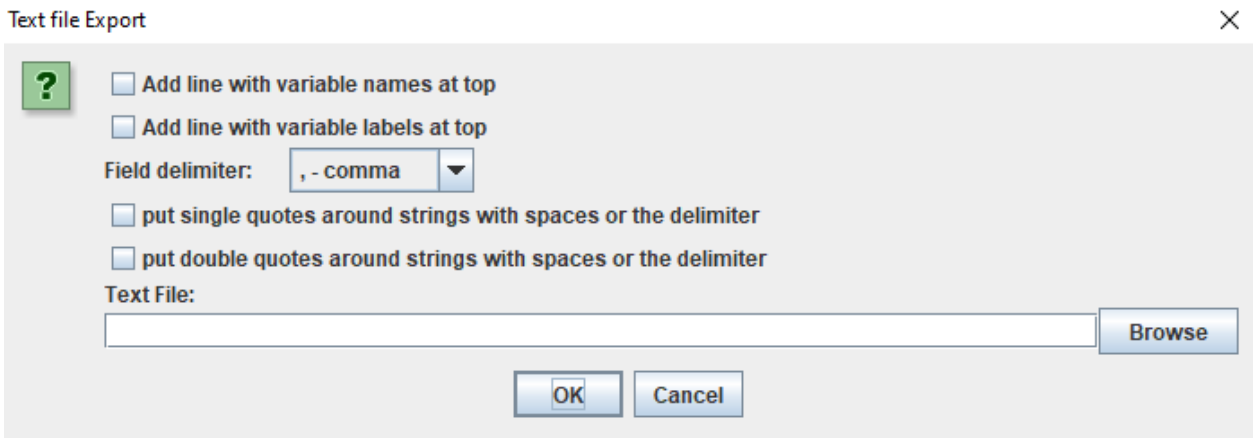
Exporting the table

In some cases, one want to export the table contents, of parts of it, as simple text or CSV files for use in other applications. Essentially, this should not be necessary, as we do already have the data in modern Dataset-JSON format, but as long as tools like Excel are used in clinical research (they shouldn't except for bookkeeping, that's what they were developed for), people will want to be able to import data from dedicated applications into such applications.

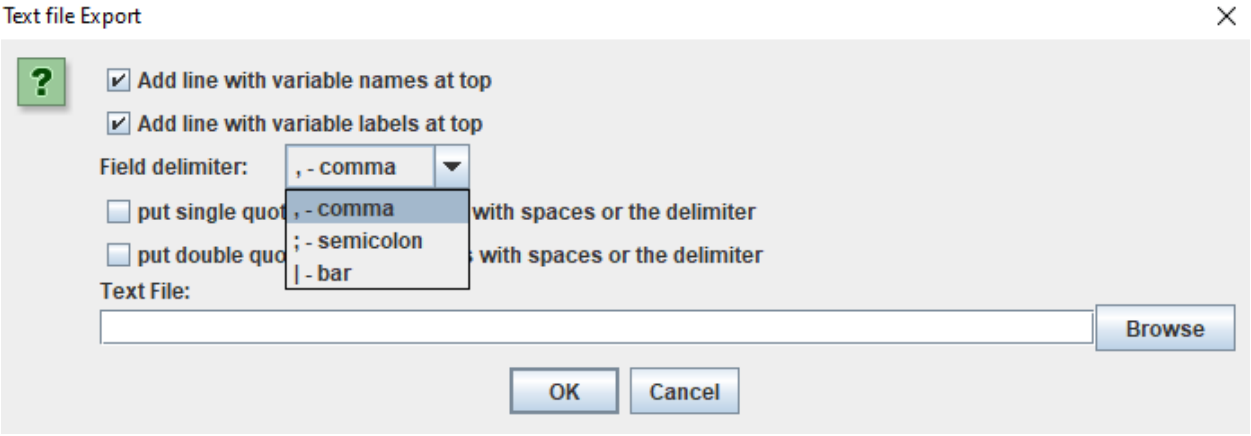
The "Smart Submission Dataset Viewer" provides several possibilities though the menu items "File - Export ...":



One can either export the full dataset of the selected table as a text file, or only the rows and columns that are currently visible, or only the set of currently selected cells. In each of these three cases, the system will ask how to export the data, by the following dialog:



The first two checkboxes allow to add a line with the variable names, and a line with the variable labels, which can serve as a header in receiving applications.
The system then asks what the delimited between the fields should be:



giving the choice between the comma, the semicolon and the vertical bar ("pipe" character).
The next two checkboxes allow to set whether either single or double quotes need to be added around the strings values of the variables. This can be important when e.g. the comma is selected as the field delimiter, but the string values themselves can contain the comma character[3].
The text field and "Browse" button then allow to define where the text (or CSV) file needs to be written to. Remark that the file will be written using UTF-8 encoding, essentialyl representing Unicode.

An example of such an export is:

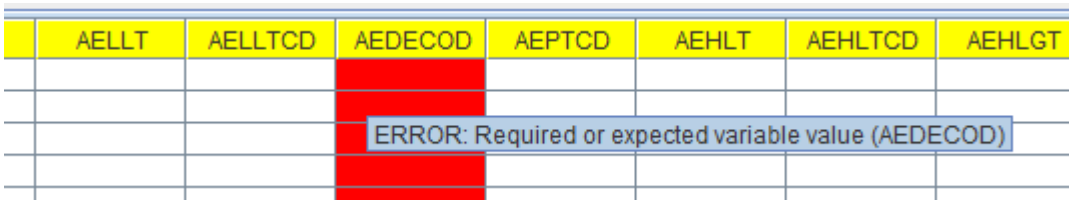
```
Rec.#,STUDYID,DOMAIN,USUBJID,LBSEQ,LBTESTCD,LBTEST,LBCAT,LBORRES,LBORRESU,LBORNRLO,LBORNRHI,LBSTRESC,LBSTRESN,LBSTRESU,LBSTNRLO,LBSTNRHI,LBNRIND,LBLOBXFL,VISITNUM,VISIT,EPOCH,LBDTC,LBDY
,"Study Identifier","Domain Abbreviation","Unique Subject Identifier","Sequence Number","Lab Test or Examination Short Name","Lab Test or Examination Name","Category for Lab Test","Result or Finding in Original Units","Original Units","Reference Range Lower Limit in Orig Unit","Reference Range Upper Limit in Orig Unit","Character Result/Finding in Std Format","Numeric Result/Finding in Standard Units","Standard Units","Reference Range Lower Limit-Std Units","Reference Range Upper Limit-Std Units","Reference Range Indicator","Last Observation Before Exposure Flag","Visit Number","Visit Name",Epoch,"Date/Time of Specimen Collection","Study Day of Specimen Collection"
1,CDISCPILOT01,LB,CDISC001,1,ALB,Albumin,CHEMISTRY,3.9,g/dL,3.5,4.6,39,39,g/L,35,46,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
2,CDISCPILOT01,LB,CDISC001,2,ALP,"Alkaline Phosphatase",CHEMISTRY,93,U/L,35,115,93,93,U/L,35,115,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
3,CDISCPILOT01,LB,CDISC001,3,ALT,"Alanine Aminotransferase",CHEMISTRY,18,U/L,6,35,18,18,U/L,6,35,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
4,CDISCPILOT01,LB,CDISC001,4,AST,"Aspartate Aminotransferase",CHEMISTRY,26,U/L,11,36,26,26,U/L,11,36,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
5,CDISCPILOT01,LB,CDISC001,5,BASO,Basophils,HEMATOLOGY,0.03,10^9/L,0,0.2,0.03,0.03,10^9/L,0,0.2,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
6,CDISCPILOT01,LB,CDISC001,6,BILI,Bilirubin,CHEMISTRY,0.5,mg/dL,0.2,1.2,8.55,8.55,umol/L,3,21,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
7,CDISCPILOT01,LB,CDISC001,7,UREAN,"Urea Nitrogen",CHEMISTRY,21,mg/dL,4,24,7.497,7.497,mmol/L,1.4,8.6,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
8,CDISCPILOT01,LB,CDISC001,8,CA,Calcium,CHEMISTRY,9.1,mg/dL,8.4,10.3,2.27045,2.27045,mmol/L,2.1,2.57,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
9,CDISCPILOT01,LB,CDISC001,9,CHOL,Cholesterol,CHEMISTRY,254,mg/dL,149,286,6.56844,6.56844,mmol/L,3.85,7.4,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
10,CDISCPILOT01,LB,CDISC001,10,CK,"Creatine Kinase",CHEMISTRY,79,U/L,22,198,79,79,U/L,22,198,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
11,CDISCPILOT01,LB,CDISC001,11,CL,Chloride,CHEMISTRY,102,mEq/L,94,112,102,102,mmol/L,94,112,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
12,CDISCPILOT01,LB,CDISC001,12,COLOR,Color,URINALYSIS,NORMAL,,,,NORMAL,,,,,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
13,CDISCPILOT01,LB,CDISC001,13,CREAT,Creatinine,CHEMISTRY,1.3,mg/dL,0.8,1.6,114.92,114.92,umol/L,71,141,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
14,CDISCPILOT01,LB,CDISC001,14,EOS,Eosinophils,HEMATOLOGY,0.08,10^9/L,0,0.57,0.08,0.08,10^9/L,0,0.57,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
15,CDISCPILOT01,LB,CDISC001,15,GGT,"Gamma Glutamyl Transferase",CHEMISTRY,31,U/L,10,50,31,31,U/L,10,50,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
16,CDISCPILOT01,LB,CDISC001,16,GLUC,Glucose,CHEMISTRY,74,mg/dL,50,250,4.10774,4.10774,mmol/L,2.8,13.9,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
17,CDISCPILOT01,LB,CDISC001,17,HCT,Hematocrit,HEMATOLOGY,46.0,%,37,51,0.46,0.46,,0.37,0.51,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
18,CDISCPILOT01,LB,CDISC001,18,HGB,Hemoglobin,HEMATOLOGY,14.9,g/dL,12.5,17,9.24694,9.24694,mmol/L,7.76,10.55,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
19,CDISCPILOT01,LB,CDISC001,19,K,Potassium,CHEMISTRY,4.3,mEq/L,3.4,5.4,4.3,4.3,mmol/L,3.4,5.4,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
20,CDISCPILOT01,LB,CDISC001,20,KETONES,Ketones,URINALYSIS,0,,,,0,0,,,,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
21,CDISCPILOT01,LB,CDISC001,21,LYM,Lymphocytes,HEMATOLOGY,1.20,10^9/L,0.8,3,1.2,1.2,10^9/L,0.8,3,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
22,CDISCPILOT01,LB,CDISC001,22,MCH,"Ery. Mean Corpuscular Hemoglobin",HEMATOLOGY,30,pg,26,34,1.8618,1.8618,fmol,1.6,2.1,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
23,CDISCPILOT01,LB,CDISC001,23,MCHC,"Ery. Mean Corpuscular HGB Concentration",HEMATOLOGY,33,g/dL,31,38,20.4798,20.4798,mmol/L,19,24,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
24,CDISCPILOT01,LB,CDISC001,24,MCV,"Ery. Mean Corpuscular Volume",HEMATOLOGY,92,fL,80,100,92,92,fL,80,100,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
25,CDISCPILOT01,LB,CDISC001,25,MONO,Monocytes,HEMATOLOGY,0.38,10^9/L,0.12,0.92,0.38,0.38,10^9/L,0.12,0.92,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
26,CDISCPILOT01,LB,CDISC001,26,SODIUM,Sodium,CHEMISTRY,139,mEq/L,135,145,139,139,mmol/L,135,145,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
27,CDISCPILOT01,LB,CDISC001,27,PH,pH,URINALYSIS,5,,5,8,5,5,,5,8,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
28,CDISCPILOT01,LB,CDISC001,28,PHOS,Phosphate,CHEMISTRY,3.4,mg/dL,2.2,5.1,1.09786,1.09786,mmol/L,0.71,1.65,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
29,CDISCPILOT01,LB,CDISC001,29,PLAT,Platelets,HEMATOLOGY,150,10^9/L,130,394,150,150,10^9/L,130,394,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
30,CDISCPILOT01,LB,CDISC001,30,PROT,Protein,CHEMISTRY,6.7,g/dL,6,8,67,67,g/L,60,80,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
31,CDISCPILOT01,LB,CDISC001,31,RBC,Erythrocytes,HEMATOLOGY,5.00,10^12/L,4,5.8,5,5,10^12/L,4,5.8,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
32,CDISCPILOT01,LB,CDISC001,32,SPGRAV,"Specific Gravity",URINALYSIS,1.017,,1.006,1.03,1.017,1.017,,1.006,1.03,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
33,CDISCPILOT01,LB,CDISC001,33,TSH,Thyrotropin,OTHER,2.63,mIU/L,.32,5,2.63,2.63,mU/L,0.32,5,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
34,CDISCPILOT01,LB,CDISC001,34,URATE,Urate,CHEMISTRY,4.9,mg/dL,2.5,7.5,291.452,291.452,umol/L,149,446,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
35,CDISCPILOT01,LB,CDISC001,35,UROBIL,Urobilinogen,URINALYSIS,0,,,,0,0,,,,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
36,CDISCPILOT01,LB,CDISC001,36,VITB12,"Vitamin B12",OTHER,641,ng/L,200,900,472.9298,472.9298,pmol/L,148,664,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
37,CDISCPILOT01,LB,CDISC001,37,WBC,Leukocytes,HEMATOLOGY,4.99,10^9/L,3.8,10.7,4.99,4.99,10^9/L,3.8,10.7,NORMAL,Y,1,"SCREENING 1",SCREENING,2012-11-23T11:20,-7
38,CDISCPILOT01,LB,CDISC001,38,ALB,Albumin,CHEMISTRY,3.6,g/dL,3.5,4.6,36,36,g/L,35,46,NORMAL,,4,"WEEK 2",TREATMENT,2012-12-13T14:05,14
```

Basic validation

During loading, one will have noticed two "progress bars", one showing the progress of the loading of the data, and the other one showing the progress of the basic validation that is always performed. This primary validation of the data is really very basic, e.g. checking whether "required" variables are populated. For example, when the column AEDECOD (Dictionary-Derived Term for the adverse event), which is a "required" variable, is not populated, the cells are colored red, and a tooltip is provided displaying the cause of the

[3] The safest is always to use the vertical bar ("pipe" character) as the field delimiter, as it is a character that seldomly appears in the table values, but it might be that the receiving application, e.g. a worksheet, is not capable to accept that.

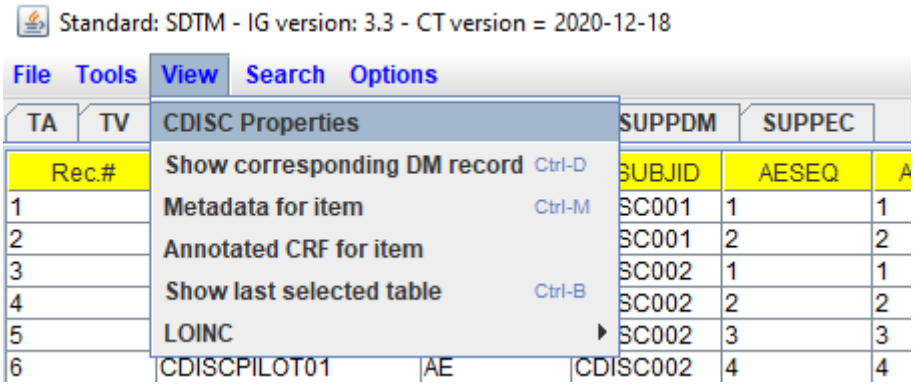
issue:



Further validation can be done by setting some options, by clicking the "Options" button in the main window, and by using the CDISC CORE (CDISC Open Rules Engine) features[4] of the Smart Submission Dataset Viewer.

Search and View

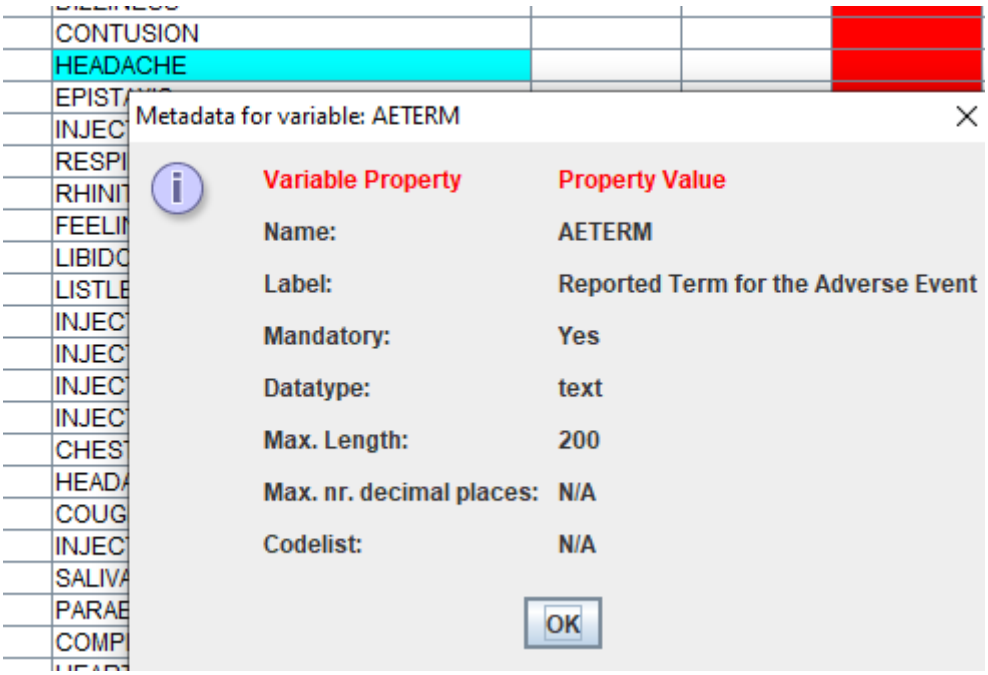
In the "tables view", there are two other menus that we did not explain so far, but are easy to understand. When using the menu "View", one can choose between:
- CDISC properties:
- Show corresponding DM record (Ctrl-D): this was already explained before
- Metadata for item
- Annotated CRF for item
- Show last selected table (Ctrl-B): allows to "toggle" between tables
- LOINC => ... : this will be explained separately later



View - CDISC properties: displays which IG was used and which version(s), and which version(s) of the CDISC controlled terminology were used. This information is retrieved from the define.xml file.

