ODM Generator

version 2.0 - User Manual

Date last update: 2018-09-02

Introduction

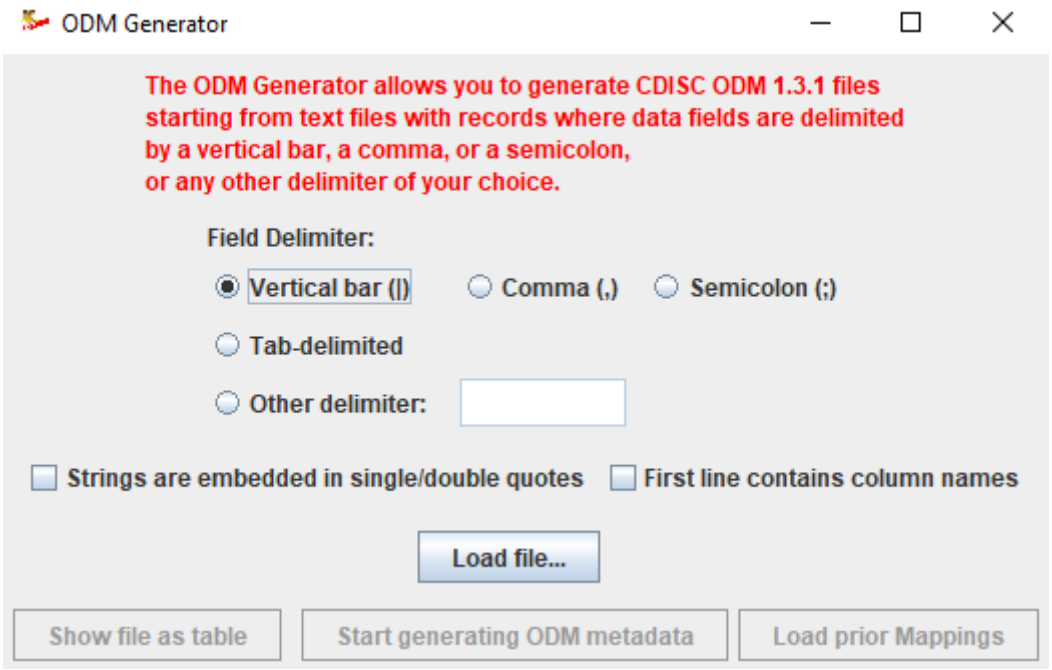
The ODM Generator is a new tool from XML4Pharma to transform data from "flat files", such as CSV files, exports from Excel and other worksheets, into CDISC ODM format.

These CDISC ODM files can then be used in other software tools, such as the popular SDTM-ETL software for the generation of CDISC SDTM or SEND datasets as well as the corresponding "define.xml" (coordinated approach). The ODM Generator has been developed on the request of SDTM-ETL users who had difficulties using non-ODM files (e.g. files send by laboratories) in the SDTM-ETL software. It enabled to move from a "puzzle of formats" of their clinical data and metadata to a single format that is at the same time the worldwide standard format for clinical data and metadata.

Starting the software and first steps

Navigate to the folder/directory where you installed the software. On windows, double-click the icon with the name "ODMGenerator.bat". On Linux or Linux-based operating systems, use "ODMGenerator.sh".

The following screen is then displayed:



The software allows you to convert "flat text files" to CDISC ODM format, where data fields are separated by a delimiter such as a vertical bar, a comma or semicolon or by a tab ("CSV" files), or any

other single character. In case the delimiter itself appears as text in a field, the text in the field (i.e. between the delimiters) should be in double quotes. In that case, it is advised that one also checks the checkbox "Strings are embedded in double quotes". The software will normally try to detect whether strings are embedded in quotes, but this cannot always be guaranteed.
For example, when you export data from an Excel file with the semicolon as the delimiter, cells that have a semicolon in the text will be exported with the cell content between quotes, such as:

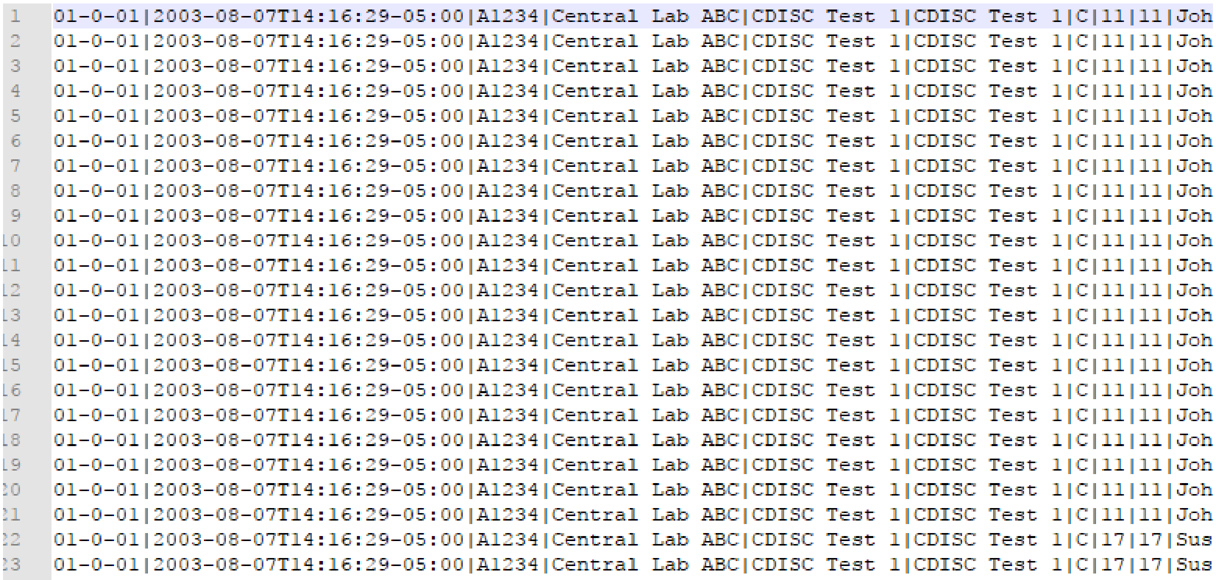
a;b;"c;d";e;f

The ODM Generator knows and understands this mechanism and will take appropriate action.

If in the data file, the first row contains the "column names", i.e. the first row should not be considered to be containing data, then one should also check the checkbox "First line contains column names".

Loading your data file

After having selected the delimiter (or having it defined in the field "other delimiter"), load the file that you want to have transformed into CDISC ODM.
In this user manual, we will use a sample "CDISC Lab v.1.0.1" file (see https://www.cdisc.org/standards/transport/lab) that uses the vertical bar "|" as delimiter. Here is a snapshot:

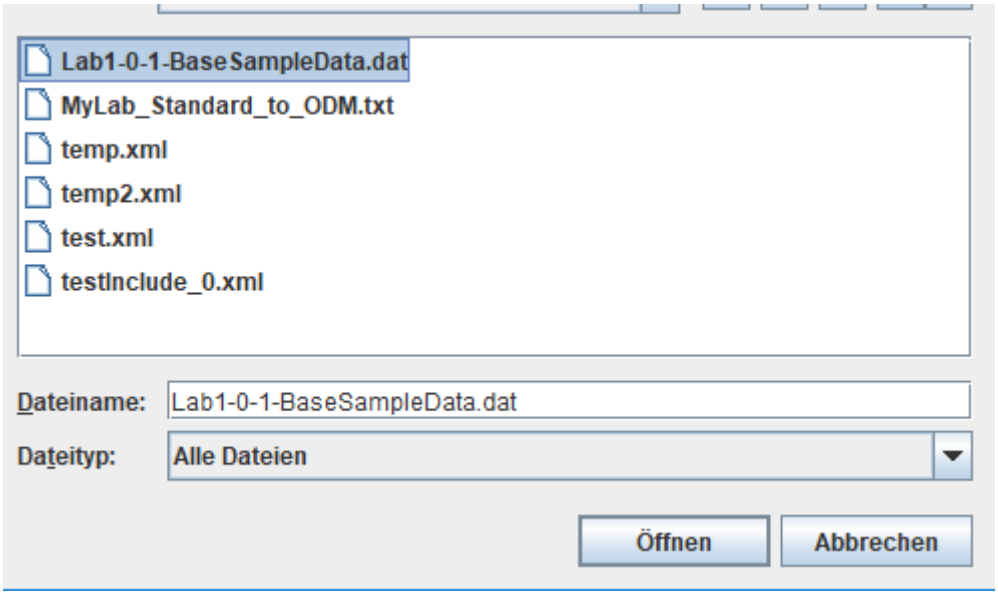


```
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|11|11|Joh
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|17|17|Sus
01-0-01|2003-08-07T14:16:29-05:00|A1234|Central Lab ABC|CDISC Test 1|CDISC Test 1|C|17|17|Sus
```

