

OpenClinica Data Importer

Software Design Document and User Guide

Opmerking [JR1]: General remarks:
 -Add a section with the data contracts and formats. (decimal separation character, the date format etc)
 -Add introduction (Marinel)
 -Add process description (Marinel)
 - Study Sites; how do we upload them?

Document History

Version	Date	Description	Who
1.0	21/05/2010	Initial version	C. Parlayan, J.A.M. Beliën
1.1	21/04/2012	Production version	C. Parlayan
2.0	08/11/2012	Production 2.0	C. Parlayan
2.0.1	01/02/2013	Added Repeating events and groups	C. Parlayan
2.0.2	26/03/2013	The GRID now keeps the previously matched items when a new CRF or Group is chosen for more matching.	C. Parlayan
2.0.3	01/04/2013	Bug solved in matching with repeated items.	C. Parlayan
2.0.4	01/05/2013	Remove white space when matching	C. Parlayan
2.0.5	21/07/2013	Do not print form data with no items in XML as this causes error in OpenClinica upload.	C. Parlayan, S. de Ridder
2.0.6	26/08/2013	Added "Limit number of characters to match" to make matching easier	C. Parlayan
2.0.7/8/9	29/08/2013	Bugs introduced in 2.0.5 solved.	C. Parlayan, S. de Ridder
2.1.1	09/09/2013	Introduced label-oid file.	C. Parlayan, J. Rousseau, R. Voorham
2.1.2	01/11/2013	Bug ItemGroupRepeatKey solved.	C. Parlayan, R. Voorham
2.1.3	18/11/2013	Possibility to not generate events if startdate is blank.	C. Parlayan, S. de Ridder
2.1.4	01/11/2013	XML escaping.	C. Parlayan, J. Rousseau
2.2	30/11/2013	Input file allows EVENT_INDEX and GROUP_INDEX to be defined to accept Repeating events and items in rows.	C. Parlayan, J. Rousseau
3.0	01/11/2013	Type and range validations.	C. Parlayan, J. Rousseau
3.0	20/12/2013	Better messaging	C. Parlayan
3.01	20-01-2014	XML parsing error while RangeCheck	J. Rousseau, C. Parlayan
3.02	20-01-2014	Warnings to status window	J. Rousseau, C. Parlayan
3.03	22-01-2014	matching now uses "ItemName" instead of ItemOID	C. Parlayan, S. de Ridder
3.03	24-01-2014	matching is done for all groups in a CRF in one go	C. Parlayan, S. de Ridder
4.0	29-01-2014	Better UI	C. Parlayan
4.01	07-02-2014	Show/hide Subj related columns bugfixes on validation	C. Parlayan, J. Rousseau, S. de Ridder
4.1	10-02-2014	Auto detection and show/hide subject related columns	C. Parlayan, J. Rousseau
4.1	10-02-2014	Added Multiple select to validations	S. de Ridder
4.1	10-02-2014	Leading zeroes in times and subjectid's in log records	C. Parlayan, J. Rousseau
			C. Parlayan, J. Rousseau

OCDatImporter			
4.1	10-02-2014	fix: "-- select --" in insert statements	C. Parlayan, S. de Ridder
4.2	17-02-2014	Show/hidden warn ONLY if show and has no value	C. Parlayan, S. de Ridder
4.2.1	19-02-2014	Ignore range check if other problems detected with data	C. Parlayan, S. de Ridder
4.3	21-02-2014	SE/GR repeating index in rows bugs fixed	C. Parlayan, J. Rousseau
4.3	21-02-2014	Better error messages if input file is open by other programs	C. Parlayan, J. Rousseau

4.4	11-03-2014	<u>Decoupling of the input parameters from the view, introduction of a utilities class, a OutputFile class to stream output to, and a main converter which is the controller of the conversion (MVC), various fixes. Project is now available on github for change history.</u>	C. Parlayan, J. Rousseau
-----	------------	---	--------------------------

Met opmaak: Lettertype: (Standaard)
Arial, 10 pt

Contents

<i>Document History</i>	<i>I</i>
1 INTRODUCTION	53
1.1 Product Identification.....	53
1.2 Purpose of the Document	53
1.3 Scope of the Document	53
1.4 Intended Audience	53
1.5 References.....	53
2 GENERAL DESCRIPTION	64
2.1 Product Perspective	64
2.2 Product Availability.....	64
2.3 Principle Product Functionality	64
2.4 General Constraints.....	64
3 USER GUIDE	64
3.1 Installation.....	64
3.2 First step: Convert study to a tab delimited text file.	75
3.3 Create the Study, events and CRF(s) in OpenClinica.	76
3.4 Create OpenClinica Meta Data File using OpenClinica.	76
3.5 Read the input files in “OCDataImporter”	86
3.5.1 Usage of label-oid translation file	87
3.6 Program Parameters	98
3.6.1 Specifying date format.....	98
3.6.2 Specifying gender codes.....	108
3.6.3 Splitting the ODM file	108
3.6.4 Specifying the location.....	108
3.6.5 Specifying if duplicate key check is needed	109
3.6.6 Options for events without starting dates	119
3.6.7 Confirming Program Parameters	119
3.7 Indicating study subject id, Subject sex, subject person’s id, subject start date.....	119
3.8 Matching data columns and OpenClinica Items.....	1210
3.9 Starting the conversion.....	1412
3.10 Creating subjects, study_subjects and study_events in PostGreSQL using “Inserts.sql”.....	1614
3.11 Importing data to OpenClinica using “DataImport_(n).xml”.....	1745
3.12 Importing CRF data containing repeating events and repeating groups	1917
3.12.1 Importing CRF data containing repeating events and repeating groups by using data file containing the repeating information on separate columns:	1917

OCDataImporter

3.12.2 Importing CRF data containing repeating events and repeating groups, by using data file containing the repeating information under columns EVENT_INDEX and GROUP_INDEX	2220
3.13 Data validation	2321
Appendix: Error messages	2324

1 INTRODUCTION

1.1 *Product Identification*

This document contains the Software Design Specifications for the OpenClinica generic data importer.

1.2 *Purpose of the Document*

The purpose of this document is to provide the detailed software design and the user guide.

1.3 *Scope of the Document*

The scope of the body of this document is to describe the implementation of “OCDatalImporter” program. Other aspects related to SPSS, MS Excel and OpenClinica can be found in SPSS, Excel and OpenClinica documents.

1.4 *Intended Audience*

Open Clinica users, data managers.

1.5 *References*

Microsoft .NET documentation, OpenClinica documents, PostGreSQL documents.

2 GENERAL DESCRIPTION

2.1 Product Perspective

This product will make it possible to convert a study data file into an OpenClinica study.

2.2 Product Availability

The product is available in May 2010.

2.3 Principle Product Functionality

This application is based on reading a study file and converting it to a OpenClinica study.

2.4 General Constraints

This product requires Microsoft .NET Framework version 3.5 distributable Package.
It is included in Windows Vista and Windows 7.

Otherwise, it is possible to download this from the following site:

<http://msdn.microsoft.com/downloads/>

The program is not yet available under other operating systems.

