

OpenClinica - Data Import & Export

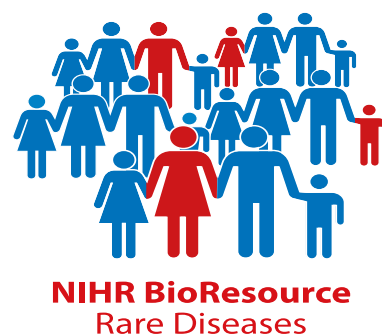
Purpose of this document: To provide instruction and best practice guidelines' for the import and export of study data into and out of OpenClinica.

Version: OpenClinica - Data Import & Export v1.1

Date: 26 January 2015

Authors: Catherine Titterton

NIHR Rare Diseases Translational Research Collaboration
Department of Haematology, University of Cambridge
National Health Service Blood and Transplant Cambridge
Long Road, Cambridge
CB2 0PT



Document control

Version	Date	Who	Changes
1.0	21/08/2014	Catherine Titterton	First draft
1.1	26/01/2015	Catherine Titterton	Inclusion of section 4 – Advanced Data Import; section 5 – CIT Data importer

Check if there is a later version of this document on the CIT website.

Table of Contents

1	Overview.....	5
1.1	Introduction.....	5
2	Installing the import tools	6
2.1	OCdataimporter.....	6
2.2	CIT Importer.....	6
2.3	TDS access database.....	6
3	Import Subjects & Events	8
3.1	Prepare your subject and event data	8
3.2	TDS access database.....	8
3.3	Import your subject records to TDS	9
3.4	Import and Schedule Events with TDS	12
4	Importing CRF data with OCDataImporter	16
4.1	OCdataimporter – manual mapping.....	17
4.2	Download your study metadata file	18
4.3	Match you study data and study metadata	19
4.4	Generate the XML CRF data import file	21
4.5	Import your CRF data XML file into OpenClinica.....	22
5	CIT data importer	24
6	Advanced Data Import Mapping	30
6.1	Files required for advanced mapping import.....	30
6.2	Extract source files for creation of mapping definitions file	30
6.3	Create Mapping Definitions file.....	32
6.3.1	In Excel - create the mapping definitions file	32
6.3.2	Creating the ODI grid.dmp file.....	33
6.3.3	Generate XML data import file	35
7	Data Export from OpenClinica	36
8	Data Cleaning.....	38
9	Test Data Generation (optional).....	38
10	Common Import Errors.....	39
10.1	Study Subject IDs and OIDs	39

11 Exporting Data 40

1 Overview

1.1 Introduction

There are a number of tools and approaches available for importing data into OpenClinica. Each can be used singly or in parallel, depending on the state of your data and level of expertise or familiarity with the tools. All tools are accessible to the end user and do not require specialist IT technical skills. Tools available include:

- Excel
- TDS data importer (developed by <http://www.trialdatasolutions.com>)
- OC Data Importer (developed by <https://community.openclinica.com/extension/ocdataimporter>)
- CIT Importer (developed by the Clinical Infrastructure team)

Which tool should I use?

There is no hard and fast rule which tool will be most appropriate for you use, they can all do the job with the same end results. The key differences between the tools are as follows:

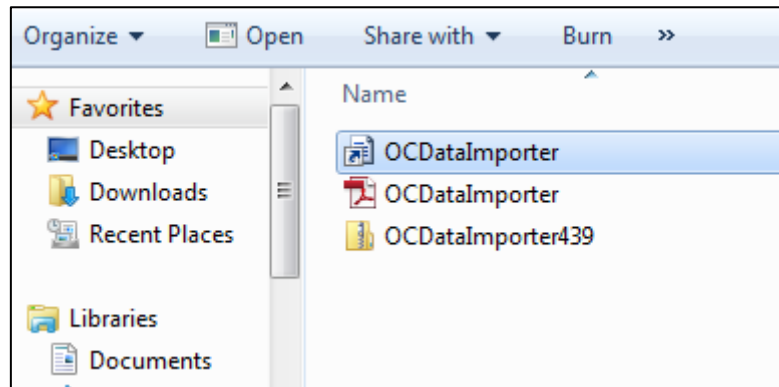
- CIT Data Importer includes an option to skip populating events and fields for which no data has been collected against a particular subject.
- OC DataImporter has good error checking prior to data load into OpenClinica. You can rerun the generation of the XML file over and over again until the report files show no errors. CIT data importer runs the import immediately and only shows errors after import which can then be corrected with a second import (OpenClinica keeps an audit log of changes).
- Both CIT and OC data importers have a manual mapping process. The advanced data mapping process outlined in section 6 can be used with both tools, although the excel formulas are currently set up to create the OC data import mapping file, not the mapping file for CIT.
- TDS Importer is useful for those comfortable working in Access and primarily used for loading subjects and scheduling events, not for loading study data.

2 Installing the import tools

2.1 OCdataimporter

OCDataImporter is a graphical interface that takes a text file as input and maps the data to the Study Meta Data XML file for conversion and import of CRF data as an XML file into OpenClinica.

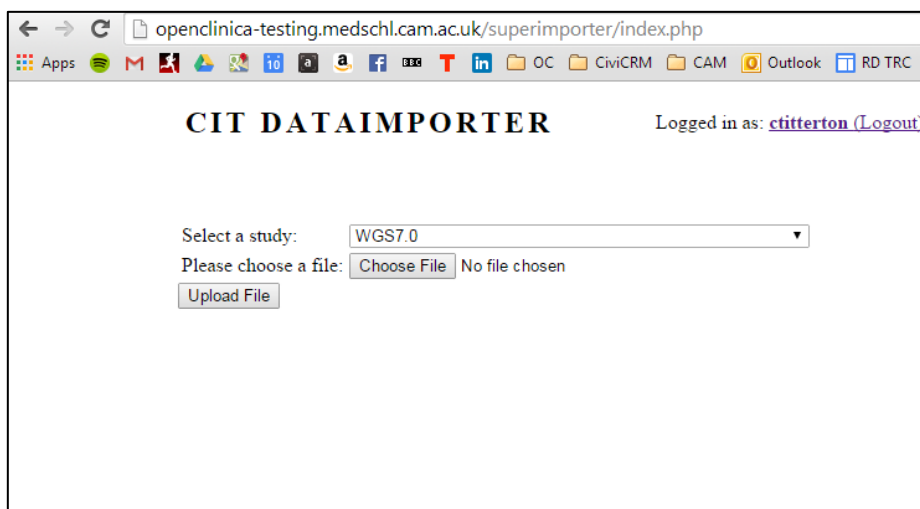
- 1.) Download OCdataimporter from <https://community.openclinica.com/extension/ocdataimporter>. Unzip the files and save to a new folder .../Desktop/OC Import.



- 2.) Click the shortcut OCDataImporter to open the application.

2.2 CIT Importer

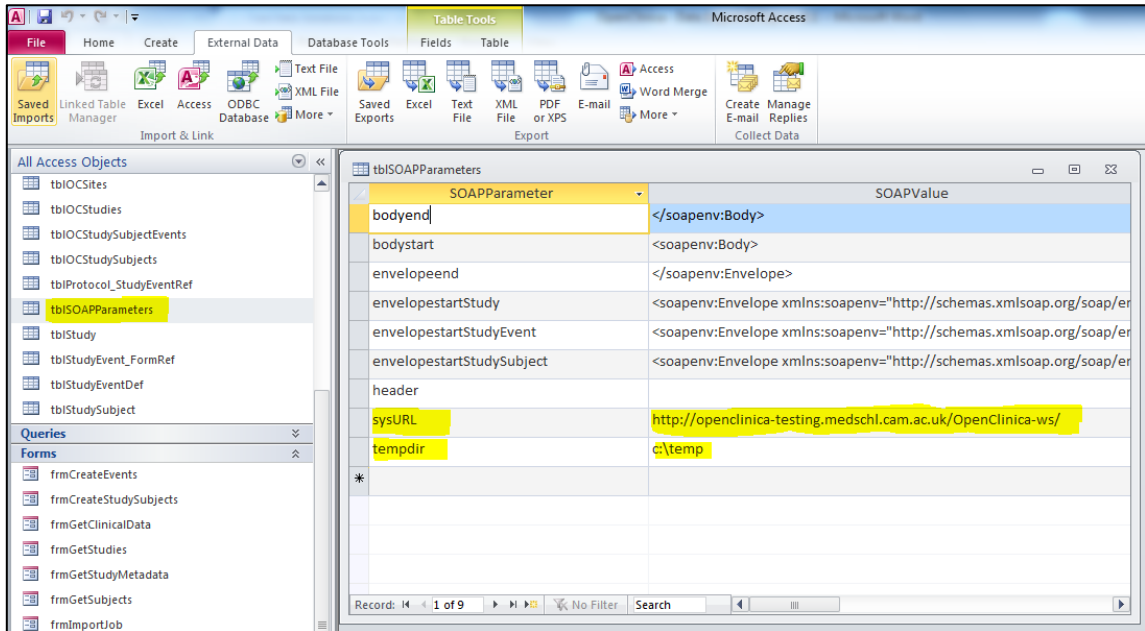
The CIT Importer does not require download. It is installed on your OpenClinica instance and available directly through a url <http://openclinica-testing.medschl.cam.ac.uk/superimporter/index.php> (the highlighted section will be specific to your instance). If you are using a locally managed instance of OpenClinica, please contact the CIT team for information on installing the CIT Data Importer on your server.



2.3 TDS access database

The TDS Access database.

1. Download the TDS Access database file [here](#).
2. Unzip the file and open the Access mdb-file.
3. Go to the tables-section and open **tbISOAPPParameters**.