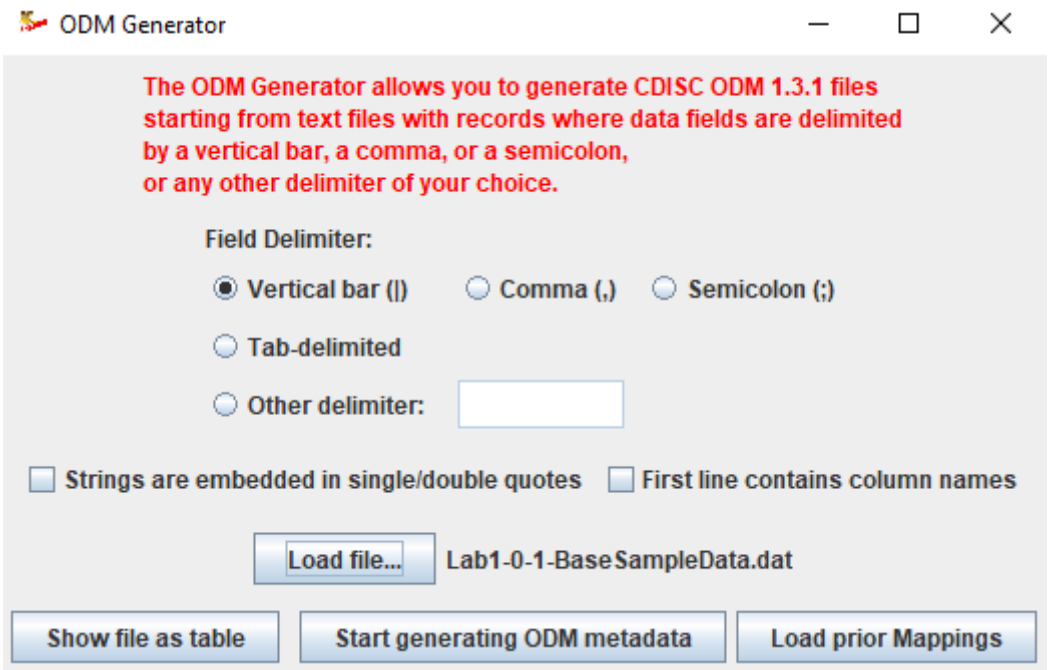
For a description of the fields, see https://www.cdisc.org/standards/transport/lab.

The application is not at all limited to Lab files, any type of "flat text" files with fields separated by a delimiter, including exports from Excel files, can be handled.

In order to load a file, use the button "Load file…". A file chooser shows up allowing you to select the file, e.g.:



After clicking "Ok", "Load" or the corresponding expression in your own language (depending on the operating system), the filename of the loaded file is displayed:



Generating metadata

You can now do one of three things:

a) Show the file as a table for inspection (recommended)
b) Start generating the ODM metadata
c) Load a file with definitions of ODM metadata for this type of loaded file

When working with a "flat text" file type, you will probably first want to see the data in order to understand what the data is about. In that case, click the button "Show file as table". In our case, this leads to:

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	11	11	John Smith, M.D.	8222			ABC
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	17	17	Susan Jones, M.D.	8277			RST
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	17	17	Susan Jones, M.D.	8277			RST
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	17	17	Susan Jones, M.D.	8277			RST
01-0-01	2003-08-07T14:16:29-05:00	A1234	Central Lab ABC	CDISC Test 1	CDISC Test 1	C	17	17	Susan Jones, M.D.	8277			RST

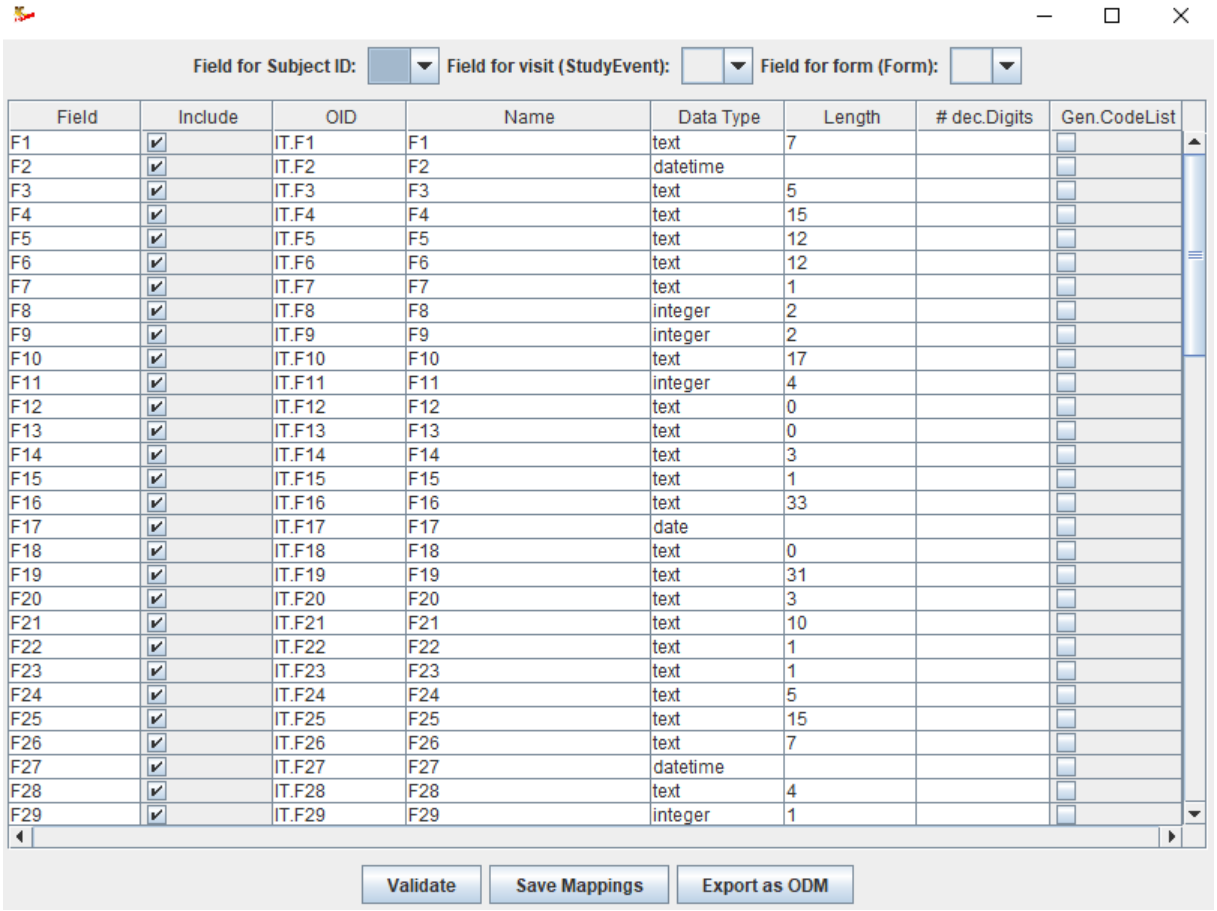
We can already see that field 2 (F2) represents a date & timestamp, and that field 4 (F4) probably represents the name of the laboratory. We need to look in the specification document for the details however.

If however, the first line in the source file contains the column names, and you have checked the checkbox "First line contains column names", one will see the original column names as a tooltip on the column header. For example:

F9	F10	F11
11	John Smith, M.D.	8222
11	John Smith, M.D.	8222

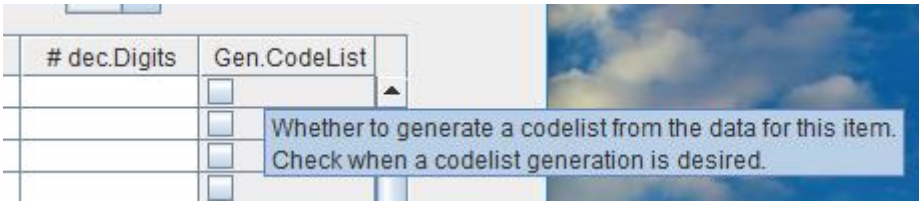
Investigator Name

When using the button "Start generating ODM metadata", the system will analyze the data and make proposals for its metadata, especially the datatype (according to the ODM standard) and the maximal field length. When clicking the button "Start generating ODM metadata", in our case this leads to a new window being opened with the following proposed metadata:



Field	Include	OID	Name	Data Type	Length	# dec.Digits	Gen.CodeList
F1	✔	IT.F1	F1	text	7		
F2	✔	IT.F2	F2	datetime			
F3	✔	IT.F3	F3	text	5		
F4	✔	IT.F4	F4	text	15		
F5	✔	IT.F5	F5	text	12		
F6	✔	IT.F6	F6	text	12		
F7	✔	IT.F7	F7	text	1		
F8	✔	IT.F8	F8	integer	2		
F9	✔	IT.F9	F9	integer	2		
F10	✔	IT.F10	F10	text	17		
F11	✔	IT.F11	F11	integer	4		
F12	✔	IT.F12	F12	text	0		
F13	✔	IT.F13	F13	text	0		
F14	✔	IT.F14	F14	text	3		
F15	✔	IT.F15	F15	text	1		
F16	✔	IT.F16	F16	text	33		
F17	✔	IT.F17	F17	date			
F18	✔	IT.F18	F18	text	0		
F19	✔	IT.F19	F19	text	31		
F20	✔	IT.F20	F20	text	3		
F21	✔	IT.F21	F21	text	10		
F22	✔	IT.F22	F22	text	1		
F23	✔	IT.F23	F23	text	1		
F24	✔	IT.F24	F24	text	5		
F25	✔	IT.F25	F25	text	15		
F26	✔	IT.F26	F26	text	7		
F27	✔	IT.F27	F27	datetime			
F28	✔	IT.F28	F28	text	4		
F29	✔	IT.F29	F29	integer	1		

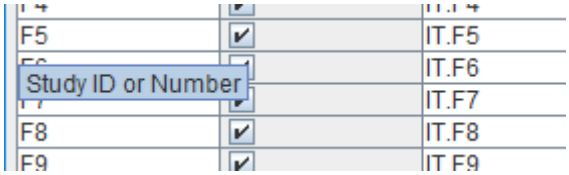
Although the column names are pretty self-explanatory, one can always obtain more information by holding the mouse over a column header. For example, for "Gen. CodeList":



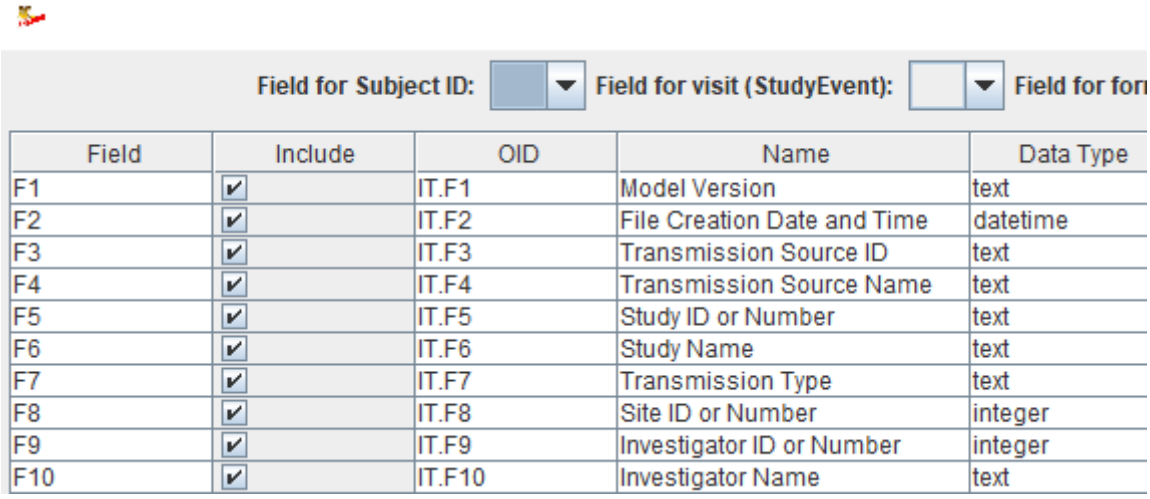
The "Generate CodeList" feature is a new feature in version 2 of the software. For each of the variables, when the checkbox is checked, the data is analyzed and a codelist is generated from those data. For example, for the field 15 (F15), from the specification and the data itself, we see that this represents the sex of the subject, with only two possible values: "F" (female) and "M" (male). If the "Generate CodeList" checkbox for F15 is checked, a codelist will be generated containing these two values, and will be associated to the item definition.

For the first field (F1) it proposes to assign the datatype "text" with a maximal length of 7 (this is the length of the longest value for this field found in the file). For the second field, it proposes to assign the datatype "datetime". In that case the maximal length should not be assigned.

Also here, when the first line of the source file contains the column headers, and you checked the checkbox "First line contains column names", one will see the original column names as a tooltip on each of the first cells in the rows. For example:

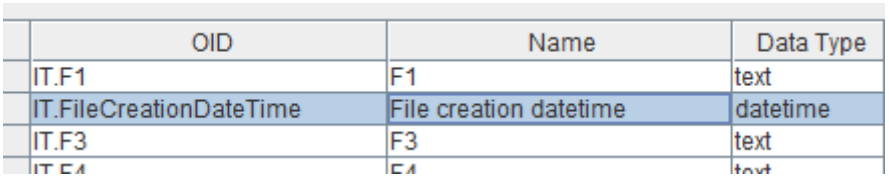


At the same time, the 4th field, field "Name" will be populated with the column name of the source file when the checkbox "First line contains column names" was checked. For example:



Field	Include	OID	Name	Data Type
F1	✔	IT.F1	Model Version	text
F2	✔	IT.F2	File Creation Date and Time	datetime
F3	✔	IT.F3	Transmission Source ID	text
F4	✔	IT.F4	Transmission Source Name	text
F5	✔	IT.F5	Study ID or Number	text
F6	✔	IT.F6	Study Name	text
F7	✔	IT.F7	Transmission Type	text
F8	✔	IT.F8	Site ID or Number	integer
F9	✔	IT.F9	Investigator ID or Number	integer
F10	✔	IT.F10	Investigator Name	text

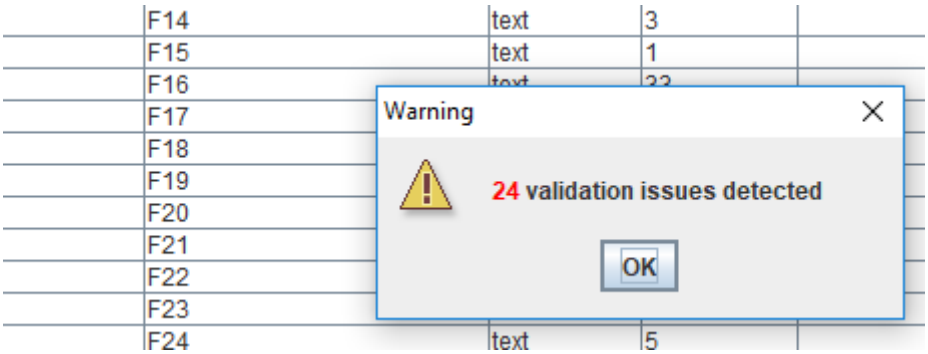
CDISC ODM uses "OIDs" which are identifiers. As a first guess, the identifier "IT.F" followed by the field number is assigned, and "Fx" (where x being the field number) is assigned for the "Name". If we e.g. know that the second field represents the "file creation datetime", we can make the changes as follows:



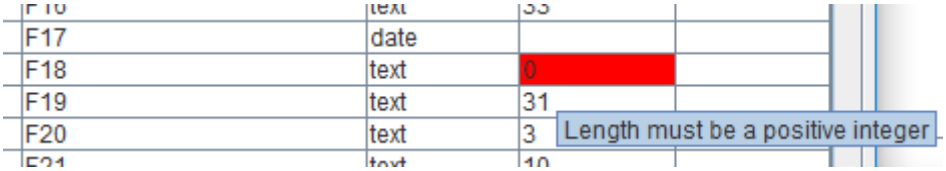
OID	Name	Data Type
IT.F1	F1	text
IT.FileCreationDateTime	File creation datetime	datetime
IT.F3	F3	text

Remark that OIDs are arbitrary and only used as identifiers in references. So, if you want to use "abracadabra" for the OID for the second field, that's OK. I personally prefer to assign something meaningful to OIDs.

You can now start assigning metadata by editing the table. As there are certain rules in ODM, you can do a validation on your assignments using the "**Validate**" button. In our example this will lead to a warning:



After clicking "OK", the assignments for which there is an issue will be highlighted, and a tooltip with give some details. For example:



In this case, the underlying reason is that there were no values at all for field 19 in the file, so a maximum length could not be assigned. CDISC ODM however always want us to have a maximum length assigned when the datatype is "text"[1].