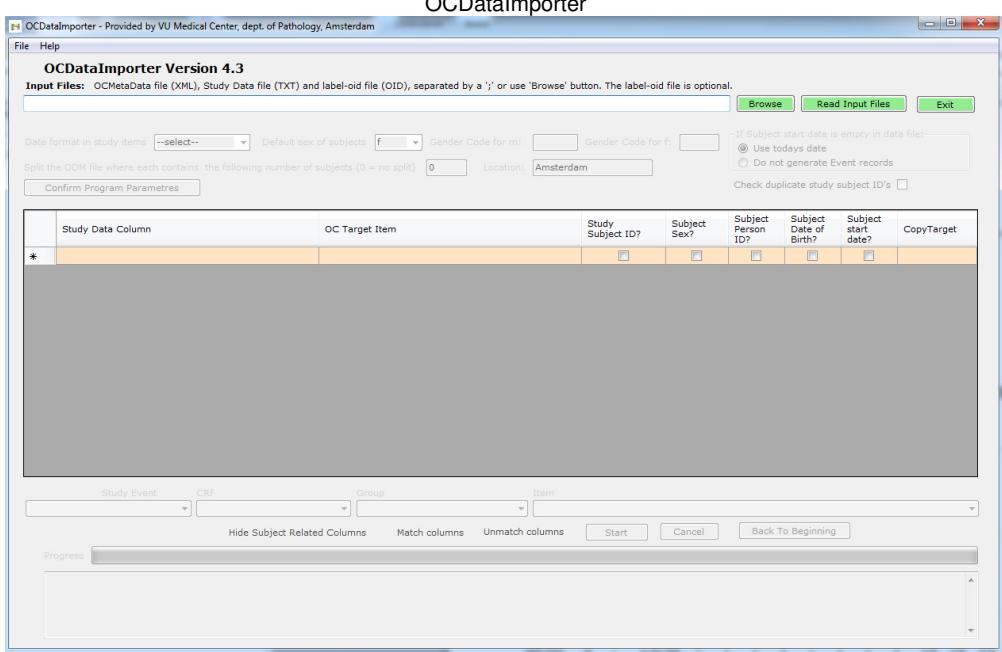
CAUTION: This program introduces a complex procedure of importing legacy data in Open Clinica.
Usage of this program requires expertise level knowledge of Open Clinica and the Postgres database.
Furthermore administrator privileges will be needed at the database server side in order to run the
database scripts to create subjects and events. Please make sure you read this document carefully
before you start using the application

3 USER GUIDE

This section describes how to use this application.

3.1 Installation

To install this program, use OCDatImporter.msi. Follow the instructions on screen.
When run, the following form should be seen:



On this screen, there are 3 steps defined: Read input files, define parameters and start conversion.
The buttons of the stage that active is are green.

3.2 First step: Convert study to a tab delimited text file.

- 1- Create a directory, for example C:\DataImporter30_test\pr1.
- 2- Start SPSS, Excel or another program to convert the study file to a tab delimited text file, named for example "test-set-v01-UTF8.txt".
- 3- Copy this file to C:\DataImporter30_test\pr1.

3.3 Data formats

The format of date and real number data in the data file should be:

- Dates: must have one of the formats YYYY-MM-DD, MM-DD-YYYY or DD-MM-YYYY
- Real numbers: must use a point (.) as the decimal separator. E.g. 2.718
- No tabs must be present in the data fields; tabs can only be used as separators between fields
- The subject ID must be contained in a column with the name 'subject_id'
- The sex of a subject must be contained in a column with the name 'gender'
- The data of birth must be contained in a column with the name 'Date_of_birth'. Studies started after the 1st of March 2014 can't use the date of birth

Opmerking [JR2]: The buttons of the active stage are green.

Opmerking [JR3]: Is het niet beter om te zeggen: '... to convert the data you want to upload to OpenClinica to a tab delimited.....'

Met opmaak: Met opsommingstekens + Niveau: 1 + Uitgelijnd op: 0,63 cm + Inspringen op: 1,27 cm
Met opmaak: Lettertype: Vet

Met opmaak: Inspringing: Links: 1,27 cm

3.3 Create the Study, events and CRF(s) in OpenClinica.

To do this, start OpenClinica, login as Datamanager, Create a study, create the necessary events and CRF's with items which matches with the items for the file generated in section 3.2. For more on using OpenClinica, refer to OpenClinica documents.

Opmerking [JR4]: ...with items that match the data in the file generated in section 3.2

3.4 Create OpenClinica Meta Data File using OpenClinica.

Once the OpenClinica objects are created, generate the meta data file.

OCDatImporter

View the study and Click "here" of the sentence "Download all of the OID's needed for data import and rules". Give the file a name, for example "DownloadStudyMetadata.xml" and save the file in C:\DataImporter30_test\pr1.

3.5 Read the input files in "OCDatImporter"

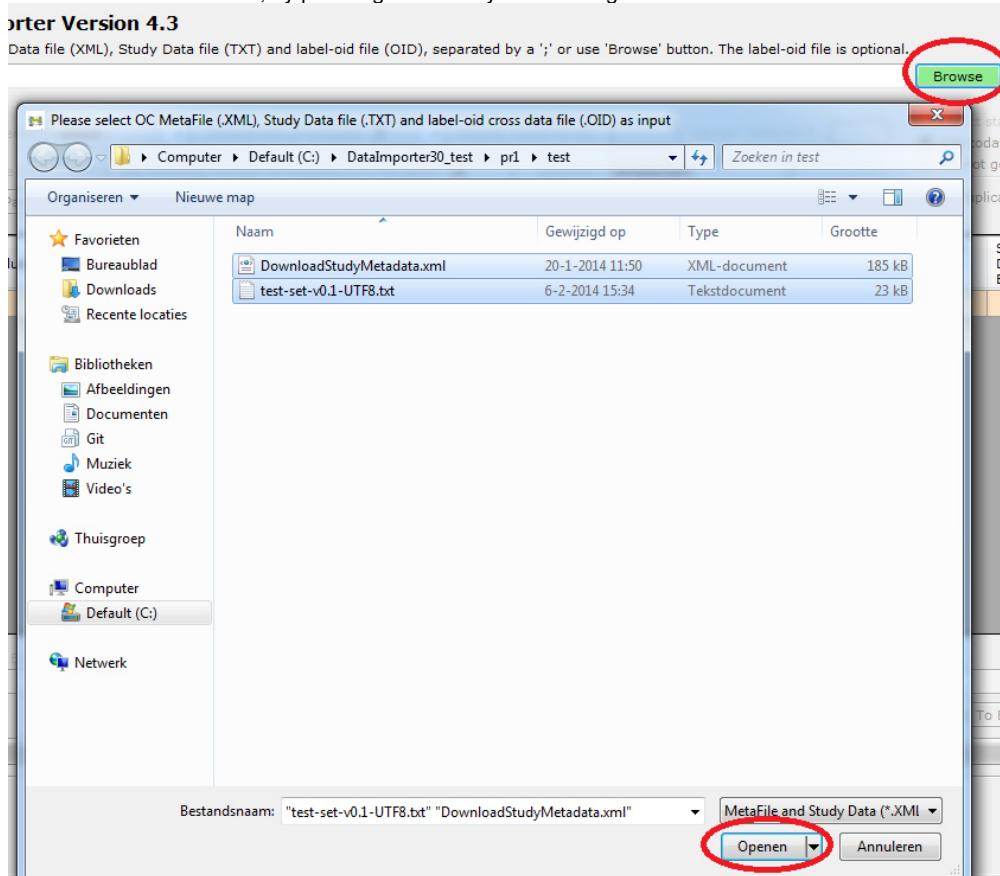
Hit the "browse" button and select the files that are generated: The meta file and the tab delimited data file.

Opmerking [JR5]: Start the OCDatImporter and hit the 'browse'

You must select both files, by pressing the ctrl key and clicking on both files.

Data Importer Version 4.3

Data file (XML), Study Data file (TXT) and label-oid file (OID), separated by a ';' or use 'Browse' button. The label-oid file is optional.



Now hit the "read input files" button on OCDatImporter screen.

The program will check if the number of columns in the data file is the same with the items in the Meta file. If so, an automatic matching is done, but mostly there are more than one CRF's defined for one patient data file so the column matching will be mostly done manually.

Opmerking [JR6]: As the number of items in the Meta file

The ideal situation is, OC data items and the source file columns all have same names. For example if the source data column name is "Gender" and the OC item is also named "Gender" an automatic matching can be made.

Opmerking [JR7]: Preferably?

3.5.1 Usage of label-oid translation file

Normally OpenClinica makes standard study subject oid's like "SS_<subjectID>", for example "SS_1200" will be the study subject oid for a subject with ID = 1200. OCDatImporter assumes this procedure is followed and generates the insert statements (explained in later chapters) accordingly.

OCDatalImporter

If the inserts will be used for creating the study subjects, you will not need the label-oid translation table and skip reading this section.