View - Metadata for item: when a single cell has been selected, and this menu is used, a dialog with metadata information is displayed. For example when "headache" in the AETERM was selected, the resulting information is:

[4] At the moment of writing (December 2024), the CORE engine is being adapted to accept Dataset-JSON 1.1 as being the input format. CORE was already able to work with Dataset-JSON 1.0, but as we are now moving to Dataset-JSON, some adaptions are necessary.

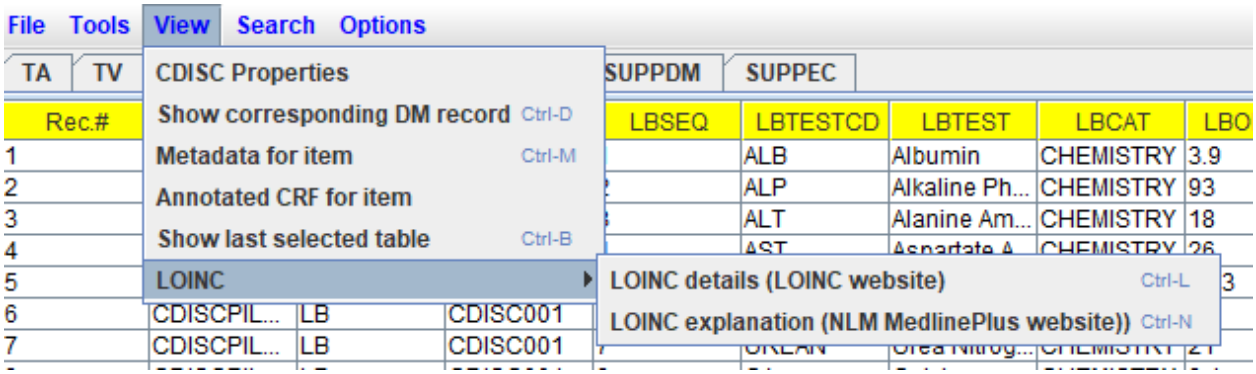


Also this information is retrieved from the define.xml

View - Annotated CRF for item: The goal of this feature is that when a single cell is selected, and the the link to the annotated CRF is present in the define.xml, the annotated CRF is opened in the default PDF viewer on the correct page, or, when there are several pages for the variable, on the first of these.
This feature has not been implemented yet - it is planned for 2025.

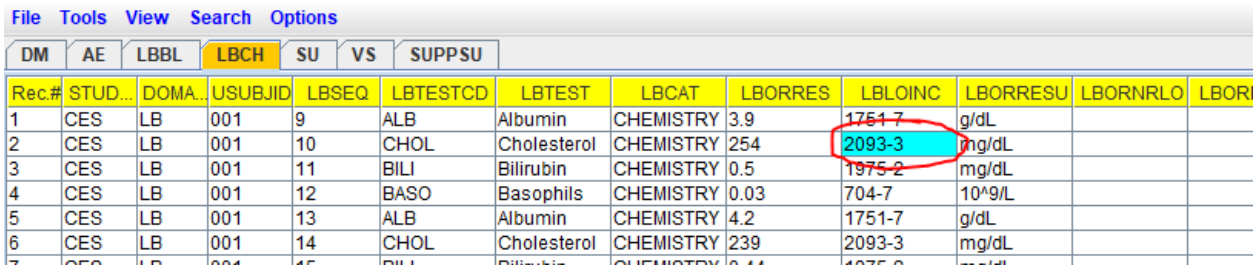
The next one "View - LOINC" is a bit special, but important due to the growing importance of LOINC as the only real identifier of the test, not only for lab tests, but also for vital signs, microbiology, virology, questionnaires, ECG, etc.[5].

When selecting the menu "View - LOINC", we get additional possibilities:

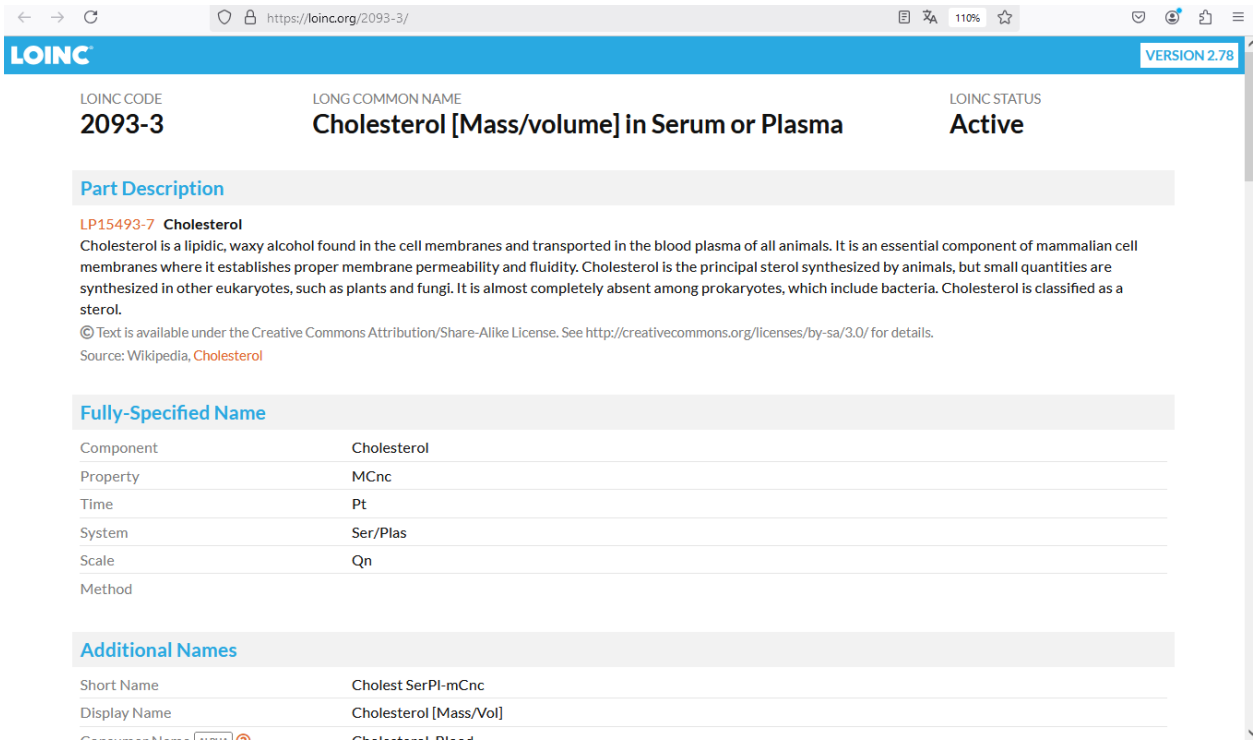


For example, when having selected the code "2093-3" in the LBLOINC column:

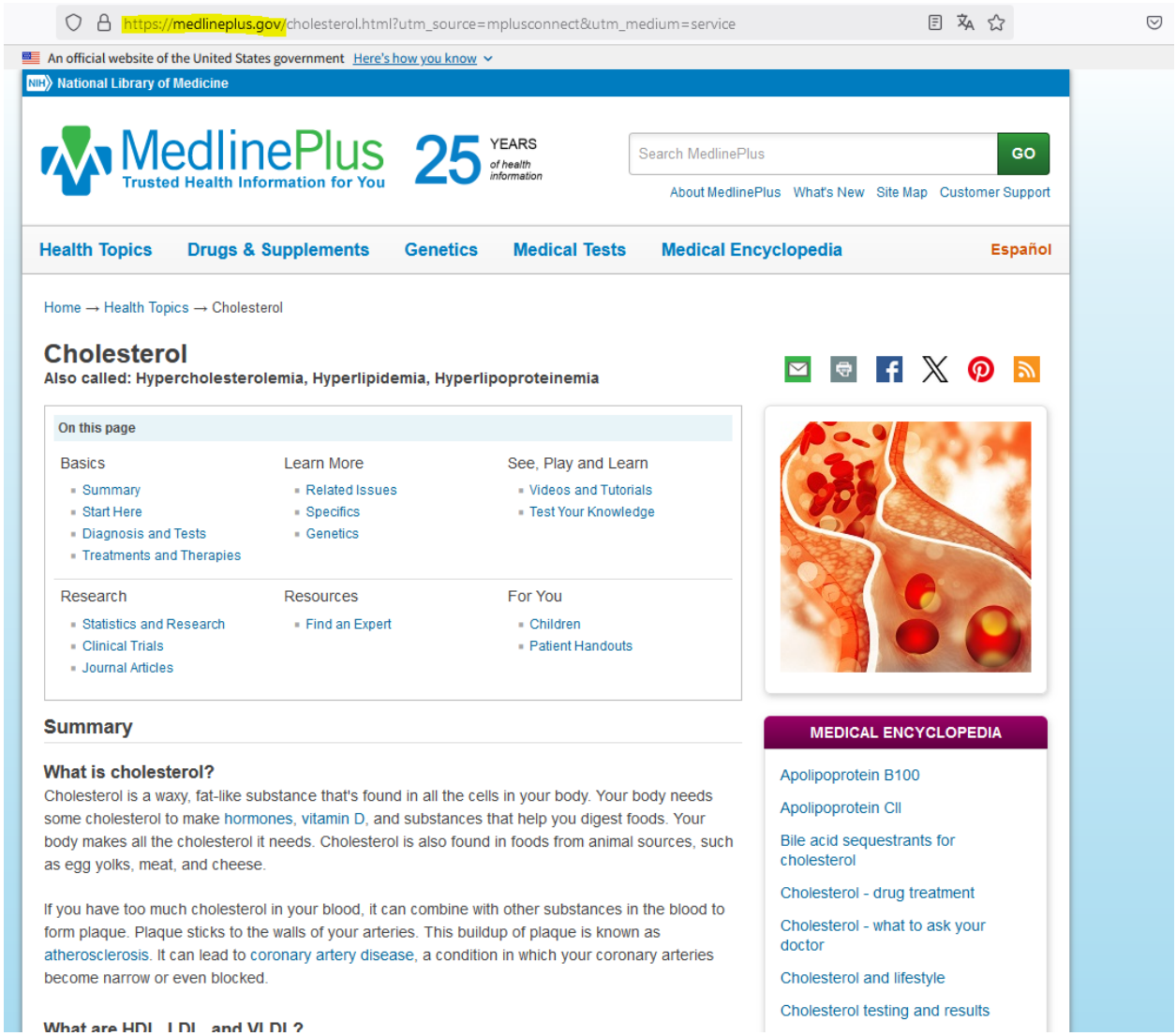
[5] CDISC controlled terminology, as a "post-coordinated" system is not capable to uniquely identify tests, also not when combining values from variables such as LBTESTCD, LBSPEC, LBMETHOD, … LOINC, as a "pre-coordinated" system makes this "uniqueness" possible. This is also the reason why LOINC is heavily used in healthcare, and is mandated by law in many countries (the number growing) for ordering lab tests and reporting lab results.



and using the menu "View - LOINC - LOINC details", this will trigger a process that opens the LOINC website page for this specific LOINC code in the user's default browser, i.e.:

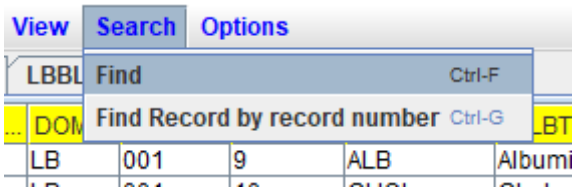


similarly, when the menu "View - LOINC - LOINC explanation" is used, the page for that code from the NLM (National Library of Medicine) is opened in the user's default browser, i.e.:

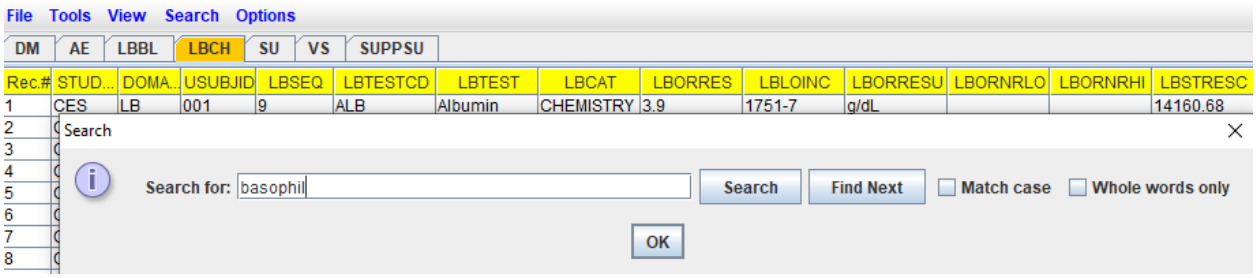


IMPORTANT REMARK: Essentially, this is an implementation of RWS (RESTful Web Services). The potential for such features is enormous! Using the mechanism, just by selecting the LOINC code, it would become possible to perform things such as literature searches, connect to specialist technical literature and other documentation, and also connect directly connect to specialist (in medicine or in clinical research) AI systems. Implementing this in the Smart Submission Dataset Viewer will often only require to add a few lines in the source code … the only limitation being … lack of creativity.

Another menu is "Search", allowing to quickly find a word, part of a word or value in the table, or to quickly jump to a specific record (by record number) in the selected table:



For the former (Search - Find), a separate dialog is then popping up:



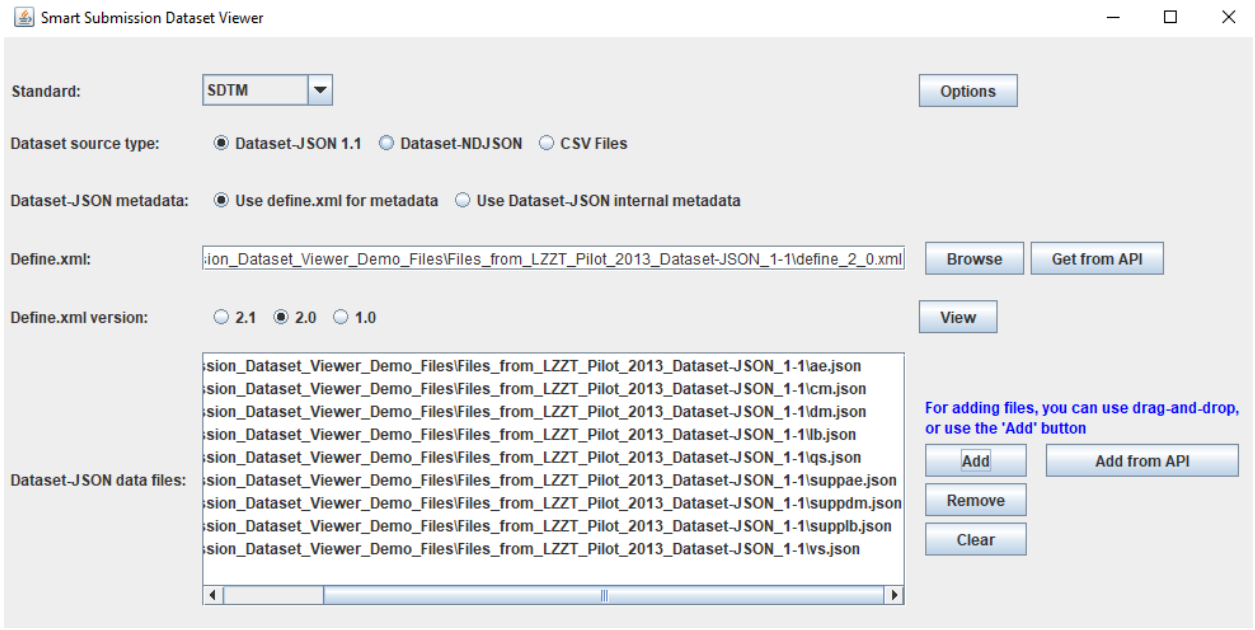
which is pretty self-explaining.