4. Change **sysURL** to match you're the location of your OpenClinica server. (Ensure that the end / is included in your url)
5. Change **tempdir** to the directory you use on your local PC for temporary files.

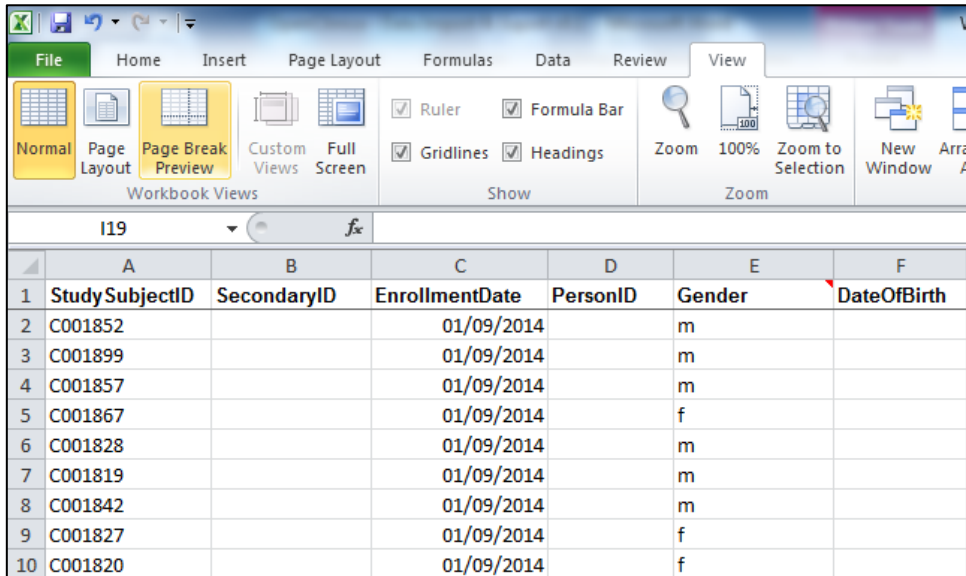
To use the TDS importer it is assumed you have your web-services running on OpenClinica and that your OpenClinica user account has been set up as an authorized user root to use the SOAP-services.

3 Import Subjects & Events

3.1 Prepare your subject and event data

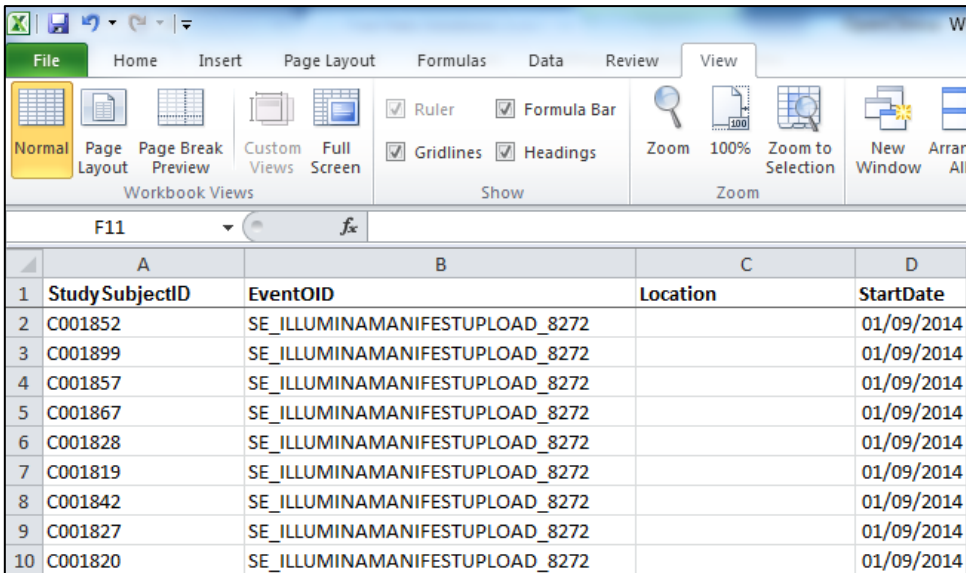
It is assumed your data is in an Excel format. The example below shows only three fields being populated. Depending on your study *Secondary ID*, *Person ID* and *Date of Birth* may also be populated.

Prepare an Excel file with both Subject and Event records as show below. (These can be in the same Excel file on different tabs). Ensure Gender is in the form of lowercase **m** and **f**. OpenClinica is case sensitive and import of M and F (or Male/Female) will fail.



	A	B	C	D	E	F
1	StudySubjectID	SecondaryID	EnrollmentDate	PersonID	Gender	DateOfBirth
2	C001852		01/09/2014		m	
3	C001899		01/09/2014		m	
4	C001857		01/09/2014		m	
5	C001867		01/09/2014		f	
6	C001828		01/09/2014		m	
7	C001819		01/09/2014		m	
8	C001842		01/09/2014		m	
9	C001827		01/09/2014		f	
10	C001820		01/09/2014		f	

Event records are shown below. Extraction of the **EventOID** is detailed below in section 3.4

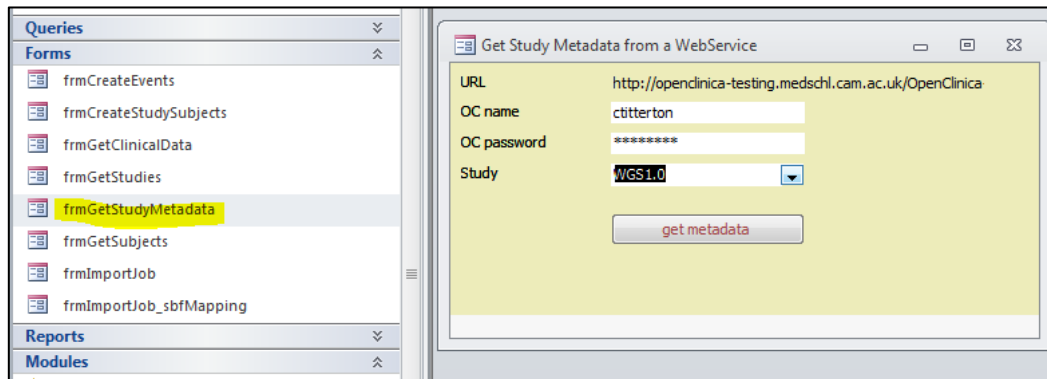


	A	B	C	D
1	StudySubjectID	EventOID	Location	StartDate
2	C001852	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
3	C001899	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
4	C001857	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
5	C001867	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
6	C001828	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
7	C001819	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
8	C001842	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
9	C001827	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
10	C001820	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014

3.2 TDS access database

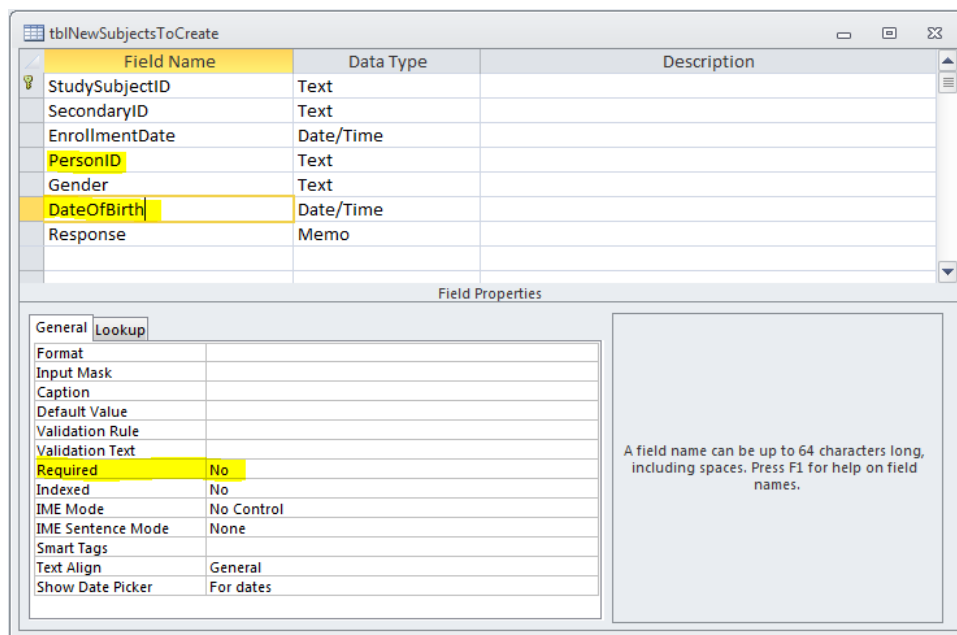
Study subjects can be added using the TDS Access database. (See section 2.3 TDS access database for installation instructions)

In the TDS Access database, open the form **frmGetStudies**. Fill in your OpenClinica *username* and *password* and click the button *get studies*. The result is written into **tblOCStudies**.