This is however not the case if the study subjects are made with OC web services. If the subject id is longer than a certain number of characters, OC generates a study subject oid with some part of the subject id + sometimes a 4 digit number to make the id unique. Example:

Subject: COCOS_22100	OC generates the study subject id as: SS_COCOS_22_9563
Subject: COCOS_22101	OC generates the study subject id as: SS_COCOS_22_8260
Subject: 20220	OC generates the study subject id as: SS_20220

Etc.

In order to use OCDatalImporter's CRF data ODM files with the study subjects generated by OC web services, a cross data file is needed to translate the subject id to study subject oid. This file should only contain the subject id's having a study subject oid **other than** SS_<subject id>, like the following:

Label	oid
COCOS_22100	SS_COCOS_22_9563
COCOS_22101	SS_COCOS_22_8260

Note that 20220 is not included because this oid is conform the structure SS_<subject id>.

This file can be generated easily by using a postgres-query, similar to this:

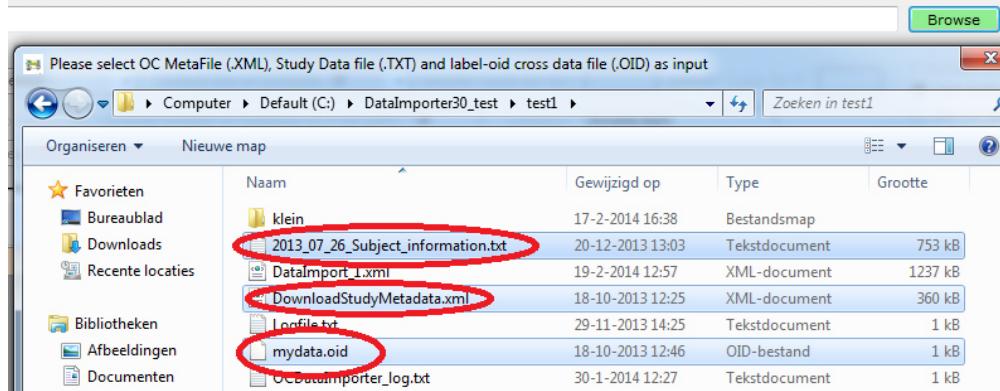
```
psql -c "select label, oc_oid from study_subject  
where study_id in (5,6,7,8) and (oc_oid NOT LIKE 'SS_` label)" -d openclinica
```

5,6,7,8 are the studies to include and oc_oid's are selected to be other than SS_<label>.

The label-oid file should be in the working directory, must have a .oid extention and can be selected as shown below:

Version 4.3

(XML), Study Data file (TXT) and label-oid file (OID), separated by a ';' or use 'Browse' button. The label-oid file is optional.



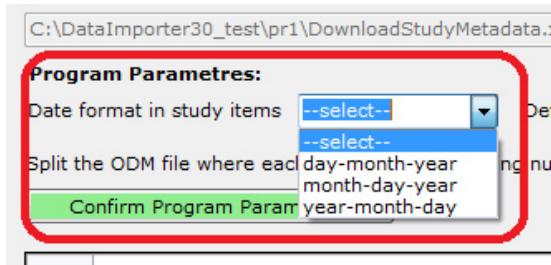
3.6 Program Parameters

3.6.1 Specifying date format

You can optionally specify the format of dates in the data file by using the "Date format in study items" combo box. The dates in OC Import XML file **must** be in YYYY-MM-DD format. The program takes

OCDatImporter

care of converting the dates from the format of your choice to OC format. If you do not want any conversion to take place, leave the selected item as “—select—”, but in this case either you must be sure that all dates are in ODM format YYYY-MM-DD or there are no dates at all in your data file.



3.6.2 Specifying gender codes

The codes for male and female in OpenClinica has to be “m” and “f” respectively. If this appears otherwise in your data file you can indicate this by using the appropriate textboxes as indicated below. If male is coded as 1 and female as 2 in your data file, below coding will make the proper translation.

Opmerking [JR8]: ...,the coding below will.....

If there is no gender code in your file and all participants are female, then you can leave the Gender code fields blank and select “f” as default sex of subjects. If the gender code fields are filled, default sex will be ignored.

3.6.3 Splitting the ODM file

Open Clinica may not be able to handle ODM files with more than 75 subject-data in one ODM file when the user interface is used for uploading. It is recommended to split the ODM file into pieces which will contain no more than a specified number of subject data. In version 3.1.2 it was able to handle 75 subject data in one file. When you enter 75, several files will be generated with 75 subject-data each, named “DataImport_1, ..._2, ..._3, etc.

Enter 0 if no splitting is desired. (We hope that there will be no splitting needed in the future versions)

3.6.4 Specifying the location

It is possible to give the location name to be used in event subjects.

3.6.5 Specifying if duplicate key check is needed

If your data file contains only one row for each subject and the subject id's must be unique, this program can check that and issue an error message if that fails. Use the checkbox below, to perform this check.

OCDatalImporter

Check duplicate study subject ID's

3.6.6 Options for events without starting dates

Use todays date: The event start date will be assumed todays date. This way it is possible to add data for that event.

Do not generate Event records: This is the logical choice, since data addition to an event without start date is of no use.

But if the event has to be generated and the data must be added, choose "use todays date".

Item

Gender Code for f: Check duplicate study subject ID's

Amsterdam

If Subject start date is empty in data file:
 Use todays date
 Do not generate Event records

Columns Unmatch columns

	Date?	Subject Sex?	Subject Person ID?	Subject Date of Birth?	Subject start date?	CopyTarget
	<input type="checkbox"/>	CopyTarget				
	<input type="checkbox"/>	CopyTarget				

3.6.7 Confirming Program Parameters

When all of the parameters are defined as explained in section 3.6, hit the "Confirm" button to proceed to next step.

Program Parameters:

Date format in study items: day-month-year Default sex of subjects: f Gender Code for m: 1 Gender Code for f: 2 If Subject start date is empty in data file:
 Use todays date
 Do not generate Event records

Split the GDM file where each contains the following number of subjects (0 = no split): 75 Location: Amsterdam

Check duplicate study subject ID's

[Confirm Program Parameters](#)

3.7 Indicating study subject id, Subject sex, subject person's id, subject start date

The program needs to know which data column is the subject **id**. Without a subject id, the process can't be made. So there has to be one (and only one) subject id checked in the whole grid. This will be used to create subject records and relate the CRF data with this subject.

Opmerking [JR9]: ID

Opmerking [JR10]: complete

There are also other items needed to create the subject like subject sex, subject person's id, subject date of birth and subject start date. These can be indicated with the related checkboxes as shown below.

	Study Data Column	OC Target Item	Study Subject ID?	Subject Sex?	Subject Person ID?	Subject Date of Birth?	Subject start date?
1	Subject_id	none	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Date_of_Birth	none	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	Gender	none	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Event_index	none	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

OCDatImporter

This will be used in insert statements to create study, event_study and event_crf records and is described later in this manual.

Once defined, the columns "Study Subject ID?", "Subject Sex?", "Subject Person ID?", "Subject Date of Birth?" and "Subject Start Date?" can be hidden (in order to make space in grid to display other data) by using link button:



3.8 Matching data columns and OpenClinica Items

To match the columns in case there are more than one study event and/or CRF's, you can choose the event and the CRF and then hit "Match columns" link. The Group and Items combobox should be left as --select--. This will check for similar names and match them.

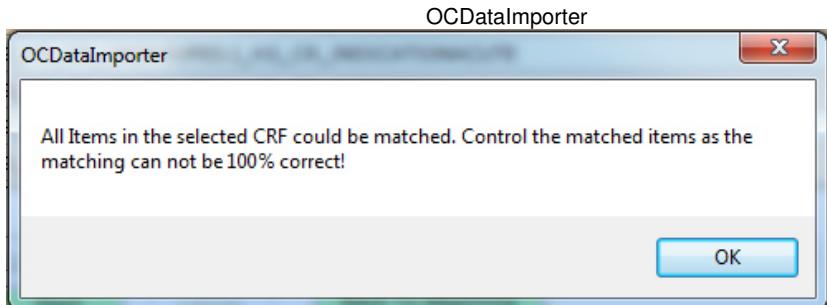
The screenshot shows a table mapping study data columns to OC target items. Below the table is a control panel with dropdowns for OC Target (Study Event and CRF), a Group dropdown set to "-- select --", and a "Match columns" button. The "Match columns" button is highlighted with a red circle.