Filtering before loading

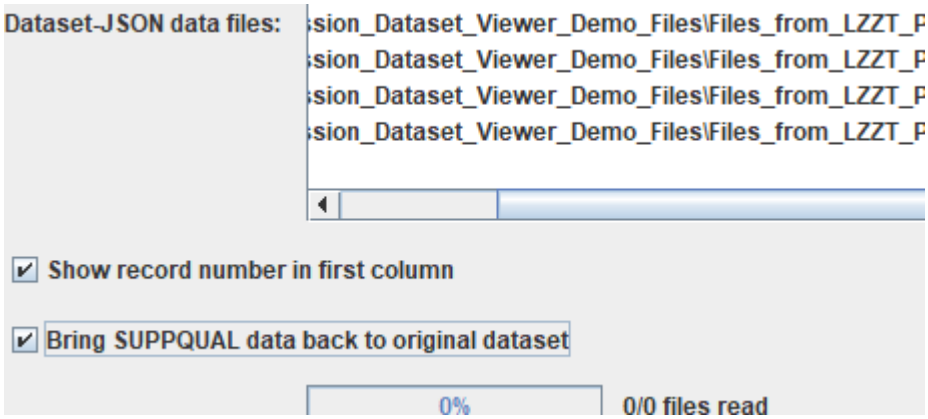
When one has large datasets (thousand or even millions of records), it is very hard for reviewers to keep oversight. Of course one can filter after the dataset has been loaded, but in such cases, classic viewers such as the "SAS Universal Viewer" need to do "paging" in order to save memory. With "paging", any filters are only applied to a single page, which can easily lead to information loss. For example, when a reviewer only wants see the Albumin lab test results of a very large dataset, with the "SAS Universal Viewer" he/she can apply the filter on the first loaded page, and needs to repeat the whole process when wanting to see the Albumin values for subjects on other pages.

The Smart Submission Dataset Viewer uses a smarter method. Before loading, one can already apply a filter, and only those records that pass the filter are really loaded. This not only saves a lot of memory, but also ensures that all the records that pass the filter will be displayed.
In the Smart Submission Dataset Viewer, filtering is usually based on the --TESTCD or --CAT ("test code" and "category") variables for "Findings" datasets. In the future, we also want to apply the same principle for "Events" datasets (e.g. based on --DECOD), and for "Intervention" datasets, e.g. based on --TRT or --CAT or --INDC. We need input from the users about which variables are the most useful ones to do pre-filtering for these classes.
To find out which e.g. test codes or categories are available, a define.xml must be present and a codelist must be attached to the given variable. This is usually already the case for the "Findings" datasets.

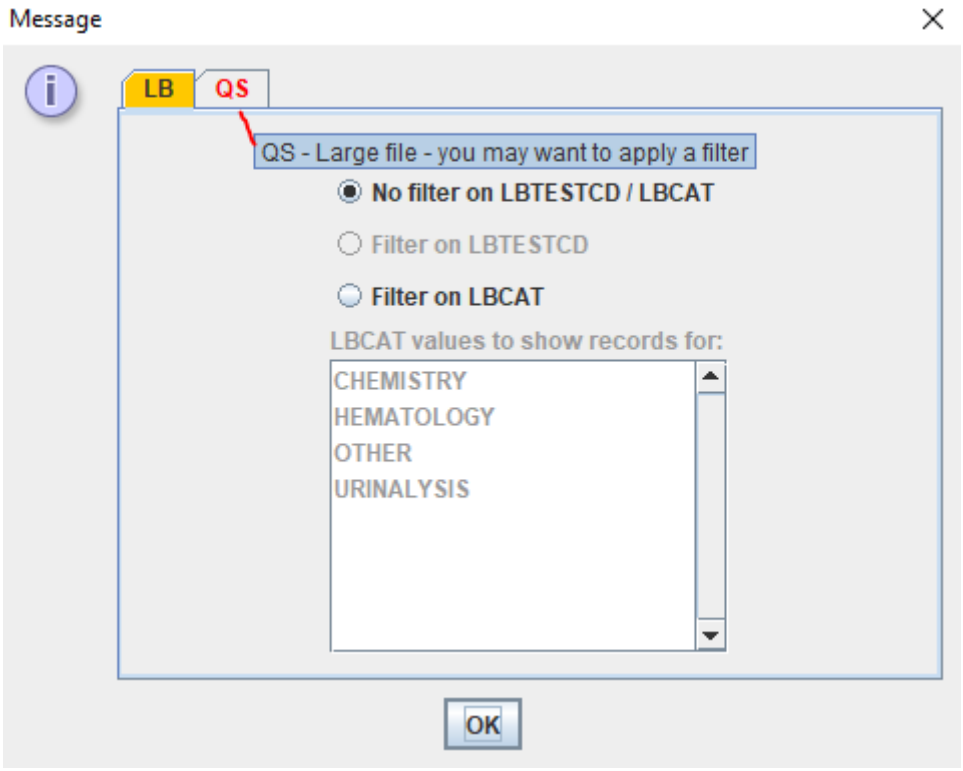
The system always offer the user to do pre-filtering on specific datasets, when there is a codelist available from the define.xml, and urge the user to do pre-filtering based on the file sizes, the default being 20MB (this threshold can be changed by the user, see later). For example, in our case, the LB.json dataset has a size of 0.7MB (which is much less than for the SAS-XPT corresponding dataset, which has 2.7MB), but still has almost 3500 records. The LB dataset from the LZZT-pilot is 10.4MB in size (the corresponding LB.xpt is 66MB), and has almost 60,000 records, which means that in such a case, pre-filtering may very well make sense, as one cannot imagine how a reviewer can handle such an amount of records just by visual inspection.
For demonstrating the pre-filtering, we take the SDTM Dataset-JSON files of the LZZT pilot, and e.g. select:



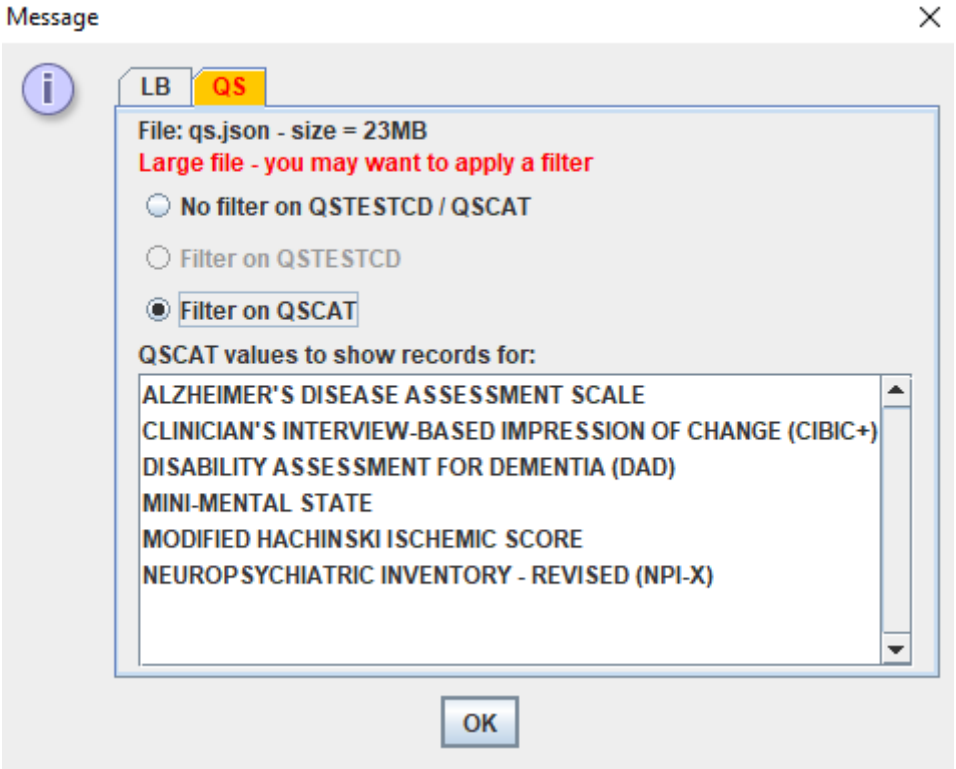
the QS dataset having the largest file size (23MB) followed by the LB dataset (10MB).
We also load the SUPPLB dataset (8MB) and the SUPPAE and SUPPDM datasets, and ask the system to "Bring SUPPQUAL data back":



When then clicking "Start", first the define.xml is loaded, and then the system shows us the following dialog:



The system found codelists for LBCAT in the define.xml, so offers to filter on it, and strongly suggests to do pre-filtering for QS, for which it presents:

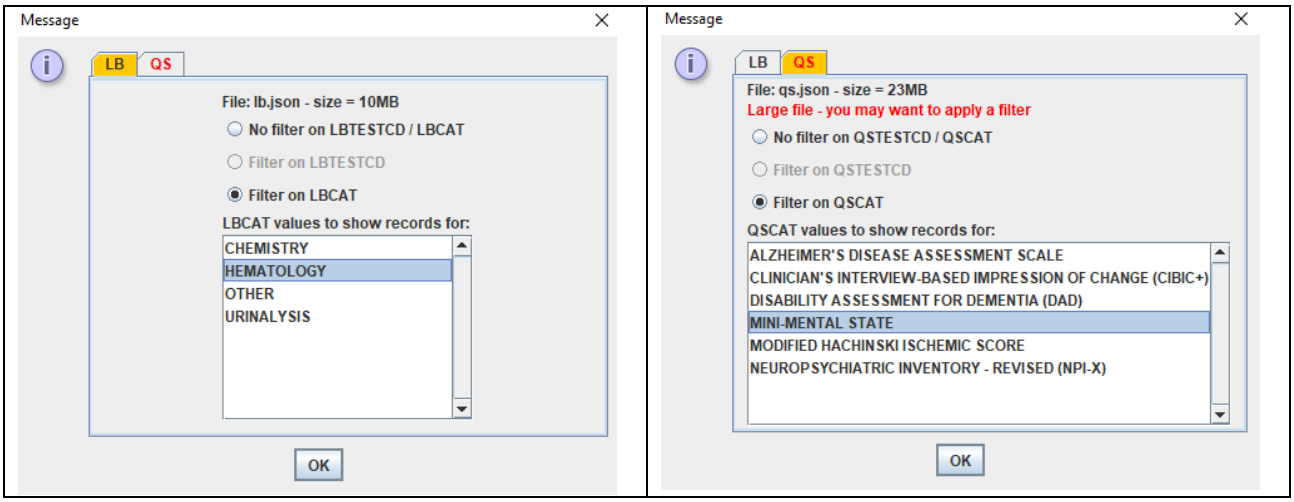


as it found a codelist for QSCAT, but none for QSTESTCD.

When we leave everything "as is", i.e. do not any pre-filtering, it takes 18 minutes to load, especially "recombining" SUPPLB with LB taking a lot of time.
Anyway, visually inspecting 60,000 LB records doesn't make sense anyway isn't it?

So, let us be smarter, and do pre-filtering.

For LB, we pre-filter on "hematology" only, and for QS on "Mini-mental state":



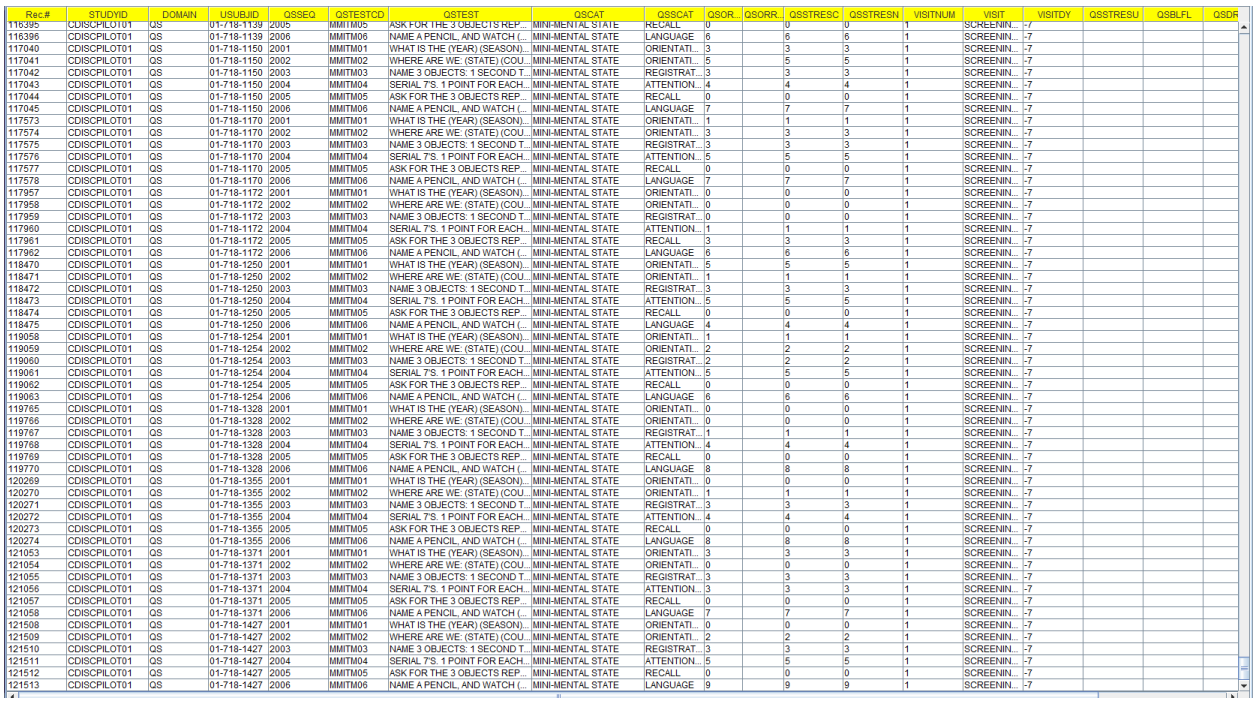
and start loading again. Loading the selection now takes 7 minutes, with the result for LB being:

Rec.#	STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	LBORRES	LBORRE...	LB...	LBO...	LBST...	LBST...	LBST...	LBST...	LBST...	LBNRIND	LBBLFL	VISITNUM	VISIT	VISITDY	LBDTC	LBDY	ENDPOINT	LBTMSHI
31	CDISCPILOT01	LB	01-701-1015	4	ANISO	Anisocytes	HEMATOLOGY	1	NO UNITS			1	1				ABNORMAL	Y	1	SCREENIN...	-7	2013-12-2...	-7		
42	CDISCPILOT01	LB	01-701-1015	6	BASO	Basophils	HEMATOLOGY	0.05	THOU/uL	0	0.2	0.05	0.05	GI/L	0	0.2	NORMAL	Y	1	SCREENIN...	-7	2013-12-2...			
43	CDISCPILOT01	LB	01-701-1015	43	BASO	Basophils	HEMATOLOGY	0.03	THOU/uL	0	0.2	0.03	0.03	GI/L	0	0.2	NORMAL		4	WEEK 2	14	2014-01-1...			
44	CDISCPILOT01	LB	01-701-1015	78	BASO	Basophils	HEMATOLOGY	0.02	THOU/uL	0	0.2	0.02	0.02	GI/L	0	0.2	NORMAL		5	WEEK 4	28	2014-01-3...			
45	CDISCPILOT01	LB	01-701-1015	108	BASO	Basophils	HEMATOLOGY	0.04	THOU/uL	0	0.2	0.04	0.04	GI/L	0	0.2	NORMAL		7	WEEK 6	42	2014-02-1...			
46	CDISCPILOT01	LB	01-701-1015	138	BASO	Basophils	HEMATOLOGY	0.05	THOU/uL	0	0.2	0.05	0.05	GI/L	0	0.2	NORMAL		8	WEEK 8	56	2014-03-0...			
47	CDISCPILOT01	LB	01-701-1015	168	BASO	Basophils	HEMATOLOGY	0.04	THOU/uL	0	0.2	0.04	0.04	GI/L	0	0.2	NORMAL		9	WEEK 12	84	2014-03-2...			
48	CDISCPILOT01	LB	01-701-1015	203	BASO	Basophils	HEMATOLOGY	0.03	THOU/uL	0	0.2	0.03	0.03	GI/L	0	0.2	NORMAL		10	WEEK 16	112	2014-05-0...			
49	CDISCPILOT01	LB	01-701-1015	233	BASO	Basophils	HEMATOLOGY	0.03	THOU/uL	0	0.2	0.03	0.03	GI/L	0	0.2	NORMAL		11	WEEK 20	140	2014-05-2...	140		0.2
50	CDISCPILOT01	LB	01-701-1015	263	BASO	Basophils	HEMATOLOGY	0.04	THOU/uL	0	0.2	0.04	0.04	GI/L	0	0.2	NORMAL		12	WEEK 24	168	2014-06-1...	168	Y	0.2
51	CDISCPILOT01	LB	01-701-1015	298	BASO	Basophils	HEMATOLOGY	0.03	THOU/uL	0	0.2	0.03	0.03	GI/L	0	0.2	NORMAL		13	WEEK 26	182	2014-07-0...	182		0.2
126	CDISCPILOT01	LB	01-701-1015	15	EOS	Eosinophils	HEMATOLOGY	0.13	THOU/uL	0	0.57	0.13	0.13	GI/L	0	0.57	NORMAL	Y	1	SCREENIN...	-7	2013-12-2...	-7		0.2
127	CDISCPILOT01	LB	01-701-1015	52	EOS	Eosinophils	HEMATOLOGY	0.08	THOU/uL	0	0.57	0.08	0.08	GI/L	0	0.57	NORMAL		4	WEEK 2	14	2014-01-1...	15		0.1
128	CDISCPILOT01	LB	01-701-1015	86	EOS	Eosinophils	HEMATOLOGY	0.08	THOU/uL	0	0.57	0.08	0.08	GI/L	0	0.57	NORMAL		5	WEEK 4	28	2014-01-3...	29		0.1
129	CDISCPILOT01	LB	01-701-1015	116	EOS	Eosinophils	HEMATOLOGY	0.08	THOU/uL	0	0.57	0.08	0.08	GI/L	0	0.57	NORMAL		7	WEEK 6	42	2014-02-1...	42		0.1
130	CDISCPILOT01	LB	01-701-1015	146	EOS	Eosinophils	HEMATOLOGY	0.09	THOU/uL	0	0.57	0.09	0.09	GI/L	0	0.57	NORMAL		8	WEEK 8	56	2014-03-0...	63		0.2
131	CDISCPILOT01	LB	01-701-1015	177	EOS	Eosinophils	HEMATOLOGY	0.13	THOU/uL	0	0.57	0.13	0.13	GI/L	0	0.57	NORMAL		9	WEEK 12	84	2014-03-2...	84		0.2
132	CDISCPILOT01	LB	01-701-1015	211	EOS	Eosinophils	HEMATOLOGY	0.09	THOU/uL	0	0.57	0.09	0.09	GI/L	0	0.57	NORMAL		10	WEEK 16	112	2014-05-0...	126		0.2
133	CDISCPILOT01	LB	01-701-1015	241	EOS	Eosinophils	HEMATOLOGY	0.06	THOU/uL	0	0.57	0.06	0.06	GI/L	0	0.57	NORMAL		11	WEEK 20	140	2014-05-2...	140		0.1
134	CDISCPILOT01	LB	01-701-1015	272	EOS	Eosinophils	HEMATOLOGY	0.09	THOU/uL	0	0.57	0.09	0.09	GI/L	0	0.57	NORMAL		12	WEEK 24	168	2014-06-1...	168	Y	0.2
135	CDISCPILOT01	LB	01-701-1015	306	EOS	Eosinophils	HEMATOLOGY	0.07	THOU/uL	0	0.57	0.07	0.07	GI/L	0	0.57	NORMAL		13	WEEK 26	182	2014-07-0...	182		0.1
156	CDISCPILOT01	LB	01-701-1015	18	HCT	Hematocrit	HEMATOLOGY	42.0	%	34	48	0.42	0.42	1	0.34	0.48	NORMAL	Y	1	SCREENIN...	-7	2013-12-2...	-7		0.9
157	CDISCPILOT01	LB	01-701-1015	55	HCT	Hematocrit	HEMATOLOGY	43.0	%	34	48	0.43	0.43	1	0.34	0.48	NORMAL		4	WEEK 2	14	2014-01-1...	15		0.9
158	CDISCPILOT01	LB	01-701-1015	89	HCT	Hematocrit	HEMATOLOGY	39.0	%	34	48	0.39	0.39	1	0.34	0.48	NORMAL		5	WEEK 4	28	2014-01-3...	29		0.8
159	CDISCPILOT01	LB	01-701-1015	119	HCT	Hematocrit	HEMATOLOGY	40.0	%	34	48	0.4	0.4	1	0.34	0.48	NORMAL		7	WEEK 6	42	2014-02-1...	42		0.8
160	CDISCPILOT01	LB	01-701-1015	149	HCT	Hematocrit	HEMATOLOGY	42.0	%	34	48	0.42	0.42	1	0.34	0.48	NORMAL		8	WEEK 8	56	2014-03-0...	63		0.9
161	CDISCPILOT01	LB	01-701-1015	180	HCT	Hematocrit	HEMATOLOGY	41.0	%	34	48	0.41	0.41	1	0.34	0.48	NORMAL		9	WEEK 12	84	2014-03-2...	84		0.9
162	CDISCPILOT01	LB	01-701-1015	214	HCT	Hematocrit	HEMATOLOGY	41.0	%	34	48	0.41	0.41	1	0.34	0.48	NORMAL		10	WEEK 16	112	2014-05-0...	126		0.9
163	CDISCPILOT01	LB	01-701-1015	244	HCT	Hematocrit	HEMATOLOGY	40.0	%	34	48	0.4	0.4	1	0.34	0.48	NORMAL		11	WEEK 20	140	2014-05-2...	140		0.8
164	CDISCPILOT01	LB	01-701-1015	275	HCT	Hematocrit	HEMATOLOGY	44.0	%	34	48	0.44	0.44	1	0.34	0.48	NORMAL		12	WEEK 24	168	2014-06-1...	168	Y	0.9
165	CDISCPILOT01	LB	01-701-1015	309	HCT	Hematocrit	HEMATOLOGY	45.0	%	34	48	0.45	0.45	1	0.34	0.48	NORMAL		13	WEEK 26	182	2014-07-0...	182		0.9
166	CDISCPILOT01	LB	01-701-1015	19	HGB	Hemoglobin	HEMATOLOGY	14.3	g/dL	11.5	15.8	8.874...	8.874...	mmo...	7.14	9.81	NORMAL	Y	1	SCREENIN...	-7	2013-12-2...	-7		0.9
167	CDISCPILOT01	LB	01-701-1015	56	HGB	Hemoglobin	HEMATOLOGY	14.3	g/dL	11.5	15.8	8.874...	8.874...	mmo...	7.14	9.81	NORMAL		4	WEEK 2	14	2014-01-1...	15		0.9
168	CDISCPILOT01	LB	01-701-1015	90	HGB	Hemoglobin	HEMATOLOGY	13.4	g/dL	11.5	15.8	8.316...	8.316...	mmo...	7.14	9.81	NORMAL		5	WEEK 4	28	2014-01-3...	29		0.8
169	CDISCPILOT01	LB	01-701-1015	120	HGB	Hemoglobin	HEMATOLOGY	13.8	g/dL	11.5	15.8	8.564...	8.564...	mmo...	7.14	9.81	NORMAL		7	WEEK 6	42	2014-02-1...	42		0.9
170	CDISCPILOT01	LB	01-701-1015	150	HGB	Hemoglobin	HEMATOLOGY	14.0	g/dL	11.5	15.8	8.6884	8.6884	mmo...	7.14	9.81	NORMAL		8	WEEK 8	56	2014-03-0...	63		0.9
171	CDISCPILOT01	LB	01-701-1015	181	HGB	Hemoglobin	HEMATOLOGY	13.4	g/dL	11.5	15.8	8.316...	8.316...	mmo...	7.14	9.81	NORMAL		9	WEEK 12	84	2014-03-2...	84		0.8
172	CDISCPILOT01	LB	01-701-1015	215	HGB	Hemoglobin	HEMATOLOGY	14.1	g/dL	11.5	15.8	8.750...	8.750...	mmo...	7.14	9.81	NORMAL		10	WEEK 16	112	2014-05-0...	126		0.9
173	CDISCPILOT01	LB	01-701-1015	245	HGB	Hemoglobin	HEMATOLOGY	13.7	g/dL	11.5	15.8	8.502...	8.502...	mmo...	7.14	9.81	NORMAL		11	WEEK 20	140	2014-05-2...	140		0.9
174	CDISCPILOT01	LB	01-701-1015	276	HGB	Hemoglobin	HEMATOLOGY	14.0	g/dL	11.5	15.8	8.6884	8.6884	mmo...	7.14	9.81	NORMAL		12	WEEK 24	168	2014-06-1...	168	Y	0.9
175	CDISCPILOT01	LB	01-701-1015	310	HGB	Hemoglobin	HEMATOLOGY	14.3	g/dL	11.5	15.8	8.874...	8.874...	mmo...	7.14	9.81	NORMAL		13	WEEK 26	182	2014-07-0...	182		0.9
190	CDISCPILOT01	LB	01-701-1015	22	LYM	Lymphocytes	HEMATOLOGY	2.11	THOU/uL	0.8	3	2.11	2.11	GI/L	0.8	3	NORMAL	Y	1	SCREENIN...	-7	2013-12-2...	-7		0.7
191	CDISCPILOT01	LB	01-701-1015	59	LYM	Lymphocytes	HEMATOLOGY	1.98	THOU/uL	0.8	3	1.98	1.98	GI/L	0.8	3	NORMAL		4	WEEK 2	14	2014-01-1...	15		0.7

Define-XML metadata:
Name: LBTMSHI
Label: LAB RESULT/UPPER LIMIT OF NORMAL
Mandatory: No
Datatype: float
Length: 8
Non-Standard Variable (NSV)

Remark that the record numbers are the same as if all the records were loaded.

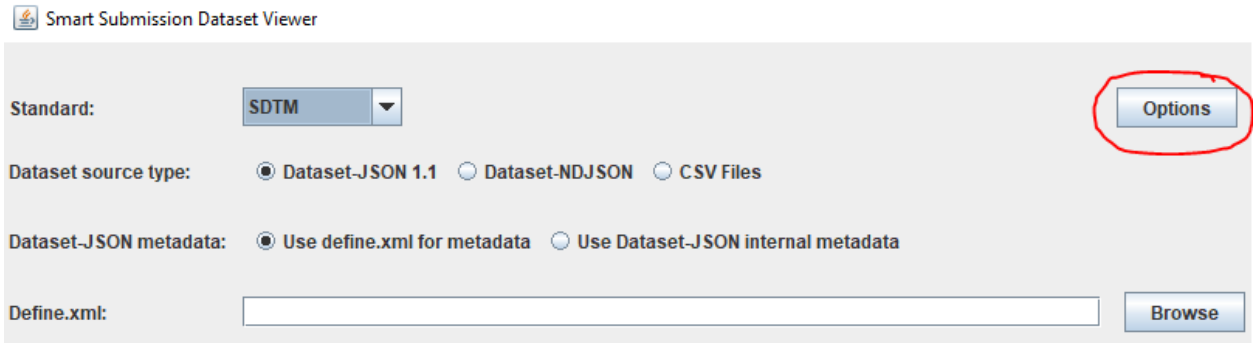
For QS, we find:



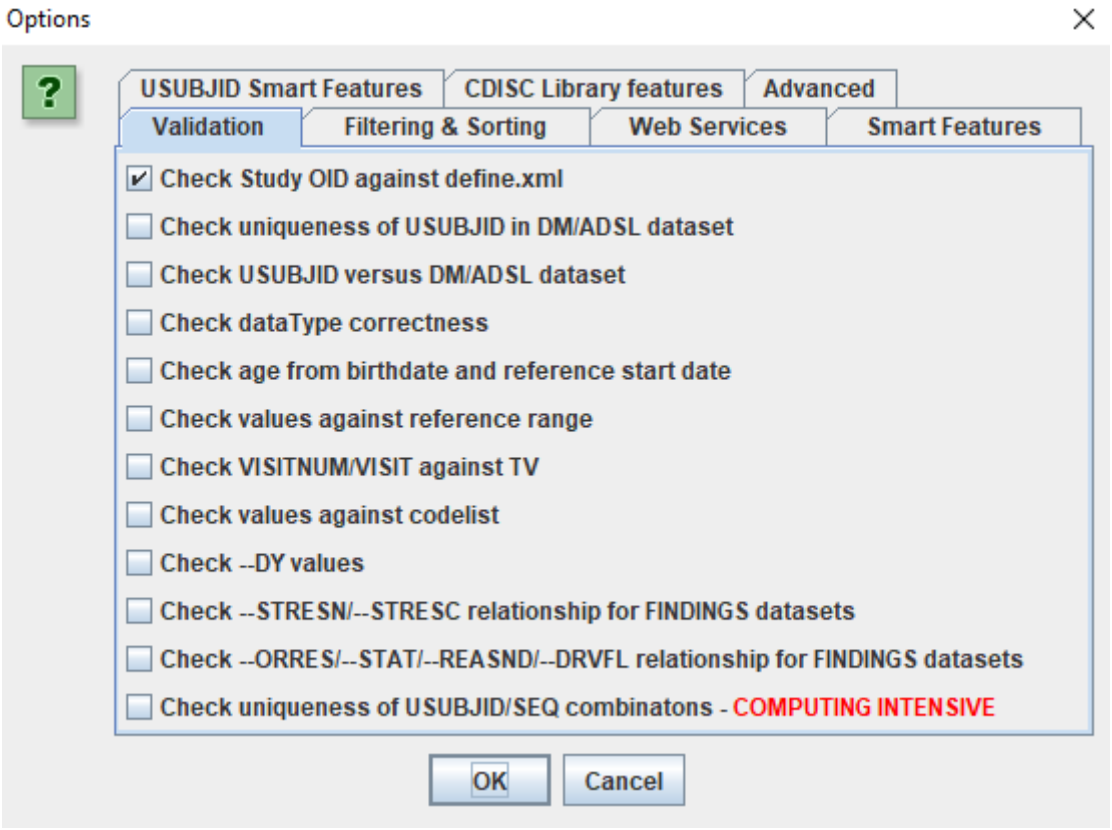
Rec.#	STUDYID	DOMAIN	USUBJID	QSSEQ	QSTESTCD	QSTEST	QSCAT	QSSCAT	QSOR...	QSORR...	QSSTRESC	QSSTRESN	VISITNUM	VISIT	VISITDY	QSSTRESU	QSBLFL	QSDR
116395	CDISCPILOT01	QS	01-718-1139	2005	MMITM05	ASK FOR THE 3 OBJECTS REP...	MINI-MENTAL STATE	RECALL	0		0	0	1	SCREENIN...	-7			
116396	CDISCPILOT01	QS	01-718-1139	2006	MMITM06	NAME A PENCIL, AND WATCH (...	MINI-MENTAL STATE	LANGUAGE	6		6	6	1	SCREENIN...	-7			
117040	CDISCPILOT01	QS	01-718-1150	2001	MMITM01	WHAT IS THE (YEAR) (SEASON)...	MINI-MENTAL STATE	ORIENTATI...	3		3	3	1	SCREENIN...	-7			
117041	CDISCPILOT01	QS	01-718-1150	2002	MMITM02	WHERE ARE WE: (STATE) (COU...	MINI-MENTAL STATE	ORIENTATI...	5		5	5	1	SCREENIN...	-7			
117042	CDISCPILOT01	QS	01-718-1150	2003	MMITM03	NAME 3 OBJECTS: 1 SECOND T...	MINI-MENTAL STATE	REGISTRAT...	3		3	3	1	SCREENIN...	-7			
117043	CDISCPILOT01	QS	01-718-1150	2004	MMITM04	SERIAL 7'S. 1 POINT FOR EACH...	MINI-MENTAL STATE	ATTENTION...	4		4	4	1	SCREENIN...	-7			
117044	CDISCPILOT01	QS	01-718-1150	2005	MMITM05	ASK FOR THE 3 OBJECTS REP...	MINI-MENTAL STATE	RECALL	0		0	0	1	SCREENIN...	-7			
117045	CDISCPILOT01	QS	01-718-1150	2006	MMITM06	NAME A PENCIL, AND WATCH (...	MINI-MENTAL STATE	LANGUAGE	7		7	7	1	SCREENIN...	-7			
117573	CDISCPILOT01	QS	01-718-1170	2001	MMITM01	WHAT IS THE (YEAR) (SEASON)...	MINI-MENTAL STATE	ORIENTATI...	1		1	1	1	SCREENIN...	-7			
117574	CDISCPILOT01	QS	01-718-1170	2002	MMITM02	WHERE ARE WE: (STATE) (COU...	MINI-MENTAL STATE	ORIENTATI...	3		3	3	1	SCREENIN...	-7			
117575	CDISCPILOT01	QS	01-718-1170	2003	MMITM03	NAME 3 OBJECTS: 1 SECOND T...	MINI-MENTAL STATE	REGISTRAT...	3		3	3	1	SCREENIN...	-7			
117576	CDISCPILOT01	QS	01-718-1170	2004	MMITM04	SERIAL 7'S. 1 POINT FOR EACH...	MINI-MENTAL STATE	ATTENTION...	5		5	5	1	SCREENIN...	-7			
117577	CDISCPILOT01	QS	01-718-1170	2005	MMITM05	ASK FOR THE 3 OBJECTS REP...	MINI-MENTAL STATE	RECALL	0		0	0	1	SCREENIN...	-7			
117578	CDISCPILOT01	QS	01-718-1170	2006	MMITM06	NAME A PENCIL, AND WATCH (...	MINI-MENTAL STATE	LANGUAGE	7		7	7	1	SCREENIN...	-7			
117957	CDISCPILOT01	QS	01-718-1172	2001	MMITM01	WHAT IS THE (YEAR) (SEASON)...	MINI-MENTAL STATE	ORIENTATI...	0		0	0	1	SCREENIN...	-7			
117958	CDISCPILOT01	QS	01-718-1172	2002	MMITM02	WHERE ARE WE: (STATE) (COU...	MINI-MENTAL STATE	ORIENTATI...	0		0	0	1	SCREENIN...	-7			
117959	CDISCPILOT01	QS	01-718-1172	2003	MMITM03	NAME 3 OBJECTS: 1 SECOND T...	MINI-MENTAL STATE	REGISTRAT...	0		0	0	1	SCREENIN...	-7			
117960	CDISCPILOT01	QS	01-718-1172	2004	MMITM04	SERIAL 7'S. 1 POINT FOR EACH...	MINI-MENTAL STATE	ATTENTION...	1		1	1	1	SCREENIN...	-7			
117961	CDISCPILOT01	QS	01-718-1172	2005	MMITM05	ASK FOR THE 3 OBJECTS REP...	MINI-MENTAL STATE	RECALL	3		3	3	1	SCREENIN...	-7			
117962	CDISCPILOT01	QS	01-718-1172	2006	MMITM06	NAME A PENCIL, AND WATCH (...	MINI-MENTAL STATE	LANGUAGE	6		6	6	1	SCREENIN...	-7			
118470	CDISCPILOT01	QS	01-718-1250	2001	MMITM01	WHAT IS THE (YEAR) (SEASON)...	MINI-MENTAL STATE	ORIENTATI...	5		5	5	1	SCREENIN...	-7			
118471	CDISCPILOT01	QS	01-718-1250	2002	MMITM02	WHERE ARE WE: (STATE) (COU...	MINI-MENTAL STATE	ORIENTATI...	1		1	1	1	SCREENIN...	-7			
118472	CDISCPILOT01	QS	01-718-1250	2003	MMITM03	NAME 3 OBJECTS: 1 SECOND T...	MINI-MENTAL STATE	REGISTRAT...	3		3	3	1	SCREENIN...	-7			
118473	CDISCPILOT01	QS	01-718-1250	2004	MMITM04	SERIAL 7'S. 1 POINT FOR EACH...	MINI-MENTAL STATE	ATTENTION...	5		5	5	1	SCREENIN...	-7			
118474	CDISCPILOT01	QS	01-718-1250	2005	MMITM05	ASK FOR THE 3 OBJECTS REP...	MINI-MENTAL STATE	RECALL	0		0	0	1	SCREENIN...	-7			
118475	CDISCPILOT01	QS	01-718-1250	2006	MMITM06	NAME A PENCIL, AND WATCH (...	MINI-MENTAL STATE	LANGUAGE	4		4	4	1	SCREENIN...	-7			
119058	CDISCPILOT01	QS	01-718-1254	2001	MMITM01	WHAT IS THE (YEAR) (SEASON)...	MINI-MENTAL STATE	ORIENTATI...	1		1	1	1	SCREENIN...	-7			
119059	CDISCPILOT01	QS	01-718-1254	2002	MMITM02	WHERE ARE WE: (STATE) (COU...	MINI-MENTAL STATE	ORIENTATI...	2		2	2	1	SCREENIN...	-7			
119060	CDISCPILOT01	QS	01-718-1254	2003	MMITM03	NAME 3 OBJECTS: 1 SECOND T...	MINI-MENTAL STATE	REGISTRAT...	2		2	2	1	SCREENIN...	-7			
119061	CDISCPILOT01	QS	01-718-1254	2004	MMITM04	SERIAL 7'S. 1 POINT FOR EACH...	MINI-MENTAL STATE	ATTENTION...	5		5	5	1	SCREENIN...	-7			
119062	CDISCPILOT01	QS	01-718-1254	2005	MMITM05	ASK FOR THE 3 OBJECTS REP...	MINI-MENTAL STATE	RECALL	0		0	0	1	SCREENIN...	-7			
119063	CDISCPILOT01	QS	01-718-1254	2006	MMITM06	NAME A PENCIL, AND WATCH (...	MINI-MENTAL STATE	LANGUAGE	6		6	6	1	SCREENIN...	-7			
119765	CDISCPILOT01	QS	01-718-1328	2001	MMITM01	WHAT IS THE (YEAR) (SEASON)...	MINI-MENTAL STATE	ORIENTATI...	0		0	0	1	SCREENIN...	-7			
119766	CDISCPILOT01	QS	01-718-1328	2002	MMITM02	WHERE ARE WE: (STATE) (COU...	MINI-MENTAL STATE	ORIENTATI...	0		0	0	1	SCREENIN...	-7			
119767	CDISCPILOT01	QS	01-718-1328	2003	MMITM03	NAME 3 OBJECTS: 1 SECOND T...	MINI-MENTAL STATE	REGISTRAT...	1		1	1	1	SCREENIN...	-7			
119768	CDISCPILOT01	QS	01-718-1328	2004	MMITM04	SERIAL 7'S. 1 POINT FOR EACH...	MINI-MENTAL STATE	ATTENTION...	4		4	4	1	SCREENIN...	-7			
119769	CDISCPILOT01	QS	01-718-1328	2005	MMITM05	ASK FOR THE 3 OBJECTS REP...	MINI-MENTAL STATE	RECALL	0		0	0	1	SCREENIN...	-7			
119770	CDISCPILOT01	QS	01-718-1328	2006	MMITM06	NAME A PENCIL, AND WATCH (...	MINI-MENTAL STATE	LANGUAGE	8		8	8	1	SCREENIN...	-7			
120269	CDISCPILOT01	QS	01-718-1355	2001	MMITM01	WHAT IS THE (YEAR) (SEASON)...	MINI-MENTAL STATE	ORIENTATI...	0		0	0	1	SCREENIN...	-7			
120270	CDISCPILOT01	QS	01-718-1355	2002	MMITM02	WHERE ARE WE: (STATE) (COU...	MINI-MENTAL STATE	ORIENTATI...	1		1	1	1	SCREENIN...	-7			
120271	CDISCPILOT01	QS	01-718-1355	2003	MMITM03	NAME 3 OBJECTS: 1 SECOND T...	MINI-MENTAL STATE	REGISTRAT...	3		3	3	1	SCREENIN...	-7			
120272	CDISCPILOT01	QS	01-718-1355	2004	MMITM04	SERIAL 7'S. 1 POINT FOR EACH...	MINI-MENTAL STATE	ATTENTION...	4		4	4	1	SCREENIN...	-7			
120273	CDISCPILOT01	QS	01-718-1355	2005	MMITM05	ASK FOR THE 3 OBJECTS REP...	MINI-MENTAL STATE	RECALL	0		0	0	1	SCREENIN...	-7			
120274	CDISCPILOT01	QS	01-718-1355	2006	MMITM06	NAME A PENCIL, AND WATCH (...	MINI-MENTAL STATE	LANGUAGE	8		8	8	1	SCREENIN...	-7			
121053	CDISCPILOT01	QS	01-718-1371	2001	MMITM01	WHAT IS THE (YEAR) (SEASON)...	MINI-MENTAL STATE	ORIENTATI...	3		3	3	1	SCREENIN...	-7			
121054	CDISCPILOT01	QS	01-718-1371	2002	MMITM02	WHERE ARE WE: (STATE) (COU...	MINI-MENTAL STATE	ORIENTATI...	0		0	0	1	SCREENIN...	-7			
121055	CDISCPILOT01	QS	01-718-1371	2003	MMITM03	NAME 3 OBJECTS: 1 SECOND T...	MINI-MENTAL STATE	REGISTRAT...	3		3	3	1	SCREENIN...	-7			
121056	CDISCPILOT01	QS	01-718-1371	2004	MMITM04	SERIAL 7'S. 1 POINT FOR EACH...	MINI-MENTAL STATE	ATTENTION...	3		3	3	1	SCREENIN...	-7			
121057	CDISCPILOT01	QS	01-718-1371	2005	MMITM05	ASK FOR THE 3 OBJECTS REP...	MINI-MENTAL STATE	RECALL	0		0	0	1	SCREENIN...	-7			
121058	CDISCPILOT01	QS	01-718-1371	2006	MMITM06	NAME A PENCIL, AND WATCH (...	MINI-MENTAL STATE	LANGUAGE	7		7	7	1	SCREENIN...	-7			
121508	CDISCPILOT01	QS	01-718-1427	2001	MMITM01	WHAT IS THE (YEAR) (SEASON)...	MINI-MENTAL STATE	ORIENTATI...	0		0	0	1	SCREENIN...	-7			
121509	CDISCPILOT01	QS	01-718-1427	2002	MMITM02	WHERE ARE WE: (STATE) (COU...	MINI-MENTAL STATE	ORIENTATI...	2		2	2	1	SCREENIN...	-7			
121510	CDISCPILOT01	QS	01-718-1427	2003	MMITM03	NAME 3 OBJECTS: 1 SECOND T...	MINI-MENTAL STATE	REGISTRAT...	3		3	3	1	SCREENIN...	-7			
121511	CDISCPILOT01	QS	01-718-1427	2004	MMITM04	SERIAL 7'S. 1 POINT FOR EACH...	MINI-MENTAL STATE	ATTENTION...	5		5	5	1	SCREENIN...	-7			
121512	CDISCPILOT01	QS	01-718-1427	2005	MMITM05	ASK FOR THE 3 OBJECTS REP...	MINI-MENTAL STATE	RECALL	0		0	0	1	SCREENIN...	-7			
121513	CDISCPILOT01	QS	01-718-1427	2006	MMITM06	NAME A PENCIL, AND WATCH (...	MINI-MENTAL STATE	LANGUAGE	9		9	9	1	SCREENIN...	-7			

Setting Viewer Options

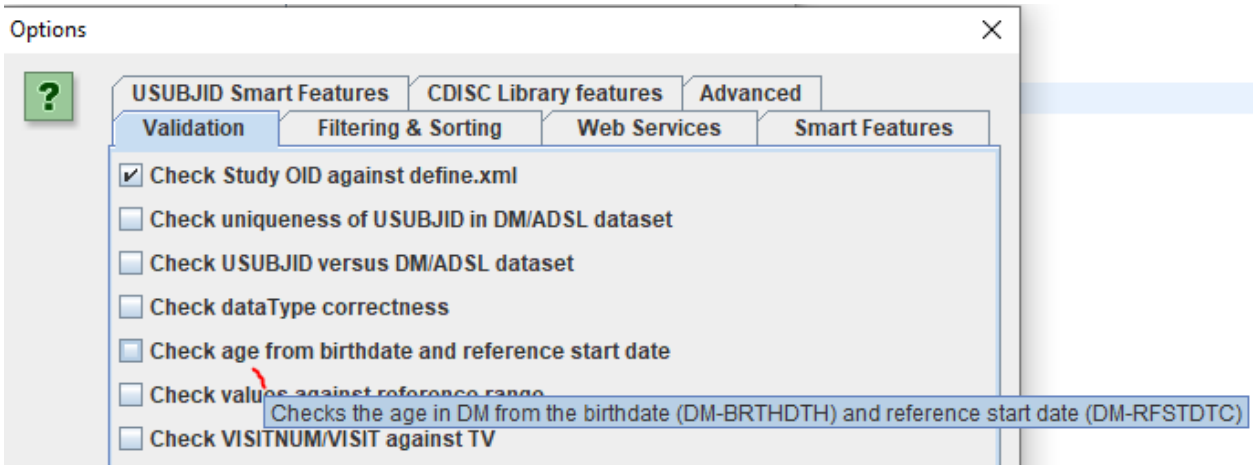
In the main window, one also sees an "Options" button on the right-top side:



When clicked, the following window with different tabs is shown:

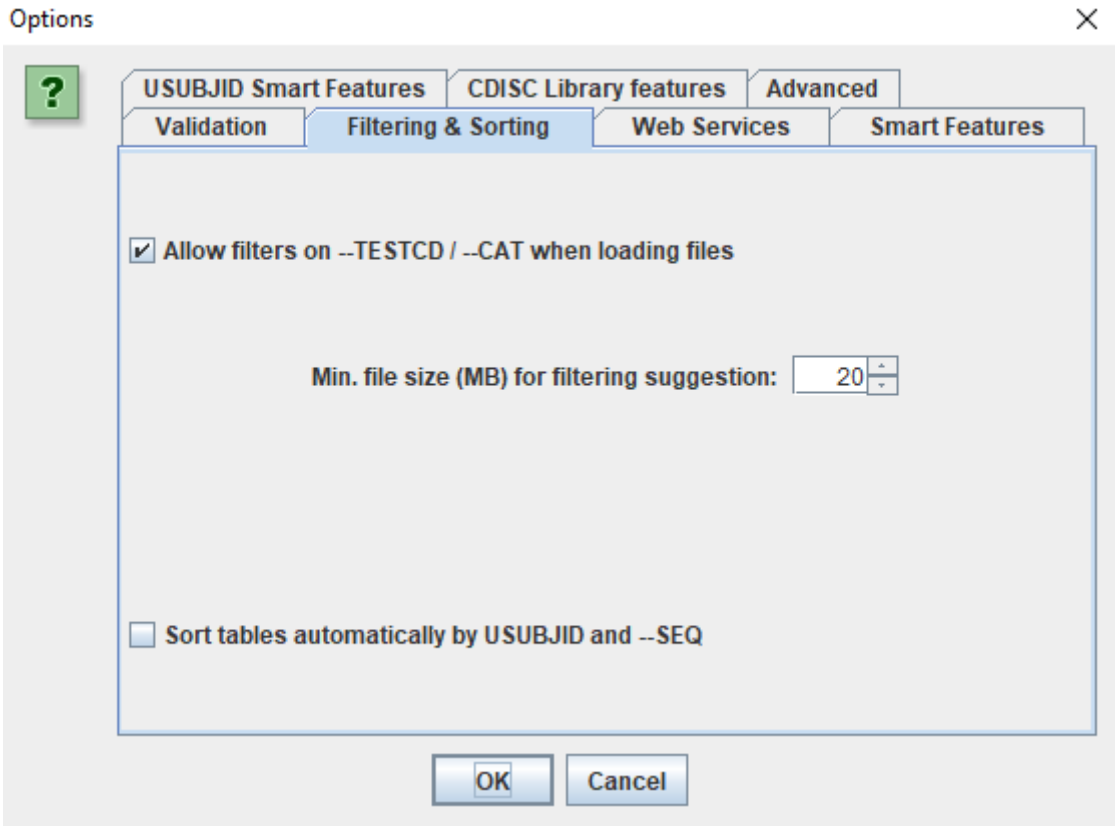


The options for the first tab "Validation" allows to switch on/off certain basic validation options. Most of them are obvious, but one can always hold the mouse over each of them to get more information. For example:



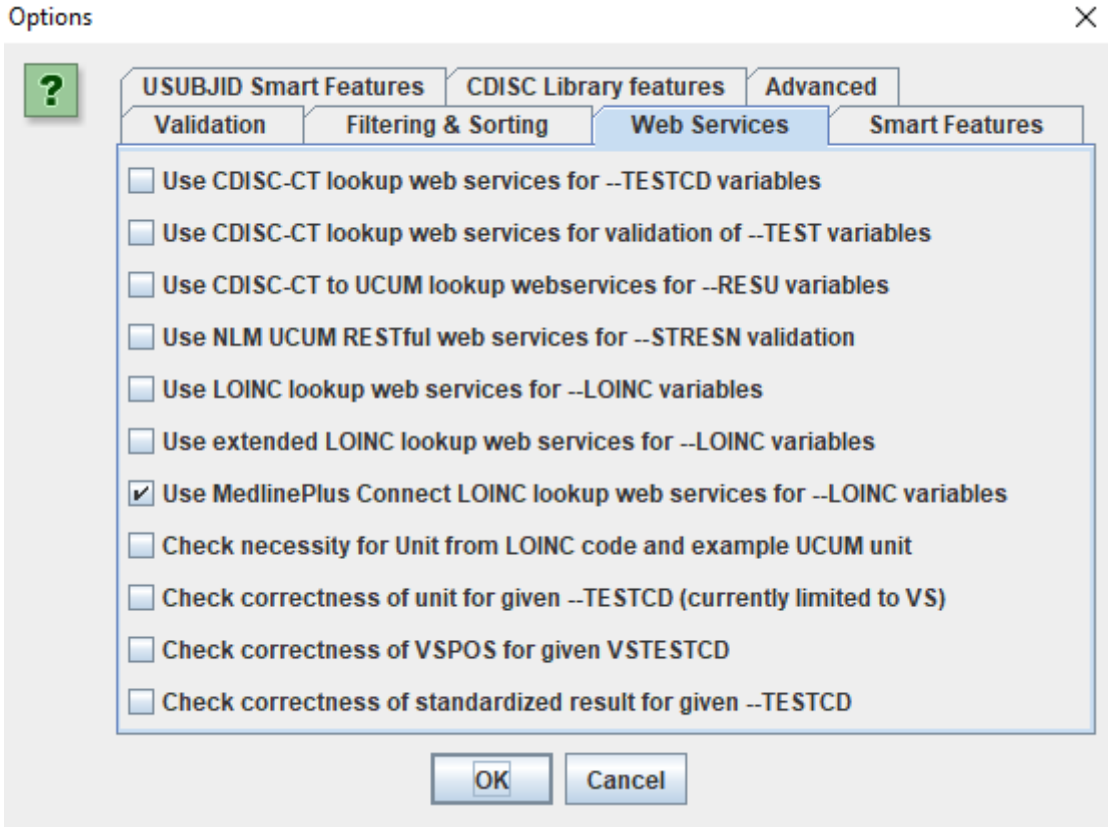
We will later, in the section "Validation" show how these can be used.

