

# Smart Dataset-XML Viewer: Sorting and Filtering

Document version: 2018-12-15

## Introduction

Filtering and sorting is one of the most important features when reviewing regulatory submissions in clinical research. The "Smart Submission Dataset Viewer" has filtering capabilities that go far beyond what most of the reviewers at the FDA and PMDA have available: due to the current restriction for SAS-XPT files, most of them still use the outdated "SAS System Viewer" which was never meant to support SDTM, SEND or ADaM CDISC standards<sup>1</sup>.

The "Smart Submission Dataset Viewer" however has been developed with the CDISC standards in mind, and can thus be said to be "CDISC-savvy".

## Example files

We will be demonstrating the filtering features of the "Smart Submission Dataset Viewer" using the SDTM submission files "Files\_from\_LZZT\_Pilot\_2013\_LBLOINC\_Dataset-XML" can be downloaded from the project website.

All the filtering features are however equally applicable to SEND and ADaM files.

## Loading a submission - initial filtering

After startup of the software, one should always start with loading a define.xml that describes the metadata of the submission. One can as well load define.xml v.1.0 as 2.0 files, but it is recommended (when possible) to use define.xml 2.0 as the FDA is currently fading out v.1.0. For example:



This is important, as all important information about the study is located in the define.xml file, like datatypes and maximal length of variables, codelists and valuelists, and much much more.

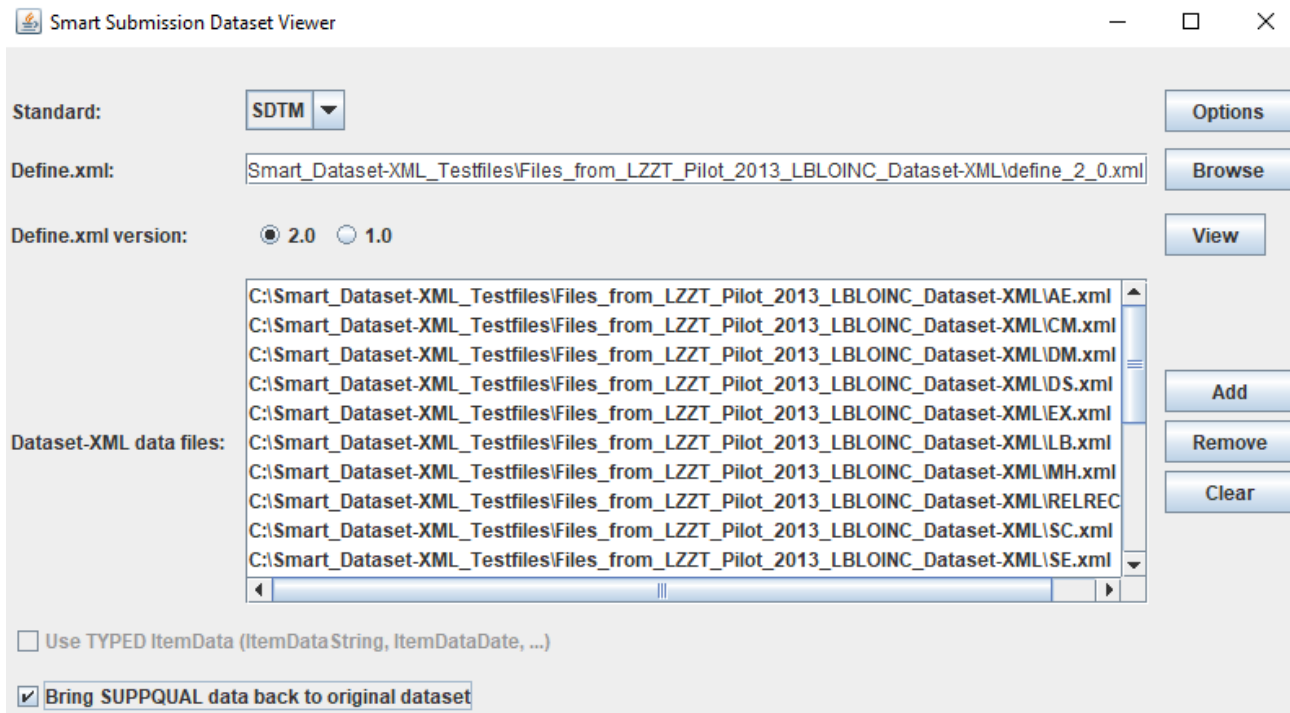
After having loaded the define.xml file, one can inspect it, either in your favorite browser, or the source XML itself. This is accomplished by the "View" button. Further details are explained in a separate document.

Once the define.xml loaded, one can start loading the submission files. These can either be the Dataset-XML files, or their zipped versions (one file per dataset). When using the zipped files, these will not be unzipped during loading: **the viewer is capable to read from zipped files without the need of unzipping them.**

If we just use the Dataset-XML files themselves, this e.g. leads to:

---

<sup>1</sup>The latest version of the "SAS System Viewer" stems from 2004!



It is always advised to at least have the DM (Demographics) dataset included, as one can then easily have information of that dataset displayed even in other datasets (usually as tooltips). This feature is also further explained in another document.

It is also advised to have the trial design datasets loaded. These can already be generated even before study start, once the study design is fixed<sup>2</sup>.

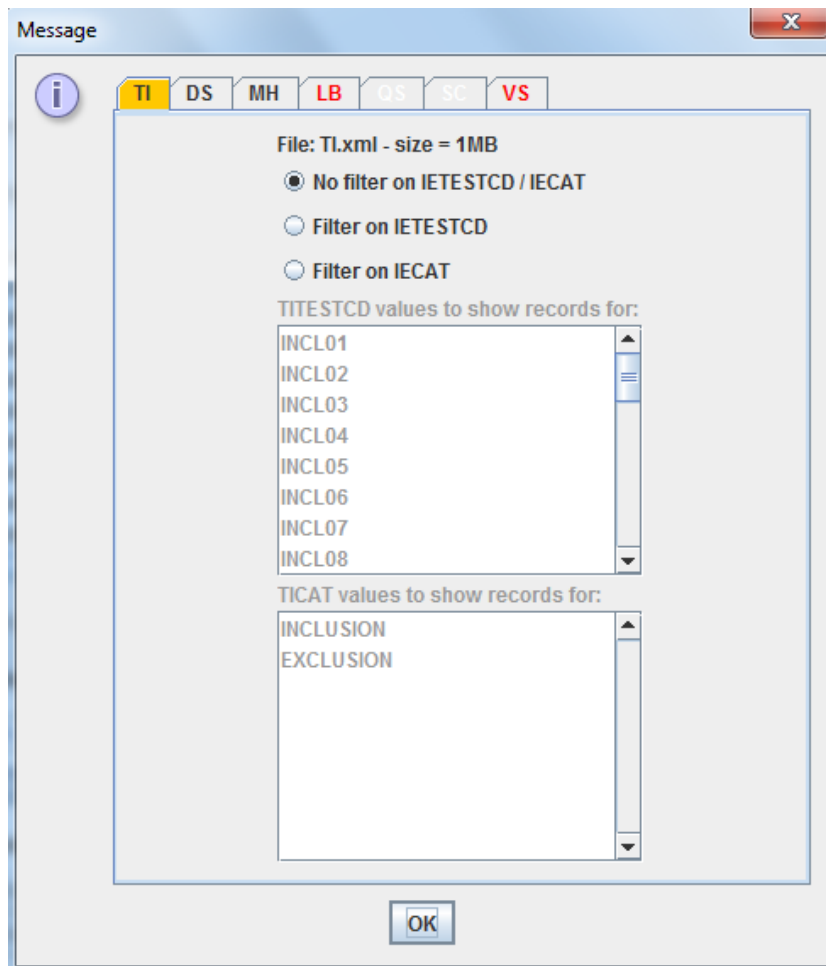
If you have "Supplemental Qualifier Datasets"<sup>3</sup> you can decide to bring them back to their "parent" domain by checking the "Bring SUPPQUAL data back to original dataset". This will make the review process considerable easier. Even if you do not recombine these datasets in the viewer, you will still be able to quickly switch between a "SUPPQUAL" datapoint and its parent record, using the CTRL-S and CTRL-B keyboard combinations (there are also menus for this). Also this is a feature that is not available by many reviewers.

There is a large amount of options that one can switch on or off using the "Options" button. Also this is explained in detail in a number of other documents. These options are independent from the filtering capabilities of the software, so we won't use any of them here.

Clicking the "Start" button near the bottom now starts the loading process. Within a few seconds, a dialog is displayed, allowing us to do the first filtering:

<sup>2</sup>We unfortunately observe that it is still common to generate the trial design datasets at the end of the process, and people than complaining that they have problems doing so!

<sup>3</sup>Unfortunately the FDA and PMDA still do not allow to have "non standard variables" (NSVs) in the main dataset itself, and describe this in the define.xml. The reason for this is probably that reviewers cannot combine information from the define.xml with the datasets themselves, due to the use of the SAS-XPT format.

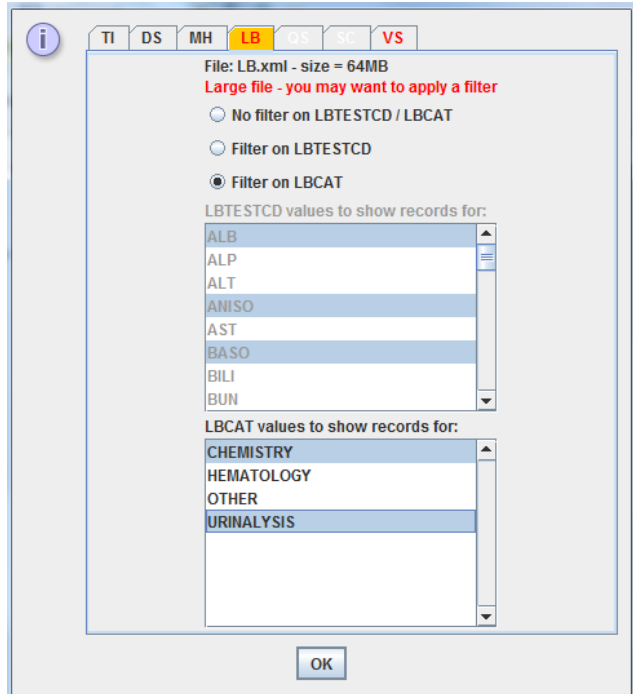
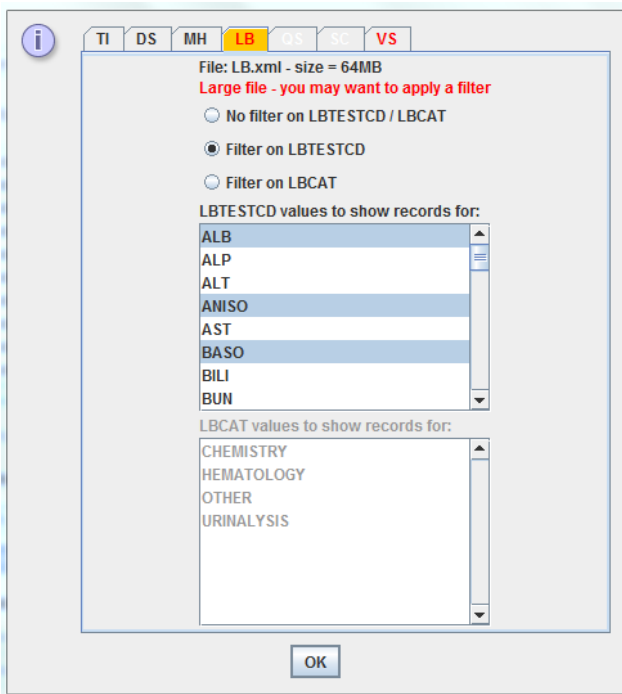


The dialog shows a list of datasets that are suitable for filtering, with the ones having a red tab for which filtering may be very useful, due to the file size (one can set the criteria for this using the "Options" button in the main frame).