	Study Data Column	OC Target Item
▶	Subejct_id	none
	Date_of_Birth	none
	Gender	none
	Event_index	none
	Group_index	none
	STRING_TYPE_E1_G1	none
	STRING_TYPE_WITH_WIDTH_E1_G1	none
	REAL_TYPE_E1_G1	none
	REAL_TYPE_WITH_WIDTH_E1_G1	none
	INTEGER_TYPE_E1_G1	none
	INTEGER_TYPE_WITH_WIDTH_E1_G1	none
	DATE_TYPE_E1_G1	none
	PDATE_TYPE_E1_G1	none
	ENUMERATION_INT_SINGLESELECT_E1_G1	none

OC Target: Study Event CRF
SE_INPUTTYPESOPENCLINICA F_TESTBED_DATA_V001 Group
-- select --

Show All Columns Match columns

This can be done for all CRF's separately, so gradually most of the items will be matched. If all columns are matched, program displays:



Even if this is the case, it is recommended to control if the matching is correct.

Opmerking [JR11]: verify

If the names are not similar or the OC item does not exist in the selected CRF (because it exists in another CRF or even doesn't exist at all) matching can also be done manually. The program will issue a message indicating this. To see the items that are NOT matched, see the Progress textbox as shown below:

Opmerking [JR12]: An example of unmatched item (subject_id) is given below.

Study Data Column	OC Target Item
Subject_id	none
Date_of_Birth	none
Gender	none
Event_index	none
Group_index	none
STRING_TYPE_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TESTB_STRING...
STRING_TYPE_WITH_WIDTH_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TESTB_STRING...
REAL_TYPE_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TESTB_REAL_T...
REAL_TYPE_WITH_WIDTH_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TESTB_REAL_T...
INTEGER_TYPE_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TESTB_INTEGER...
INTEGER_TYPE_WITH_WIDTH_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TESTB_INTEGER...
DATE_TYPE_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TESTB_DATE_T...
PDATE_TYPE_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TESTB_PDATE_...
ENUMERATION_INT_SINGLESELECT_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_SINGLE_SELECT.I_TESTB_ENU...

OC Target: Study Event CRF: SE_INPUTTYPESOPENCLINICA Group: -- select -- Item: F_TESTBED_DATA_V001

Progress: No match for: subject_id

Match columns Unmatch columns Start Cancel

OCDatImporter

Not all Items in the selected CRF could be matched. For the list of UNMATCHED items, see the progress textbox below. You can match those items by using the checkboxes above. Control the matched items too, as the matching can not be 100% correct!

OK

When no matching could be made, the program will display the "study data columns" of the grid, but it will leave the "OC Target Item" blank or fill it as "none". This can be filled in by using the "CopyTarget" link. To use this link, first chose the target item using the Study event, CRF, Group and Items combo boxes, then hit the CopyTarget of the row which has to be matched with that item.

Suppose you want to match "ENUMERATION_INT_MULTIPLESELECT" with

"SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V002.IG_TESTB_MULTIPLE_SELECT.I_TESTB_ENUMERATION_INT_MULTIPLESEL"

Use the combo boxes so that the above item appears in the combo boxes and hit "CopyTarget" of the row of "ENUMERATION_INT_MULTIPLESELECT".

Study Data Column	OC Target Item	CopyTarget
REAL_TYPE_WITH_WIDTH_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TESTB_REAL_T...	CopyTarget
INTEGER_TYPE_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TESTB_INTEGE...	CopyTarget
INTEGER_TYPE_WITH_WIDTH_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TESTB_INTEGE...	CopyTarget
DATE_TYPE_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TESTB_DATE_T...	CopyTarget
PDATE_TYPE_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TESTB_PDATE_...	CopyTarget
ENUMERATION_INT_SINGLESELECT_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_SINGLE_SELECT.I_TESTB_ENU...	CopyTarget
ENUMERATION_STRING_SINGLESELECT_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_SINGLE_SELECT.I_TESTB_ENU...	CopyTarget
ENUMERATION_REAL_SINGLESELECT_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_SINGLE_SELECT.I_TESTB_ENU...	CopyTarget
▶ ENUMERATION_INT_MULTIPLESELECT_E1_G1	none	CopyTarget
ENUMERATION_STRING_MULTIPLESELECT...	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_MULTIPLE_SELECT.I_TESTB_EN...	CopyTarget
ENUMERATION_REAL_MULTIPLESELECT_E1...	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_MULTIPLE_SELECT.I_TESTB_EN...	CopyTarget
ENUMERATION_INT_RADIO_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_SINGLE_SELECT.I_TESTB_ENU...	CopyTarget
ENUMERATION_STRING_RADIO_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_SINGLE_SELECT.I_TESTB_ENU...	CopyTarget
ENUMERATION_REAL_RADIO_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_SINGLE_SELECT.I_TESTB_ENU...	CopyTarget

OC Target: Study Event CRF Group Item

SE_INPUTTYPESOPENCLINICA F_TESTBED_DATA_V002 IG_TESTB_MULTIPLE_SELECT.I_TESTB_ENUMERATION_INT_MULTIPLESEL

The program will copy the target item into the “OC target Item” field.

▶ ENUMERATION_INT_MULTIPLESELECT_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V002.IG_TESTB_MULTIPLE_SELECT.I_TESTB_ENUMERATION_INT_MULTIPLESEL	CopyTarget
--	---	----------------------------

3.9 Starting the conversion

After matching columns, the program is ready to process the data and generate the files needed to upload to OpenClinica. Hit the start button, as shown below.

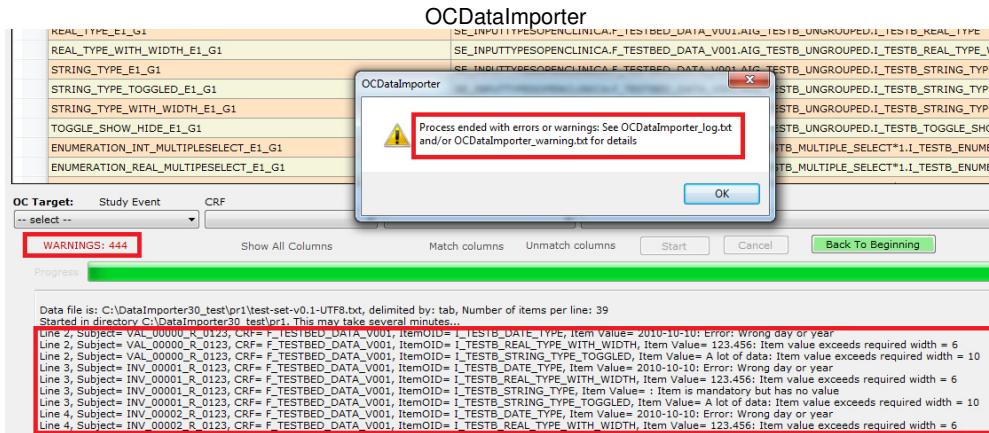
Study Data Column	OC Target Item
REAL_TYPE_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TE
REAL_TYPE_WITH_WIDTH_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TE
INTEGER_TYPE_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TE
INTEGER_TYPE_WITH_WIDTH_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TE
DATE_TYPE_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TE
PDATE_TYPE_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_UNGROUPED.I_TE
ENUMERATION_INT_SINGLESELECT_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_SINGLE_SELECT.I
ENUMERATION_STRING_SINGLESELECT_E1...	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_SINGLE_SELECT.I
ENUMERATION_REAL_SINGLESELECT_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_SINGLE_SELECT.I
▶ ENUMERATION_INT_MULTIPLESELECT_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_MULTIPLE_SELECT.I
ENUMERATION_STRING_MULTIPLESELECT...	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_MULTIPLE_SELECT.I
ENUMERATION_REAL_MULTIPLESELECT_E1...	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_MULTIPLE_SELECT.I
ENUMERATION_INT_RADIO_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_SINGLE_SELECT.I
ENUMERATION_STRING_RADIO_E1_G1	SE_INPUTTYPESOPENCLINICA.F_TESTBED_DATA_V001.IG_TESTB_SINGLE_SELECT.I