3.3 Import your subject records to TDS

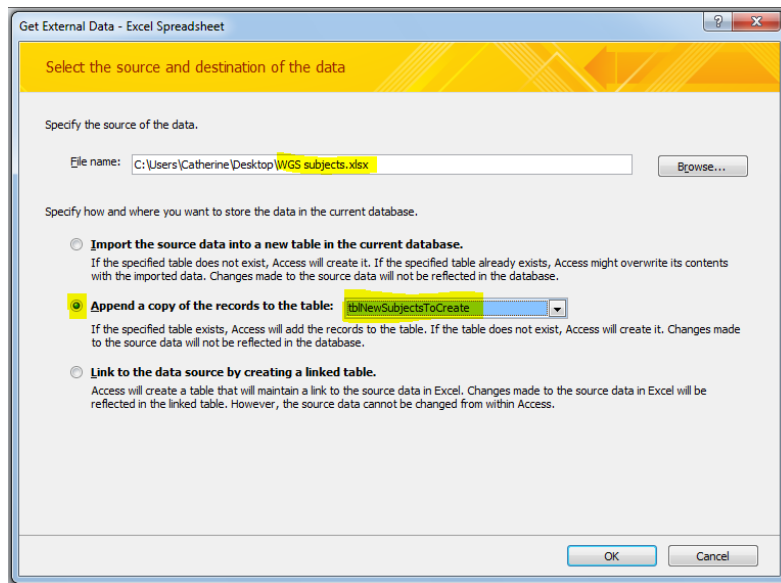
1. OPTIONAL: If you are not importing Date of Birth or Person ID then rightmouse click the **tblNewSubjectsToCreate** open in design mode and change the attribute of fields *PersonID* and *DateOfBirth* to Required: No



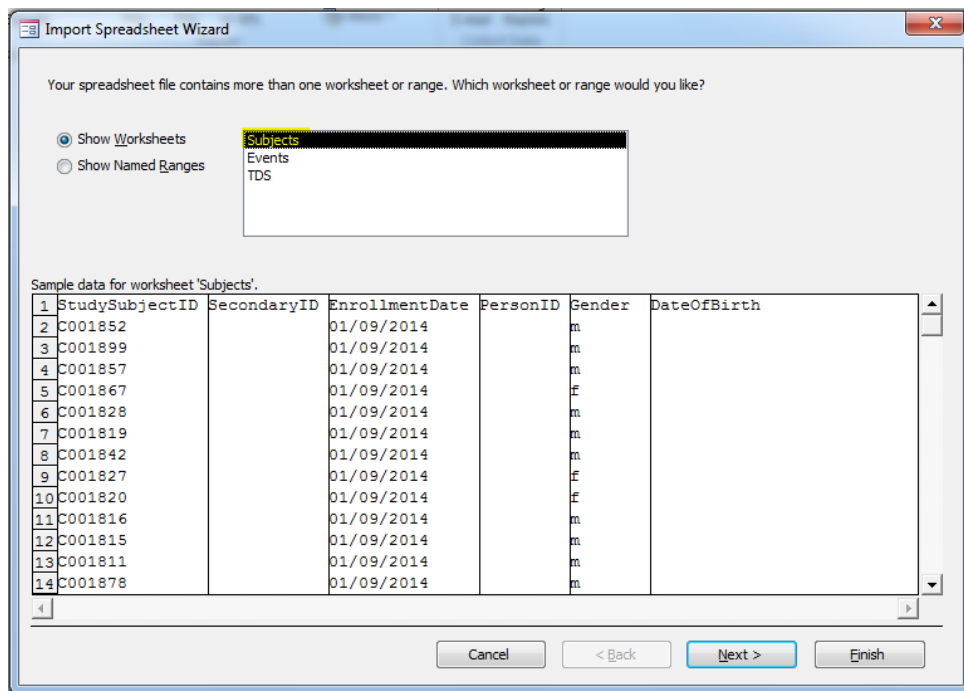
2. On the External data tab of the Access toolbar, click Excel (to import)



3. which opens the Get External Data dialog. Browse to your excel file containing your subject data.
4. select the option to Append a copy of records to the table and select from the dropdown **tblNewSubjectsToCreate**



5. Click OK
6. Select the worksheet which contains subject data. (If the column headers in your spread sheet are named the same as the Access table columns you will not need to match the fields)



7. Click Next
8. Click Finish
9. Open the table **tblNewSubjectsToCreate** to show the subject records imported into Access.

StudySubjectID	SecondaryII	EnrollmentDate	PersonID	Gende	DateOfBirth	Response
C001807		01/09/2014		F		
C001808		01/09/2014		F		
C001809		01/09/2014		M		
C001810		01/09/2014		M		
C001811		01/09/2014		M		
C001812		01/09/2014		F		
C001813		01/09/2014		M		
C001814		01/09/2014		F		
C001815		01/09/2014		M		
C001816		01/09/2014		M		

10. Open the *form frmCreateStudySubjects*, fill in your *username* and *password* and choose the appropriate Study. If a Study has one or more Sites, you can choose one of the Sites and the Subjects will be assigned to that Site. If you leave this blank, the Subjects will be assigned to the Study.
11. Click the button *create study subjects*
12. Once the process has run, you can open the *table tblNewSubjectsToCreate* and the response field will contain details of success or error. If there are errors, review your input data, selection of study etc.

StudySubjectID	SecondaryII	EnrollmentDate	PersonID	Gende	DateOfBirth	Response
C001807		01/09/2014		f		<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"><SOAP-ENV:Header/><SOAP-ENV:Body><createResponse xmlns="http://openclinica.org/ws/studySubject/v1"><result xmlns="http://openclinica.org/ws/studySubject/v1"><Success</result></createResponse></SOAP-ENV:Body></SOAP-ENV:Envelope>
C001808		01/09/2014		f		<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"><SOAP-ENV:Header/><SOAP-ENV:Body><createResponse xmlns="http://openclinica.org/ws/studySubject/v1"><result xmlns="http://openclinica.org/ws/studySubject/v1"><Success</result></createResponse></SOAP-ENV:Body></SOAP-ENV:Envelope>

13. View the imported subject records in OpenClinica in the subject matrix

Study Subject ID	Illumina Manifest Upload	Illumina QC	Sequencing files check	Actions
C001807				
C001808				
C001809				

3.4 Import and Schedule Events with TDS

When preparing your Excel import of event records you can find the StudyEventOIDs in the table **tblStudyEventDef**.

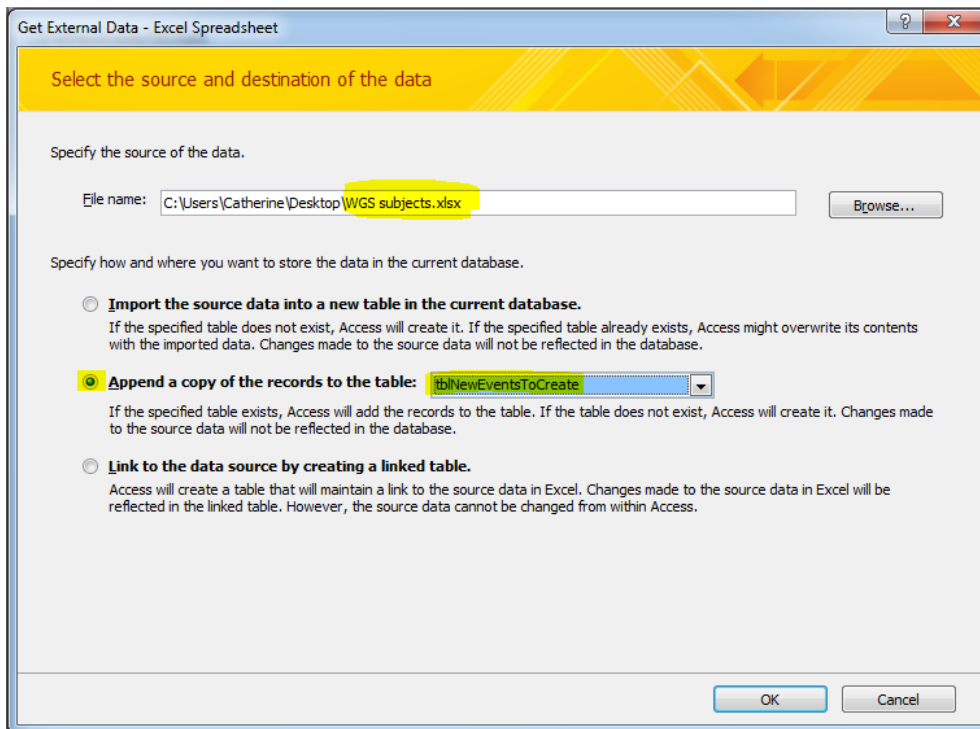
StudyEventOID	Name	Repeating	Type
SE_HPCPROCESSING_8939	HPC processing	No	Common
SE_ILLUMINAPROCESSING_3391	Illumina Processing	Yes	Common
SE_PLATECREATION_8479	Plate creation	No	Common

1. Copy and paste the event OID against your subject records in Excel. Note in the above example there are three events. For multiple events you need to create a row for each event against each subject (so three rows for each subject)

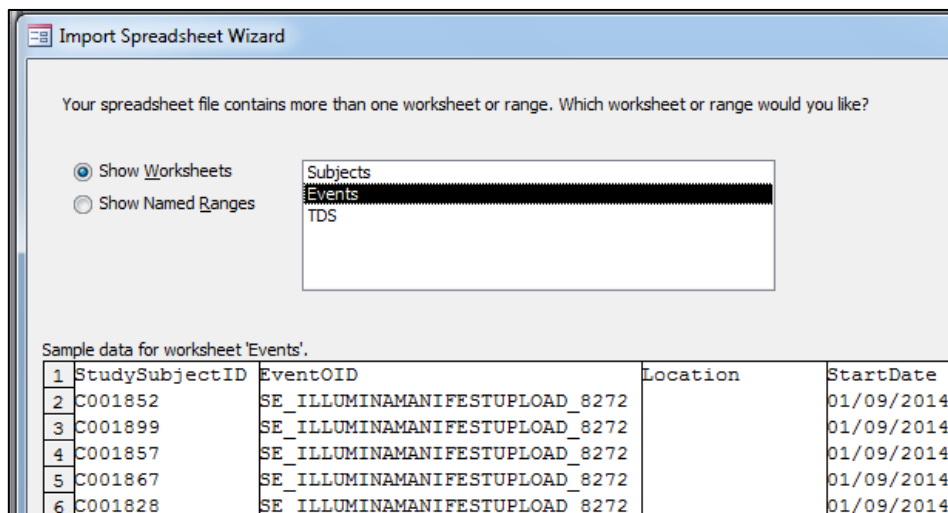
StudySubjectID	EventOID	Location	StartDate
C001807	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
C001807	SE_ILLUMINAQC_8933		07/09/2014
C001807	SE_SEQUENCINGFILESHECK_2213		18/09/2014