In this case, it is reported that both the "LB" and "VS" dataset contain a (very) large amount of records, so that filtering may be useful<sup>4</sup> in order to keep good oversight when inspecting the data. Primary filtering (before actual loading) can always be done based on the test code (--TESTCD) or on the test category (--CAT). For example for the LB (Laboratory) dataset:

---

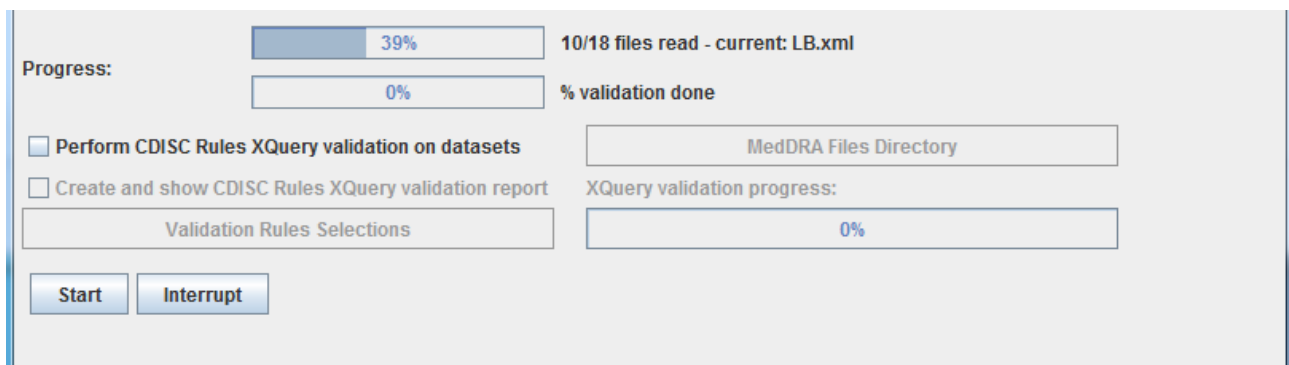
<sup>4</sup>I have never understood how reviewers can load hundred thousand of records at the same time and then still retrieve information from it by simple inspection.



In our case, let us filter on "category" (LBCAT) where we select to only load the chemistry data and the urinalysis data.

For VS (vital signs) we decide to leave everything as is and to no do any initial filtering (we can still apply filtering later).

After clicking the "OK" button, the system starts loading the files, with the progress displayed for each file:



In case you decided the merge SUPPQUAL datasets with their parent dataset, it will take somewhat longer (but still very acceptable, due to the "smart" algorithm that was implemented). As a result, both the "parent" dataset including additional columns, as well as the SUPPQUAL dataset (e.g. SUPPLB) will be displayed. Our result is then:

TA	TE	TI	TS	TV	DM	AE	CM	DS	EX	LB	MH	SV	VS	SUPPAE	SUPPDM	SUPPDS	SUPPLB
DRRESU	LBORNRLD	LBORNRI	LBSTRESC	LBSTRESN	LBSTRESU	LBSTNRLO	LBSTNRHI	LBNRIND	LBLOINC	LBLFL	VISITNUM	VISIT	VISITDY	LBOTC	LBDY	ENDPOINT	LBTMSHI
3.3	4.9	38	38	g/L	33	49	NORMAL	1751-7		Y	1	SCREENIN...	-7	2013-12-2...	-7		0.8
3.3	4.9	39	39	g/L	33	49	NORMAL	1751-7			4	WEEK 2	14	2014-01-1...	15		0.8
3.3	4.9	38	38	g/L	33	49	NORMAL	1751-7			5	WEEK 4	28	2014-01-3...	29		0.8
3.3	4.9	37	37	g/L	33	49	NORMAL	1751-7			7	WEEK 6	42	2014-02-1...	42		0.8
3.3	4.9	38	38	g/L	33	49	NORMAL	1751-7			8	WEEK 8	56	2014-03-0...	63		0.8
3.3	4.9	38	38	g/L	33	49	NORMAL	1751-7			9	WEEK 12	84	2014-03-2...	84		0.8
3.3	4.9	37	37	g/L	33	49	NORMAL	1751-7			10	WEEK 16	112	2014-05-0...	126		0.8
3.3	4.9	37	37	g/L	33	49	NORMAL	1751-7			11	WEEK 20	140	2014-05-2...	140		0.8
3.3	4.9	38	38	g/L	33	49	NORMAL	1751-7			12	WEEK 24	168	2014-06-1...	168	Y	0.8
3.3	4.9	38	38	g/L	33	49	NORMAL	1751-7			13	WEEK 26	182	2014-07-0...	182		0.8
35	115	34	34	U/L	35	115	LOW	6768-6		Y	1	SCREENIN...	-7	2013-12-2...	-7		0.3
35	115	50	50	U/L	35	115	NORMAL	6768-6			4	WEEK 2	14	2014-01-1...	15		0.4
35	115	41	41	U/L	35	115	NORMAL	6768-6			5	WEEK 4	28	2014-01-3...	29		0.4
35	115	43	43	U/L	35	115	NORMAL	6768-6			7	WEEK 6	42	2014-02-1...	42		0.4
35	115	47	47	U/L	35	115	NORMAL	6768-6			8	WEEK 8	56	2014-03-0...	63		0.4
35	115	53	53	U/L	35	115	NORMAL	6768-6			9	WEEK 12	84	2014-03-2...	84		0.5
35	115	41	41	U/L	35	115	NORMAL	6768-6			10	WEEK 16	112	2014-05-0...	126		0.4
35	115	45	45	U/L	35	115	NORMAL	6768-6			11	WEEK 20	140	2014-05-2...	140		0.8
35	115	46	46	U/L	35	115	NORMAL	6768-6			12	WEEK 24	168	2014-06-1...	168	Y	0.4
35	115	46	46	U/L	35	115	NORMAL	6768-6			13	WEEK 26	182	2014-07-0...	182		0.4

with a tab for each dataset. For the LB dataset, one sees that there are two extra columns (data points have a blue foreground color) "ENDPOINT" and "LBTMSHI" which originate from the SUPPLB dataset:

DY	LBOTC	LBDY	ENDPOINT	LBTMSHI
2013-12-2...	-7			0.8
2014-01-1...				
2014-01-3...				
2014-02-1...				
2014-03-0...				
2014-03-2...	84			0.8
2014-05-0...	126			0.8

with additional information displayed when holding the mouse over the column header<sup>5</sup>.

In the LB dataset, one also sees that only records for the categories (LBCAT) "CHEMISTRY" and "URINANALYSIS" have been loaded.

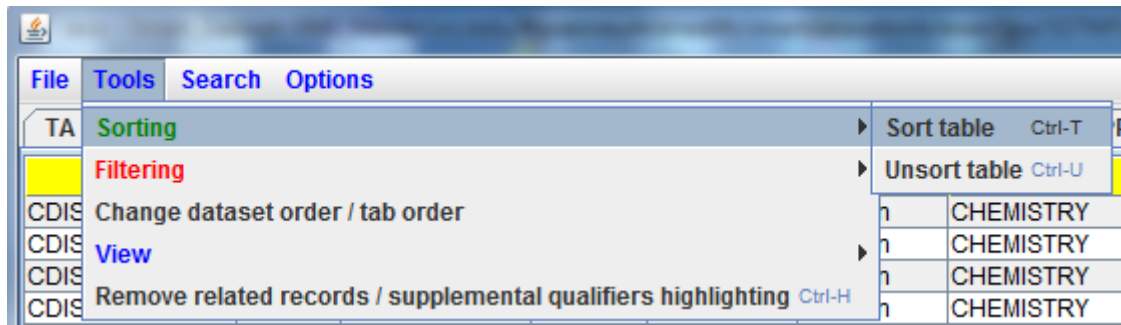
During loading, one may also have observed that basic validation has been performed. Additional validation can also be done by using the "Options" button in the main frame again, and choosing those validations that one wants to perform. It is even possible to perform the FDA and PMDA validation checks as well CDISC's own validation checks in the case of ADaM, using the checkbox "Perform CDISC Rules XQuery Validation on datasets". This is extensively treated in the document "The Smart Dataset-XML Viewer: SDTM Dataset XQuery Validation".

### Sorting and Filtering once the datasets are loaded

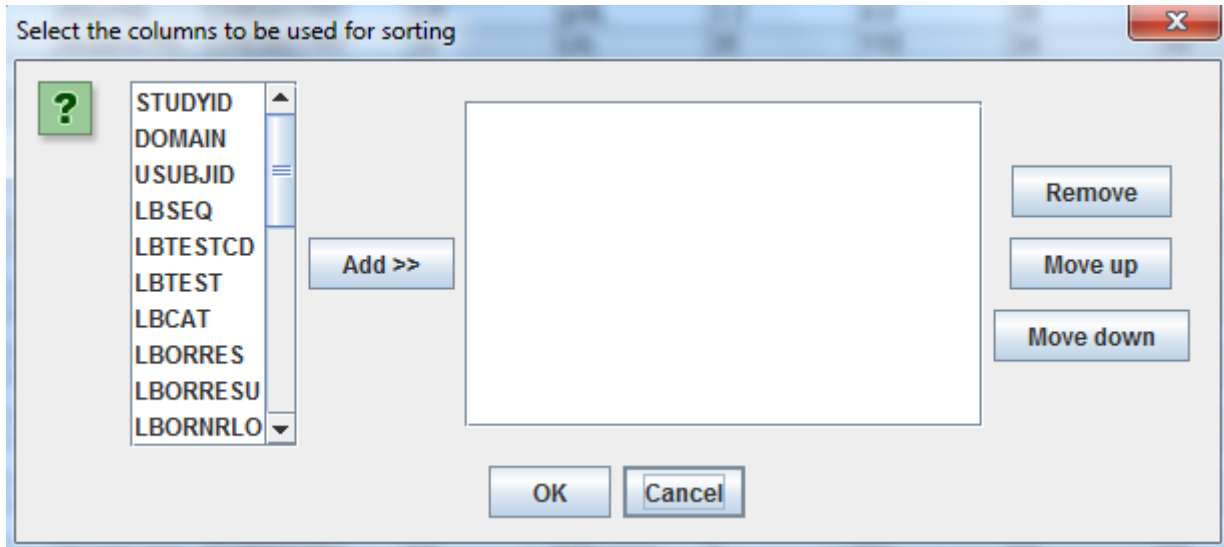
Once the datasets are loaded, one can do all the things that is expected from a good table implementation like resizing and moving (i.e. changing the order of) the columns, and simple sorting by clicking the column header.