OC Target: Study Event CRF Group Item

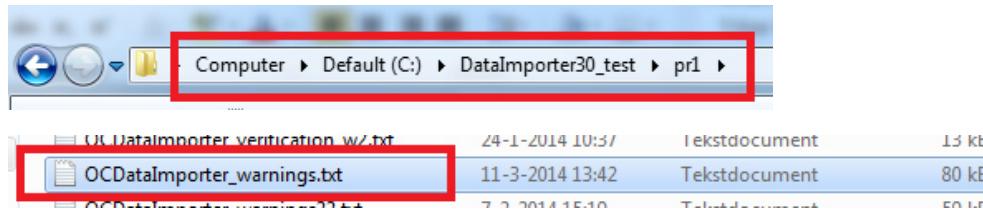
-- select --

Show All Columns Match columns Unmatch columns [Start](#)

This action can result like the following:

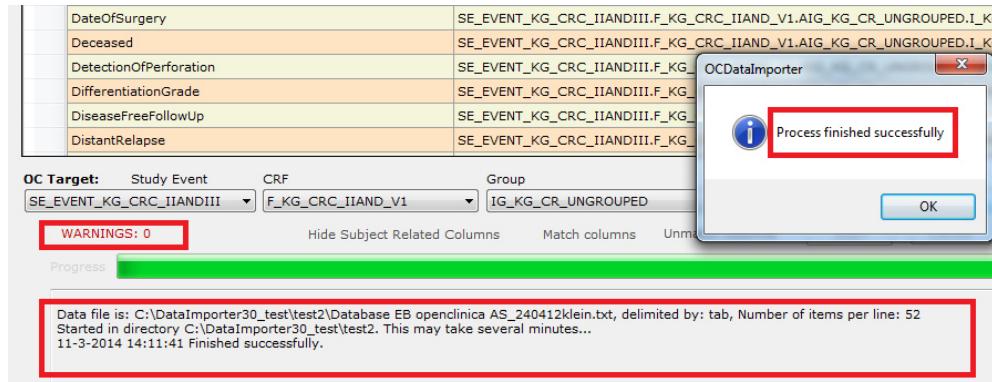


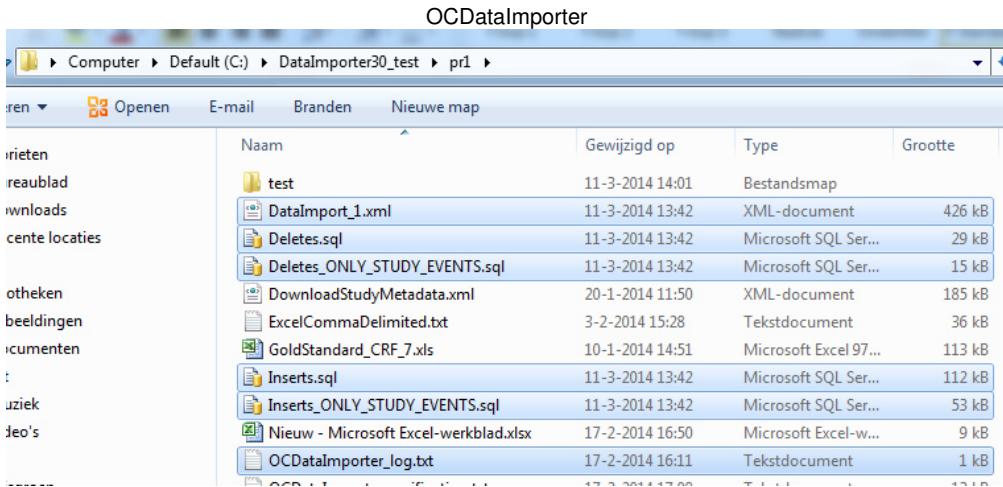
In this run, there are 444 errors detected, for example “Item Value= 123.456: Item value exceeds required width = 6”. These warnings can be found in the file called “OCDataImporter_warnings.txt” in the working directory.



Each warning/error should be corrected before upload as they will cause errors in OpenClinica.
See appendix for the list of possible things that can go wrong.

When the process ends with no errors or warnings, the program will issue a “success” message and the output files will be seen in the working directory:





3.10 Creating subjects, study_subjects and study_events in PostGreSQL using "Inserts.sql".

The file "Inserts.sql" will be used to create subject, study_subject and study_event rows in Postgres. It contains insert statements for those three tables.

For example, for the first line of the study data:

```
NIJM_CAIRO_CRC_002      0          No distant metastasis, number of affected
organs is 0 (array CGH data is available)      0      5-mrt-1942      1      1      1
13-jan-2003    60      1      4      -1      -1      1      -1      1
      -1      -1      0      1      3      0      -1      -1      -1      -1      0
      -3      -1      -1      -1      0      1      1      0      18      -1
      1      435      -1      1      0      -1      -1      435      -1
      1      0      -1      -1      435      -1      1      1454      1
      1      1      0
```

The following will be created:

```
INSERT INTO subject(status_id, gender, unique_identifier, date_created, owner_id, dob_collected,
date_of_birth)
VALUES (1, 'm', 'NIJM_CAIRO_CRC_002', '2012-11-05', 1, '1', '1942-03-05');
INSERT INTO study_subject(label, study_id, status_id, enrollment_date, date_created, date_updated,
owner_id, oc_oid, subject_id)
VALUES ('NIJM_CAIRO_CRC_002', (SELECT study_id FROM study WHERE oc_oid =
'S_CAIRO'),
1, '2012-11-05', '2012-11-05', 1, 'SS_NIJM_CAIRO_CRC_002', (SELECT
subject_id FROM subject where unique_identifier = 'NIJM_CAIRO_CRC_002'));
INSERT INTO study_event(study_event_definition_id, study_subject_id, location, sample_ordinal,
date_start, owner_id, status_id, date_created, subject_event_status_id, start_time_flag,
end_time_flag)
VALUES ((SELECT study_event_definition_id FROM study_event_definition WHERE oc_oid =
'SE_CAIRO_EVENT'),
(SELECT study_subject_id FROM study_subject WHERE oc_oid =
'SS_NIJM_CAIRO_CRC_002'), 'Amsterdam', 1, '2012-11-05 12:00:00', 1, 1, '2012-11-05', 3, '0', '0');
```

To run this SQL file, start PGAdmin at the OpenClinica database server, select the openclinica database, start Query tool from "Tools" pulldown menu, read "Inserts.sql" file and run this with "execute pgscript". See Postgres user manuals for more information on how this is made, if necessary. **CAUTION: Make sure the OpenClinica instance is NOT running at this time and don't forget to make a full backup of the database before you start!**

OCDataImporter

The screenshot shows the pgAdmin III interface. In the Object browser, under 'PostgreSQL 8.4 (localhost:5432)', there is a database named 'opendinica'. The SQL Editor tab is active, displaying a script named 'C:\tmp\Inserts.sql' which contains several INSERT statements for creating records in tables such as 'subject', 'study_subject', and 'study_event'. The SQL code includes various parameters and values, including dates like '2010-04-18' and unique identifiers like 'B000-00372'.

When this is done, proceed with the following section.

Note: Deletes.sql can be used to undo the above operation.