We can always store our assignments to a file for later use and reuse. So, these assignments need to be done only once for a specific file type, and can be reused over and over again. When clicking the button "**Save mappings**", a file chooser will pop up to choose a file allowing you to save everything to a text file. In our case, the contents of this file look like:

```
NumFields=92
SkipFirstLine=false
Field=F1 Include=true OID="IT.F1" Name="F1" DataType=text Length=7 GenerateCodeList=false
Field=F2 Include=true OID="IT.FileCreationDateTime" Name="File creation datetime" DataType=datetime GenerateCodeList=false
Field=F3 Include=true OID="IT.F3" Name="F3" DataType=text Length=5 GenerateCodeList=false
Field=F4 Include=true OID="IT.F4" Name="F4" DataType=text Length=15 GenerateCodeList=false
Field=F5 Include=true OID="IT.F5" Name="F5" DataType=text Length=12 GenerateCodeList=false
Field=F6 Include=true OID="IT.F6" Name="F6" DataType=text Length=12 GenerateCodeList=false
Field=F7 Include=true OID="IT.F7" Name="F7" DataType=text Length=1 GenerateCodeList=false
Field=F8 Include=true OID="IT.F8" Name="F8" DataType=integer Length=2 GenerateCodeList=false
Field=F9 Include=true OID="IT.F9" Name="F9" DataType=integer Length=2 GenerateCodeList=false
Field=F10 Include=true OID="IT.F10" Name="F10" DataType=text Length=17 GenerateCodeList=false
Field=F11 Include=true OID="IT.F11" Name="F11" DataType=integer Length=4 GenerateCodeList=false
Field=F12 Include=true OID="IT.F12" Name="F12" DataType=text Length=0 GenerateCodeList=false
Field=F13 Include=true OID="IT.F13" Name="F13" DataType=text Length=0 GenerateCodeList=false
Field=F14 Include=true OID="IT.F14" Name="F14" DataType=text Length=3 GenerateCodeList=false
Field=F15 Include=true OID="IT.F15" Name="F15" DataType=text Length=1 GenerateCodeList=true
Field=F16 Include=true OID="IT.F16" Name="F16" DataType=text Length=33 GenerateCodeList=false
Field=F17 Include=true OID="IT.F17" Name="F17" DataType=date GenerateCodeList=false
Field=F18 Include=true OID="IT.F18" Name="F18" DataType=text Length=0 GenerateCodeList=false
Field=F19 Include=true OID="IT.F19" Name="F19" DataType=text Length=31 GenerateCodeList=false
Field=F20 Include=true OID="IT.F20" Name="F20" DataType=text Length=3 GenerateCodeList=false
Field=F21 Include=true OID="IT.F21" Name="F21" DataType=text Length=10 GenerateCodeList=false
```

It has a very simple format, so can also be edited offline, outside the tool.

In our case, the sample file is a "flat text" file that obtains the "CDISC Lab v.1.0.1" standard. In your case, this may be a company-internal standard, or a standardized format from your provider.

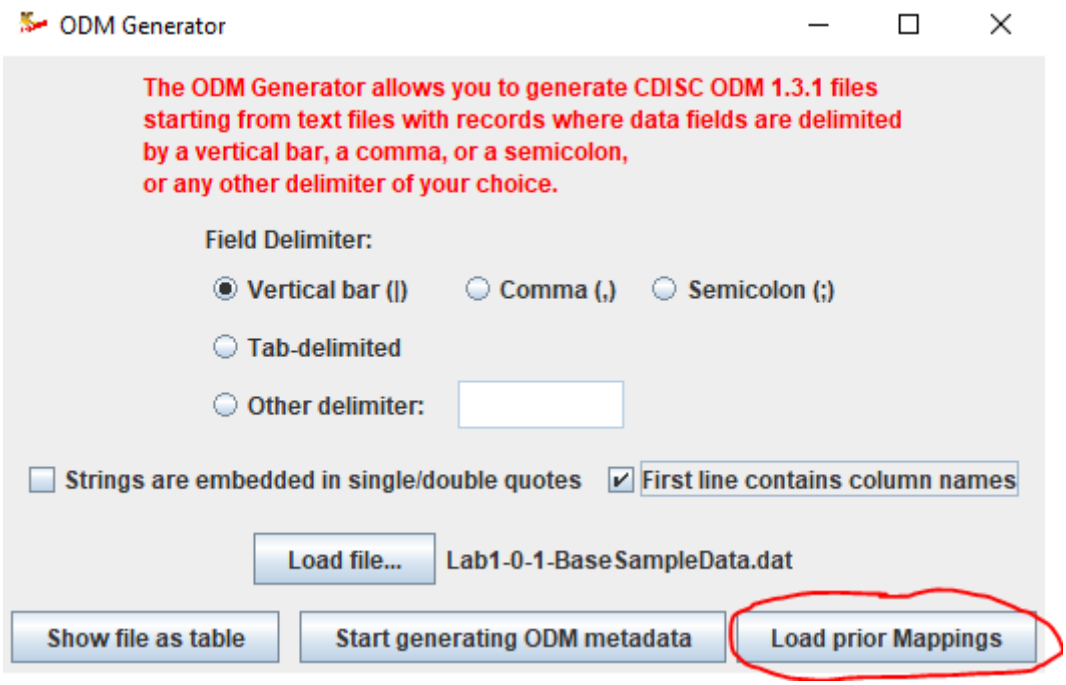
We will now continue with a "mappings" file where all the mappings with "CDISC Lab v.1.0.1" standard. As already stated, these assignments can either be done within the tool (followed by saving the mappings to file), or by editing an exported "mappings file".

In our case, our file with assignments and mappings looks like:

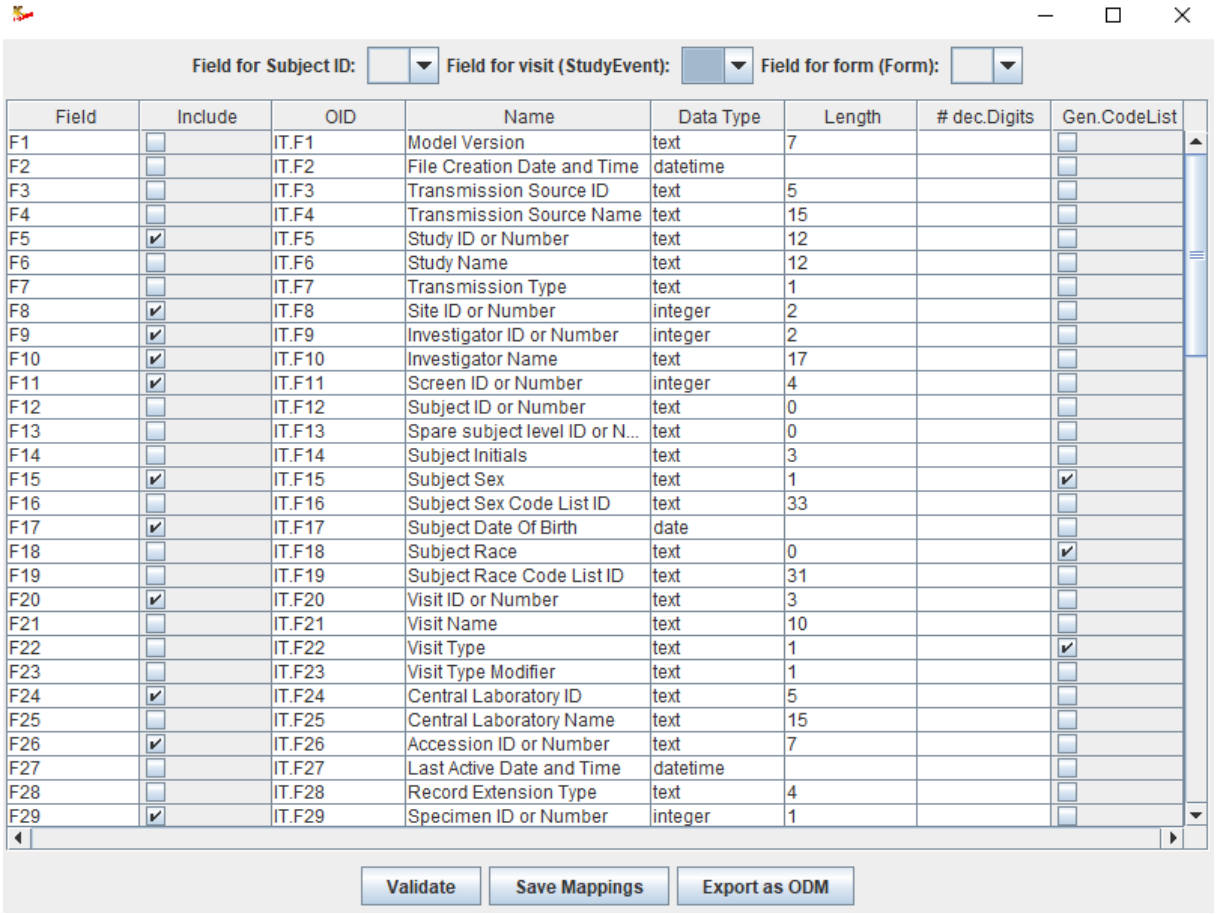
[1] This has to do with the ODM metadata being used as a specification for relational database table, i.e. assigning the VARCHAR(n) datatype.

```
Field=F1 Include=false OID="IT.F1" Name="Model Version" DataType=text Length=7 GenerateCodeList=false
Field=F2 Include=false OID="IT.F2" Name="File Creation Date and Time" DataType=datetime GenerateCodeList=false
Field=F3 Include=false OID="IT.F3" Name="Transmission Source ID" DataType=text Length=5 GenerateCodeList=false
Field=F4 Include=false OID="IT.F4" Name="Transmission Source Name" DataType=text Length=15 GenerateCodeList=false
Field=F5 Include=true OID="IT.F5" Name="Study ID or Number" DataType=text Length=12 GenerateCodeList=false
Field=F6 Include=false OID="IT.F6" Name="Study Name" DataType=text Length=12 GenerateCodeList=false
Field=F7 Include=false OID="IT.F7" Name="Transmission Type" DataType=text Length=1 GenerateCodeList=false
Field=F8 Include=true OID="IT.F8" Name="Site ID or Number" DataType=integer Length=2 GenerateCodeList=false
Field=F9 Include=true OID="IT.F9" Name="Investigator ID or Number" DataType=integer Length=2 GenerateCodeList=false
Field=F10 Include=true OID="IT.F10" Name="Investigator Name" DataType=text Length=17 GenerateCodeList=false
Field=F11 Include=true OID="IT.F11" Name="Screen ID or Number" DataType=integer Length=4 GenerateCodeList=false
Field=F12 Include=false OID="IT.F12" Name="Subject ID or Number" DataType=text Length=0 GenerateCodeList=false
Field=F13 Include=false OID="IT.F13" Name="Spare subject level ID or Number" DataType=text Length=0 GenerateCodeList=false
Field=F14 Include=false OID="IT.F14" Name="Subject Initials" DataType=text Length=3 GenerateCodeList=false
Field=F15 Include=true OID="IT.F15" Name="Subject Sex" DataType=text Length=1 GenerateCodeList=true
Field=F16 Include=false OID="IT.F16" Name="Subject Sex Code List ID" DataType=text Length=33 GenerateCodeList=false
Field=F17 Include=true OID="IT.F17" Name="Subject Date Of Birth" DataType=date GenerateCodeList=false
Field=F18 Include=false OID="IT.F18" Name="Subject Race" DataType=text Length=0 GenerateCodeList=true
Field=F19 Include=false OID="IT.F19" Name="Subject Race Code List ID" DataType=text Length=31 GenerateCodeList=false
Field=F20 Include=true OID="IT.F20" Name="Visit ID or Number" DataType=text Length=3 GenerateCodeList=false
Field=F21 Include=false OID="IT.F21" Name="Visit Name" DataType=text Length=10 GenerateCodeList=false
Field=F22 Include=false OID="IT.F22" Name="Visit Type" DataType=text Length=1 GenerateCodeList=true
```