One can also sort the data in the columns based on multiple columns. For this, use the menu "Tools - Sorting - Sort table":

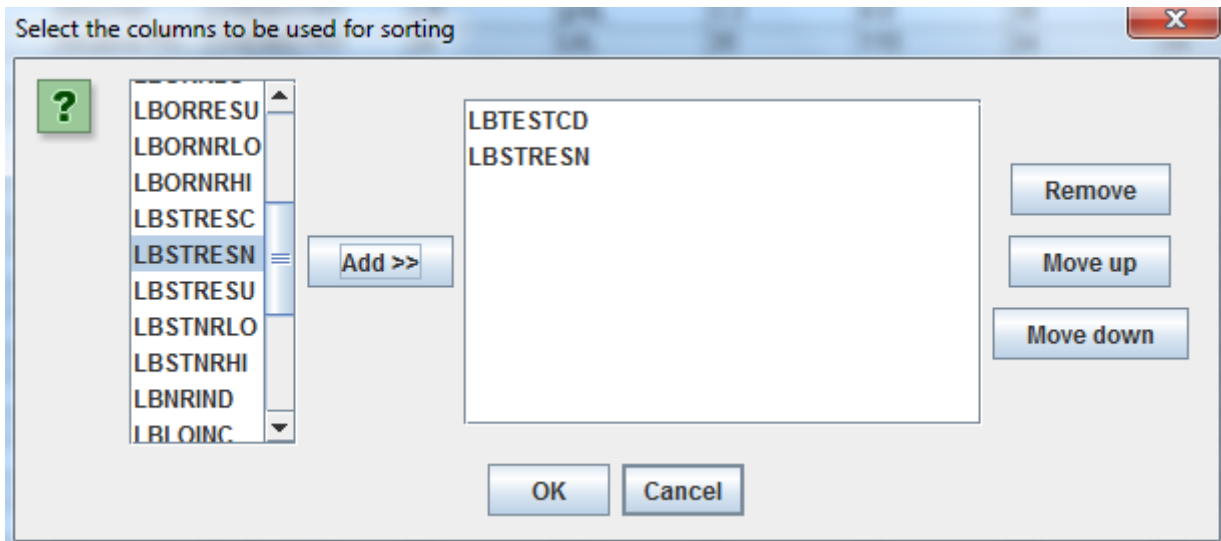
<sup>5</sup>This is exactly some of the information taken from the define.xml



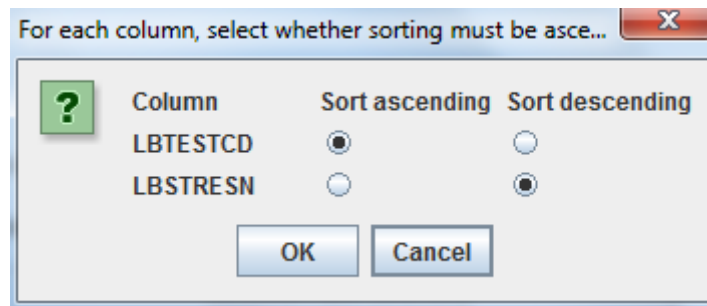
A dialog is then displayed allowing the user to define the sorting criteria:



which can be filled like e.g.:



stating that the user wants to sort first on LBTESTCD (the lab test code), and within that, on LBSTRESN (numeric standardized result). After clicking "OK", a new dialog is displayed, allowing the user to decide whether the sorting should be done ascending (A-Z in the case of character values) or descending. For example:

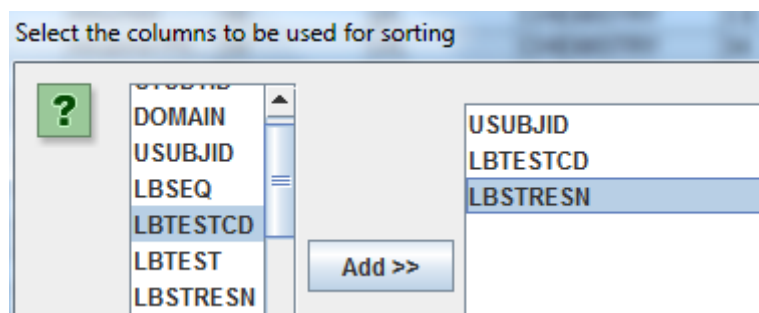


The final result being (after moving the LBSTRESN column to the left):

TA	TE	TI	TS	TV	DM	AE	CM	DS	EX	LB	MH	SV	VS	SUPPAE	SUPP
STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBSTRESN	LBSTRESU								
CDISCPIL01	LB	01-716-1103	133	ALB	Albumin	51	g/L								
CDISCPIL01	LB	01-703-1076	1	ALB	Albumin	49	g/L								
CDISCPIL01	LB	01-703-1210	115	ALB	Albumin	49	g/L								
CDISCPIL01	LB	01-703-1076	73	ALB	Albumin	48	g/L								
CDISCPIL01	LB	01-716-1024	200	ALB	Albumin	48	g/L								
CDISCPIL01	LB	01-701-1345	73	ALB	Albumin	47	g/L								
CDISCPIL01	LB	01-704-1351	155	ALB	Albumin	47	g/L								
CDISCPIL01	LB	01-708-1342	134	ALB	Albumin	47	g/L								
CDISCPIL01	LB	01-708-1342	278	ALB	Albumin	47	g/L								
CDISCPIL01	LB	01-716-1026	193	ALB	Albumin	47	g/L								
CDISCPIL01	LB	01-701-1341	1	ALB	Albumin	46	g/L								
CDISCPIL01	LB	01-701-1341	38	ALB	Albumin	46	g/L								
CDISCPIL01	LB	01-701-1442	1	ALB	Albumin	46	g/L								
CDISCPIL01	LB	01-704-1164	38	ALB	Albumin	46	g/L								
CDISCPIL01	LB	01-704-1266	1	ALB	Albumin	46	g/L								

showing that the highest albumin value found was 51 g/L for subject 01-716-1103.

Later we will see how one can select on a single test (or group of tests) and do the same kind of sorting. If one e.g. wants to have the values for albumin subject 01-716-1103 in a sorted order, one just adds "USUBJID" to the list of sorting criteria, by:

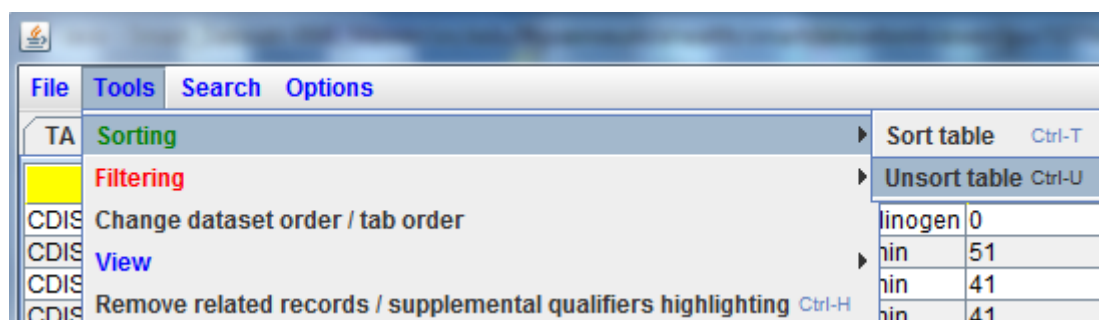


leading to:

TA	TE	TI	TS	TV	DM	AE	CM	DS	EX	LB	MH	SV	VS	SUPPAE	SUPPDM	SUPPDS	SUPPI
STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBSTRESN	LBSTRESU	LBSTRESU	LBSTRESU	LBSTRESU	LBSTRESU	LBSTRESU	LBSTRESU	LBSTRESU	LBSTRESU	LBSTRESU	LBSTRESU
CDISCPIL0T01	LB	01-716-1094	136	UROBIL	Urobilinogen	0											2013-01-26T11:20
CDISCPIL0T01	LB	01-716-1103	133	ALB	Albumin	51	g/L										2014-05-22T12:30
CDISCPIL0T01	LB	01-716-1103	1	ALB	Albumin	41	g/L										2014-03-07T12:30
CDISCPIL0T01	LB	01-716-1103	198	ALB	Albumin	41	g/L										2014-07-23T11:20
CDISCPIL0T01	LB	01-716-1103	73	ALB	Albumin	38	g/L										2014-04-22T10:25
CDISCPIL0T01	LB	01-716-1103	163	ALB	Albumin	38	g/L										2014-06-21T11:05
CDISCPIL0T01	LB	01-716-1103	38	ALB	Albumin	37	g/L										2014-04-05T10:00
CDISCPIL0T01	LB	01-716-1103	103	ALB	Albumin	37	g/L										2014-05-08T11:40
CDISCPIL0T01	LB	01-716-1103	270	ALB	Albumin	37	g/L										2014-09-09T07:40
CDISCPIL0T01	LB	01-716-1103	305	ALB	Albumin	37	g/L										2014-09-20T12:10
CDISCPIL0T01	LB	01-716-1103	228	ALB	Albumin	36	g/L										2014-08-19T10:40
CDISCPIL0T01	LB	01-716-1103	258	ALP	Alkaline Ph...	49	U/L										2014-08-30T12:00

Of course, one can also first filter on the subject (or on a group of subjects) first, and then do the same kind of sorting again.

To remove the sorting, use the menu "Tools - Sorting - Unsort table":



Also remark that there are keyboard shortcuts for both sorting and unsorting, using CTRL-T and CTRL-U.

### Filtering

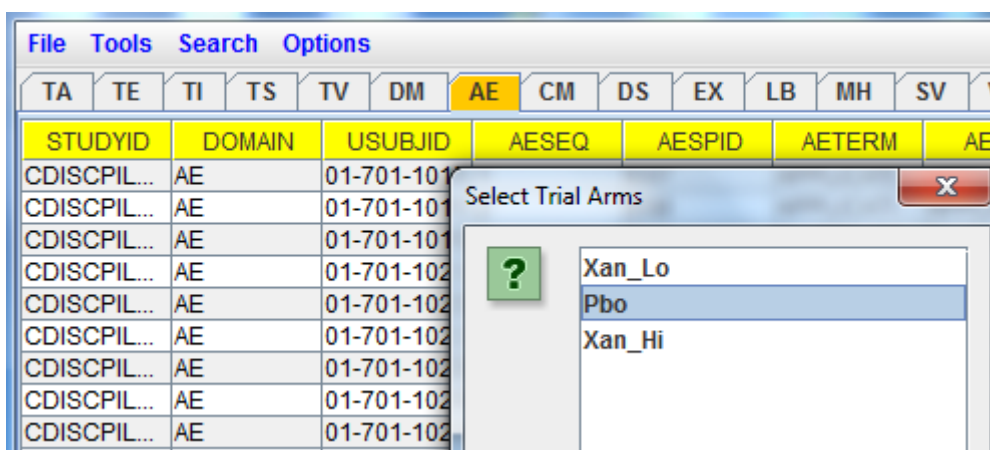
The second important feature once the dataset tables are loaded is filtering. The filtering is organized in such a way that one can apply a filter, then apply a second filter on the first one, a third one on the second one, etc. (filter stack). One can always undo either all filters, or the last one applied (i.e. returning to a lower filter on the stack).

Filtering is possible based on the values in the columns of the table itself, or even on subject properties that are not directly available in the loaded table.

Currently the following filters can be applied:

- filter by subject (on basis of USUBJID)
- filter on the topic variable (--TESTCD in case of Findings domain, --TERM in case of Events domains, and --TRT in case of Interventions domains)
- filter on the category (--CAT)
- filter on subject demographics properties (ACTARMCD, AGE, SEX in DM), in any table
- filter on any column values in the dataset

If you think that an important filter is missing in this list, please let us know, and we will see whether it can be added.