3.11 Importing data to OpenClinica using “DataImport_(n).xml”.

The file DataImport_(n).xml contains a fixed begin section + import statements for each line in the file created in section 3.2. + fixed end section.

```
<?xml version="1.0" encoding="UTF-8"?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.3"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.cdisc.org/ns/odm/v1.3 ODM1-3.xsd"
      ODMVersion="1.3" FileOID="1D20080412202420" FileType="Snapshot"
      Description="First dataset for testing of ODM" CreationDateTime="2008-04-12T20:24:20" >
  <ClinicalData StudyOID="S_VUMCDEPT" MetaDataVersionOID="v1.0.0">
    <SubjectData SubjectKey="SS_NIJM_CAIRO_CRC_002">
      <StudyEventData StudyEventOID="SE_CAIRO_EVENT" StudyEventRepeatKey="1">
        <FormData FormOID="F_CAIRO_GEN_IN_V12">
          <ItemGroupData ItemGroupOID="IG_CAIRO_UNGROUPED" TransactionType="Insert" >
            <ItemData ItemOID="I_CAIRO_AGE" Value="60" />
            <ItemData ItemOID="I_CAIRO_DATE_OF_BIRTH" Value="1942-03-05" />
            <ItemData ItemOID="I_CAIRO_DATE_OF_RANDOMISATION" Value="2003-01-13" />
            <ItemData ItemOID="I_CAIRO_ELIGIBLE" Value="0" />
            <ItemData ItemOID="I_CAIRO_GENDER" Value="1" />
            <ItemData ItemOID="I_CAIRO_PATIENT_ID" Value="NIJM_CAIRO_CRC_002" />
            <ItemData ItemOID="I_CAIRO_WHO_PERFORMANCE_STATUS" Value="1" />
          </ItemGroupData>
        </FormData>
      </StudyEventData>
    </SubjectData>
  </ClinicalData>
```

```

        OCDatImporter
    </SubjectData>
    ....
    ....
    </ClinicalData>
</ODM>

```

The above subject data corresponds with the first line of the data file:

NIJM_CAIRO_CRC_002	0	No distant metastasis, number of affected organs is 0 (array CGH data is available)
13-jan-2003	60	0 5-mrt-1942 1 1 1
-1	-1	-1 -1 -1 -1 0
-3	-1	-1 0 18 -1
1	435	-1 -1 435 -1
1	0	1 1454 1
1	1	0

Depending on the given split factor (section 3.11), a number of Dataimport_(n).xml files will be created.

When this file is (these files are) uploaded in OpenClinica, the data conversion process will be completed.

Opmerking [JR13]: ... conversion and upload process

To do this, start OpenClinica, click ok "Import Data" of the tasks menu.

The screenshot shows the OpenClinica interface for the study 'POBASCAM_demo'. The main menu bar includes 'Home', 'Subject Matrix', 'Notes & Discrepancies', 'Study Audit Log', 'Tasks', 'Report Issue', 'Support', and 'Study Subject Id'. The 'Tasks' menu is open, displaying various options like 'Submit Data', 'Extract Data', and 'Administration'. The 'Import Data' option is circled in red. The central workspace displays study details such as 'Site: VUmc, dept of Pathology', 'Enrolled: 3006', 'Expected Enrollment: 3006', and a 100% completion bar. Below this is a table of event statuses: scheduled (0, 0%), data entry started (4, 0%), completed (3002, 100%), signed (0, 0%), locked (0, 0%), skipped (0, 0%), and stopped (0, 0%).

In the next screen, you can enter the path to Dataimport_(n).xml:

The screenshot shows the OCDataImporter interface within the OpenClinica system. The main title is 'OCDataImporter'. The top navigation bar includes links for 'POBASCAM_demo : VUmc, dept of Pathology (VUmc, dept of Pathology)', 'Change Study/Site', 'cparlayan (Data Manager)', 'Log Out', 'Home', 'Subject Matrix', 'Notes & Discrepancies', 'Study Audit Log', 'Tasks', 'Report Issue', 'Support', 'Study Subject Id', and 'Go'. On the left, there's a sidebar with 'Alerts & Messages' and 'Info' sections. The 'Info' section shows 'Study: POBASCAM_demo', 'Site: VUmc, dept of Pathology', and 'Start Date: N/A'. The main content area is titled 'Import CRF Data' with a question mark icon. It contains instructions about uploading XML files and a file upload form with a 'Browse...' button. Below the form are 'Continue' and 'Cancel' buttons.

Click then "continue" to start data import. Repeat this for all Dataimport files.

3.12 Importing CRF data containing repeating events and repeating groups

Importing CRF data containing repeating events and repeating groups are in general very similar to importing simple CRF data as explained in the above sections. The way of designing the column names for repeating events and/or repeating groups in the data file must be like one of the following formats:

3.12.1 Importing CRF data containing repeating events and repeating groups by using data file containing the repeating information on separate columns:

columnName_Ex_Gy Where E stands for Repeating Events, x = StudyEventRepeatKey,
G stands for Repeating groups, y = ItemGroupRepeatKey.

Therefore valid examples are: col_E1, col_E1_G1, col_G1, col_E2, etc.

Example data file:

```
subject_id      item_before_E1 item_before_E2 Adverse_event_E1_G1 Adverse_event_E1_G2
                 Adverse_event_E1_G3 Adverse_event_E2_G1 Adverse_event_E2_G2
                 Adverse_event_E2_G3 Date_onset_E1_G1   Date_onset_E1_G2   Date_onset_E1_G3
                 date_onset_E2_G1   date_onset_E2_G2   date_onset_E2_G3
KG_CRC_IlandIII_V2_002    bef1     bef2     ae11    ae12     ae13    ae21     ae22    ae23    1-1
1960      1-2-1960        1-3-1960       2-1-1960      2-2-1960        2-3-1960
```

Will result in:

```
<?xml version="1.0" encoding="UTF-8"?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.3"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.cdisc.org/ns/odm/v1.3 ODM1-3.xsd"
  ODMVersion="1.3" FileOID="1D20080412202420" FileType="Snapshot"
  Description="Dataset ODM" CreationDateTime="2012-11-01T10:00:00" >
  <ClinicalData StudyOID="S_CUNEYTTE" MetaDataVersionOID="v1.0.0">
    <SubjectData SubjectKey="SS_KG_CRC_IlandIII_V2_007">
      <StudyEventData StudyEventOID="SE_TEST_MEDICATIE_GROUPS_REPEAT"
        StudyEventRepeatKey="1">
        <FormData FormOID="F_ADVE_6538_V11">
          <ItemGroupData ItemGroupOID="IG_ADVE_UNGROUPED_4947"
            TransactionType="Insert" >
            <ItemData ItemOID="I_ADVE_ITEM_BEFORE" Value="bef1" />
          </ItemGroupData>
        </FormData>
      </StudyEventData>
    </SubjectData>
  </ClinicalData>
</ODM>
```

```

OCDatImporter
<ItemGroupData ItemGroupOID="IG_ADVE_ADVERSEEVENTS_7407"
ItemGroupRepeatKey="1" TransactionType="Insert" >
    <ItemData ItemOID="I_ADVE_ADVERSE_EVENT_678" Value="ae11" />
    <ItemData ItemOID="I_ADVE_DATE_ONSET_7692" Value="1960-01-01" />
</ItemGroupData>
    <ItemGroupData ItemGroupOID="IG_ADVE_ADVERSEEVENTS_7407"
ItemGroupRepeatKey="2" TransactionType="Insert" >
        <ItemData ItemOID="I_ADVE_ADVERSE_EVENT_678" Value="ae12" />
        <ItemData ItemOID="I_ADVE_DATE_ONSET_7692" Value="1960-02-01" />
</ItemGroupData>
    <ItemGroupData ItemGroupOID="IG_ADVE_ADVERSEEVENTS_7407"
ItemGroupRepeatKey="3" TransactionType="Insert" >
        <ItemData ItemOID="I_ADVE_ADVERSE_EVENT_678" Value="ae13" />
        <ItemData ItemOID="I_ADVE_DATE_ONSET_7692" Value="1960-03-01" />
</ItemGroupData>

<StudyEventData StudyEventOID="SE_TEST_MEDICATIE_GROUPS_REPEAT"
StudyEventRepeatKey="2"ItemGroupRepeatKey="1" TransactionType="Insert" >
        <ItemData ItemOID="I_ADVE_ADVERSE_EVENT_678" Value="ae21" />
        <ItemData ItemOID="I_ADVE_DATE_ONSET_7692" Value="1960-01-02" />
</ItemGroupData>
    <ItemGroupData ItemGroupOID="IG_ADVE_ADVERSEEVENTS_7407"
ItemGroupRepeatKey="2" TransactionType="Insert" >
        <ItemData ItemOID="I_ADVE_ADVERSE_EVENT_678" Value="ae22" />
        <ItemData ItemOID="I_ADVE_DATE_ONSET_7692" Value="1960-02-02" />
</ItemGroupData>
    <ItemGroupData ItemGroupOID="IG_ADVE_ADVERSEEVENTS_7407"
ItemGroupRepeatKey="3" TransactionType="Insert" >
        <ItemData ItemOID="I_ADVE_ADVERSE_EVENT_678" Value="ae23" />
        <ItemData ItemOID="I_ADVE_DATE_ONSET_7692" Value="1960-03-02" />
</ItemGroupData>

```

All repeating data columns should be mapped with the same OC item, as to be seen below.