	A	B	C	D
1	StudySubjectID	EventOID	Location	StartDate
2	C001852	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
3	C001899	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
4	C001857	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
5	C001867	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
6	C001828	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
7	C001819	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
8	C001842	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
9	C001827	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014
10	C001820	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014

2. Follow the steps to import to Excel as per subject data > open the excel import option
3. Browse to the Excel file holding event data
4. Choose the append option and select the dropdown table **tblNewEventsToCreate**



5. Ensure you select the worksheet Events



6. Click Next
7. Click Continue
8. Check the table **tblNewEventsToCreate** to see your imported records

StudySubjectID	EventOID	Location	StartDate	Response
C001807	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014	
C001807	SE_ILLUMINAQC_8933		07/09/2014	
C001807	SE_SEQUENCINGFILESHECK_2213		18/09/2014	
C001808	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014	
C001808	SE_ILLUMINAQC_8933		07/09/2014	
C001808	SE_SEQUENCINGFILESHECK_2213		18/09/2014	
C001809	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014	
C001809	SE_ILLUMINAQC_8933		07/09/2014	

9. Open the form **frmCreateEvents** type in your *username* and *password* and select the appropriate study

10. Check the table **tblNewEventsToCreate** and check the response success – if there are errors review your data, table and study selected is correct.

StudySubjectID	EventOID	Location	StartDate	Response
C001807	SE_ILLUMINAMANIFESTUPLOAD_8272		01/09/2014	<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"><SOAP-ENV:Header/><SOAP-ENV:Body><scheduleResponse xmlns="http://openclinica.org/ws/event/v1"><result xmlns="http://openclinica.org/ws/event/v1"><Success/></result></eventDefinitionOID xmlns="http://openclinica.org/ws/event/v1">SE_ILLUMINAMANIFESTUPLOAD_8272</eventDefinitionOID><studySubjectOID xmlns="http://openclinica.org/ws/event/v1">SS_C001807</studySubjectOID><studyEventOrdinal xmlns="http://openclinica.org/ws/event/v1">1</studyEventOrdinal/></scheduleResponse></SOAP-ENV:Body></SOAP-ENV:Envelope>
C001807	SE_ILLUMINAQC_8933		07/09/2014	<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"><SOAP-ENV:Header/><SOAP-ENV:Body><scheduleResponse xmlns="http://openclinica.org/ws/event/v1"><result xmlns="http://openclinica.org/ws/event/v1"><Success/></result></eventDefinitionOID xmlns="http://openclinica.org/ws/event/v1">SE_ILLUMINAQC_8933</eventDefinitionOID><studySubjectOID xmlns="http://openclinica.org/ws/event/v1">SS_C001807</studySubjectOID><studyEventOrdinal/></scheduleResponse></SOAP-ENV:Body></SOAP-ENV:Envelope>

11. Review your scheduled events inOpenClinica

Subject Matrix for WGS1.0 🔗

⏪ ⏩ ⏴ ⏵ | 15 ▾ | Show More | Select An Event ▾ | Add New Subject

Study Subject ID	Illumina Manifest Upload	Illumina QC	Sequencing files check	Actions
				Apply Filter Clear Filter
C001822				
C001823				
C001824				
C001825				
C001826				
C001827				

4 Importing CRF data with OCDataImporter

Once you have added your subject and event records to OpenClinica you are now ready to import the actual CRF data collected by clinicians.

It is expected that you will already have gone through the process of creating your CRFs and that these have been tested and loaded into OpenClinica and assigned to the event records you created above. Development of CRFs is detailed in a separate document.

For simplicity the example below shows all data to be imported held in a single spreadsheet, although spread over three events/CRFs to be imported against. (Blue, green and yellow each represent a separate CRF)

Sample ID	Study Subject ID	Species	Gender (M/F/U)	Volume (ul)	Concentration (ng/ul)	OD 260/280	Tissue Source	Extraction Method	Ethnicity	Parent 1 ID	Parent 2 ID	Replicate(s) ID	ILMN GT	ILMN quant	insert size	le250	Comment	Date sent from ILMN	Date sent at
2	C001852_A	C001852	Homo sapien	M	105	50	1.82	Blood	Chloroform	White	British		Male	104.56		1.967364337		7.8.14	
3	C001899_A	C001899	Homo sapien	M	105	50	1.9	DNA	Chloroform	White	British		Male	85.1		2.269740478		7.8.14	
4	C001857_A	C001857	Homo sapien	M	105	50	1.85	Blood	Chloroform	White	British		Male	81.24		1.767426184		7.8.14	
5	C001867_A	C001867	Homo sapien	F	105	50	1.85	Blood	Chloroform	White	British		Female	142.93		1.403163689		7.8.14	
6	C001828_A	C001828	Homo sapien	M	105	50	1.88	Blood	Chloroform	White	British	C002031_/_C002026_A	Male	110.38		5.146985758		7.8.14	
7	C001819_A	C001819	Homo sapien	M	105	50	1.93	DNA	Chloroform	White	British	C002053_/_C002052_A	Male	85.39		3.563283199		7.8.14	
8	C001842_A	C001842	Homo sapien	M	105	50	1.87	Blood	Chloroform	White	British	C002044_/_C002043_A	Male	104.42		5.195884573		7.8.14	
9	C001827_A	C001827	Homo sapien	F	105	50	1.87	Blood	Chloroform	White	British	C002029_/_C002030_A	Female	151.92		3.554628989		7.8.14	
10	C001820_A	C001820	Homo sapien	F	105	50	1.87	DNA	Chloroform	White	British	C002017_/_C002016_A	Female	63.89		1.592303816		7.8.14	
11	C001816_A	C001816	Homo sapien	M	105	50	1.88	Blood	Chloroform	White	British	C002012_/_C002013_A	Male	78.18		1.083694248		7.8.14	
12	C001815_A	C001815	Homo sapien	M	105	50	1.85	Blood	Chloroform	White	British	C002010_/_C002011_A	Male	77.8		6.558420784		7.8.14	
13	C001811_A	C001811	Homo sapien	M	105	50	1.88	Blood	Chloroform	White	British	C002005_/_C002004_A	Male	85.35		5.073945236		7.8.14	
14	C001878_A	C001878	Homo sapien	M	105	50	1.775	Blood	Chloroform	Black	British	C002062_/_C002061_A	Male	126.96		3.53819762		7.8.14	
15	C001856_A	C001856	Homo sapien	M	105	50	1.84	Blood	Chloroform	Black	British	C002056_/_C002057_A	Male	77.74		5.809789736		7.8.14	
16	C001855_A	C001855	Homo sapien	F	105	50	1.86	Blood	Chloroform	Black	British	C002055_/_C002054_A	Female	98.02		1.540665929		7.8.14	
17	C001880_A	C001880	Homo sapien	F	105	50	1.78	Blood	Chloroform	Black	British		Female	89.6		2.461096416		7.8.14	
18	C001884_A	C001884	Homo sapien	F	105	50	1.77	Blood	Chloroform	Black	British	C002058_/_	Female	108.99					
19	C001866_A	C001866	Homo sapien	M	105	50	1.84	Blood	Chloroform	Black	British	C002059_/_	Male	80.65					
20	C001894_A	C001894	Homo sapien	M	105	50	1.88	DNA	Chloroform	Black	British	C002082_/_	Male	63.54					
21	C001879_A	C001879	Homo sapien	M	105	50	1.79	Blood	Chloroform	Indian		C002067_/_	Male	100.01					
22	C001823_A	C001823	Homo sapien	F	105	50	1.85	Blood	Chloroform	Indian		C002022_/_	Female	144.29					
23	C001822_A	C001822	Homo sapien	M	105	50	1.87	Blood	Chloroform	Indian		C002023_/_	Male	106.68					
24	C001869_A	C001869	Homo sapien	M	105	50	1.85	Blood	Chloroform	Indian			Male	92.3					
25	C001874_A	C001874	Homo sapien	F	105	50	1.78	Blood	Chloroform	Indian			Female	87.05					
26	C001875_A	C001875	Homo sapien	M	105	50	1.76	Blood	Chloroform	Indian			Male	109.94					
27	C001818_A	C001818	Homo sapien	M	105	50	1.86	Blood	Chloroform	Indian		C002042_/_	Male	101.99					
28	C001896_A	C001896	Homo sapien	M	105	50	1.88	DNA	Chloroform	Indian		C002087_/_	Male	55.45					
29	C001882_A	C001882	Homo sapien	M	105	50	1.77	Blood	Chloroform	Indian		C002063_/_	Male	90.17					

The three events are:

Study Subject ID	Illumina Manifest Upload	Illumina QC	Sequencing files check
C001807			
C001808			
C001809			
C001810			
C001811			