This file can now be loaded in the software using the button "Load prior mappings":



Which will display our assignments as:



Field	Include	OID	Name	Data Type	Length	# dec.Digits	Gen.CodeList
F1	☐	IT.F1	Model Version	text	7		☐
F2	☐	IT.F2	File Creation Date and Time	datetime			☐
F3	☐	IT.F3	Transmission Source ID	text	5		☐
F4	☐	IT.F4	Transmission Source Name	text	15		☐
F5	☑	IT.F5	Study ID or Number	text	12		☐
F6	☐	IT.F6	Study Name	text	12		☐
F7	☐	IT.F7	Transmission Type	text	1		☐
F8	☑	IT.F8	Site ID or Number	integer	2		☐
F9	☑	IT.F9	Investigator ID or Number	integer	2		☐
F10	☑	IT.F10	Investigator Name	text	17		☐
F11	☑	IT.F11	Screen ID or Number	integer	4		☐
F12	☐	IT.F12	Subject ID or Number	text	0		☐
F13	☐	IT.F13	Spare subject level ID or N...	text	0		☐
F14	☐	IT.F14	Subject Initials	text	3		☐
F15	☑	IT.F15	Subject Sex	text	1		☑
F16	☐	IT.F16	Subject Sex Code List ID	text	33		☐
F17	☑	IT.F17	Subject Date Of Birth	date			☐
F18	☐	IT.F18	Subject Race	text	0		☑
F19	☐	IT.F19	Subject Race Code List ID	text	31		☐
F20	☑	IT.F20	Visit ID or Number	text	3		☐
F21	☐	IT.F21	Visit Name	text	10		☐
F22	☐	IT.F22	Visit Type	text	1		☑
F23	☐	IT.F23	Visit Type Modifier	text	1		☐
F24	☑	IT.F24	Central Laboratory ID	text	5		☐
F25	☐	IT.F25	Central Laboratory Name	text	15		☐
F26	☑	IT.F26	Accession ID or Number	text	7		☐
F27	☐	IT.F27	Last Active Date and Time	datetime			☐
F28	☐	IT.F28	Record Extension Type	text	4		☐
F29	☑	IT.F29	Specimen ID or Number	integer	1		☐

Showing that some for some fields (15, 18, 22), a codelist will be generated from the data, and that for some fields (1-4. 6-7, …) the information will not be included in the ODM file later. See the section "Filtering" for further information.

It is always a good idea to use the "Validate" button at this time.

As already stated, you will only need to develop these assignments only once for each "flat text" file type in your organization.

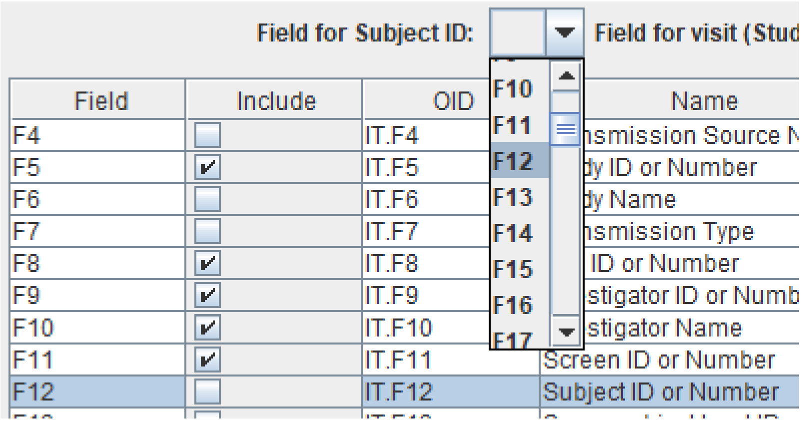
Assigning fields for visit ID or number, form ID or number, and subform ID or number

ODM organizes the data per visit (in ODM named "StudyEvent"), per form, and per subform (named "ItemGroup"). We now need to assign (some of these) from the fields in the source file.
Adding this information essentially allows to go from a flat 2-dimensional representation of the data (our source data) to a multidimensional representation: the CDISC ODM format.

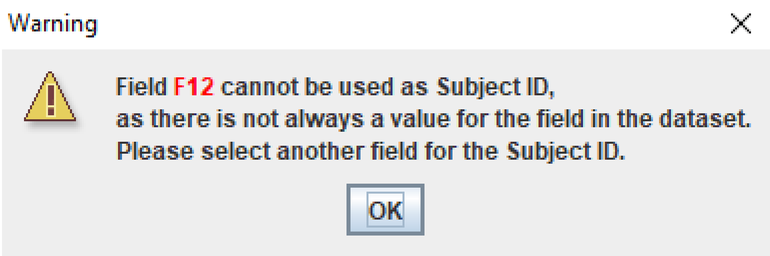
First of all, we must assign one of the fields to the "Subject ID". If this is not done, it will not be possible to generate an ODM file. In our lab data example, the Subject ID can be found in field 12:

F10	☑	IT.F10	Investigator Name	text	17
F11	☑	IT.F11	Screen ID or Number	integer	4
F12	☐	IT.F12	Subject ID or Number	text	0
F13	☐	IT.F13	Spare subject level ID or N...	text	0
F14	☐	IT.F14	Subject Initials	text	3

So, in the graphical user interface, we select "F12" for the field to be used for the Subject ID:



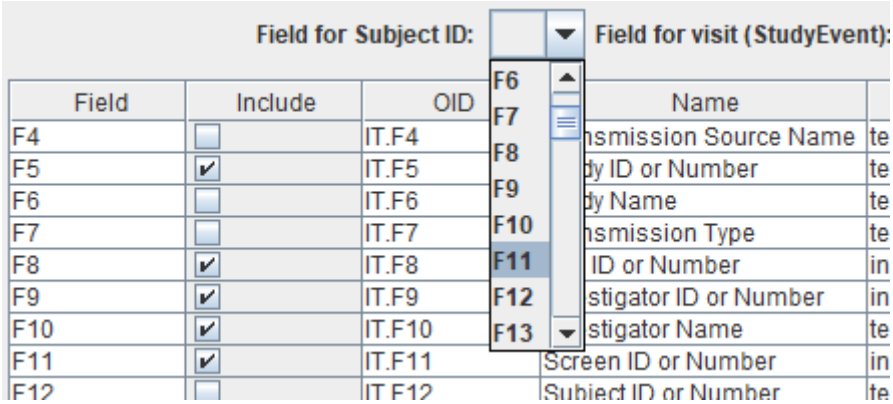
We do however get a warning:



Telling us that there is not always data in this field.
Inspection of the data reveals that only field 11 ("Screen ID") is populated:

F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	11	11	John S...	8222			ABC	M
CDISC...	CDISC...	C	17	17	Susan...	8277			RST	F
CDISC...	CDISC...	C	17	17	Susan...	8277			RST	F
CDISC...	CDISC...	C	17	17	Susan...	8277			RST	F

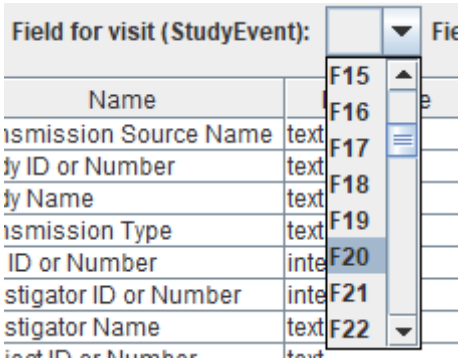
So it looks here as the "Screen ID" was used as the subject identifier, and the "Subject ID" was left empty. So, we will use field 11 for the identifier of the subject (Subject ID):



When done, we will notice that the row for F11 disappears from the table, as it is "promoted" to be the first level organizer in the data:

F8	☑	IT.F8	Site ID or Number	integer
F9	☑	IT.F9	Investigator ID or Number	integer
F10	☑	IT.F10	Investigator Name	text
F12	☐	IT.F12	Subject ID or Number	text
F13	☐	IT.F13	Spare subject level ID or N...	text
F14	☐	IT.F14	Subject Initials	text
F15	☑	IT.F15	Subject Sex	text

Secondly, we want to assign a field for the "visit ID or number". If we don't, all the data will be assigned to a single, "default" visit. In some seldom cases, where there was only one visit, this can indeed be the case. In the CDISC Lab Standard however, field 20 "Visit ID or number" is however exact meant for storing this information. So in the graphical user interface, we select "F20" for the "Field for visit (StudyEvent)":

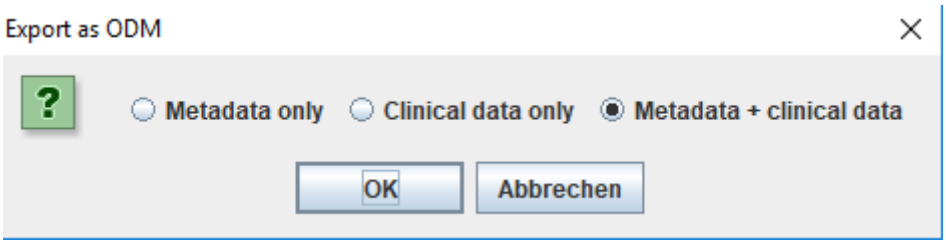


Also here, we then find that the row F12 disappears from the table, as it has been "promoted".

We would now also need to assign a field for the "form" and for the "subform" (ItemGroup). However, we do not find a suitable field for these! The reason is simply that all data are supposed to come from a single form, the "laboratory form". The same applies for the "subform".
So we will leave the assignments for these unused. The software will then later generate a single default form and subform in the metadata, which we will be enabled to give an identifier and a name.

After assign a field for the subject ID and for the visit ID, we are now ready to generate the ODM file. We can first save the mappings using the "Save Mappings" button, so that all this information can be reloaded later, e.g. when new, additional data comes in and we either want to revisit the process and mappings, or generate new or additional clinical data in ODM format, for which we need this information.

After clicking the "Export as ODM" button, the following dialog appears:

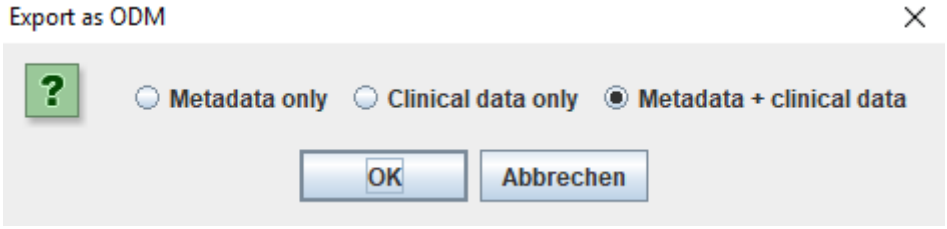


Remark that the software will first do a validation. If issues are reported, one may choose to correct these first, or ignore these (which essentially might lead to a non-conform ODM file), and continue.

In our case, we will generate as well metadata as clinical data. The selection will usually depend on what one will use the ODM file for. When later new data comes in, one can use the mappings again and then e.g. only export clinical data.

Exporting to CDISC ODM

You can now start exporting the data and metadata as CDISC ODM using the "Export as ODM" button. The following dialog is displayed:

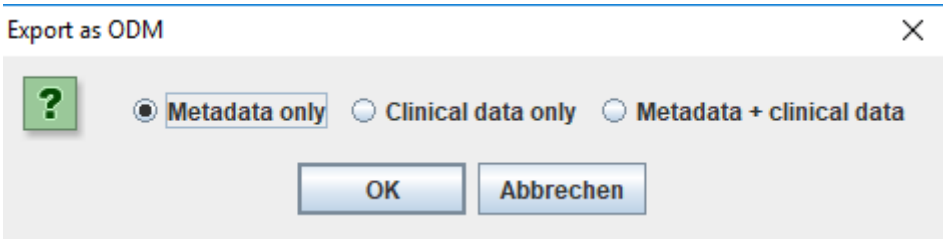


Depending on what you would like to do with the ODM data, you may choose between:

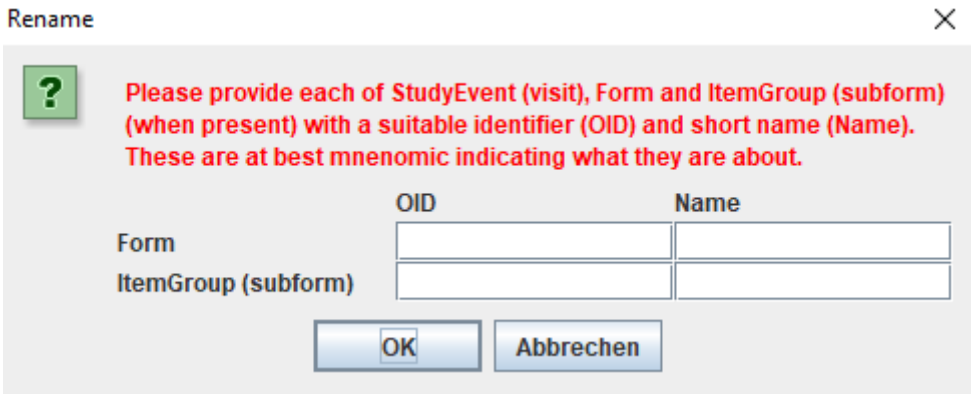
a) Metadata only: the ODM file will be a file with study design information only
b) Clinical data only: the data from your "flat text" file will be converted to CDISC "ODM ClinicalData" without the metadata. This is an option that is recommended only in the case that you have generated a "metadata" file separately.
c) Metadata + clinical data: An ODM file will be generated containing as well the data as metadata (this is never a bad choice)

Remark that the software will first do a validation. If issues are reported, one may choose to correct these first, or ignore these (which essentially might lead to a non-conform ODM file), and continue.

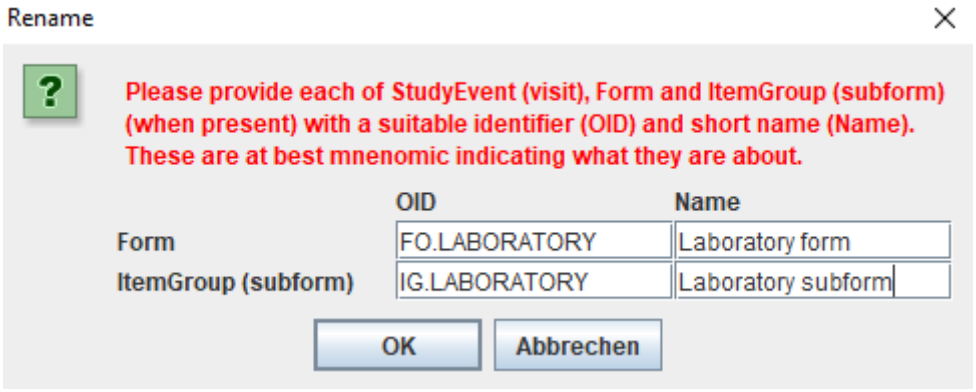
Let us first generate a file with "metadata only":



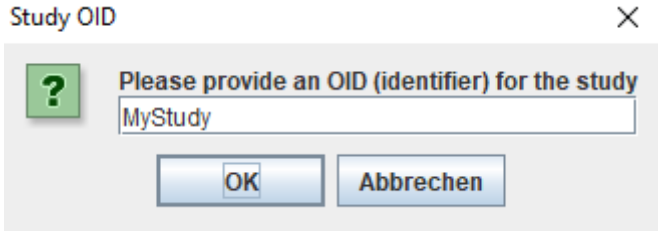
After clicking "OK", you will first need to add some extra information:



As we did not assign fields for the form and subform levels, the software will generate a single "default" form, and a single "default" subform. We are now asked to provide an identifier (OID) and name for these. It is recommended to use mnemonic values for this. As our data is about lab data, we e.g. use "FO.LABORATORY" and "Laboratory form", and "IG.LABORATORY" and "Laboratory subform" for these, were "FO." and "IG." are pretty usual prefixes used for the identifiers in ODM. So we enter:



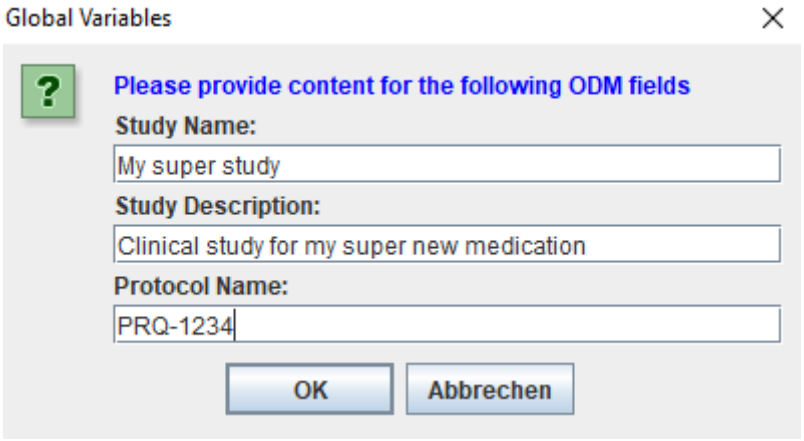
After clicking OK, the system asks us for some additional information:



We give "MyStudy" as the identifier (code) for the study. It is recommended to use the company internal code for the study for this. Remark that it is a very good idea to use the "STUDYID" for SDTM or SEND as the OID in this dialog. This will make mapping using SDTM-ETL even easier.

After clicking OK, a new dialog is displayed:

Which must be filled as these are required fields in the ODM, but is not often used afterwards. It is recommended to fill these fields with information from the protocol. For example:

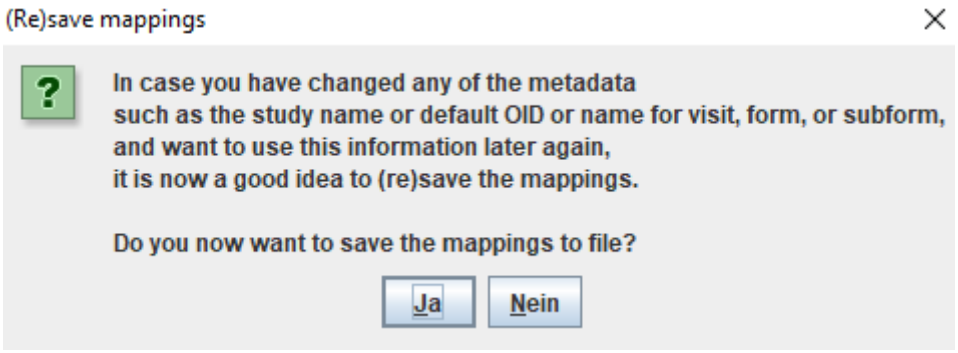


After clicking OK, the system will start generating the ODM, and will then display a file chooser, allowing you to assign and name of the ODM file. The contents of the latter will look like:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.3" xmlns:xlink="http://www.w3.org/1999/xlink"
    CreationDateTime="2018-09-02T19:03:50.613"
    Description="Generated by the XML4Pharma ODMCreator from file C:\CDISC_Standards\CDISC_Lab_1_0_1_
    FileOID="MyStudy" FileType="Snapshot" Granularity="Metadata" ODMVersion="1.3.1"
    SourceSystem="XML4Pharma ODMCreator" SourceSystemVersion="2017R1">
    <Study OID="MyStudy">
        <GlobalVariables>
            <StudyName>My super study</StudyName>
            <StudyDescription>Clinical study for my super new medication</StudyDescription>
            <ProtocolName>PRQ-1234</ProtocolName>
        </GlobalVariables>
        <MetaDataVersion OID="MV.TestStudy" Name="Test study metadata version 1">
            <Protocol>
                <StudyEventRef StudyEventOID="SE.01" Mandatory="No"/>
                <StudyEventRef StudyEventOID="SE.01R" Mandatory="No"/>
                <StudyEventRef StudyEventOID="SE.02" Mandatory="No"/>
            </Protocol>
            <StudyEventDef OID="SE.01" Name="01" Repeating="No" Type="Scheduled">
                <FormRef FormOID="FO.LABORATORY" Mandatory="No"/>
            </StudyEventDef>
            <StudyEventDef OID="SE.01R" Name="01R" Repeating="No" Type="Scheduled">
                <FormRef FormOID="FO.LABORATORY" Mandatory="No"/>
            </StudyEventDef>
            <StudyEventDef OID="SE.02" Name="02" Repeating="No" Type="Scheduled">
                <FormRef FormOID="FO.LABORATORY" Mandatory="No"/>
            </StudyEventDef>
            <FormDef OID="FO.LABORATORY" Name="Laboratory form" Repeating="No">
                <ItemGroupRef ItemGroupOID="IG.LABORATORY" Mandatory="No"/>
            </FormDef>
            <ItemGroupDef OID="IG.LABORATORY" Name="Laboratory subform" Repeating="No">
                <ItemRef ItemOID="IT.F5" Mandatory="No"/>
                <ItemRef ItemOID="IT.F8" Mandatory="No"/>
                <ItemRef ItemOID="IT.F9" Mandatory="No"/>
```

You can always validate this ODM file for compliance with the standard, using the "XML4Pharma CDISC ODM Checker", which is free of charge for CDISC members.

In case you have changed (or added new) any of the metadata, such as the study name, or any of the "default" OID or Name for visit (StudyEvent), form or subform (ItemGroup), you may now want to save the mappings (again) to use them in future. So the software will ask you:

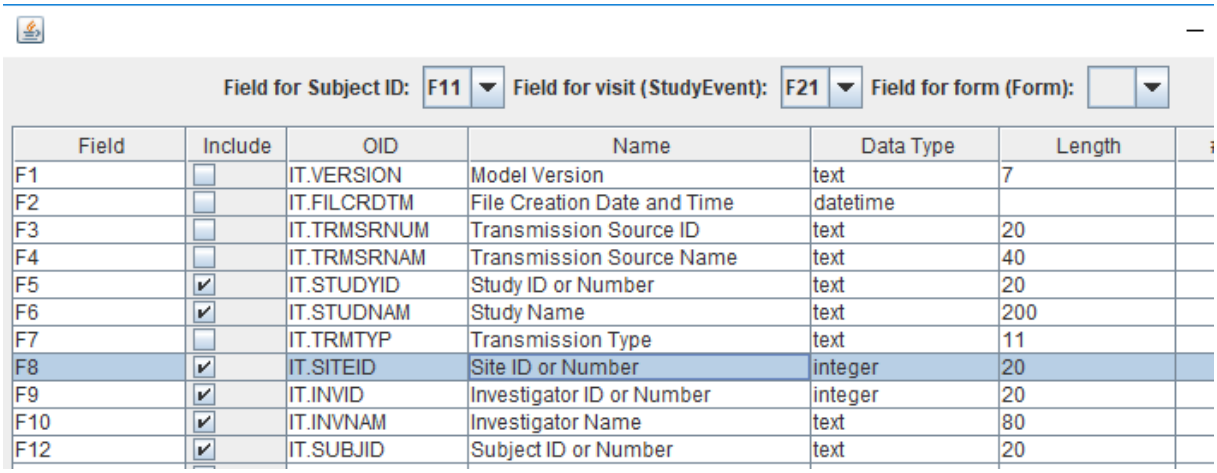


If you click "Yes", the software will ask you where (in which file) you want to save the mappings, just like in the case when you use the "save mappings" button.

Filtering

In many cases, you will not want to export the content of all the fields of your "flat text" file to CDISC ODM. In our case, there are 92 fields, and we only want to retain these that are relevant for a mapping to the SDTM LB (Laboratory Test Results) domain.

In order to select the fields that need to be exported as CDISC ODM, one should use the "Include" checkboxes on the left of the metadata mapping table. For example:



Field for Subject ID: F11 Field for visit (StudyEvent): F21 Field for form (Form):

Field	Include	OID	Name	Data Type	Length
F1	☐	IT.VERSION	Model Version	text	7
F2	☐	IT.FILCRDTM	File Creation Date and Time	datetime	
F3	☐	IT.TRMSRNUM	Transmission Source ID	text	20
F4	☐	IT.TRMSRNAM	Transmission Source Name	text	40
F5	☑	IT.STUDYID	Study ID or Number	text	20
F6	☑	IT.STUDNAM	Study Name	text	200
F7	☐	IT.TRMTYP	Transmission Type	text	11
F8	☑	IT.SITEID	Site ID or Number	integer	20
F9	☑	IT.INVID	Investigator ID or Number	integer	20
F10	☑	IT.INVNAM	Investigator Name	text	80
F12	☑	IT.SUBJID	Subject ID or Number	text	20

In this case, we do not want any export for the 4 first fields, as this information does not go into the SDTM "LB" (Laboratory Test Results) datasets, and is not relevant for review by the regulatory authorities. We however do want to keep the information about site and investigator.

We then select those fields only that are relevant. Whether a field is selected for export is also kept in the file with assignments and mappings that can be exported and imported again ("Load prior mappings"), so that the same mappings and assignment, including the filtering, can be used over and over again when new "flat text" files are received.

Running the software in batch mode

Once a file with "mappings" has been generated, it is also possible to use this file in batch mode, i.e. without the use of the graphical user interface. This is especially useful when new clinical data is generated in the source format, and an ODM file with clinical data needs to be generated from that and the previously developed mappings between source data and ODM data.

At this moment, running in batch mode only support generation of "clinical data only", as this is the most common use case for running in batch. We are considering to also implement the two other use cases "metadata only" and "metadata + clinical data" in the future. These are however only minor use cases.

An example of the processing instructions in a "batch file" is found in the file "ODMGenerator2_batch.bat".

Here are the contents:

```
1  @ECHO OFF
2  set CLASSPATH=.
3  set CLASSPATH=%CLASSPATH%;c:\ODMGenerator2\ODMGenerator.jar
4  java com.xml4pharma.odmgenerator.batch.ODMGeneratorBatch
5      -inputfilelocation=C:\CDISC_Standards\CDISC_Lab_1_0_1_final\Lab1-0-1-BaseSampleData.dat
6      -odmoutpufilelocation=C:\temp\ODM_test_export.xml
7      -mappingsfilelocation=C:\ODMGenerator2\Mappings\Lab_example_mappings_v2_new.dat
8
```

Lines 4 to 7 should essentially be one single line, but we have split them here for better display.

The first line is essentially an old DOS command (for those who remember the pre-Windows era) stating that the lines from the batch file need not be repeated in the output.
Lines 2 and 3 tells the system where to find the executables of the software. You may need to adapt the second part of the third line depending on where you installed the software.
Line 4 than contains the command to do the execution. Every statement that starts with a dash ("-") is a parameter-value pair, and all of these are necessary.

The parameters are the following:

Parameter	Meaning
-inpufilelocation	location (path) and name of the source data (usually CSV or other "flat" data), e.g. exported from Excel
-outputfilelocation	location (path) and name of the ODM file that will be generated
-mappingsfilelocation	location (path) and name of the file with the mappings that were stored when using the graphical user interface

When then executing this "batch" file, a window will open and display the progress of the execution. For example:

```
PS C:\ODMGenerator2\batch> .\ODMGenerator2_batch.bat
Using input file = C:\CDISC_Standards\CDISC_Lab_1_0_1_final\Lab1-0-1-BaseSampleData.dat
Using mappings file = C:\ODMGenerator2\Mappings\Lab_example_mappings_v2_new.dat
Now starting executing mappings using follow input:
Input file = C:\CDISC_Standards\CDISC_Lab_1_0_1_final\Lab1-0-1-BaseSampleData.dat
Output ODM file = C:\temp\ODM_test_export.xml
Mappings file = C:\ODMGenerator2\Mappings\Lab_example_mappings_v2_new.dat
First line in input file will be skipped = false
NumFields=92
SubjectIDField=F11
StudyEventField=F20
Delimiter=|
SkipFirstLine=true
StudyOID=TestStudy
MetaDataVersionOID=MV.TestStudy
MetaDataVersionName=Test study metadata version 1
StudyName=My test study
StudyDescription=test study description
ProtocolName=test study protocol name
FormOIDDefault=FO.Laboratory
FormNameDefault=Laboratory form
ItemGroupOIDDefault=IG.LABORATORY
ItemGroupNameDefault=Laboratory subform
Field=F1 Include=false OID="IT.F1" Name="Model Version" DataType=text Length=7 GenerateCodeList=false
Field=F2 Include=false OID="IT.F2" Name="File Creation Date and Time" DataType=datetime GenerateCodeList=false
Field=F3 Include=false OID="IT.F3" Name="Transmission Source ID" DataType=text Length=5 GenerateCodeList=false
Field=F4 Include=false OID="IT.F4" Name="Transmission Source Name" DataType=text Length=15 GenerateCodeList=false
Field=F5 Include=true OID="IT.F5" Name="Study ID or Number" DataType=text Length=12 GenerateCodeList=false
Field=F6 Include=false OID="IT.F6" Name="Study Name" DataType=text Length=12 GenerateCodeList=false
Field=F7 Include=false OID="IT.F7" Name="Transmission Type" DataType=text Length=1 GenerateCodeList=false
Field=F8 Include=true OID="IT.F8" Name="Site ID or Number" DataType=integer Length=2 GenerateCodeList=false
Field=F9 Include=true OID="IT.F9" Name="Investigator ID or Number" DataType=integer Length=2 GenerateCodeList=false
Field=F10 Include=true OID="IT.F10" Name="Investigator Name" DataType=text Length=17 GenerateCodeList=false
Field=F11 Include=true OID="IT.F11" Name="Screen ID or Number" DataType=integer Length=4 GenerateCodeList=false
Field=F12 Include=false OID="IT.F12" Name="Subject ID or Number" DataType=text Length=0 GenerateCodeList=false
Field=F13 Include=false OID="IT.F13" Name="Spare subject level ID or Number" DataType=text Length=0 GenerateCodeList=false
Field=F14 Include=false OID="IT.F14" Name="Subject Initials" DataType=text Length=3 GenerateCodeList=false
Field=F15 Include=true OID="IT.F15" Name="Subject Sex" DataType=text Length=1 GenerateCodeList=true
Field=F16 Include=false OID="IT.F16" Name="Subject Sex Code List ID" DataType=text Length=33 GenerateCodeList=false
Field=F17 Include=true OID="IT.F17" Name="Subject Date Of Birth" DataType=date GenerateCodeList=false
Field=F18 Include=false OID="IT.F18" Name="Subject Race" DataType=text Length=0 GenerateCodeList=true
Field=F19 Include=false OID="IT.F19" Name="Subject Race Code List ID" DataType=text Length=31 GenerateCodeList=false
```

The generated ODM file with clinical data then looks like:

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <ODM xmlns="http://www.cdisc.org/ns/odm/v1.3" CreationDateTime="2018-09-03T19:17:07.088" Description="Ge
3  <ClinicalData StudyOID="TestStudy" MetaDataVersionOID="MV.TestStudy">
4  <SubjectData SubjectKey="8222">
5  <StudyEventData StudyEventOID="SE.01">
6  <FormData FormOID="FO.Laboratory">
7  <ItemGroupData ItemGroupOID="IG.LABORATORY">
8  <ItemData ItemOID="IT.F5" Value="CDISC Test 1"/>
9  <ItemData ItemOID="IT.F8" Value="11"/>
10 <ItemData ItemOID="IT.F9" Value="11"/>
11 <ItemData ItemOID="IT.F10" Value="John Smith, M.D."/>
12 <ItemData ItemOID="IT.F15" Value="M"/>
13 <ItemData ItemOID="IT.F17" Value="1968-08-12"/>
14 <ItemData ItemOID="IT.F24" Value="C1234"/>
15 <ItemData ItemOID="IT.F26" Value="C434382"/>
16 <ItemData ItemOID="IT.F29" Value="6"/>
17 <ItemData ItemOID="IT.F30" Value="2001-05-09T10:55:00-05:00"/>
18 <ItemData ItemOID="IT.F40" Value="Urine"/>
19 <ItemData ItemOID="IT.F41" Value="32"/>
20 <ItemData ItemOID="IT.F42" Value="Y"/>
21 <ItemData ItemOID="IT.F46" Value="L1234"/>
22 <ItemData ItemOID="IT.F47" Value="Central Lab ABC - Chicago NA"/>
23 <ItemData ItemOID="IT.F48" Value="CMT5"/>
24 <ItemData ItemOID="IT.F49" Value="Urine Glucose"/>
25 <ItemData ItemOID="IT.F52" Value="2349-9"/>
26 <ItemData ItemOID="IT.F59" Value="Neg"/>
27 <ItemData ItemOID="IT.F67" Value="Neg"/>
28 <ItemData ItemOID="IT.F75" Value="Neg"/>
29 <ItemData ItemOID="IT.F91" Value="2001-05-10T10:19:32-05:00"/>
30 </ItemGroupData>
31 <ItemGroupData ItemGroupOID="IG.LABORATORY">
32 <ItemData ItemOID="IT.F5" Value="CDISC Test 1"/>
```

You may then want to use the for CDISC members freely available "ODMChecker" to validate the contents of this file with clinical data for ODM compliance and against the file with metadata that you generated using the graphical user interface.