Study Data Column	OC Target Item
subject_id	Use link button 'CopyTarget' to fill this cell with the selected target item
item_before_E1	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_ADVE_6538_V11.IG_ADVE_UNGROUPED_4947.I_ADVE_ITEM_BEFORE
item_before_E2	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_ADVE_6538_V11.IG_ADVE_UNGROUPED_4947.I_ADVE_ITEM_BEFORE
Adverse_event_E1_C1	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_ADVE_6538_V11.IG_ADVE_ADVERSEEVENTS_7407.I_ADVE_ADVERSE_EVENT_678
Adverse_event_E1_C2	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_ADVE_6538_V11.IG_ADVE_ADVERSEEVENTS_7407.I_ADVE_ADVERSE_EVENT_678
Adverse_event_E1_C3	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_ADVE_6538_V11.IG_ADVE_ADVERSEEVENTS_7407.I_ADVE_ADVERSE_EVENT_678
Adverse_event_E2_C1	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_ADVE_6538_V11.IG_ADVE_ADVERSEEVENTS_7407.I_ADVE_ADVERSE_EVENT_678
Adverse_event_E2_C2	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_ADVE_6538_V11.IG_ADVE_ADVERSEEVENTS_7407.I_ADVE_ADVERSE_EVENT_678
Adverse_event_E2_C3	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_ADVE_6538_V11.IG_ADVE_ADVERSEEVENTS_7407.I_ADVE_ADVERSE_EVENT_678
Date_onset_E1_G1	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_ADVE_6538_V11.IG_ADVE_ADVERSEEVENTS_7407.I_ADVE_DATE_ONSET_7692

OCDataImporter

Study Data Column	OC Target Item	Study Subject ID?
subject_id	Use link button 'CopyTarget' to fill this cell with the selected target item	<input checked="" type="checkbox"/>
item_before_E1	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_UNGROUPED_4947.I_Adve_Item_Before	<input type="checkbox"/>
item_before_E2	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_UNGROUPED_4947.I_Adve_Item_Before	<input type="checkbox"/>
Adverse_event_E1_C1	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Adverse_Event_678	<input type="checkbox"/>
Adverse_event_E1_C2	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Adverse_Event_678	<input type="checkbox"/>
Adverse_event_E1_C3	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Adverse_Event_678	<input type="checkbox"/>
Adverse_event_E2_C1	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Adverse_Event_678	<input type="checkbox"/>
Adverse_event_E2_C2	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Adverse_Event_678	<input type="checkbox"/>
Adverse_event_E2_C3	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Adverse_Event_678	<input type="checkbox"/>
Date_onset_E1_G1	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Date_Onset_7692	<input type="checkbox"/>

Study Data Column	OC Target Item	Study Subject ID?
Adverse_event_E1_C3	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Adverse_Event_678	<input type="checkbox"/>
Adverse_event_E2_C3	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Adverse_Event_678	<input type="checkbox"/>
Adverse_event_E2_C2	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Adverse_Event_678	<input type="checkbox"/>
Adverse_event_E2_C1	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Adverse_Event_678	<input type="checkbox"/>
Date_onset_E1_G1	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Date_Onset_7692	<input type="checkbox"/>
Date_onset_E1_G2	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Date_Onset_7692	<input type="checkbox"/>
Date_onset_E1_G3	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Date_Onset_7692	<input type="checkbox"/>
Date_onset_E2_C1	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Date_Onset_7692	<input type="checkbox"/>
Date_onset_E2_C2	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Date_Onset_7692	<input type="checkbox"/>
Date_onset_E2_C3	SE_TEST_MEDICATIE_GROUPS_REPEAT.F_Adve_6538_V11.IG_Adve_AdverseEvents_7407.I_Adve_Date_Onset_7692	<input type="checkbox"/>

Note: Please do not use data column names containing “_Ex” or “_Gx” where x is a digit; unless you are referring to a repeating event and/or group item, as this will not work and produce an error.

Example: Adding subjects, study subjects and repeating study events with start dates:

StartDateT0_E1 is the start date for event number 1 and StartDateT0_E2 for event 2. The suffixes E1 and E2 represents this.

Data file:

PersonID	StudySubjectID	Geboortedatum	Geslacht	StartDateT0_E1	StartDateT0_E2
TR-20-0086	ESRA-20-0112	02-07-1945	f	09-02-2012	
TR-20-0146	ESRA-20-0175	16-01-1953	f	17-02-2011	18-02-2011
TR-20-0050	ESRA-20-0068	24-10-1932	m	14-10-2010	
TR-20-0066	ESRA-20-0090	28-08-1976	f	19-05-2011	

OCDataImporter:

OC TARGET: Study Event
CRF
Group
Item

SE_T0_3737
F_TRACERRETROE_001
IG_TRACE_UNGROUPED
-- select --

Date format in study items dd-mm-yyyy
Default sex of subjects f
Gender Code for m: m
Gender Code for f: f

Split the ODM file where each contains the following number of subjects (0 = no split) 0
Location: Amsterdam
Check

Limit matching number of characters to: 0 (0 = no limit, recommended)
Match columns
Unmatch columns

	Study Data Column	OC Target Item	Study Subject ID?	Date?	Subject Sex?	Subject Person	Subject Date of Birth?	Subject start date?	CopyTarget
PersonID	none		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	CopyTarget
StudySubjectID	none		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	CopyTarget
Geborendatum	none		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	CopyTarget
Geslacht	none		<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	CopyTarget
StartDateT0_E1	none		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	CopyTarget
StartDateT0_E2	none		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	CopyTarget
*			<input type="checkbox"/>						

OCDatalImporter

Insert statements generated for TR-20-0146:

```
INSERT INTO subject(status_id, gender, unique_identifier, date_created, owner_id, dob_collected, date_of_birth)
VALUES (1, 'f', 'TR-20-0146', '2013-08-28', 1, '1', '1953-01-16');
INSERT INTO study_subject(label, study_id, status_id, enrollment_date, date_created, date_updated, owner_id, oc_oid, subject_id)
VALUES ('ESRA-20-0175', (SELECT study_id FROM study WHERE oc_oid = 'S_TRACERUM'),
1, '2013-08-28', '2013-08-28', 1, 'SS_ESRA-20-0175', (SELECT subject_id
FROM subject where unique_identifier = 'TR-20-0146'));
INSERT INTO study_event(study_event_definition_id, study_subject_id, location, sample_ordinal, date_start, owner_id, status_id, date_created, subject_event_status_id, start_time_flag, end_time_flag)
VALUES ((SELECT study_event_definition_id FROM study_event_definition WHERE oc_oid =
'SE_T0_3737'),
(SELECT study_subject_id FROM study_subject WHERE oc_oid = 'SS_ESRA-20-0175'), 'Amsterdam', 1, '2011-02-17 12:00:00', 1, 1, '2013-08-28', 3, '0', '0');
INSERT INTO study_event(study_event_definition_id, study_subject_id, location, sample_ordinal, date_start, owner_id, status_id, date_created, subject_event_status_id, start_time_flag, end_time_flag)
VALUES ((SELECT study_event_definition_id FROM study_event_definition WHERE oc_oid =
'SE_T0_3737'),
(SELECT study_subject_id FROM study_subject WHERE oc_oid = 'SS_ESRA-20-0175'), 'Amsterdam', 2, '2011-02-17 12:00:00', 1, 1, '2013-08-28', 3, '0', '0');
```

Result in OpenClinica after running inserts in PGAdmin:

Subject Matrix for TRACER-UMCU-BIOLOGICALS [?](#)

15 | Show More | Select An Event

Study Subject ID	T=0	Actions
ESRA-20-0175		Apply Filter Clear Filter
ESRA-20-0175	x2	

Results 1 - 1 of 1.