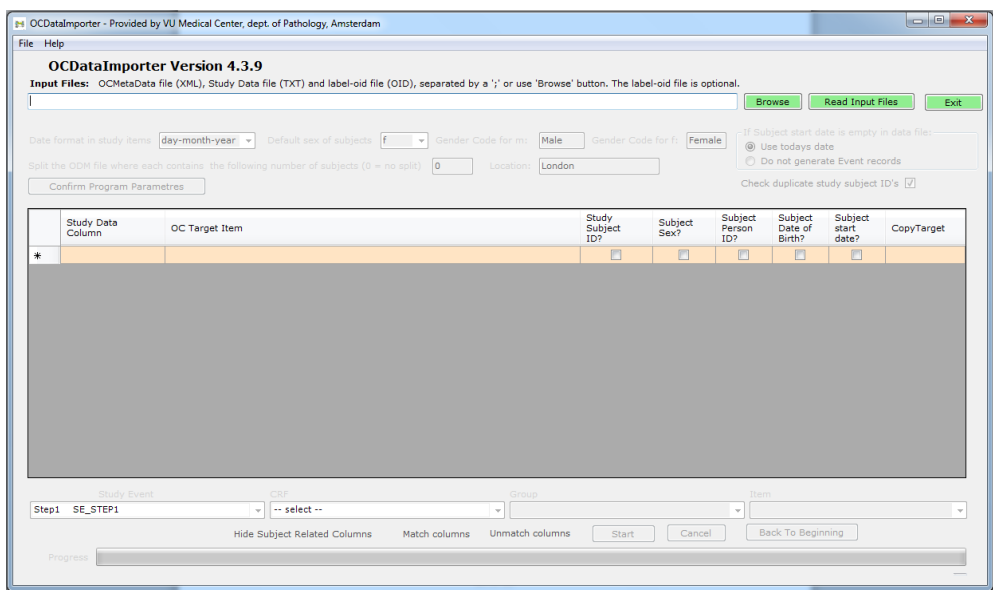
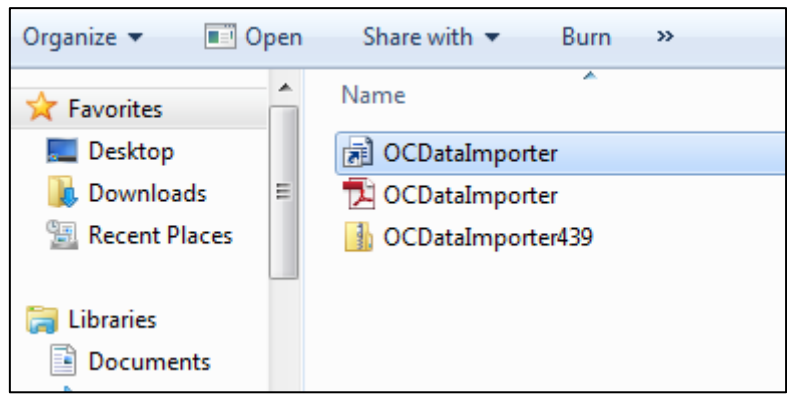
Order	Name	OID	Repeating	Type
1	Illumina Manifest Upload	SE_ILLUMINAMANIFESTUPLOAD_8272	Yes	Common
2	Illumina QC	SE_ILLUMINAQC_8933	Yes	Common
3	Sequencing files check	SE_SEQUENCINGFILESCHECK_2213	Yes	Common

And the three associated CRFs are:

CRF Name	Date Updated	Last Updated by	CRF_OID	Version(s)	Version_OID	Date Created
A-WGS1	01-Sep-2014	ctitterton	F_AWGS1	(original)		01-Sep-2014
				1.0	F_AWGS1_10	01-Sep-2014
A-WGS2	01-Sep-2014	ctitterton	F_AWGS2	(original)		01-Sep-2014
				1.0	F_AWGS2_10	01-Sep-2014
A-WGS3	01-Sep-2014	ctitterton	F_AWGS3	(original)		01-Sep-2014
				1.0	F_AWGS3_10	01-Sep-2014

4.1 OCdataimporter – manual mapping

Open OCDataImporter from the shortcut created when you downloaded OCDataImporter (See above Section 2 for instructions to download)



4.2 Download your study metadata file

- 1.) In OpenClinica in go to Tasks > View Study and click the link “Download the study metadata [here](#)”. This will open the study metadata XML file.

WGS1.0

Download the study metadata here. Click to open in your browser, or right click (option click for Mac users) and save to your computer. (Please note, you will still need to get the Study Subject OIDs from the Subject Matrix by selection the 'Show' table.)

Overview

Name: WGS1.0
 Unique Protocol ID: WGS1.0
 OID: openclinica-testing.medschl.cam.ac.uk/OpenClinica/DownloadStudyMetadata?studyId=46 - ...
 Principal Investigator: openclinica-testing.medschl.cam.ac.uk/OpenClinica/DownloadStudyMe...
 Brief Summary:
 Owner:
 Date Created: This XML file does not appear to have any style information associated with it. The below.

View Study

- View Study
- View Study
- View Study
- View Study
- View Study
- Sites: (10 Sites)
- Event Definition