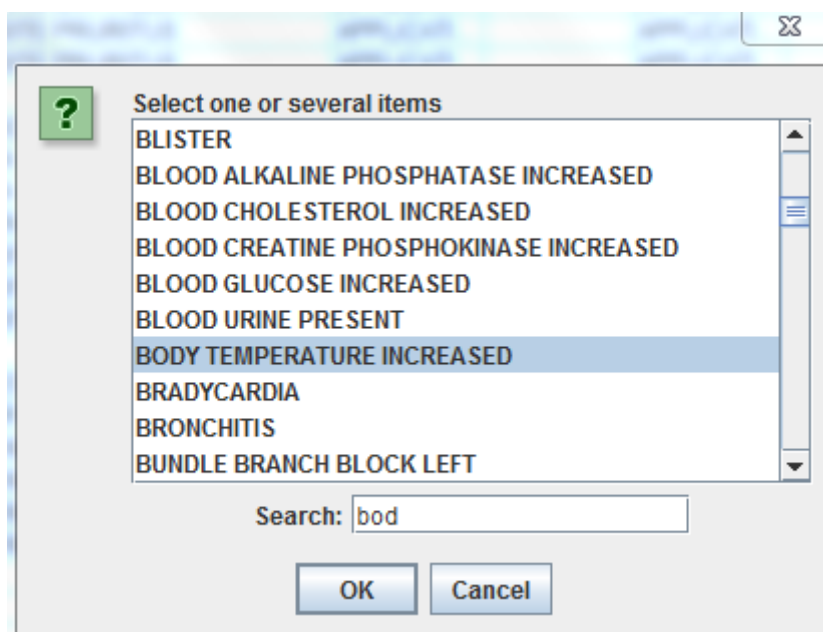


An obvious filter is one based on values for the topic variable (--TESTCD in "findings" domains, "-TERM" in "events" domains and "--TRT" in "interventions" domains). As we can sort on any column values, the latter can also be achieved by selecting the "--TESTCD", "--TERM" or "--TRT" variable when using the menu "Tools - Filtering - Filter on variable value".

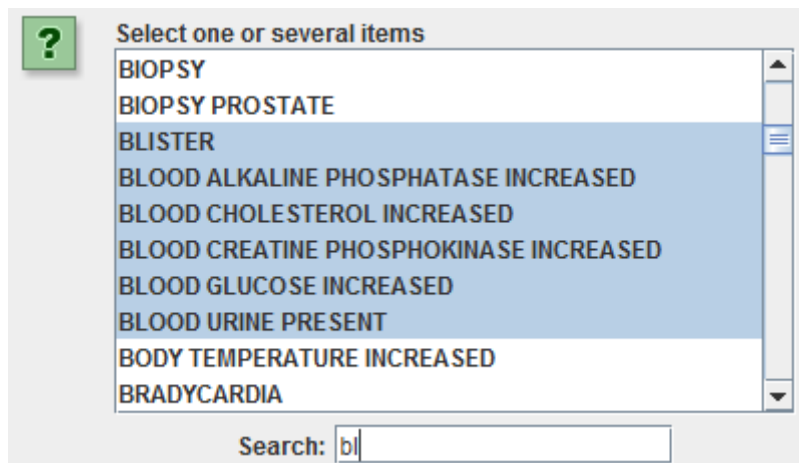
In the following, we will apply a number on filters on the records in the AE domain. First we will apply a filter on the AETERM (Reported Term for the Adverse Event).

So we can choose between using the menu "Tools - Filtering - Filter on topic variable" and "Tools - Filtering - Filter on variable value".

If we choose the former (using "Filter on topic variable"), the following dialog appears:

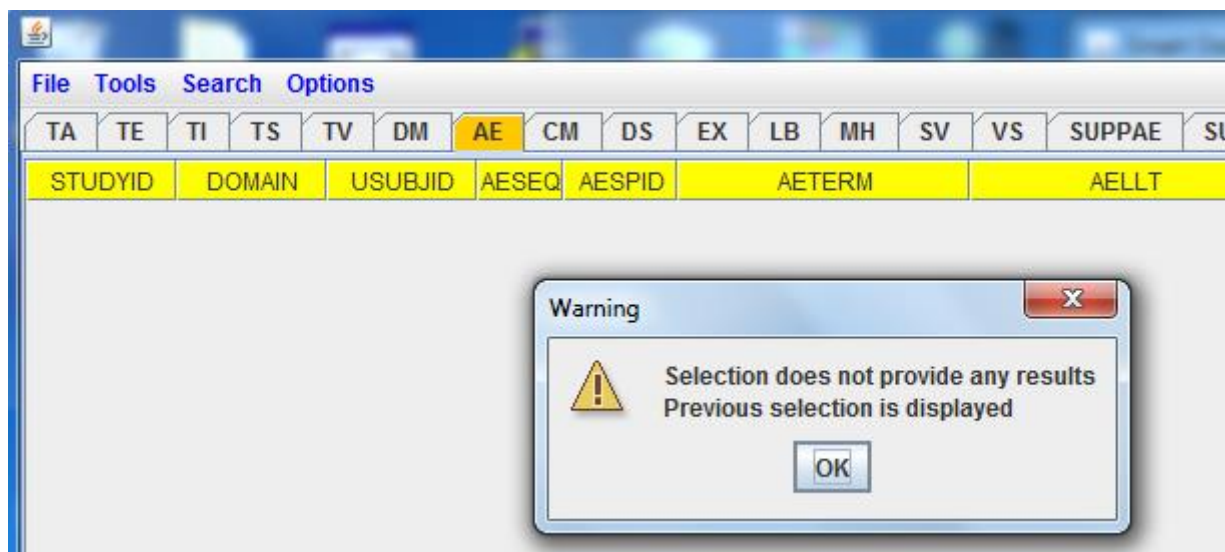


from which we can select "BODY TEMPERATURE INCREASED". One can easily find it, as the terms are sorted alphabetically and one can use the "Search" field: each time you enter a character, the possible terms are updated. For example, if one types "bl" in the search field, the selection updates to:



Remark that the list is either generated of all available values of AETERM in the original (unfiltered) table, or from the codelist in the define.xml in case the variable has an associated codelist.

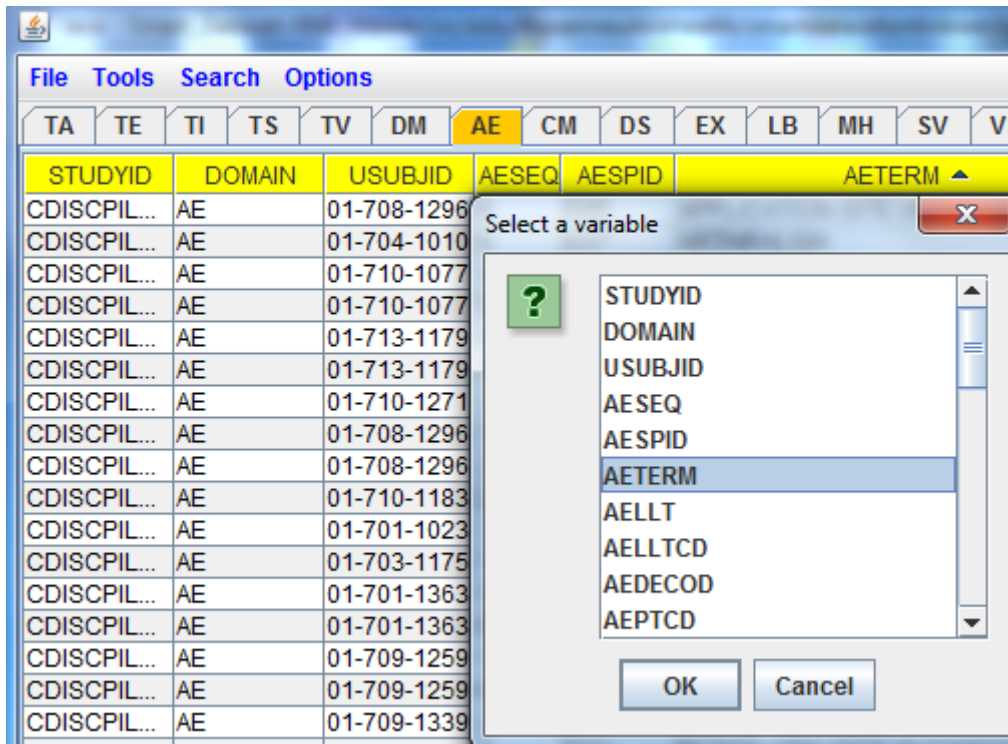
For the selection "BODY TEMPERATURE INCREASED", clicking "OK" leads to:



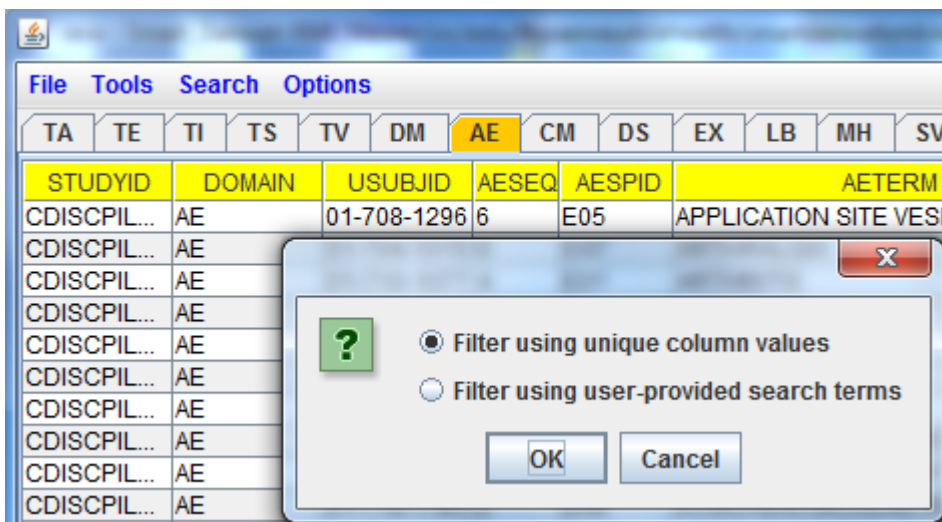
stating that none of the placebo-treated subjects had a increased body temperature AE reported. In such a case, the filter is removed, and the table is displayed again in its previous state.

So let us take another AETERM value, e.g. "blood urine present".