A second important tab is the "Filtering and Sorting" tab:

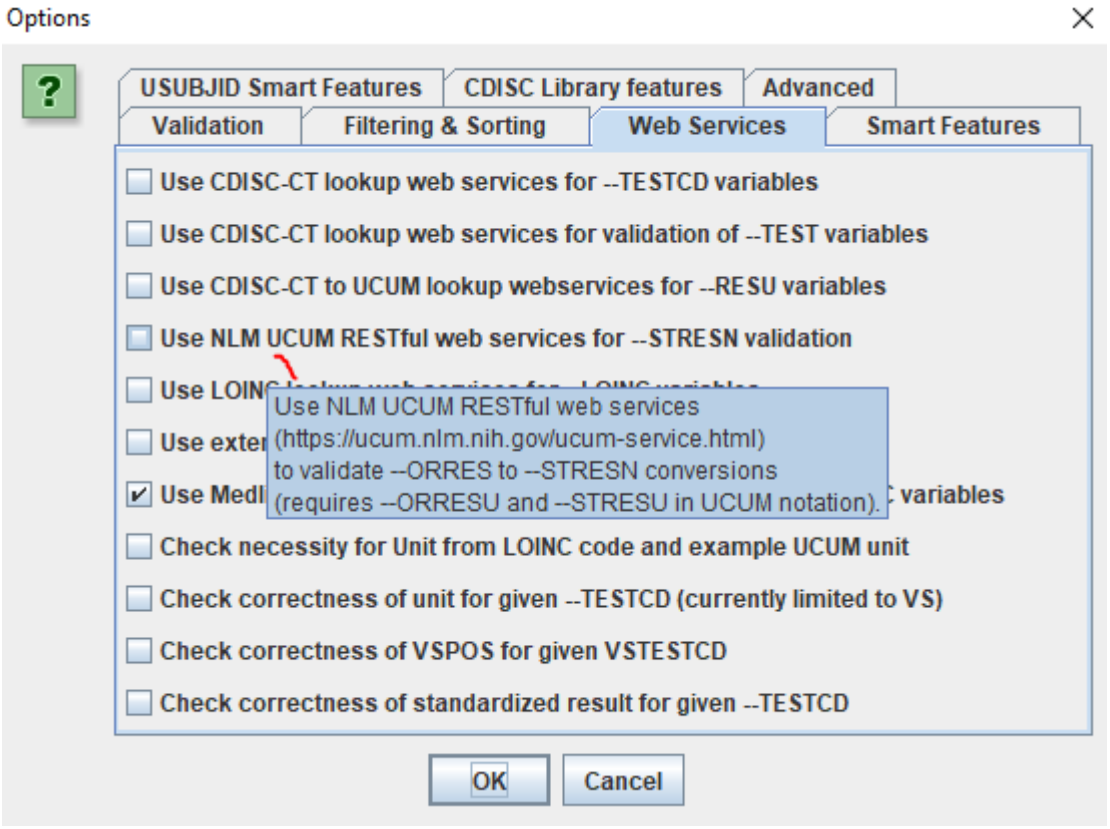


It allows the user to specify whether the system will propose to add "pre-loading filters" and if so, from what dataset size, the system will urge the user to apply a filter - see the section "Filtering before loading".

The next tab is about "Web Services":



The Smart Submission Dataset Viewer is unique in the ability to use (RESTful) web services to aid in validation. Also here, hovering the mouse over an item shows more details about the feature, e.g.:

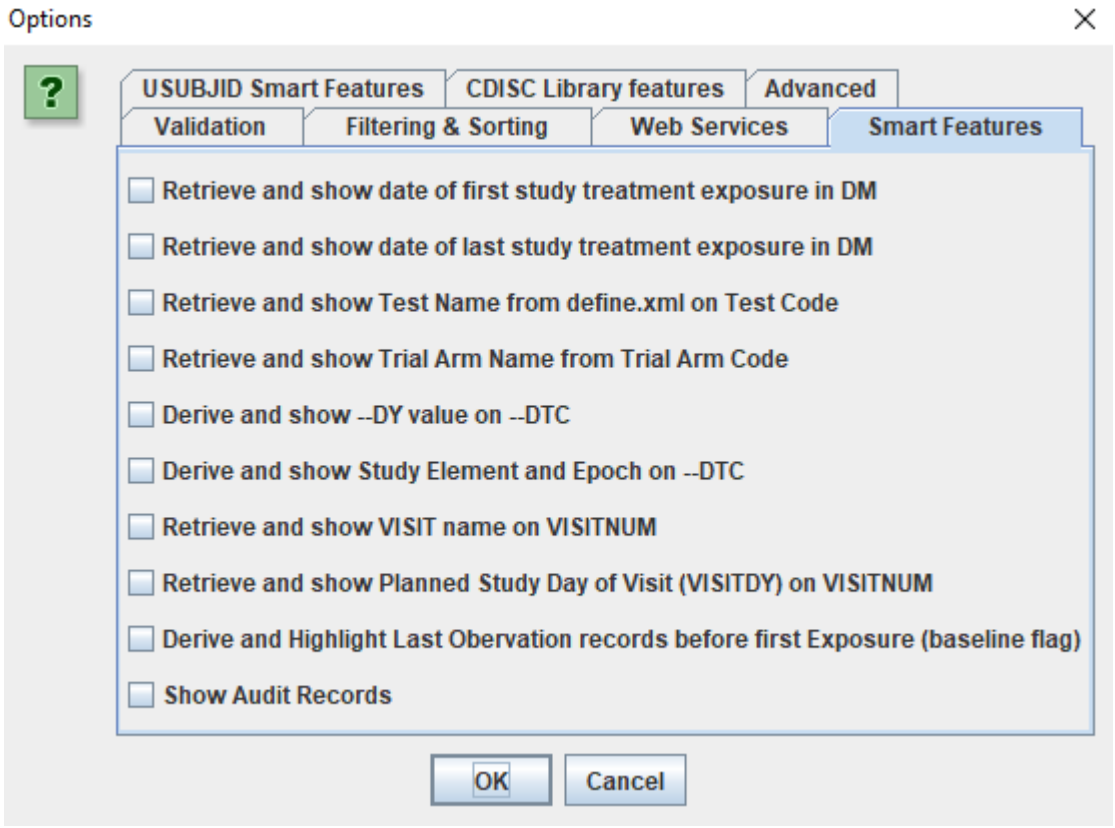


This feature allowing to use the NLM (National Library of Medicine) RESTful web services for automated validation whether the conversion the --ORRES values to the --STRESN values have been done correctly.

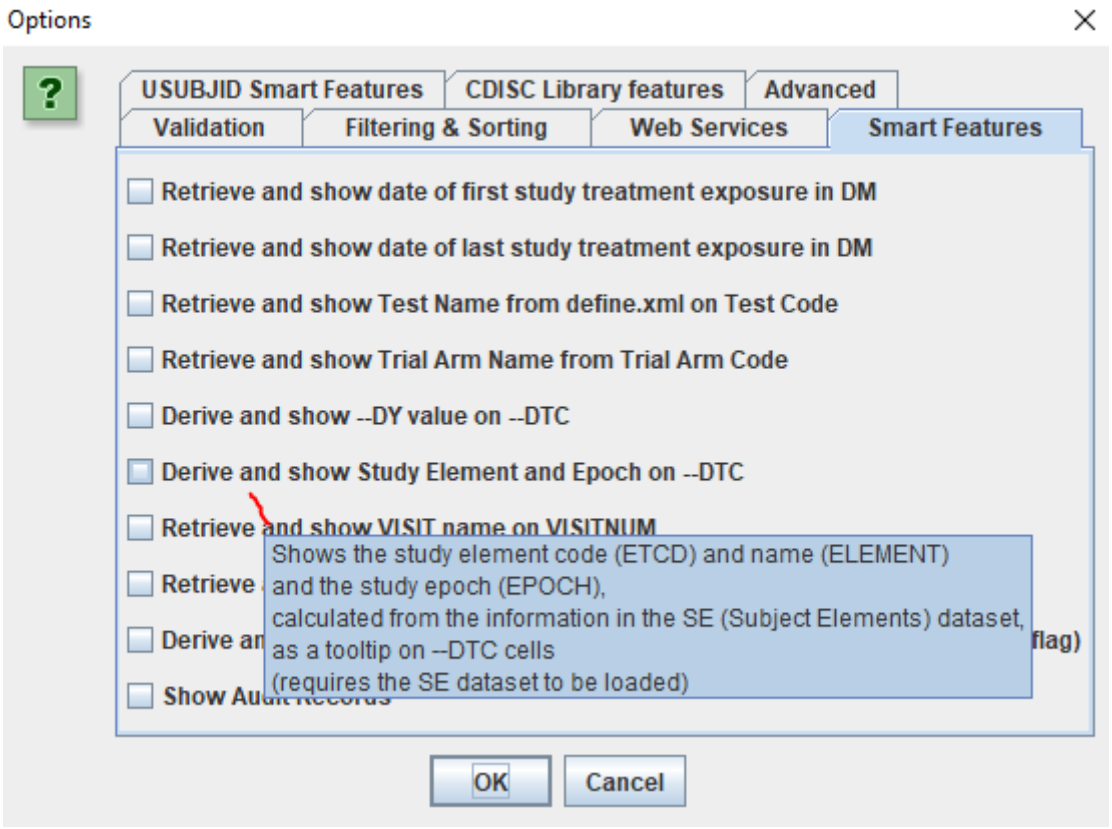
Some of these web service validation features even go far beyond what is available in CDISC CORE (see later) and especially what is offered by Pinnacle21 validation.

More explanation can be found in the separate document "Using RESTful web services (NLM, FDA, XML4Pharma)". Also this document will be updated in the near future.

The next tab is "Smart Features":



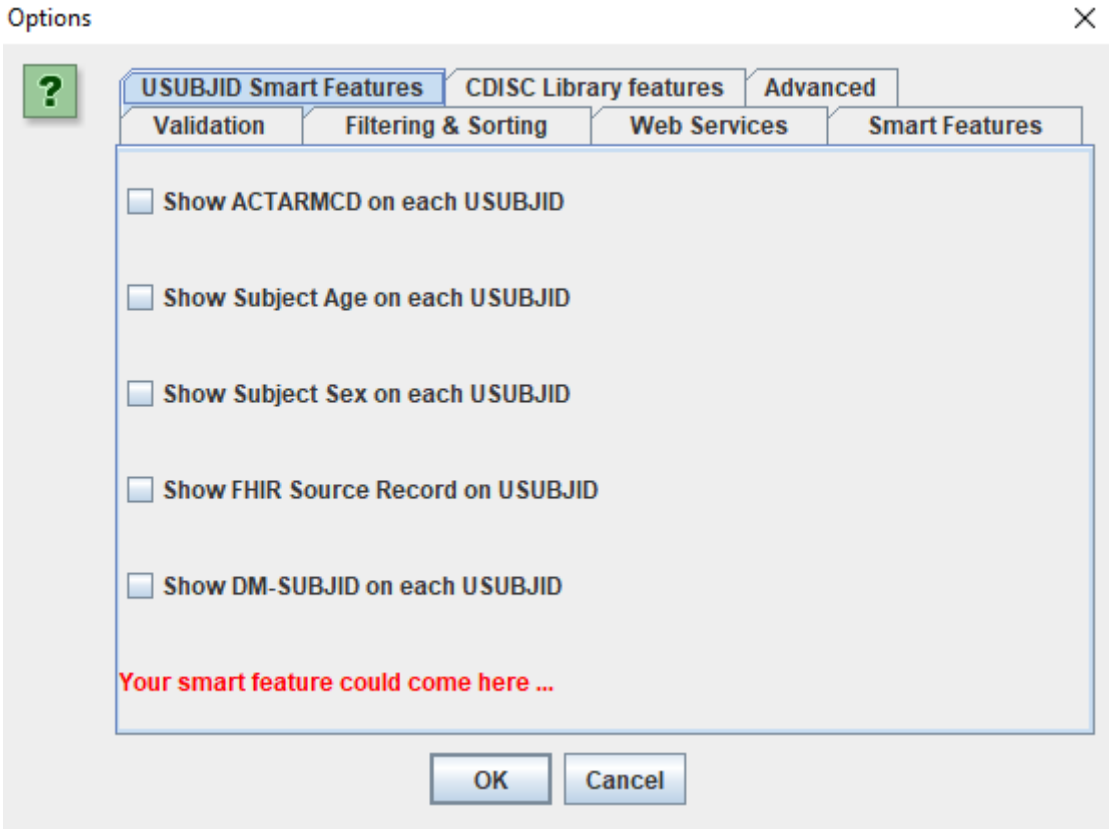
Also these are pretty self-explaining, with the possibility of obtaining additional information by howevering the mouse over an item, for example:



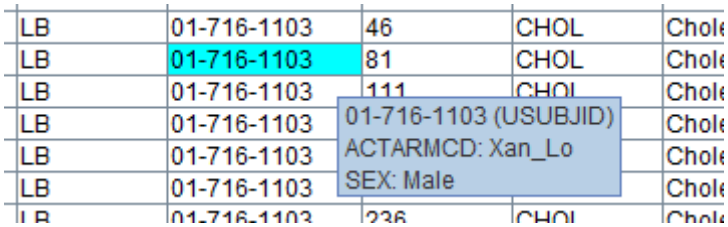
which, when checked, will show the value of EDTC and ELEMENT (study element code and name) as a

tooltip on each --DTC cell in each table, which can be very useful[6]. As one can expect, this requires the SE (Subject Elements) to have been loaded.

Furthermore, there is the tab "USUBJID Smart Features", which is mostly about getting information from the DM dataset (when loaded) in the tables for other datasets:



For example, when the checkboxes "Show Subject Sex on each USUBJID" and "Show ACTARMCD on each USUBJID" are checked, and one hovers the mouse over a "USUBJID" cell e.g. in the LB dataset, one obtains the information about sex and actual arm of that subject in the study. For example:



immediately showing that the subject is a male person and the actual arm was "Xanomelin low".
The same can of course be obtained by "toggling" between the current table and the DM table (using Ctrl-D and Ctrl-B), but just using hovering is of course faster.

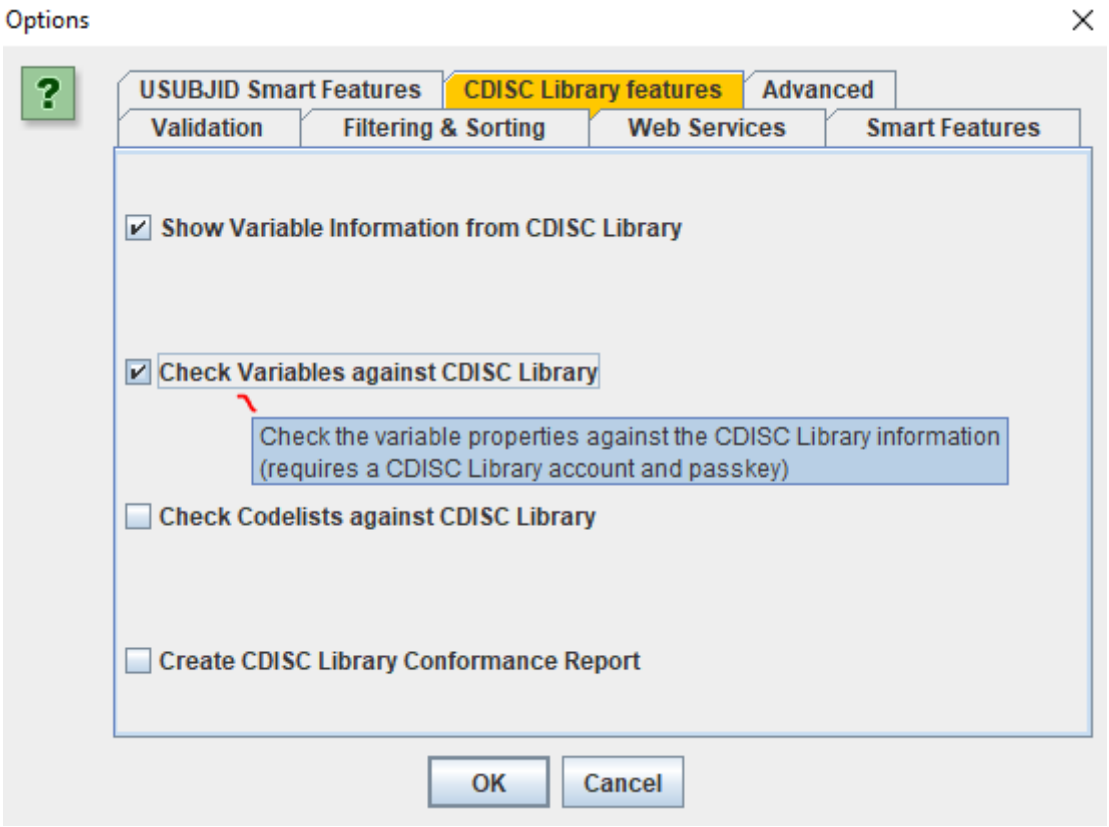
As the Smart Submission Dataset Viewer is Open Source software (member of COSA - CDISC Open Source Alliance), additional "smart features" can always be added[7] by users, the only limitation being lack of

[6] This shows again that many "derived" variables in SDTM should not be in SDTM at all. They are only there as the reviewers at the regulatory authorities do not have access to modern viewers that understand CDISC submission standards. So they asked CDISC to add them in SDTM itself.
The main reason for this is the mandated use of SAS-XPT, and the use of primitive tools such as the "SAS Universal Viewer". Essentially, the Smart Submission Dataset Viewer lifts these limitations, so that in future, many of the "derived" variables could be removed from the SDTM standard again.
[7] Of course, suggestions for new "smart features" are always welcome.

imagination ...