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<ODM xmlns="http://www.odisc.org/ns/odm/v1.3"
xmlns:OpenClinica="http://www.openclinica.org/ns/odm_ext_v130/"
xmlns:OpenClinicaRules="http://www.openclinica.org/ns/rules/v3/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" FileOID=
MetaD20140902094348+0100" Description="Study Metadata" Creatio
02T09:43:48+01:00" FileType="Snapshot" ODMVersion="1.3"
xsi:schemaLocation="http://www.odisc.org/ns/odm/v1.3 OpenClini
0.xsd">
  <Study OID="S_WGS10">
```

- 2.) Right mouse click the XML file and do a ‘Save As’ to the folder .../Desktop/OC Import. You may want to rename the file to reflect your study (here it is called StudyMetadataWGS1.0)
- 3.) Save your data file as a tab delimited txt file to the same location.

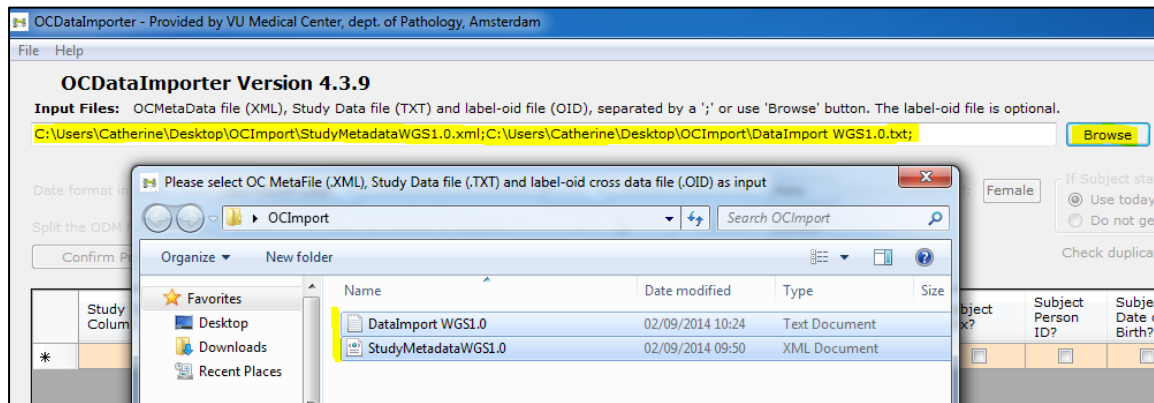
Sample ID	Study ID	Subject ID	Species	Gender (M/F/U)	Volume (ul)	Concentration (ng/ul)	OD 260/280	Tissue Source	Extraction Method	Ethnicity	Parent 1 ID	Parent 2 ID	Replicate (e) ID	ILMN GT	ILMN quant	Insert size	le250	Comment	Date sent from ILMN at
1	C001852_A	C001852	Homo sapien	M	105	50	1.82	Blood	Chloroform	White British				Male	104.56	1.967364337		7.8.14	
2	C001899_A	C001899	Homo sapien	M	105	50	1.9	DNA	Chloroform	White British				Male	85.1	2.269740478		7.8.14	
3	C001857_A	C001857	Homo sapien	M	105	50	1.85	Blood	Chloroform	White British				Male	81.24	1.767426184		7.8.14	
4	C001867_A	C001867	Homo sapien	F	105	50	1.85	Blood	Chloroform	White British				Female	142.93	1.403163689		7.8.14	
5	C001828_A	C001828	Homo sapien	M	105	50	1.88	Blood	Chloroform	White British	C002031_J	C002026_A		Male	110.38	5.146685758		7.8.14	
6	C001819_A	C001819	Homo sapien	M	105	50	1.93	DNA	Chloroform	White British	C002051_J	C002052_A		Male	85.39	3.563283199		7.8.14	
7	C001842_A	C001842	Homo sapien	M	105	50	1.87	Blood	Chloroform	White British	C002044_J	C002043_A		Male	104.42	5.155884573		7.8.14	
8	C001827_A	C001827	Homo sapien	F	105	50	1.87	Blood	Chloroform	White British	C002029_J	C002030_A		Female	151.92	3.554620989		7.8.14	
9	C001820_A	C001820	Homo sapien	F	105	50	1.87	DNA	Chloroform	White British	C002017_J	C002016_A		Female	63.89	1.592303816		7.8.14	
10	C001816_A	C001816	Homo sapien	M	105	50	1.88	Blood	Chloroform	White British	C002012_J	C002013_A		Male	78.18	1.083684248		7.8.14	
11	C001815_A	C001815	Homo sapien	M	105	50	1.85	Blood	Chloroform	White British	C002010_J	C002011_A		Male	77.8	6.558420784		7.8.14	
12	C001811_A	C001811	Homo sapien	M	105	50	1.88	Blood	Chloroform	White British	C002005_J	C002004_A		Male	85.35	5.073945236		7.8.14	
13	C001878_A	C001878	Homo sapien	M	105	50	1.79	Blood	Chloroform	Black British	C002065_J	C002061_A		Male	126.96	3.53819762		7.8.14	
14	C001856_A	C001856	Homo sapien	M	105	50	1.84	Blood	Chloroform	Black British	C002056_J	C002057_A		Male	77.74	5.809789736		7.8.14	
15	C001855_A	C001855	Homo sapien	F	105	50	1.86	Blood	Chloroform	Black British	C002055_J	C002054_A		Female	98.02	1.540665929		7.8.14	
16	C001880_A	C001880	Homo sapien	F	105	50	1.78	Blood	Chloroform	Black British				Female	89.6	2.461096416		7.8.14	
17	C001864_A	C001864	Homo sapien	F	105	50	1.77	Blood	Chloroform	Black British	C002058_J			Female	108.99				
18	C001866_A	C001866	Homo sapien	M	105	50	1.84	Blood	Chloroform	Black British	C002059_J			Male	80.65				
19	C001894_A	C001894	Homo sapien	M	105	50	1.88	DNA	Chloroform	Black British	C002082_J			Male	63.54				
20	C001879_A	C001879	Homo sapien	M	105	50	1.79	Blood	Chloroform	Indian	C002087_J			Male	100.01				
21	C001823_A	C001823	Homo sapien	F	105	50	1.85	Blood	Chloroform	Indian	C002022_J			Female	144.29				
22	C001822_A	C001822	Homo sapien	M	105	50	1.87	Blood	Chloroform	Indian	C002023_J			Male	106.66				
23	C001869_A	C001869	Homo sapien	M	105	50	1.85	Blood	Chloroform	Indian				Male	92.3				
24	C001874_A	C001874	Homo sapien	F	105	50	1.78	Blood	Chloroform	Indian				Female	87.05				
25	C001875_A	C001875	Homo sapien	M	105	50	1.76	Blood	Chloroform	Indian				Male	109.94				
26	C001818_A	C001818	Homo sapien	M	105	50	1.86	Blood	Chloroform	Indian	C002042_J			Male	101.99				
27	C001896_A	C001896	Homo sapien	M	105	50	1.88	DNA	Chloroform	Indian	C002087_J			Male	93.45				
28	C001882_A	C001882	Homo sapien	M	105	50	1.77	Blood	Chloroform	Indian	C002063_J			Male	90.17				

You should then have two files – your study metadata XML file and your data import txt file

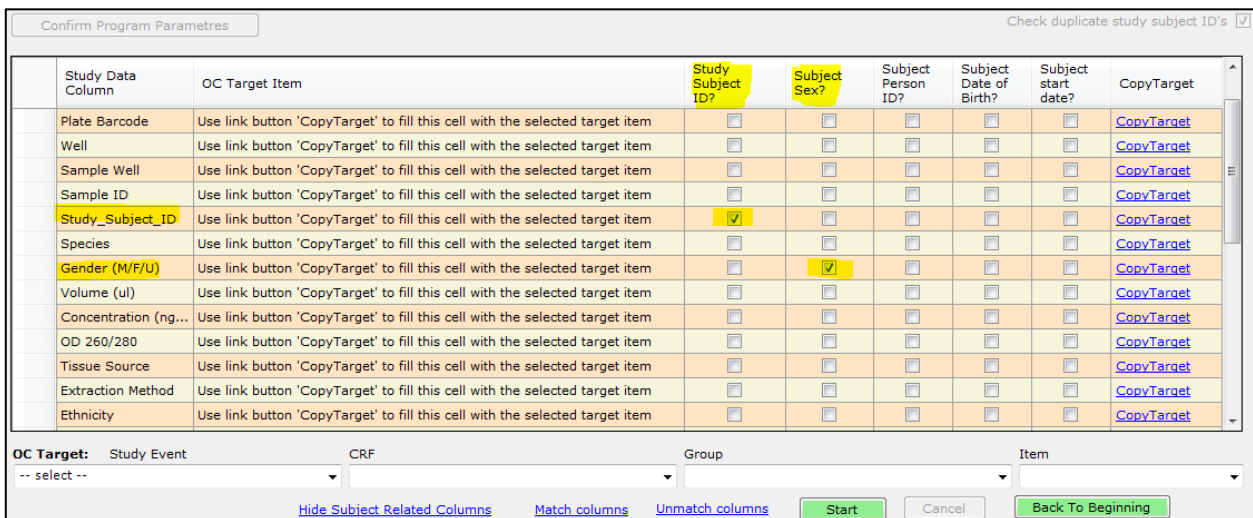
Name	Date modified	Type
OCDataImporter439	09/07/2014 16:49	Compressed (zip)
OCDataImporter	09/07/2014 16:51	Shortcut
OCDataImporter	03/08/2014 16:37	Adobe Acrobat D.
StudyMetadataWGS1.0	02/09/2014 09:50	XML Document
DataImport WGS1.0	02/09/2014 10:24	Text Document

4.3 Match you study data and study metadata

- 1.) Click Browse and select both files



- 2.) Click 'Read Input files'
- 3.) Set the program parameters for date format, gender (OpenClinica expects lowercase **m** and **f**), subject start date, duplicate subject IDs, location.
- 4.) Leave ODM file split as '0'
- 5.) Click 'Confirm Program Parameters'



- 6.) Your excel data columns will be loaded into the first column 'StudyData Column'. OCdataimporter will attempt to match again the core subject fields as show above.
- 7.) Select the **OC target: Study Event** (in the example below there are three study events)

OD 260/280	Use link button 'CopyTarget' to fill this cell with the selected target item	<input type="checkbox"/>	<input type="checkbox"/>
Tissue Source	Use link button 'CopyTarget' to fill this cell with the selected target item	<input type="checkbox"/>	<input type="checkbox"/>
Extraction Method	Use link button 'CopyTarget' to fill this cell with the selected target item	<input type="checkbox"/>	<input type="checkbox"/>
Ethnicity	Use link button 'CopyTarget' to fill this cell with the selected target item	<input type="checkbox"/>	<input type="checkbox"/>

OC Target: Study Event
 Illumina Manifest Upload SE_ILLUMINAMANI

CRF
 A-WGS1 - 1.0 F_AWGS1_10

Group
 IG_AWGS1_UNGROUPED

-- select --
 Illumina Manifest Upload SE_ILLUMINAMANIFESTUPLOA 8272 mns **Start**
 Illumina QC SE_ILLUMINAQC_8933
 Sequencing files check SE_SEQUENCINGFILES

- 8.) Select the correct CRF and Group (if groups exist in your CRF)
- 9.) The last drop down will then present you with all the fields present in the CRF

et' to fill this cell with the selected target item	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	CopyTarget
et' to fill this cell with the selected target item	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	CopyTarget

Group
 S1 - 1.0 F_AWGS1_10 IG_AWGS1_UNGROUPED

Item
 -- select --
 -- select --
 CANCER I_AWGS1_CANCER
 CONCENTRATION I_AWGS1_CONCENTRATION
 DATE I_AWGS1_DATE
 ETHNICITY I_AWGS1_ETHNICITY
 EXTRACTION_METHOD I_AWGS1_EXTRACTION_METH
 MATCHED_SAMPLE_ID I_AWGS1_MATCHED_SAMPLE_
 MATCHED_SAMPLE_TYPE I_AWGS1_MATCHED_SAMPL
 NOTES I_AWGS1_NOTES
 OD I_AWGS1_OD
 PARENT_1_ID I_AWGS1_PARENT_1_ID
 PARENT_2_ID I_AWGS1_PARENT_2_ID
 PLATE_BARCODE I_AWGS1_PLATE_BARCODE
 REPLICATE_ID I_AWGS1_REPLICATE_ID
 SAMPLE_ID I_AWGS1_SAMPLE_ID
 SAMPLE_WELL I_AWGS1_SAMPLE_WELL
 SPECIES I_AWGS1_SPECIES

- 10.) Select the appropriate CRF field and then click the 'CopyTarget' link against the related row. (Note: to clear the screen you can 'show/hide the subject related fields' by clicking the link at the bottom of the screen)

Confirm Program Parameters Check duplicate study

Study Data Column	OC Target Item	CopyTarget
▶ Plate Barcode	SE_ILLUMINAMANIFESTUPLOA 8272.F_AWGS1_10.IG_AWGS1_UNGROUPED.I_AWGS1_PLATE_BARCODE	CopyTarget
Well	Use link button 'CopyTarget' to fill this cell with the selected target item	CopyTarget
Sample Well	Use link button 'CopyTarget' to fill this cell with the selected target item	CopyTarget
Sample ID	Use link button 'CopyTarget' to fill this cell with the selected target item	CopyTarget
Study_Subject_ID	Use link button 'CopyTarget' to fill this cell with the selected target item	CopyTarget
Species	Use link button 'CopyTarget' to fill this cell with the selected target item	CopyTarget
Gender (M/F/U)	Use link button 'CopyTarget' to fill this cell with the selected target item	CopyTarget
Volume (ul)	Use link button 'CopyTarget' to fill this cell with the selected target item	CopyTarget
Concentration (ng/ul)	Use link button 'CopyTarget' to fill this cell with the selected target item	CopyTarget
OD 260/280	Use link button 'CopyTarget' to fill this cell with the selected target item	CopyTarget
Tissue Source	Use link button 'CopyTarget' to fill this cell with the selected target item	CopyTarget
Extraction Method	Use link button 'CopyTarget' to fill this cell with the selected target item	CopyTarget