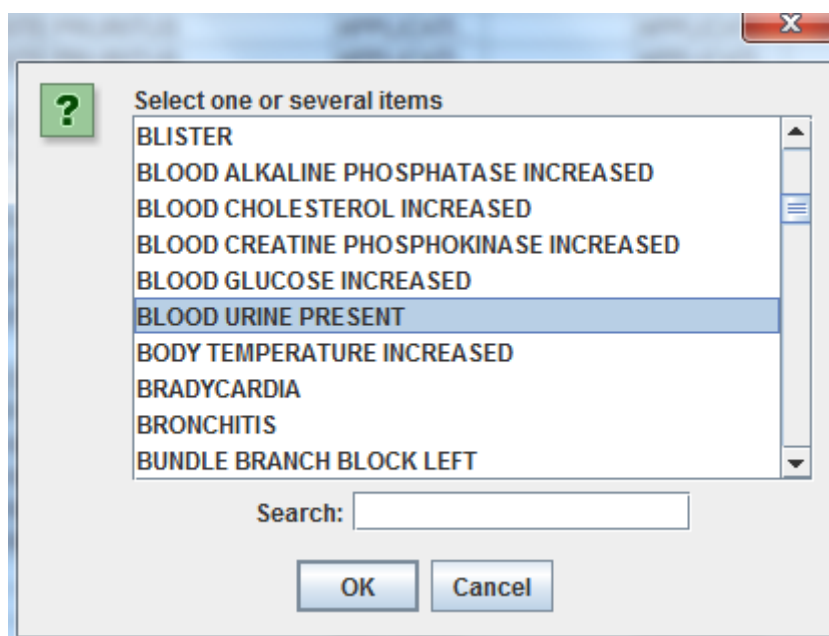
But this time, lets us start from the menu "Tools - Filtering - Filter on variable value". This leads to the following dialog:



showing us all the columns in the table. We then select "AETERM" leading to:



We can either filter on the unique values in the AETERM column, or by providing user-defined search criteria (not implemented yet). Selecting the former leads to a similar dialog as before:

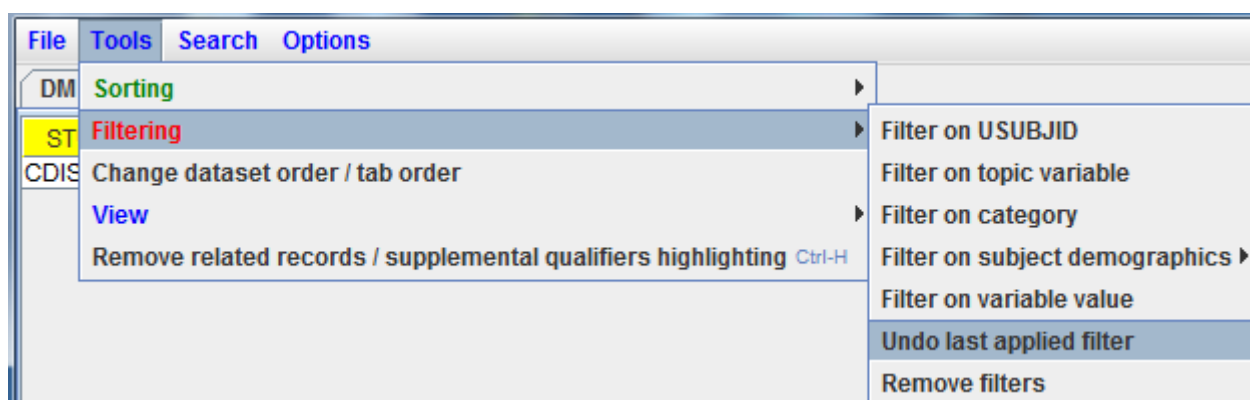


where we this time select "BLOOD URINE PRESENT". Also remark that one can select several items<sup>6</sup>. After clicking "OK", we get:

TA	TE	TI	TS	TV	DM	AE	CM	DS	EX	LB	MH	SV	VS	SUPPAE	SUPPDM	SUP
STUDYID	DOMAIN	USUBJID	AESEQ	AESPID	AETERM						AELLT					
CDISCPIL...	AE	01-709-1259	3	E06	BLOOD URINE PRESENT						BLOOD IN URINE					

showing that this was the case for one record only, for subject 01-709-1259.

We can now "undo" this filter again, using the menu "Tools - Filtering - Undo last applied filter":



The last applied filter is then removed, and the table is displayed in its previous state, which is the state where only records for placebo treated subjects are displayed.

<sup>6</sup>In future this becomes very interesting once we use "linked data", so that one can e.g. select a group of diseases that are related to each other, e.g. diseases that are known to be able to occur for a specific class of chemical entities. This will however require the regulatory authorities to move away from SAS-XPT, as there is no much chance to be able to implement such advanced features using the outdated XPT format.

If we filter on "ERYTHEMA" as AETERM value, we get:

STUDYID	DOMAIN	USUBJID	AESQ	AESPID	AETERM	AELLT
CDISCPIL...	AE	01-701-1023	1	E08	ERYTHEMA	ERYTHEMA
CDISCPIL...	AE	01-701-1023	2	E09	ERYTHEMA	LOCALIZED ERYTHEMA
CDISCPIL...	AE	01-701-1023	4	E08	ERYTHEMA	ERYTHEMA
CDISCPIL...	AE	01-705-1059	1	E02	ERYTHEMA	ERYTHEMA
CDISCPIL...	AE	01-708-1158	2	E07	ERYTHEMA	ERYTHEMA
CDISCPIL...	AE	01-710-1315	4	E05	ERYTHEMA	ERYTHEMA
CDISCPIL...	AE	01-710-1315	7	E05	ERYTHEMA	ERYTHEMA
CDISCPIL...	AE	01-714-1375	6	E09	ERYTHEMA	ERYTHEMA
CDISCPIL...	AE	01-714-1375	8	E09	ERYTHEMA	ERYTHEMA
CDISCPIL...	AE	01-716-1160	2	E03	ERYTHEMA	ERYTHEMA
CDISCPIL...	AE	01-716-1177	1	E03	ERYTHEMA	ERYTHEMA
CDISCPIL...	AE	01-716-1308	1	E01	ERYTHEMA	ERYTHEMA
CDISCPIL...	AE	01-716-1441	1	E06	ERYTHEMA	ERYTHEMA

We can now add another filter, e.g. we want to see which erythema events were observed in the first 10 days of the treatment (although it was a placebo treatment). So we use "Tools - Filtering - Filter on variable value" and this time select "AESTDY" (Adverse Event Start Day):

STUDYID	DOMAIN	USUBJID	AESQ	AESPID	AETERM	AESTDY	AELLT
CDISCPIL...	AE	01-701-1023	1	E08	ERYTHEMA		
CDISCPIL...	AE	01-701-1023	2	E09	ERYTHEMA		
CDISCPIL...	AE	01-701-1023	4	E08	ERYTHEMA		
CDISCPIL...	AE	01-705-1059	1	E02	ERYTHEMA		
CDISCPIL...	AE	01-708-1158	2	E07	ERYTHEMA		
CDISCPIL...	AE	01-710-1315	4	E05	ERYTHEMA		
CDISCPIL...	AE	01-710-1315	7	E05	ERYTHEMA		
CDISCPIL...	AE	01-714-1375	6	E09	ERYTHEMA		
CDISCPIL...	AE	01-714-1375	8	E09	ERYTHEMA		
CDISCPIL...	AE	01-716-1160	2	E03	ERYTHEMA		
CDISCPIL...	AE	01-716-1177	1	E03	ERYTHEMA		
CDISCPIL...	AE	01-716-1308	1	E01	ERYTHEMA		
CDISCPIL...	AE	01-716-1441	1	E06	ERYTHEMA		

Select a variable

- STUDYID
- DOMAIN
- USUBJID
- AESQ
- AESPID
- AETERM
- AESTDY**
- AELLT
- AELLTCD
- AEDECOD

OK Cancel

The system knows (from the define.xml) that AESTDY has the datatype "integer" and reacts accordingly by providing a dialog to select an upper and lower value of AESTDY:

Exact integer value:  
  
 Minimum-Maximum integer value:  
 Minimum  
 Maximum

One can either decide on filtering on an exact integer value, or on a range. In the latter case, leaving the "Minimum" field empty means "no minimum" and leaving the field "Maximum" empty means "no maximum". We however want to know which erythema adverse events occurred in the first 10 days of treatment, so we set:

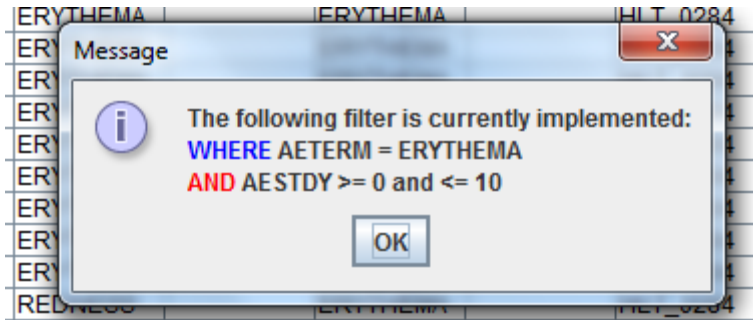
Exact integer value:  
  
 Minimum-Maximum integer value:  
 Minimum  
 Maximum

Remark that the fields are "idiot proof": it is not possible to add anything else in it than an integer number. Clicking "OK" applies our third filter on top of the prior filters, leading to:

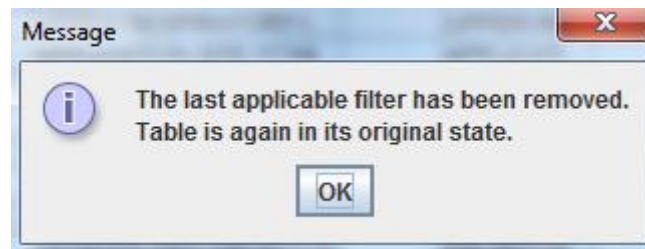
TA	TE	TI	TS	TV	DM	AE	CM	DS	EX	LB	MH	SV	VS	SUPPAE	SUPPDM	SUPPDS	SUPPLB
STUDYID	DOMAIN	USUBJID	AESSEQ	AESPID	AETERM	AESTDY	AELLT	AELLTCD	AEDECOD								
CDISCPIL01	AE	01-701-1023	1	E08	ERYTHEMA	3	ERYTHEMA		ERYTHEMA								
CDISCPIL01	AE	01-701-1023	2	E09	ERYTHEMA	3	LOCALIZE...		ERYTHEMA								
CDISCPIL01	AE	01-701-1023	4	E08	ERYTHEMA	3	ERYTHEMA		ERYTHEMA								
CDISCPIL01	AE	01-701-1097	1	E04	ERYTHEMA	3	ERYTHEMA		ERYTHEMA								
CDISCPIL01	AE	01-701-1188	1	E04	ERYTHEMA	2	ERYTHEMA		ERYTHEMA								
CDISCPIL01	AE	01-701-1188	3	E04	ERYTHEMA	2	ERYTHEMA		ERYTHEMA								
CDISCPIL01	AE	01-701-1239	1	E11	ERYTHEMA	2	ERYTHEMA		ERYTHEMA								
CDISCPIL01	AE	01-701-1239	6	E11	ERYTHEMA	2	ERYTHEMA		ERYTHEMA								
CDISCPIL01	AE	01-704-1266	2	E10	ERYTHEMA	5	ERYTHEMA		ERYTHEMA								
CDISCPIL01	AE	01-704-1266	9	E10	ERYTHEMA	5	ERYTHEMA		ERYTHEMA								
CDISCPIL01	AE	01-704-1266	12	E10	ERYTHEMA	5	ERYTHEMA	E10	ERYTHEMA								
CDISCPIL01	AE	01-706-1384	1	E03	ERYTHEMA	1	ERYTHEMA		ERYTHEMA								
CDISCPIL01	AE	01-706-1384	6	E03	ERYTHEMA	1	ERYTHEMA		ERYTHEMA								
CDISCPIL01	AE	01-706-1384	4	E01	ERYTHEMA	9	ERYTHEMA		ERYTHEMA								
CDISCPIL01	AE	01-706-1384	9	E01	ERYTHEMA	9	ERYTHEMA		ERYTHEMA								
CDISCPIL01	AE	01-710-1137	2	E05	ERYTHEMA	3	REDNESS		ERYTHEMA								
CDISCPIL01	AE	01-716-1418	2	E03	ERYTHEMA	1	ERYTHEMA		ERYTHEMA								

leading to a small number of records only.