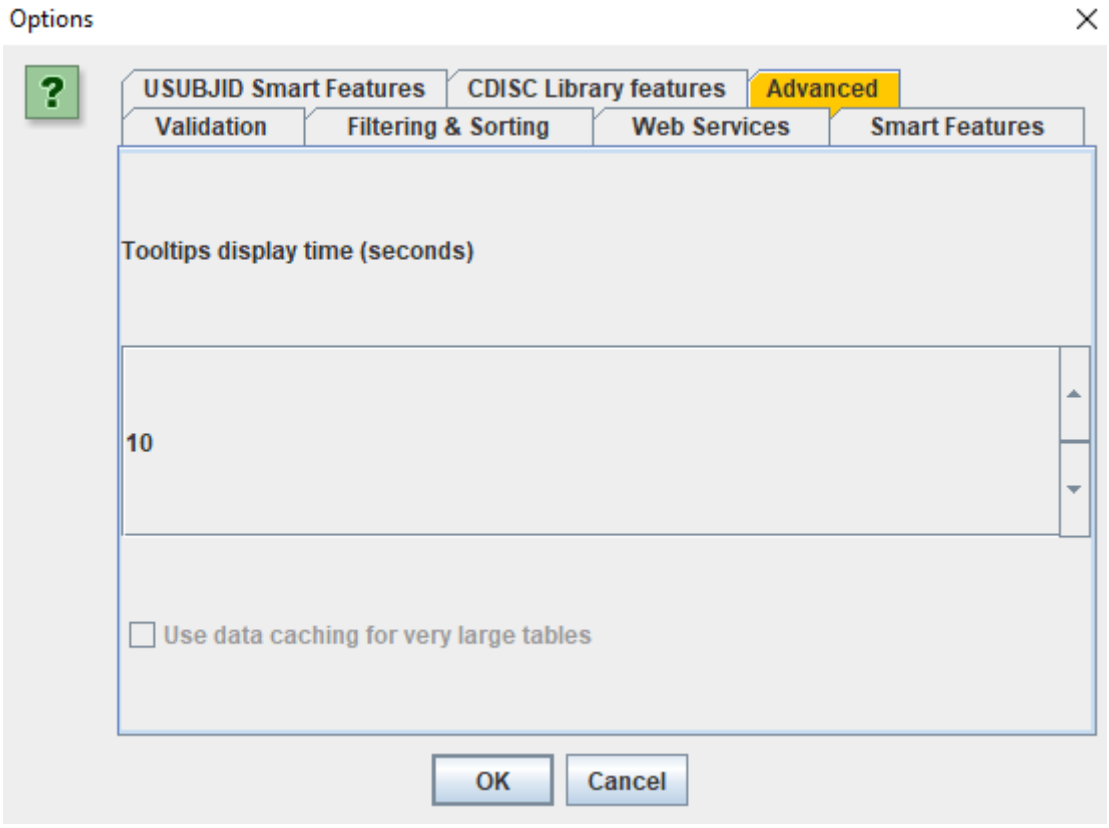
Very special is the the option "Show FHIR Source Record on USUBJID". This is an experimental feature showing how SDTM in Dataset-JSON format can include source records as e.g. from Electronic Health Records", and can then be visualized by a viewer. Currently, this has only be implemented for FHIR records. See the document "Support for embedded EHRs (FHIR)" for more details. The document still describes this feature for Dataset-XML as the transport format, but the same is of course true for JSON, as the FHIR standard has full support for JSON as the transport format. Remark that embedding source records in SDTM records is impossible when using SAS-XPT. Another reason to get rid of it as soon as possible.

Also the "CDISC Library" and its API have been implemented, tab "CDISC Library Features":



allowing to do validation by checking variable and codelist properties (from the define.xml) against the CDISC Library. This feature of course not only requires an internet connect. but also a CDISC Library account and API key. The latter must be stored in the "properties.dat" file - see section "The 'properties.dat' file".

The last tab is "Advanced":



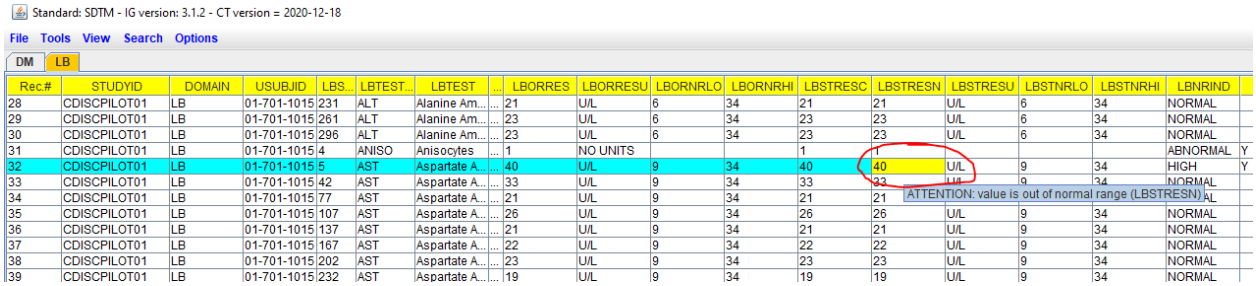
For example, normally tooltips on cells or other items show up for 10 seconds by default, but the user can set this to another time duration.

The option "Use data caching for very large tables" is currently still in development.

Extended validation

When none of the options in the "Validation" tab has been selected, only very basic validation will be performed (e.g. on "required" variables in the case of SDTM and SEND).
By switching on checkboxes in the "Validation" tab, additional validation checks are performed. This of course comes at the cost of loading time.

For example, when the checkbox "Check values against reference range" is checked, values that fall outside the --STNRLO / --STNRHI range (in any of the SDTM/SEND dataset) are highlighted. For example:

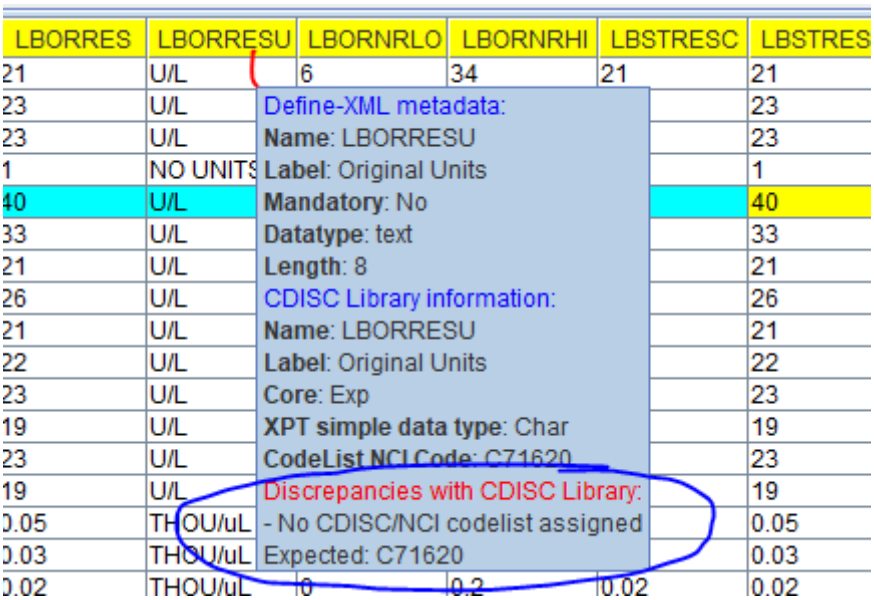


Rec.#	STUDYID	DOMAIN	USUBJID	LBS...	LBTEST...	LBTEST	...	LBORRES	LBORRESU	LBORNRLO	LBORNRHI	LBSTRESC	LBSTRESN	LBSTRESU	LBSTNRLO	LBSTNRHI	LBNRIND	
28	CDISCPILOT01	LB	01-701-1015	231	ALT	Alanine Am...	...	21	U/L	6	34	21	21	U/L	6	34	NORMAL	
29	CDISCPILOT01	LB	01-701-1015	261	ALT	Alanine Am...	...	23	U/L	6	34	23	23	U/L	6	34	NORMAL	
30	CDISCPILOT01	LB	01-701-1015	296	ALT	Alanine Am...	...	23	U/L	6	34	23	23	U/L	6	34	NORMAL	
31	CDISCPILOT01	LB	01-701-1015	4	ANISO	Anisocytes	...	1	NO UNITS			1	1				ABNORMAL	Y
32	CDISCPILOT01	LB	01-701-1015	5	AST	Aspartate A...	...	40	U/L	9	34	40	40	U/L	9	34	HIGH	Y
33	CDISCPILOT01	LB	01-701-1015	42	AST	Aspartate A...	...	33	U/L	9	34	33	33	U/L	9	34	NORMAL	
34	CDISCPILOT01	LB	01-701-1015	77	AST	Aspartate A...	...	21	U/L	9	34	21	21	U/L	9	34	NORMAL	
35	CDISCPILOT01	LB	01-701-1015	107	AST	Aspartate A...	...	26	U/L	9	34	26	26	U/L	9	34	NORMAL	
36	CDISCPILOT01	LB	01-701-1015	137	AST	Aspartate A...	...	21	U/L	9	34	21	21	U/L	9	34	NORMAL	
37	CDISCPILOT01	LB	01-701-1015	167	AST	Aspartate A...	...	22	U/L	9	34	22	22	U/L	9	34	NORMAL	
38	CDISCPILOT01	LB	01-701-1015	202	AST	Aspartate A...	...	23	U/L	9	34	23	23	U/L	9	34	NORMAL	
39	CDISCPILOT01	LB	01-701-1015	232	AST	Aspartate A...	...	19	U/L	9	34	19	19	U/L	9	34	NORMAL	

where the value of LBSTRESN for record 32, being 40 falls outside the range 9-34 U/L.
This feature can be used by reviewers to quickly find "out-of-range" values without needing to rely on the --NRIND (Reference Range Indicator), provided by the sponsor, and which can have been incorrectly

assigned[8]. This means that the reviewer is now easily able to have an independent assignment, independent of what is provided by the sponsor. Many of the validation features of the viewer do allow the same independency.
The same e.g. applies to the feature "Check --DY values" (--DY = "Study day"). It is one of this variables where we see that very often errors are made, which pass unseen when e.g. using the "SAS Universal Viewer".

Another example is the use of the "CDISC Library". When in that tab, the checkbox "Check Variables against CDISC Library" is checked, and when hovers the mouse over a column header, additional information is provided, such as:



also stating that it was expected that the CDISC codelist C71620 is assigned to variable LBORRESU, which isn't in this case.

We will not explain every validation feature further here as there are a lot of them, and new ones are regularly added. Important to know is that, except for the very basic ones, the default is that they are switched off.

Validation using CDISC CORE

Also CDISC CORE (CDISC Open Rules Engine) have been implemented in the "Smart Submission Dataset Viewer". At the moment of writing (December 2024), the feature has been disabled, as CORE is currently being updated for Dataset-JSON 1.1 (version 1.0 is now deprecated and should not be used anymore). As soon as v.1.1 of Dataset-JSON is implemented in CORE, we will re-add the features to use CDISC CORE validation.

A separate document explaining all the features of validation using CDISC CORE, as implemented in the Smart Submission Dataset Viewer, is currently in development.

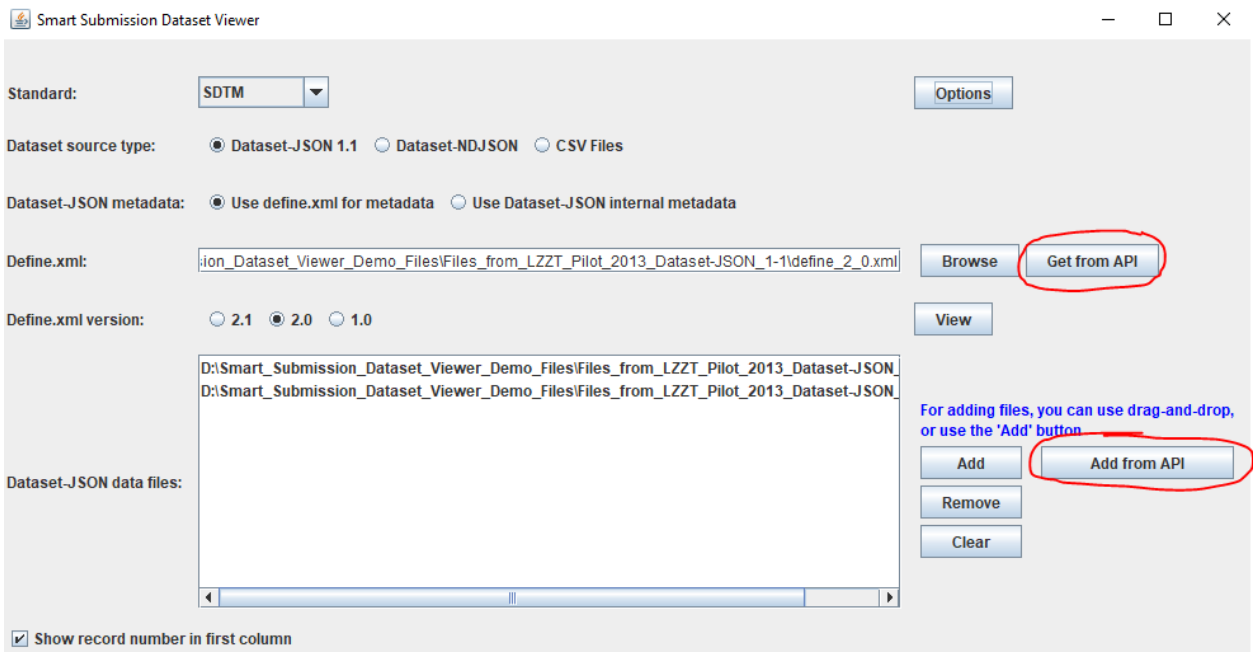
[8] This shows again that such variables such as --NRIND (Reference Range Indicator) should not be part of SDTM (as they are derived). Such variables have only be added on request by regulatory authorities, as the tools of the reviewers are not "smart" enough to derive them theirselves.

Using APIs

IMPORTANT REMARK: everything described in this section is purely experimental!
It is meant for demonstration purposes to show how APIs and RESTful Web Services (RWS) can be used to retrieve information from other or remote systems, not only full datasets, but also selected information, making the review process considerably more easy.

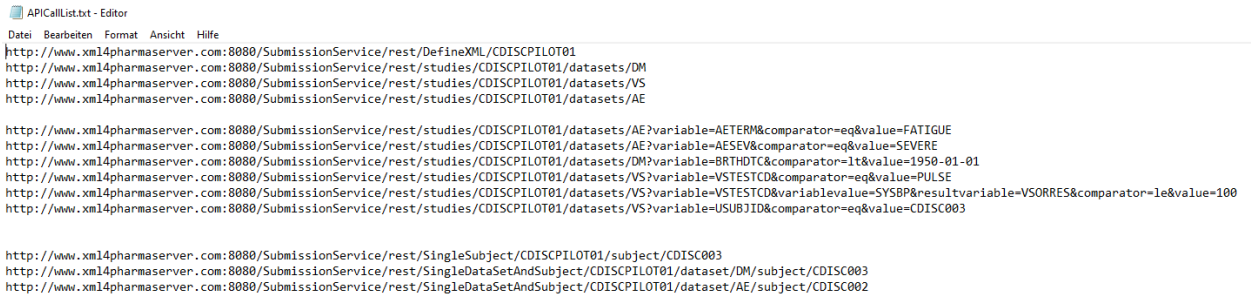
It also demonstrates that with the use of Dataset-JSON, it becomes in principle possible to evolve to "rolling submissions", in which SDTM/SEND and possibly also ADaM data already becomes available as soon as the first data points have been collected, as also the process of conversion from source data to especially SDTM and SEND can easily be made "rolling", thanks to modern conversion tools.
Such "rolling submissions" also would enable regulatory authorities to make final decisions within a short time after the last data point was collected, which would mean that new therapies can become available at least 1 to 2 years earlier than is currently the case.

In the main window, one may have noticed the buttons "Get from API" for define.xml, and "Add from API" for the datasets themselves:

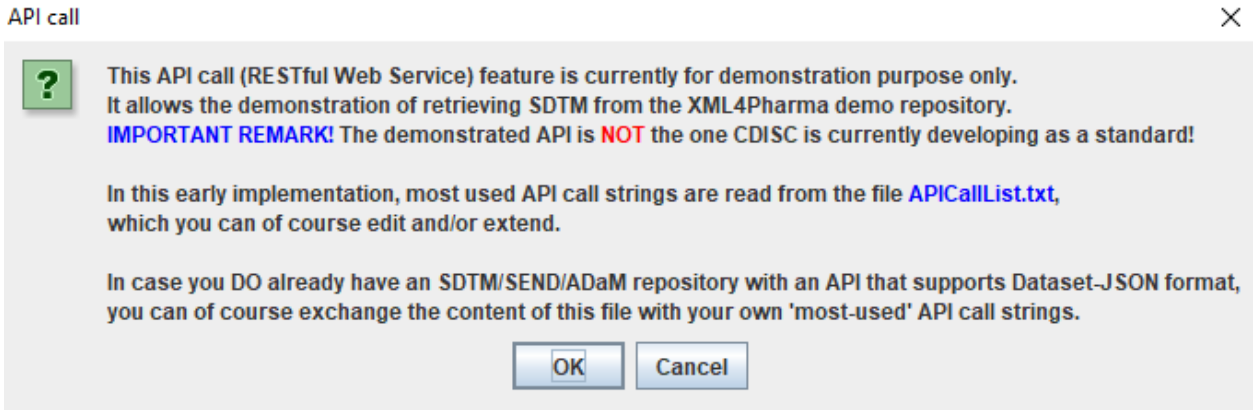


Just for demonstration purposes, we have build an SDTM dataset containing the information of the "LZZT Pilot" (no, not in XPT format …), and developed a simple API, again, just for demo purposes, so nothing formal, to retrieve information from that database and load it into the viewer.