Subject: ESRA-20-0175	
Event: T=0	
Occurrence#1 of 2 17-Feb-2011 Status: data entry started	Occurrence#2 of 2 17-Feb-2011 Status: data entry started

[Click for more options](#)

3.12.2 Importing CRF data containing repeating events and repeating groups, by using data file containing the repeating information under columns EVENT_INDEX and GROUP_INDEX

In this approach a subject must be repeated in rows, rather than columns and that is the only difference between the method described in 3.18.1.

The file in this format which was described in 3.18.1 would now look like:

OCDataImporter

PersonID	StudySubjectID	Geboortedatum	Geslacht	StartDateT0	EVENT_INDEX
TR-20-0086	ESRA-20-0112	02-07-1945	f	09-02-2012	1
TR-20-0146	ESRA-20-0175	16-01-1953	f	17-02-2011	1
TR-20-0146	ESRA-20-0175	16-01-1953	f	18-02-2011	2
TR-20-0050	ESRA-20-0068	24-10-1932	m	14-10-2010	1
TR-20-0066	ESRA-20-0090	28-08-1976	f	19-05-2011	1

A more general example is:

subject_id	item_before	Adverse_event	EVENT_INDEX	GROUP_INDEX
1000	text1	event	1	1
1000	text1	event	2	1
1000	text1	event	1	2
1000	text1	event	2	2

Etc.

Column names EVENT_INDEX and GROUP_INDEX must be used only for above purpose.

Do not combine EVENT_INDEX, GROUP_INDEX and _E1, _G1 column names, use one approach for the entire input file.

Both methods have their advantages and disadvantages.

3.13 Data validation

The data in the input file is validated against the CRF definitions during the process. For each **wrong** data a warning is created and logged in file "OCDataImporter_warnings.txt" in the working directory.

List of possible warnings are:

- 1- Item is mandatory but has no value
- 2- Item is mandatory and hidden; and has no value. This might cause errors if the item gets shown by another condition
- 3- Item type is real but contains non numeric characters:
- 4- Item contains more numbers than allowed after the decimal point
- 5- Item type is integer but contains non integer characters
- 6- Item length <n> is too small
- 7- Value not in the specified code list
- 8- Range Check fail: Value <value> is not GT|GE|LT|LE|NE|EQ <value specified in CRF>
- 9- Input data file format incorrect at line = <n> Expecting: <m> columns, found <k>
- 10- "Duplicate key <value> at line <n>
- 11- Subject sex can be only 'f' or 'm'. You have <tekst> at line <n>
- 12- Invalid subject birth date <date> at line <n>
- 13- Invalid start date <date> at line <n>

Opmerking [JR14]: Invalid data field

Opmerking [JR15]: Possible warnings are:

Opmerking [JR16]: text

It is **recommended** that the data file should be warning-free before data upload.

Opmerking [JR17]: mandatory

Appendix: Error messages

- 1- User Manual not found: User manual can't be found in the installation directory. See section 3.1
- 2- Can't open selected data file "thedatafile.txt", can't continue. Delimiter = **Items per line = 20**: The data file can't be opened, it may not exist or you don't have enough privileges to open it.
- 3- Failed to start Acrobat reader: acrord32 doesn't exist to open the document file. Go to the installation directory and double click on the document file to open it manually.
- 4- Do you want to load your previous grid?: The program saves the last grid used; if you want to reload the last saved grid, click yes.

OCDatImporter

- 5- Can't generate grid dump file (see log file for details - Do you have enough permissions to write in target folder?): You probably have no write permissions; log file is created in the working directory.
- 6- DataImport_* files will be overwritten. Do you want to delete the old files?: If you run the program on a directory which older DataImport files exist, they will be overwritten. Click No if you don't want to lose your old files, save them elsewhere and try again.
- 7- Please select (only) one field as STUDY SUBJECT ID by using check box; You have 2 selected.: Only one and only one study subject id may be selected.
- 8- "Please select at most one field as STUDY SUBJECT SEX/PERSON ID/SUBJECT DATE OF BIRTH/STUDY START DATE by using check box; You have 2 selected.: These can only be checked for one row maximum.
- 9- Please enter location: Location name is mandatory.
- 10- Input data file format incorrect at line = 26 Expecting: 12; found: 13 items; this is the faulty line: At this line there is a mismatch with number of columns and number of data items. (12 columns and 13 items)
- 11- Duplicate key "thekeyvalue" at line = 34: If the duplicate key check is performed (section 3.6.5) and a duplicate key is detected, this message will be displayed.
- 12- Subject sex can be only 'f' or 'm'. You have "'MALE'" at line 12. Index: 5. Exiting...The generated files ARE INCOMPLETE AND CAN NOT BE USED : At the 5th column there should be the gender code and this must be conform section 3.6.2.
- 13- Invalid subject birth date "12061998" at line 23. Index: 7. Exiting...The generated files ARE INCOMPLETE AND CAN NOT BE USED: At the 7th column there should be the birth date but that can't be converted to ODM format. See section 3.6
- 14- Invalid subject start date "12061998" at line 23. Index: 7. Exiting...The generated files ARE INCOMPLETE AND CAN NOT BE USED: At the 7th column there should be the start date but that can't be converted to ODM format. See section 3.6
- 15- Exception while reading data file: Unexpected error; see the log file for details.
- 16- Error while getting STUDYEVENT Repeating Key: Cant resolve the DataItemColumnName Adverse_event_3 + ". The proper name should look like 'DataItem_E2_G3 Where E2 means Event repeating key = 2 and G3 means Group repeating key = 3. Exiting...The generated files ARE INCOMPLETE AND CAN NOT BE USED: The name of the data item column for repeating events and groups has a format of DataItem_Ex_Cy where x=repeat key study event and y=repeating group. This is not the case with Adverse_Event_3. See section 3.12.
- 17- Error while getting GROUP Repeating Key: See 16.
- 18- Wrong index at: 67. Exiting...The generated files ARE INCOMPLETE AND CAN NOT BE USED"; The program is unable to get the location of the item at the specified line.
- 19- Can't get study event/group/event/CRF/item definitions; please check the format of the file: The meta file created by OpenClinica is probably corrupt. Use an XML editor to see what is wrong; eventually regenerate the file.
- 20- Please enter or select correct input files: Either type two file names separated by a semicolon (;) or use browse button.: Specify the 2 input files correctly.
- 21- Can't open selected data file, PROBABLY file doesn't belong to you or is read only. Please make sure you are the owner, then try again: This happens mostly when the data file is saved on another computer or by someone else than the user who is running the program now. Make a copy of the data file and use that one as input.
- 22- Can't open selected OC meta file, can't continue: See log file.
- 23- Not all Items in the selected CRF could be matched. For the list of UNMATCHED Items, see the progress textbox below. You can match those items by using the comboboxes above. Control the matched items too, as the matching can not be 100% correct!: See section 3.8
- 24- All Items in the selected CRF could be matched. Control the matched items as the matching can not be 100% correct!: See section 3.8
- 25- Process is not finished yet. Are you sure you want to stop this program?: This is issued when the cancel button is hit. The process has to be rerun after this.
- 26- Please use all of the comboboxes above to define a target item. (there are still -- select --'s up there): When the "copy target" link button is clicked, all of the target item comboboxes must be selected. See section 3.8
- 27- Please read input files first: Hit the "Read input files" button to do this.
- 28- Data file contains _E and/or _G columns while EVENT_INDEX and/or GROUP_INDEX columns also exist. Please stick to one method of file construction; having both is ambiguous: See section 3.12