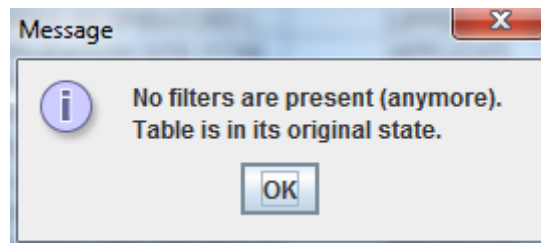
We can always see what filters have been applied using the menu "Tools - Filtering - Show current filters" (shortcut: CTRL-F). In this case, this leads to:



We can now always return to a prior state by using the menu "Tools - Filtering - Undo last applied filter". We can do several times, until we get to the original state of the table. When this occurs, the system displays a message:



If one then tries to "undo" again, the following message is displayed:



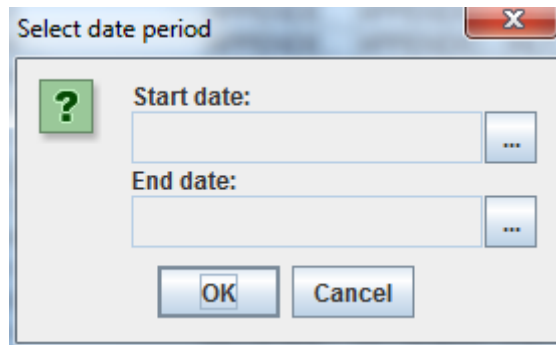
When applying a filter on a specific table column, the system knows about the datatype of that column (from the define.xml) and acts accordingly. For example in MH (medical history), there is the column "MHSTDTC" (Start Date/Time of Medical History Event) which usually is of datatype "partialDate". Some example values in our dataset are:

USUBJID	MHSEQ	MHSPID	MHTERM	MHSTDTC	MHLLT
01-701-1015	9		ALZHEIMER'S DISEASE	2010-04-30	
01-701-1015	1	E01	VERBATIM_0135		PALPITATI...
01-701-1015	8		VERBATIM_0140	1986	SUBTOTAL...
01-701-1015	2	E03	VERBATIM_0301		HEADACHE
01-701-1015	10		VERBATIM_0539	2012	GALLBLAD...
01-701-1015	3	E05	VERBATIM_0825		TINNITUS
01-701-1015	4	E02	VERBATIM_0841		HEARTBU...
01-701-1015	6		VERBATIM_1004	1973	THYROIDE...
01-701-1015	11		VERBATIM_1230	2013-12	SORE THR...
01-701-1015	7		VERBATIM_1716	1973	TONSILLE...
01-701-1015	5	E04	VERBATIM_1779		NUMBNES...
01-701-1023	16		ALZHEIMER'S DISEASE	2006-03-11	

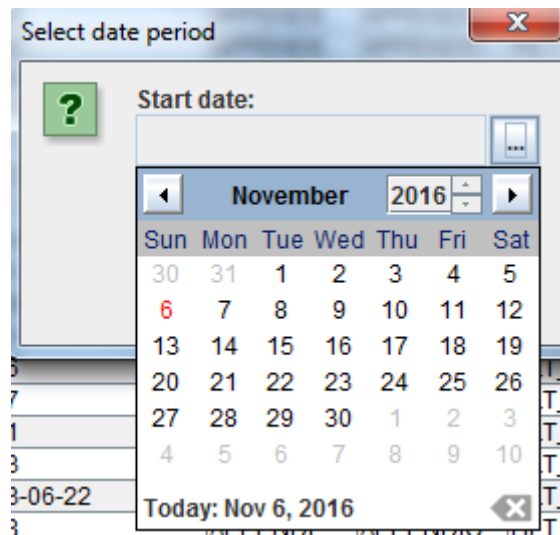
As one sees, in some case the exact data is known, whereas in other cases only the year and month or even only the year of the start of the medical history is known. If we e.g. now filter on "VERBATIM\_1775" and "VERBATIM\_1769" for "MHTERM", we obtain:

USUBJID	MHSEQ	MHSPID	MHTERM	MHSTDTC	MHLLT
01-701-1023	11		VERBATIM_1769	1978	SINUSITIS
01-701-1023	9		VERBATIM_1775	1956	APPENDE...
01-701-1028	4		VERBATIM_1775	1968	APPENDE...
01-701-1034	8		VERBATIM_1775	1949	APPENDE...
01-701-1047	7	E05	VERBATIM_1769		SINUSITIS
01-701-1097	3	E03	VERBATIM_1769		SINUSITIS
01-701-1111	13	E09	VERBATIM_1769		SINUSITIS
01-701-1111	14		VERBATIM_1775	1946	APPENDE...
01-701-1118	5	E04	VERBATIM_1769		SINUSITIS
01-701-1146	8	E04	VERBATIM_1769		SINUSITIS
01-701-1148	21	E13	VERBATIM_1769		SINUSITIS
01-701-1153	10		VERBATIM_1769	2011	SINUSITIS
01-701-1234	3		VERBATIM_1775	1949	APPENDE...
01-701-1239	11		VERBATIM_1775	1980	APPENDE...
01-701-1294	5	E04	VERBATIM_1769		SINUSITIS
01-701-1345	4		VERBATIM_1769	1963-06-22	SINUSITIS
01-701-1383	12		VERBATIM_1769	2006	SINUSITIS
01-701-1383	6		VERBATIM_1775	1969	APPENDE...
01-703-1258	7		VERBATIM_1775	1952	APPENDE...
01-704-1008	4		VERBATIM_1775	1956	APPENDE...
01-704-1114	3		VERBATIM_1775	1977-05	APPENDE...

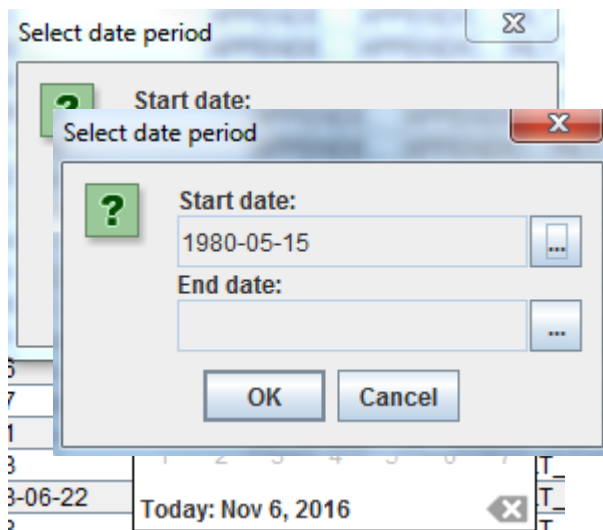
Suppose we now want to look for all start dates starting from mid-1980. So we generate a new filter on the column "MHSTDTC" (Medical History Start Date/Time). As the system knows from the define.xml that the values in the column are partial dates, it represents a dialog allowing us to pick start and end dates:



We only want to provide a start date (leaving "End date" empty again means "no limitation"), so we click on the "..." button under "Start date". A calendar is shown:



were we e.g. pick (using the available buttons), the date of May 15<sup>th</sup>, 1980:



and after clicking "OK" we get:

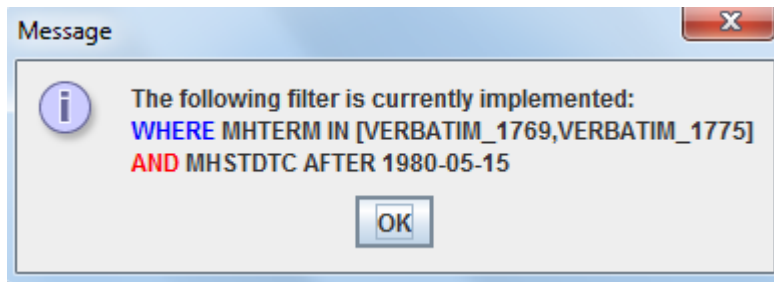
TV	DM	AE	CM	DS	EX	LB	MH	SV	VS	SUPPAE	SUPPDM	SUPPD:
USUBJID	MHSEQ	MHSPID	MHTERM	MHSTDTC	MHLLT							
01-701-1153	10		VERBATIM_1769	2011	SINUSITIS							
01-701-1239	11		VERBATIM_1775	1980	APPENDE...							
01-701-1383	12		VERBATIM_1769	2006	SINUSITIS							
01-708-1087	6		VERBATIM_1775	2008-08	APPENDE...							
01-714-1288	8		VERBATIM_1775	1988	APPENDE...							

Remark that the value "1980" is included although we set the start date on 1980-05-15. The reason is that we only know that the start date was "somewhere in 1980", so the selection is "forgiving" on this, and the record is included.

Additional datatype-sensitive filters are available for all the most important datatypes, a small number still being in development for some less used datatypes (like "time").

Using CTRL-F or the menu "Tools - Filtering - Show Filters" again show us the currently applied

filters:



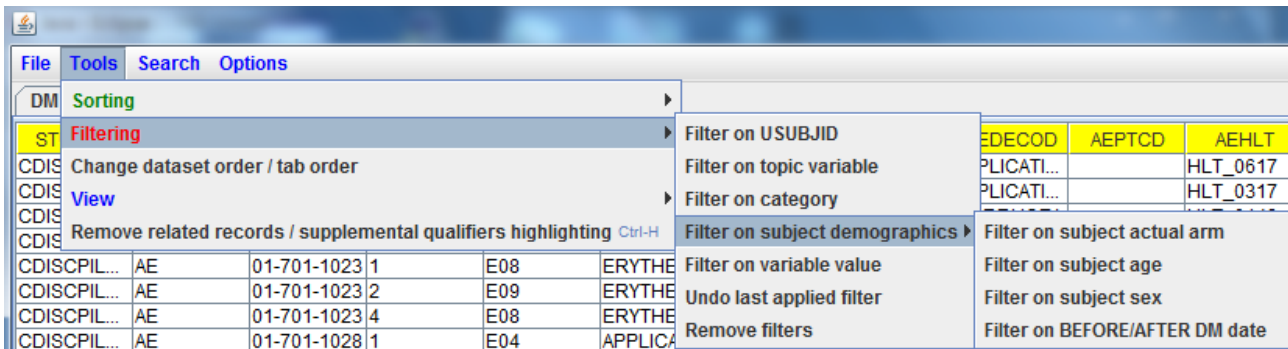
Again we can add additional filters, undo filters, sorting, etc..

To remove all filters, use the menu "Tools - Filtering - Remove filters", and to remove all sorting, use the menu "Tools - Sorting - Unsort table".

## Filtering on Subject Demographics Information

It is also possible to perform filtering based on demographics information of the subjects. Currently the filtering can be applied on the subject's actual trial arm (ACTARMCD), subject age (AGE), subject sex (SEX), and on any of the reference dates: RFSTDTC (study reference start date), RFENDTC (reference end date), RFXSTDTC (first study treatment), RFXENDTC (last study treatment), and RFICTDTC (date/time of informed consent).