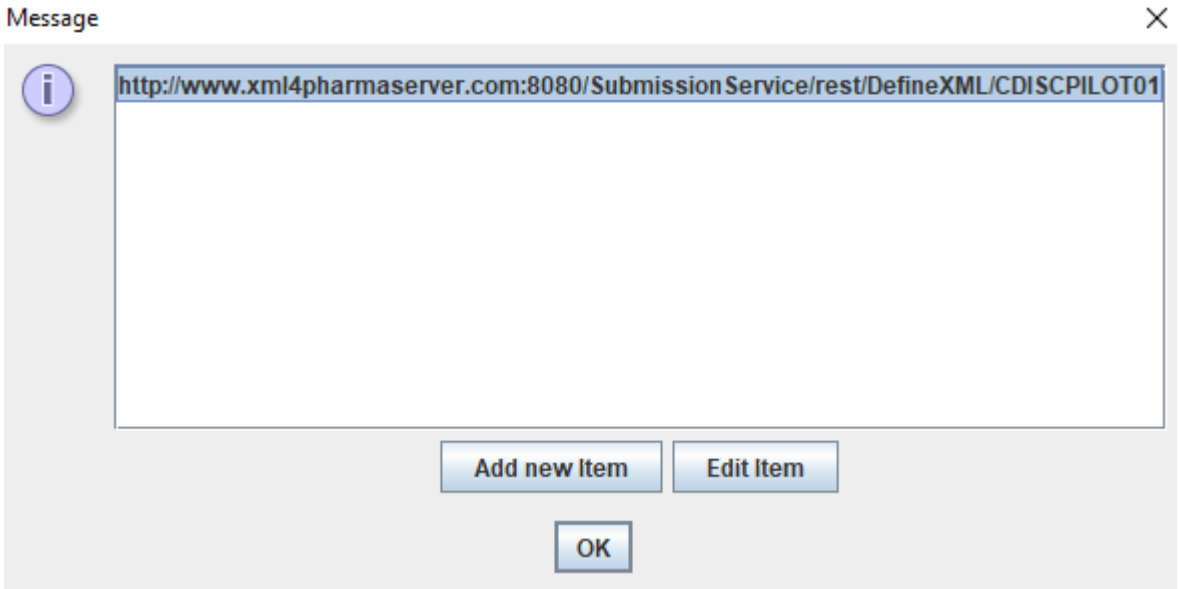
A number of "GET" queries have been added to a file "APICallList.txt", which comes with the distribution, and which can be edited/extended by the user:



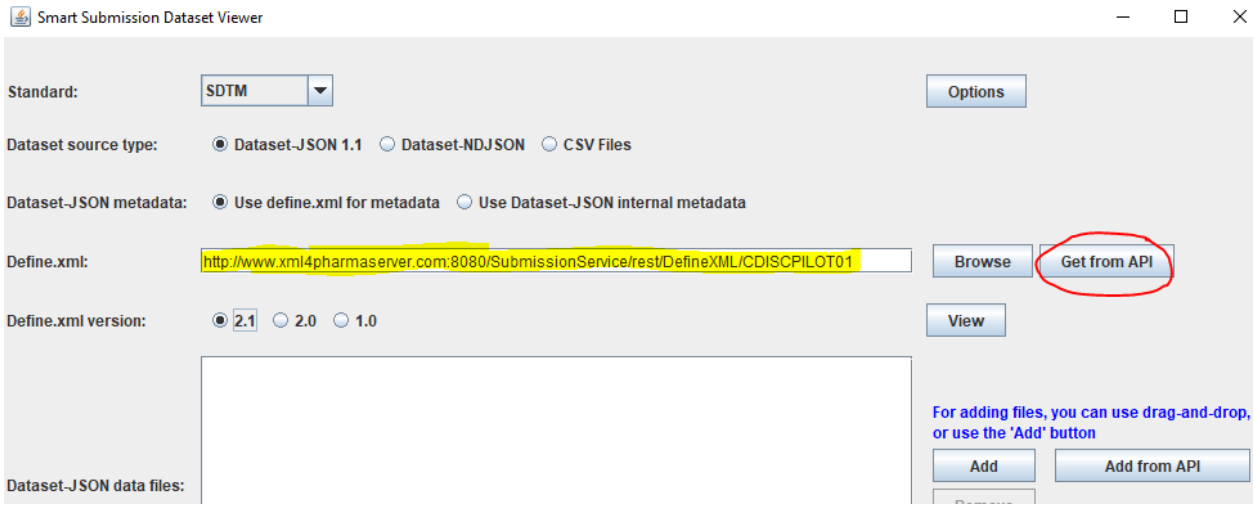
When then clicking the button "Get from API" for the define.xml, first some explanation is displayed:



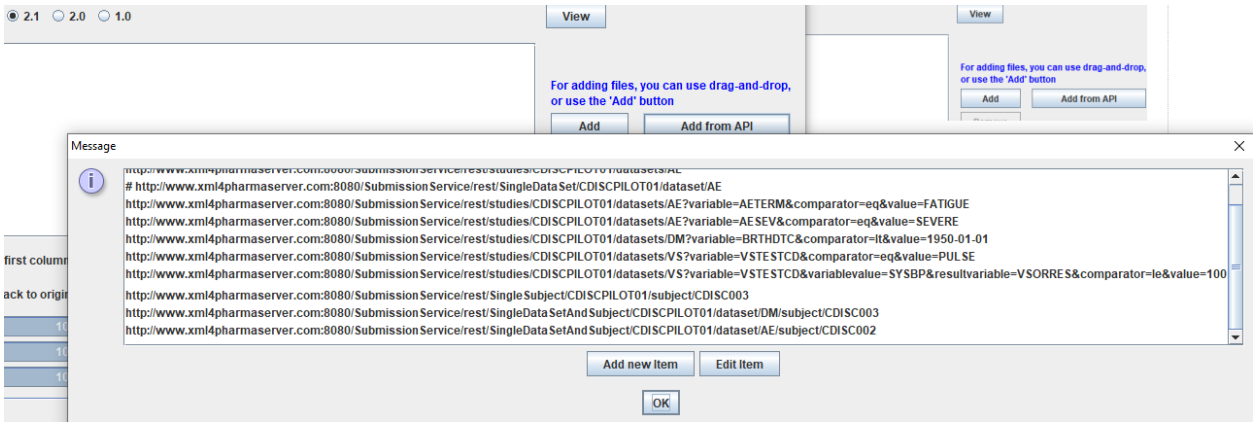
and then this list is read, and the available "GET" commands regarding the define.xml is displayed:



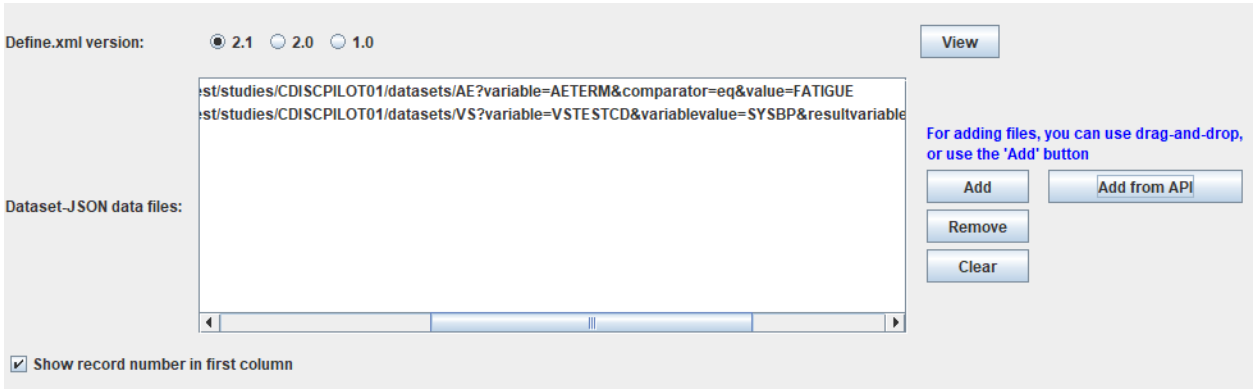
to which one can add new ones, or edit the existing ones, though this currently doesn't make sense, or it must be that one has implemented a similar API on an own server that has an SDTM database.
When selecting the first one, and clicking "OK", the text field for define.xml is populated:



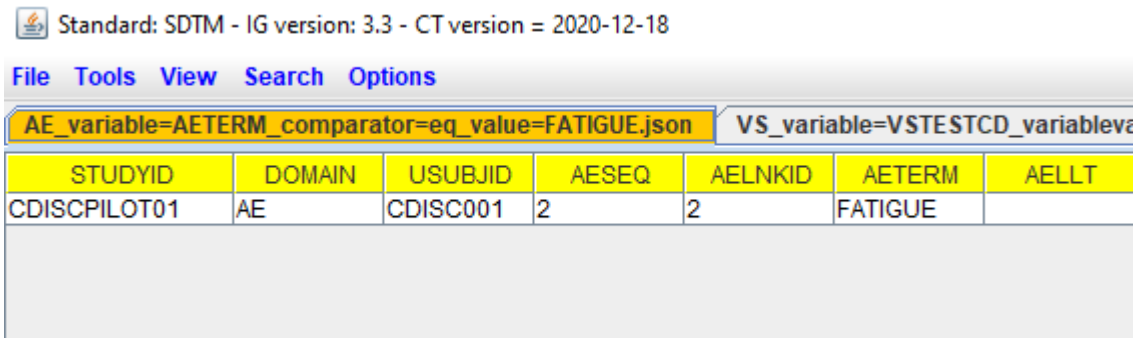
Similarly, when one clicks the "Add from API" button for the datasets themselves, the list is read and the "GET" instructions made available:



Just for the demo, let us select the one about "AE with AETERM is fatigue" and the one "VSTESTCD where VSTESTCD=SYSBP and VSORRES less or equal 100", leading to:

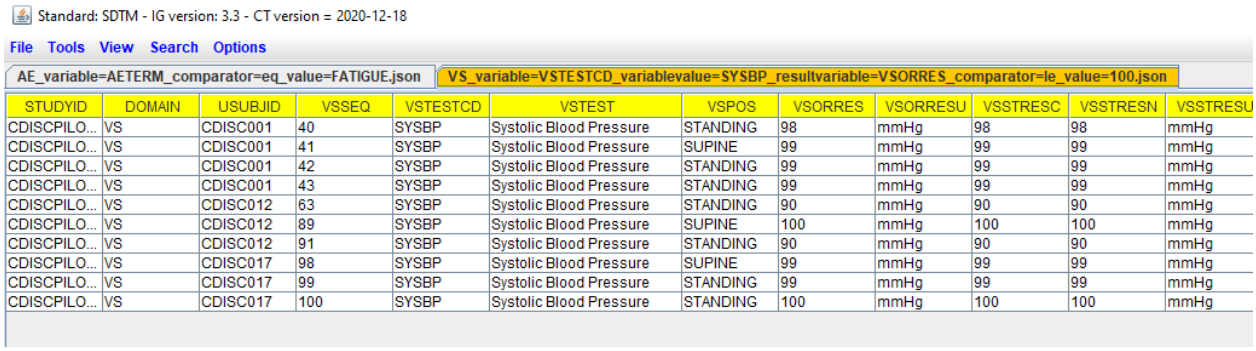


When then hitting "Start", the query is submitted to the server, and some Dataset-JSON is generated by the server and send back to our application as the "response", and then used by the viewer, leading to:



STUDYID	DOMAIN	USUBJID	AESEQ	AELNKID	AETERM	AELLT
CDISCPILOT01	AE	CDISC001	2	2	FATIGUE	

showing that the database has only 1 record with AETERM=FATIGUE, and



STUDYID	DOMAIN	USUBJID	VSSEQ	VSTESTCD	VSTEST	VSPOS	VSORRES	VSORRESU	VSSTRESC	VSSTRESN	VSSTRESU
CDISCPILO...	VS	CDISC001	40	SYSBP	Systolic Blood Pressure	STANDING	98	mmHg	98	98	mmHg
CDISCPILO...	VS	CDISC001	41	SYSBP	Systolic Blood Pressure	SUPINE	99	mmHg	99	99	mmHg
CDISCPILO...	VS	CDISC001	42	SYSBP	Systolic Blood Pressure	STANDING	99	mmHg	99	99	mmHg
CDISCPILO...	VS	CDISC001	43	SYSBP	Systolic Blood Pressure	STANDING	99	mmHg	99	99	mmHg
CDISCPILO...	VS	CDISC012	63	SYSBP	Systolic Blood Pressure	STANDING	90	mmHg	90	90	mmHg
CDISCPILO...	VS	CDISC012	89	SYSBP	Systolic Blood Pressure	SUPINE	100	mmHg	100	100	mmHg
CDISCPILO...	VS	CDISC012	91	SYSBP	Systolic Blood Pressure	STANDING	90	mmHg	90	90	mmHg
CDISCPILO...	VS	CDISC017	98	SYSBP	Systolic Blood Pressure	SUPINE	99	mmHg	99	99	mmHg
CDISCPILO...	VS	CDISC017	99	SYSBP	Systolic Blood Pressure	STANDING	99	mmHg	99	99	mmHg
CDISCPILO...	VS	CDISC017	100	SYSBP	Systolic Blood Pressure	STANDING	100	mmHg	100	100	mmHg

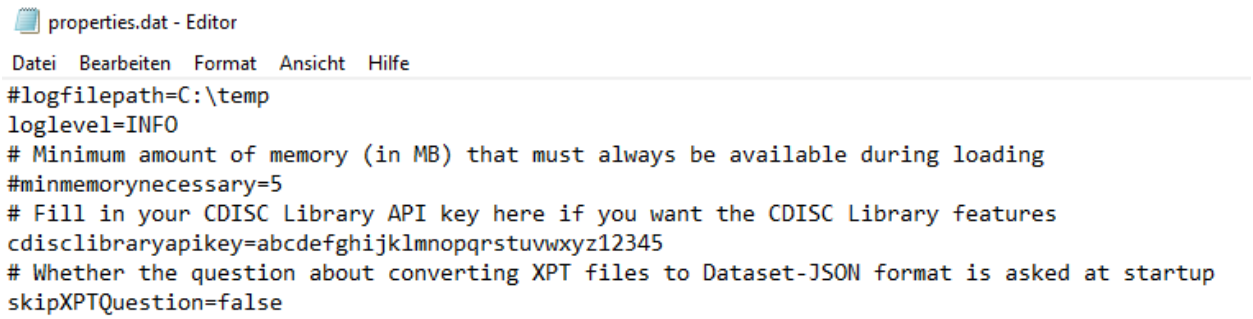
providing all the vital signs record for which the systolic blood pressure is less or equal to 100 mmHg.

Using such RESTful Web Services is much more efficient than having to load full datasets (or pre-filtered ones) and then having the reviewer to have to start filtering. It also does not require a high quality define.xml as is the case for the "pre-loading filtering" feature, which already can save a lot of time.
One can easily imagine a simple application on top of the viewer that has a user-friendly graphical user interface and takes care of assembling the queries. And of course, one can also imagine AI to be used to do so, where the user simply talks to and says (or types): "Viewer, give me vital signs records for all the subjects with a systolic blood pressure lower than 100 for study XYZ …".

Once again, please notice that what is presented in this section is not part of any formal standard, is not endorsed by CDISC, but was just developed to demonstrate the power of RESTful Web Services and API for reviewing submissions.

The "properties.dat" file

The "properties.dat" file allows to steer the behavior of the software. This file is immediately read after starting the software. The software comes with a sample properties.dat file in which some settings are "commented out". These are lines that start with a "#" character.
For example:



In the first line (here commented out), the user can set to which folder the log files will be written.
The default is that they are written to a folder named "logs" that is under the folder to which the software was installed.
The second line allows to set the "logging level". Under normal circumstances, one will set it to "INFO". If one encounters problems, best it to set it to "DEBUG". In that case, much more information is written into the log file, which you can then send to us to understand the issue, and helps the developers to further improve the software. Other possible settings are "WARNING" and "ERROR". For both of these, the log file will be much smaller, as it will only be written when a warning or error is thrown.
The next line is something that one mostly will have commented out. It defines the minimum amount of memory that must always be free during loading of the dataset. If, during loading, the amount of memory threatens to fall under this minimum, the system will issue a warning and try to do some "garbage collection". Change this setting only when you are a developer and have access to the source code.
The "Smart Submission Dataset Viewer" has some validation options that use the "CDISC Library". In order to be able to use these, you will require a "CDISC Library API Key" which you can obtain from CDISC for free. See the CDISC Library API documentation for more details. When you have obtained one, copy-paste it into the "properties.dat" file.
When you start the software, you will be asked whether you first want to convert SAS-XPT files to Dataset-JSON, or whether you already have such, and wants to start with these.
When setting skipXPTQuestion=true", the system will skip this first dialog and immediately jump to the main window.