Ethnicity	Use link button 'CopyTarget' to fill this cell with the selected target item	CopyTarget
Parent 1 ID	Use link button 'CopyTarget' to fill this cell with the selected target item	CopyTarget

OC Target: Study Event
 Illumina Manifest Upload SE_ILLUMINAMANI

CRF
 A-WGS1 - 1.0 F_AWGS1_10

Group
 IG_AWGS1_UNGROUPED

Item
 PLATE_BARC

[Show All Columns](#) [Match columns](#) [Unmatch columns](#) **Start** **Cancel** **Back To Beginning**

- 11.) Continue matching fields until all dropdown CRF items have been associated with a row in the main part of the screen.

4.4 Generate the XML CRF data import file

- 1.) Click 'Start'

Confirm Program Parameters Check duplicate study subject ID's

Study Data Column	OC Target Item
Study_Subject_ID	Use link button 'CopyTarget' to fill this cell with the selected target item
Species	SE_ILLUMINAMANIFESTUPLOAD_8272.F_AWGS1_10.IG_AWGS1_UNGROUPED.I_AWGS1_SPECIES
Gender (M/F/U)	Use link button 'CopyTarget' to fill this cell with the selected target item
Volume (ul)	SE_ILLUMINAMANIFESTUPLOAD_8272.F_AWGS1_10.IG_AWGS1_UNGROUPED.I_AWGS1_VOLUME
Concentration (ng/ul)	SE_ILLUMINAMANIFESTUPLOAD_8272.F_AWGS1_10.IG_AWGS1_UNGROUPED.I_AWGS1_CONCENTRATION
OD 260/280	SE_ILLUMINAMANIFESTUPLOAD_8272.F_AWGS1_10.IG_AWGS1_UNGROUPED.I_AWGS1_OD
Tissue Source	SE_ILLUMINAMANIFESTUPLOAD_8272.F_AWGS1_10.IG_AWGS1_UNGROUPED.I_AWGS1_TISSUE_SOURCE
Extraction Method	SE_ILLUMINAMANIFESTUPLOAD_8272.F_AWGS1_10.IG_AWGS1_UNGROUPED.I_AWGS1_EXTRACTION_METHOD
Ethnicity	SE_ILLUMINAMANIFESTUPLOAD_8272.F_AWGS1_10.IG_AWGS1_UNGROUPED.I_AWGS1_ETHNICITY
Parent 1 ID	SE_ILLUMINAMANIFESTUPLOAD_8272.F_AWGS1_10.IG_AWGS1_UNGROUPED.I_AWGS1_PARENT_1_ID
Parent 2 ID	SE_ILLUMINAMANIFESTUPLOAD_8272.F_AWGS1_10.IG_AWGS1_UNGROUPED.I_AWGS1_PARENT_2_ID
Replicate(s) ID	SE_ILLUMINAMANIFESTUPLOAD_8272.F_AWGS1_10.IG_AWGS1_UNGROUPED.I_AWGS1_REPLICATE_ID
Cancer sample (Y/N)	SE_ILLUMINAMANIFESTUPLOAD_8272.F_AWGS1_10.IG_AWGS1_UNGROUPED.I_AWGS1_CANCER
Matched Sample ID(s)	SE_ILLUMINAMANIFESTUPLOAD_8272.F_AWGS1_10.IG_AWGS1_UNGROUPED.I_AWGS1_MATCHED_SAMPLE_ID

OC Target: Study Event CRF Group Item

Illumina Manifest Upload SE_ILLUMINAMANI A-WGS1 - 1.0 F_AWGS1_10 IG_AWGS1_UNGROUPED NOTES I_AWGS1_NOTES

[Show All Columns](#) [Match columns](#) [Unmatch columns](#) **Start** Cancel [Back To Beginning](#)

Progress

- 2.) If all records were successfully processed the below screen will show. If not review the errors which will be displayed in the lower pane as well as in the file OCdataImporter_warnings in .../Desktop/OC Import.

Comment	SE_ILLUMINAMANIFESTUPLOAD_8272*1.F_AWGS1_10.ATG_AWGS1_UNGROUPED.I_AWGS1_NOTES
OD 260/280	SE_ILLUMINAMANIFESTUPLOAD_8272*1.F_AWGS1_10.ATG_AWGS1_UNGROUPED.I_AWGS1_OD
Parent 1 ID	SE_ILLUMINAMANIFESTUPLOAD_8272*1.F_AWGS1_10.ATG_AWGS1_UNGROUPED.I_AWGS1_PARENT_1_ID
Parent 2 ID	SE_ILLUMINAMANIFESTUPLOAD_8272*1.F_AWGS1_10.ATG_AWGS1_UNGROUPED.I_AWGS1_PARENT_2_ID
Plate Barcode	SE_ILLUMINAMANIFESTUPLOAD_8272*1.F_AWGS1_10.ATG_AWGS1_UNGROUPED.I_AWGS1_PLATE_BARCODE
Replicate(s) ID	SE_ILLUMINAMANIFESTUPLOAD_8272*1.F_AWGS1_10.ATG_AWGS1_UNGROUPED.I_AWGS1_REPLICATE_ID
Sample ID	SE_ILLUMINAMANIFESTUPLOAD_8272*1.F_AWGS1_10.ATG_AWGS1_UNGROUPED.I_AWGS1_SAMPLE_ID

OCDataImporter

Process finished successfully

OC Target: Study Event CRF Group Item

Illumina Manifest Upload SE_ILLUMINAMANI A-WGS1 - 1.0 F_AWGS1_10 IG_AWGS1_UNGROUPED NOTES

WARNINGS: 0 [Show All Columns](#) [Match columns](#) [Unmatch columns](#) Start Cancel [Back To Beginning](#)

Progress

Data file is: C:\Users\Catherine\Desktop\OCImport\DataImport WGS1.0.txt, delimited by: tab, Number of items per line: 29
 Started in directory C:\Users\Catherine\Desktop\OCImport. This may take several minutes...
 02/09/2014 11:04:36 Finished successfully.

- 3.) OCdataimporter generates a set of files and places them in the .../Desktop/OC Import folder. The SQL files can be ignored (this is an alternate method for importing data for which there are detailed notes in the original OCDataImporter.pdf help file found in the folder below).

The file that is important in this process is the **DataImport_1.xml** file, which is the result of the matched study meta data and your excel data file.

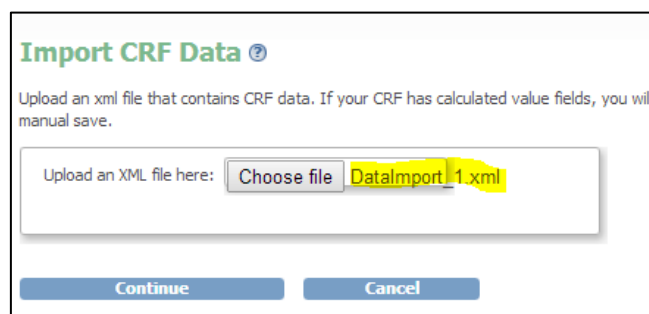
Name	Date modified	Type
OCDataImporter439	09/07/2014 16:49	Compressed (zipp...
OCDataImporter	09/07/2014 16:51	Shortcut
OCDataImporter	03/08/2014 16:37	Adobe Acrobat D...
StudyMetadataWGS1.0	02/09/2014 09:50	XML Document
DataImport WGS1.0	02/09/2014 10:24	Text Document
DataImport WGS1.0_grid.dmp	02/09/2014 11:04	DMP File
DataImport WGS1.0_parameters.dmp	02/09/2014 11:04	DMP File
DataImport_1	02/09/2014 11:04	XML Document
Deletes.sql	02/09/2014 11:04	SQL File
Deletes_ONLY_STUDY_EVENTS.sql	02/09/2014 11:04	SQL File
Inserts.sql	02/09/2014 11:04	SQL File
Inserts_ONLY_STUDY_EVENTS.sql	02/09/2014 11:04	SQL File
OCDataImporter_warnings	02/09/2014 11:04	Text Document

4.5 Import your CRF data XML file into OpenClinica

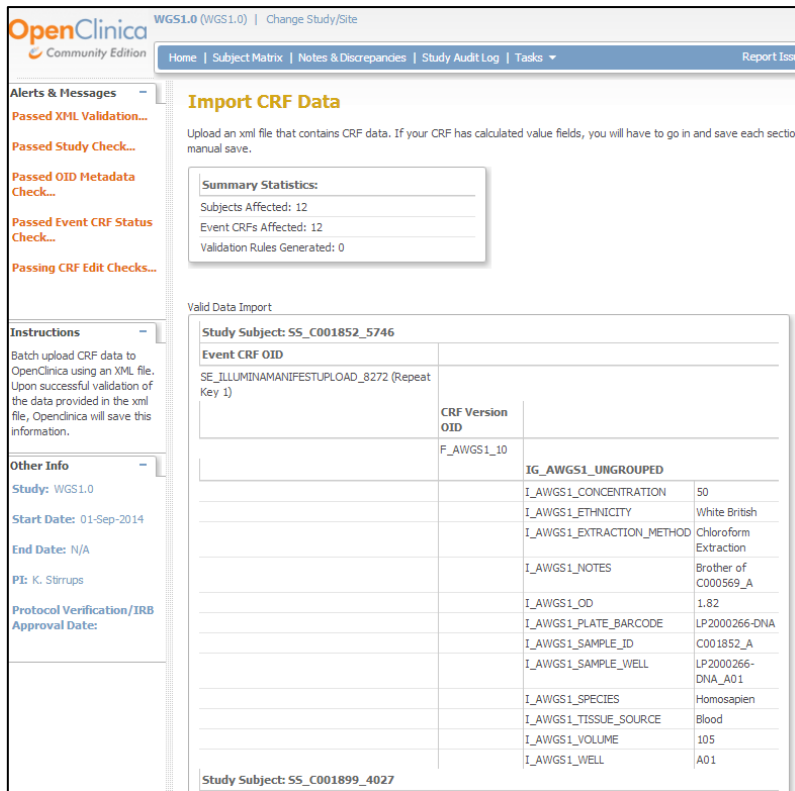
1.) In OpenClinica click Tasks > Import Data



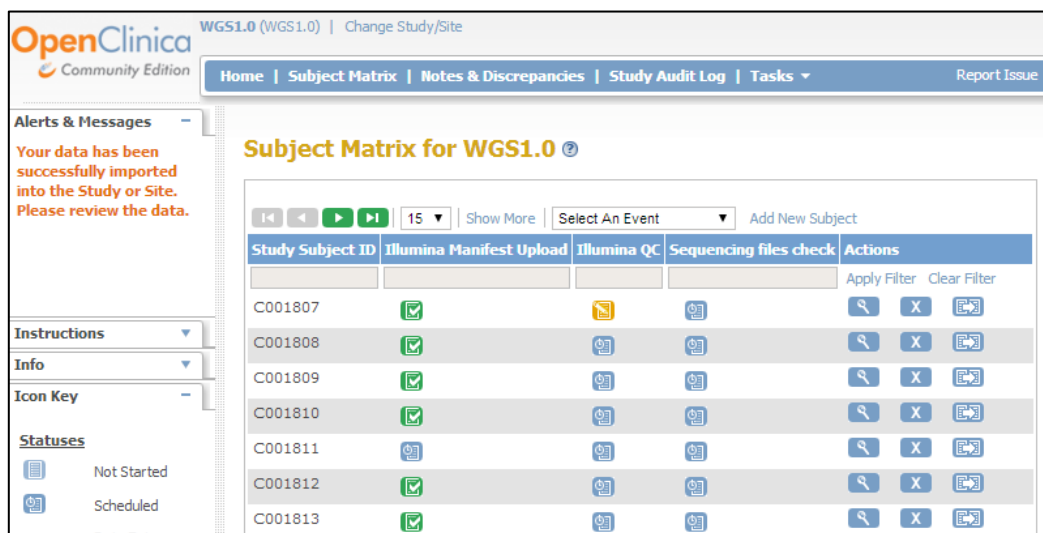
2.) Browse to the **DataImport_1.xml** file and Click continue.



3.) OpenClinica checks XML validation, study checks,OID metada, event CRF status and CRF edit checks and present the following screen (or a screen showing errors and which points it failed on)



- 4.) Scroll to the bottom of the screen and click continue to apply the records
- 5.) Data successfully imported!



The event records are then flagged as complete on data import and the icon changes to green in the subject matrix. You can see on subject C001811 on the first event that the icon remains as scheduled. This is because no data was imported for C001811.

5 CIT data importer

The CIT data importer can be used independently for the end to end import process including:

- Subject creation
- Event scheduling
- Study data import

Assuming you have already gone through the process of cleaning you data, the next steps are to:

1. Save your data file as [studyname].csv
2. Columns A to F must contain the column headings *SubjectID*, *SecondaryID*, *EnrollmentDate*, *PersonID*, *Gender* and *DOB* as shown below. The import is case sensitive so ensure column headings are exactly as show.
3. If you study is not collecting date of birth or is not using PersonID, SecondaryID or Gender the import file still expects to see the column headings.

	A	B	C	D	E	F
1	SubjectId	SecondaryID	EnrollmentDate	PersonId	Gender	DOB
2	C029AQ	ADD000120B	2012-06-29		F	
3	C02ADG	BRI000039Y	2012-07-05		F	
4	C02AFC	BRI0000363	2012-07-05		M	
5	C02AGA	BRI000014F	2012-07-05		F	
6	C02AH8	BRI000032B	2012-07-05		M	
7	C02AJ4	BRI0000275	2012-07-05		F	
8	C02AK2	BRI0000283	2012-07-05		M	
9	C02AMZ	BRI0000347	2012-07-05		M	
10	C02ANX	BRI0000355	2012-07-05		F	
11	C02APT	BRI000038*	2012-07-05		F	
12	C02AQR	BRI0000371	2012-07-05		M	
13	C02ARP	BRI0000291	2012-07-05		F	

Open the CIT data importer at the url provided to you by the CIT team e.g. <http://openclinica-testing.medschl.cam.ac.uk/superimporter/index.php> . You will be presented with the screen shown below

← → ↻ openclinica-testing.medschl.cam.ac.uk/superimporter/index.php

Apps

CIT DATAIMPORTER

Logged in as: [ctitterton](#) ([Logout](#))

Select a study:

Please choose a file: No file chosen

4. Select the study you are importing to from the dropdown list.
5. Click "Choose file" and navigate to the file [studyname].csv
6. Click "Upload file"

CIT DATAIMPORTER

The file Sample_import_test.csv has been uploaded.
[Go back](#) or [Continue to validation](#)

7. Click "Continue to validation" of the uploaded file

CIT DATAIMPORTER

Datafile structure must be the following:
SubjectId SecondaryId EnrollmentDate PersonId Gender DOB + followed by the data fields.

There were no errors in the data file.
[Continue to import subjects](#)

8. If there are errors, review the data in the first 6 columns. Check if the expected data (DOB, Gender etc) match to the study parameters. For gender, M and F are expected, not Male and Female.
9. If there are no errors, you can click "Continue to import subjects".
10. If the subjects have already been loaded to the study, new subject records will not be created, instead the below message will show "subject X exists for study Y". New records will be show "successful" with a summary of new records show at the bottom of the screen. Continue

CIT DATAIMPORTER

Connection successful

createSubject:Subject with label "C029AQ" exists for "WGS8" study.
Subject name in xlsx: C029AQ
SOID = SS_C029AQ label = C029AQ

createSubject:Subject with label "C02ADG" exists for "WGS8" study.
Subject name in xlsx: C02ADG
SOID = SS_C02ADG label = C02ADG