These filters can be applied using the menu "Tools - Filtering - Filter on subject demographics":



When using the first option "Filter on subject actual arm" and the dataset "TA" (Trial Arms) is not loaded, the following dialog is displayed:



meaning that the system will look the possible trial arms from the DM table, instead of picking them up from the TA table. This also means that in the former case, the "arms" "screen failure" (SCRNFAIL) will appear in the list (if any), whereas in the latter case, only the real planned arms will appear.

The following dialog then appears:



allowing single or multiple selection (TODO: add tooltips with the arm full name). For example, when selecting "Pbo" (placebo arm) and the current table is the AE (Adverse Events) table, the filter

will only retain those subjects in the placebo arm for which an "adverse event" was reported<sup>7</sup>.

The checkbox "Apply to all tables" allows us to immediately also apply this filter to all other subject-related tables, so that we have the filters on all tables "synchronized".

The filtered table then looks like:

DM	AE	DS	EX	MH	SUPPAE	SUPPDM
STUDYID	DOMAIN	USUBJID	AESEQ	AESPID	AETERM	AELLT
CDISCPIL...	AE	01-701-1015	1	E07	APPLICATION SITE ERYTHEMA	APPLICATI...
CDISCPIL...	AE	01-701-1015	2	E08	APPLICATION SITE PRURITUS	APPLICATI...
CDISCPIL...	AE	01-701-1015	01-701-1015 (USUBJID)		DIARRHOEA	DIARRHEA
CDISCPIL...	AE	01-701-1023	ACTARMCD: Pbo		ATRIOVENTRICULAR BLOCK SECON...	AV BLOCK ...
CDISCPIL...	AE	01-701-1023	AGE: 61 years		ERYTHEMA	ERYTHEMA
CDISCPIL...	AE	01-701-1023	SEX: Female		ERYTHEMA	LOCALIZE...
CDISCPIL...	AE	01-701-1023	4	E08	ERYTHEMA	ERYTHEMA
CDISCPIL...	AE	01-701-1047	4	E09	BUNDLE BRANCH BLOCK LEFT	LEFT BUN...

In this case the "smart features" "show actual arm on USUBJID", "show age on USUBJID" and "show sex on USUBJID" were switched on before starting loading, so we also see by the tooltip that this subject was in the placebo arm after unblinding.

As we used the "Apply to all tables" checkbox, also the MH table displays the subjects of the placebo arm:

DM	AE	CM	DS	MH	SUPPAE	SUPPDM	SUPPDS
STUDYID	DOMAIN	USUBJID	MHSEQ	MHSPID			
CDISCPIL...	MH	01-701-1015	9				
CDISCPIL...	MH	01-701-1015	1	E01			
CDISCPIL...	MH	01-701-1015	8				
CDISCPIL...	MH	01-701-1015	01-701-1015 (USUBJID)				
CDISCPIL...	MH	01-701-1015	ACTARMCD: Pbo				
CDISCPIL...	MH	01-701-1015	AGE: 61 years				
CDISCPIL...	MH	01-701-1015	SEX: Female				
CDISCPIL...	MH	01-701-1015	6				

We can now apply a second filter on top of the first. For example, we want to filter on all subjects that were in the placebo arm and that are 65 years or older. So we use the second "demographics filter" by using the menu "Tools - Filtering - Filter on subject demographics" and then selecting "subject age":

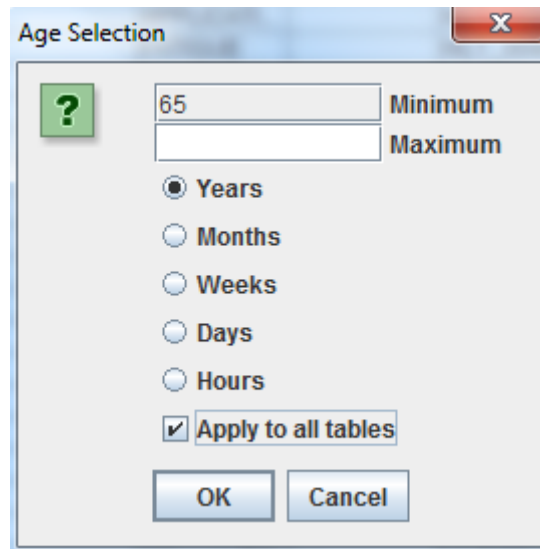
DM	AE	DS	EX	MH	SUPPAE	SUPPDM	SUPPDS
CDISCPIL...	AE	01-701-1023	1	E08	ERYT		
CDISCPIL...	AE	01-701-1023	2	E09	ERYT		
CDISCPIL...	AE	01-701-1023	4	E08	ERYT		
CDISCPIL...	AE	01-701-1047	4	E09	BUND		

Filtering	Filter on USUBJID	AELLT	AELLTCD	AEDECC
Filter on topic variable		PPLICATI...		APPLICAT
Filter on category		PPLICATI...		APPLICAT
Filter on subject demographics	Filter on subject actual arm			
Filter on variable value	Filter on subject age			
Undo last applied filter	Filter on subject sex			
Remove filters	Filter on BEFORE/AFTER DM date			

<sup>7</sup>This was a blinded study, so one cannot know whether this is an AE due to the study medication or not.

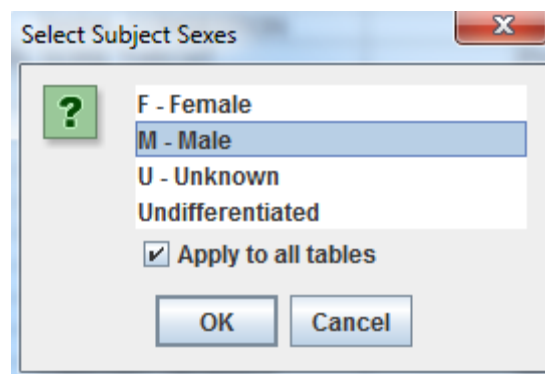
The following dialog then appears:



where we already typed in "65" for the minimum age (which we want to have in years). We leave the maximum age empty meaning that we do not want an upper limit. Also here, we check the checkbox "Apply to all tables" to keep the tables "synchronized". The result after clicking "OK" is:

DM	AE	DS	EX	MH	SUPPAE	SUPPDM
STUDYID	DOMAIN	USUBJID	AESEQ	AESPID	AETERM	
CDISCPIL...	AE	01-701-1047	4	E09	BUNDLE BRANCH BLOCK LEFT	
CDISCPIL...	AE	01-701-1047	1	E06	HIATUS HERNIA	
CDISCPIL...	AE	01-701-1047	01-701-1047 (USUBJID)		HIATUS HERNIA	
CDISCPIL...	AE	01-701-1047	ACTARMCD: Pbo		UPPER RESPIRATORY TRACT INFEC...	
CDISCPIL...	AE	01-701-1130	AGE: 85 years		EYE ALLERGY	
CDISCPIL...	AE	01-701-1130	SEX: Female		EYE PRURITUS	
CDISCPIL...	AE	01-701-1130	4	E08	EYE SWELLING	
CDISCPIL...	AE	01-701-1130	8	E12	NASAL CONGESTION	

Again, we can add a third filter on top of the second, e.g. when we would only like to see all male subjects of 65 year and older that were in the placebo arm. So, this time we choose for "Filter on subject sex", leading to:



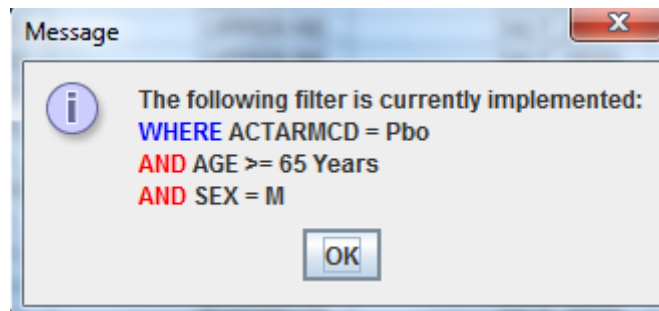
Also here, we could select more than one choice, but in this case, it does not make sense. Also remark that the choices come from the codelist in the define.xml that is associated with the DM variable "SEX".

After clicking OK, the additional filter is applied and the result is displayed:

DM	AE	DS	EX	MH	SUPPAE	SUPPDM
STUDYID	DOMAIN	USUBJID	AESQ	AESPID	AETERM	
CDISCPIL...	AE	01-701-1130	5	E09	EYE ALLERGY	
CDISCPIL...	AE	01-701-1130	6	E10	EYE PRURITUS	
CDISCPIL...	AE	01-701-1130	01-701-1130 (USUBJID)		EYE SWELLING	
CDISCPIL...	AE	01-701-1130	ACTARMCD: Pbo		NASAL CONGESTION	
CDISCPIL...	AE	01-701-1130	AGE: 84 years		PRURITUS	
CDISCPIL...	AE	01-701-1130	SEX: Male		PYREXIA	
CDISCPIL...	AE	01-701-1130	1	E06	URINARY TRACT INFECTION	

Showing us all the records for the male subjects in the placebo arm of age 65 and more.

Asking for how the complete filter looks now (CTRL-F or "Tools - Filtering - Show current filters") leads to:



One can also filter on any of the reference dates from the DM dataset: RFSTDTC (study reference start date), RFENDTC (reference end date), RFXSTDTC (first study treatment), RFXENDTC (last study treatment), and RFICTDTC (date/time of informed consent), by using the menu "Tools - Filtering - Filter on subject demographics - Filter on BEFORE/AFTER DM date". This leads to:

Filter Options

**?** The filter will take one of the following DM fields as a base and value of the selected date/time column in the present table will be compared with it. One can select whether the selected date/time must be before or on the selected DM variable date/time, or that it must be after the selected DM variable date/time.  
**REMARK: if the datetimes are not really comparable (e.g. 2006-12-13T09 with 2006-12-13T09:08) they are filtered out too**

Date/Time variable in current table:

AEDTC  
AESTDTC  
AEENDTC

DM variable to compare with:

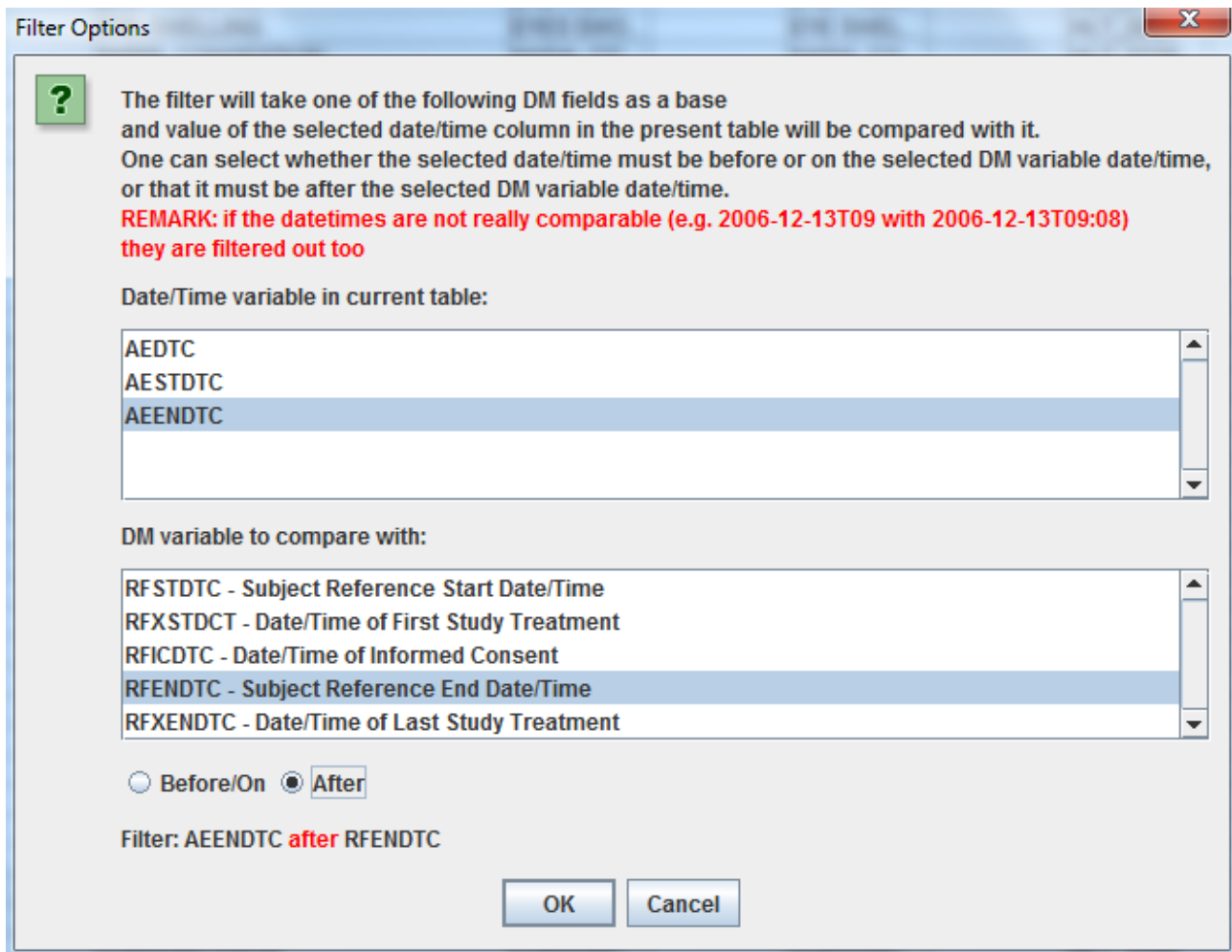
RFSTDTC - Subject Reference Start Date/Time  
RFXSTDCT - Date/Time of First Study Treatment  
RFICDTC - Date/Time of Informed Consent  
RFENDTC - Subject Reference End Date/Time  
RFXENDTC - Date/Time of Last Study Treatment

Before/On  After

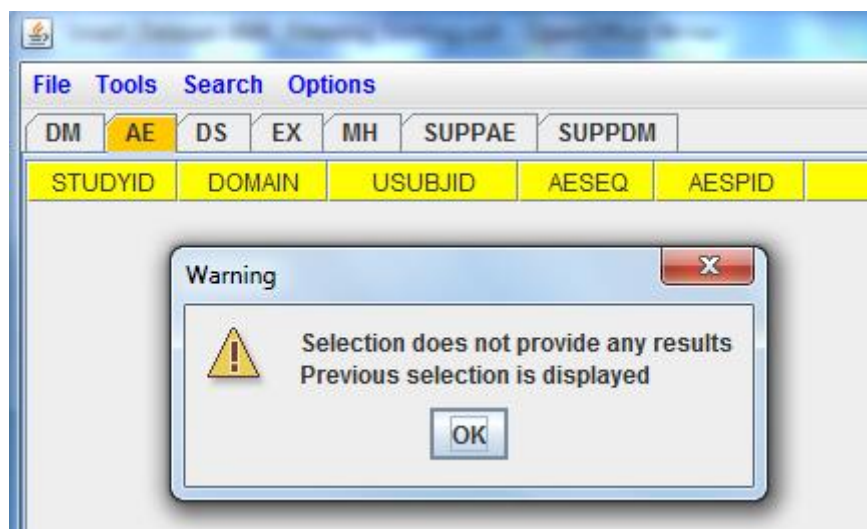
PLEASE SELECT YOUR CHOICES

OK Cancel

For example, we would like to (additionally) filter on those records where the adverse event was reported to have ended after the reference end date (RFENDTC). So we choose:



The label on the bottom also shows the "filter condition" in "human-readable language".  
 Clicking "OK" then leads to:



OK - there have been no reports of adverse events ending after the reference end date for male subjects of 65 years or older in the placebo arm.  
 After clicking "OK", the prior table is displayed again (i.e. showing all records for male subjects of 65 years or older in the placebo arm).  
 We could also ask for those records where the start of the adverse event was before or on the reference start date:

**Filter Options**

**?** The filter will take one of the following DM fields as a base and value of the selected date/time column in the present table will be compared with it. One can select whether the selected date/time must be before or on the selected DM variable date/time, or that it must be after the selected DM variable date/time.  
**REMARK: if the datetimes are not really comparable (e.g. 2006-12-13T09 with 2006-12-13T09:08) they are filtered out too**

Date/Time variable in current table:

- AEDTC
- AESTDTC**
- AEENDTC

DM variable to compare with:

- RFSTDTC - Subject Reference Start Date/Time**
- RFXSTDCT - Date/Time of First Study Treatment
- RFICDTC - Date/Time of Informed Consent
- RFENDTC - Subject Reference End Date/Time
- RFXENDTC - Date/Time of Last Study Treatment

Before/On  After

Filter: AESTDTC **before/on** RFSTDTC

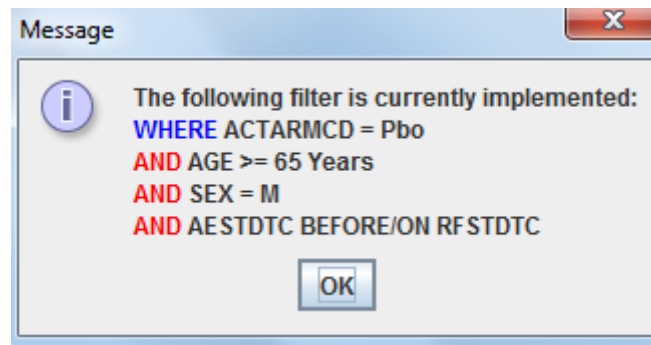
OK Cancel

leading to:

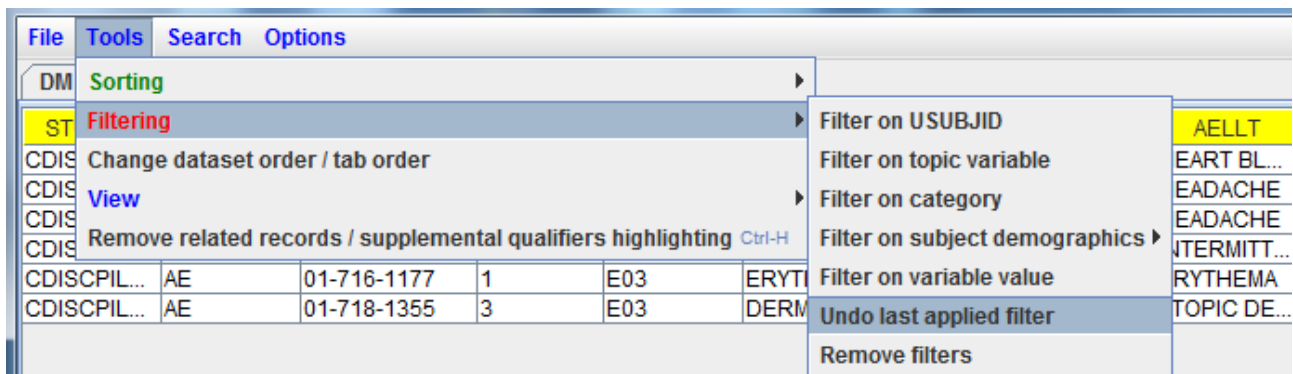
DM	AE	DS	EX	MH	SUPPAE	SUPPDM
STUDYID	DOMAIN	USUBJID	AESEQ	AESPID	AETERM	
CDISCPIL...	AE	01-703-1175	1	E01	ATRIOVENTRICULAR BLOCK SECON...	
CDISCPIL...	AE	01-704-1388	1	E01	HEADACHE	
CDISCPIL...	AE	01-709-1339	1	E04	HEADACHE	
CDISCPIL...	AE	01-711-1036	1	E03	HEADACHE	
CDISCPIL...	AE	01-716-1177	1	E03	ERYTHEMA	
CDISCPIL...	AE	01-718-1355	01-711-1036 (USUBJID)	DERMATITIS ATOPIC		
			ACTARMCD: Pbo			
			AGE: 70 years			
			SEX: Male			

with only 6 records that apply to the filter "all male subjects in the placebo arm of 65 years and older for which an adverse event was reported before or on the reference start date".

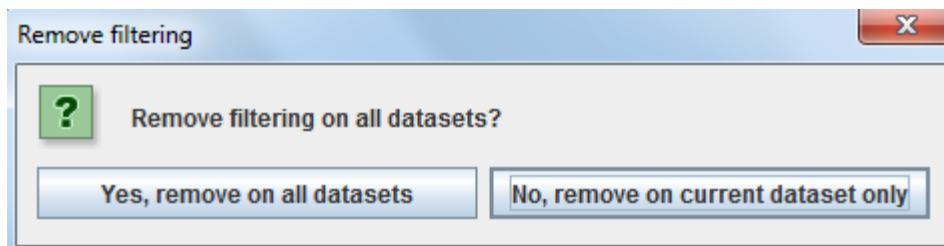
The filter description (using CTRL-F) now is:



Like before, one can now "undo" each filter separately, starting from the last implemented one, using the menu "Tools - Filtering - Undo last applied filter":

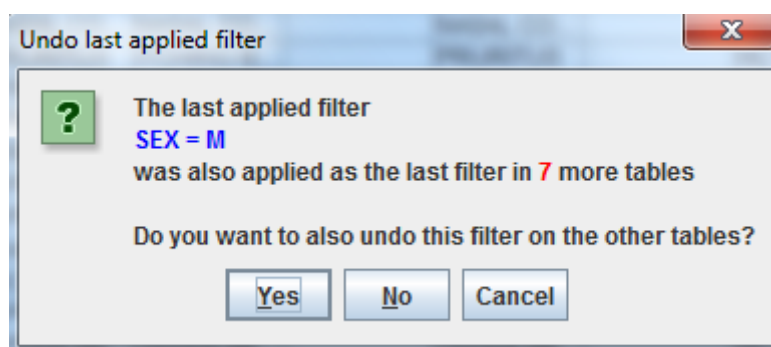


or remove all filters, thus returning to the original state by using the menu "Tools - Filtering - Remove filters", leading to the dialog:



The first choice is for the case that one has applied (different) filters on other tables, allowing to remove these too.

If one has applied a filter with the checkbox "Apply to all tables" checked, and one wants to "undo" the filter, a dialog like the following dialog will appear:



When clicking "Yes", this filter (if it was the last applied one) will be removed for all subject-

related tables (including the currently displayed one). If "No" is clicked, the last applied filter will only be removed for the currently displayed table.

### Conclusion

The "Smart Submission Dataset Viewer" has a very nice number of filtering features that (due to the build-in SDTM/SEND/ADaM knowledge) can be applied without any programming efforts. These features go considerably beyond what reviewers at regulatory authorities with their own tools.