```

createSubject:Success
Subject name in xlsx: C02ASN
SOID = SS_C02ASN_9730 label = C02ASN

createSubject:Success
Subject name in xlsx: C02ATL
SOID = SS_C02ATL_6432 label = C02ATL

Subjects import finished.

Successful imports: 14
New subjects: 14
Errors: 0

Continue to scheduling

```

11. Click “Continue to scheduling”. The next screen shows all the events in your study, select the ones you want to schedule, accept the default date or add a specific date. (Note you cannot schedule individual dates for each subject – if you need to schedule specific dates, your data import file will need to be split according to event date)

Event name	Schedule? <input checked="" type="checkbox"/>	Date	Time	Forms
Receipt	<input checked="" type="checkbox"/>	2015-01-26	14:11	<ul style="list-style-type: none"> A_WGS1 Sample Receipt - 1.2 A_WGS1 Sample Receipt - 1.1 A_WGS1 Sample Receipt - 1.0
CATGO QC	<input checked="" type="checkbox"/>	2015-01-26	14:11	<ul style="list-style-type: none"> A_WGS2 CATGO QC - 1.2 A_WGS2 CATGO QC - 1.1 A_WGS2 CATGO QC - 1.0
GEL OK	<input checked="" type="checkbox"/>	2015-01-26	14:11	<ul style="list-style-type: none"> A_WGS2 GEL PASS - 1.0
Plating	<input checked="" type="checkbox"/>	2015-01-26	14:11	<ul style="list-style-type: none"> A_WGS3 Plating - 1.4 A_WGS3 Plating - 1.3 A_WGS3 Plating - 1.2
Illumina QC	<input checked="" type="checkbox"/>	2015-01-26	14:11	<ul style="list-style-type: none"> A_WGS4 Illumina QC - 1.2 A_WGS4 Illumina QC - 1.1
Receive Files	<input checked="" type="checkbox"/>	2015-01-26	14:11	<ul style="list-style-type: none"> A_WGS5 Files Receipt - 1.2 A_WGS5 Files Receipt - 1.1
Delete Files	<input checked="" type="checkbox"/>	2015-01-26	14:11	<ul style="list-style-type: none"> A_WGS6 Files Destroy - 1.0
Submit EGA	<input checked="" type="checkbox"/>	2015-01-26	14:11	<ul style="list-style-type: none"> A_WGS8 EGA - 1.0
<input type="button" value="Schedule selected events!"/>				

12. Click “Schedule selected events”. The next screen will show a line for every event per subject as scheduled with “success”. If the event has already been scheduled for a particular subject, a warning will show. (You cannot reschedule dates within CIT data importer, if they have already been scheduled. You can manually reschedule events through the interface, but this needs to be done on an event / by subject record, prior to data import)

Scheduling event (SE_PLATING_7476) for subject (C02ATL): **Success**
Scheduling event (SE_ILLUMINAQC_5372) for subject (C02ATL): **Success**
Scheduling event (SE_RECEIVEFILES) for subject (C02ATL): **Success**
Scheduling event (SE_DELETEFILES) for subject (C02ATL): **Success**
Scheduling event (SE_SUBMITEGA) for subject (C02ATL): **Success**

[Upload a mapping file](#) or [Continue to CRF definitions](#)

If the event has already been scheduled for a particular subject, a warning will show. (You cannot reschedule dates within CIT data importer, if they have already been scheduled. You can manually reschedule events through the interface, but this needs to be done on an event / by subject record, prior to data import)

Scheduling event (SE_DELETEFILES) for subject (C02ATL): **Fail**
Cannot schedule an event for this Subject.

Scheduling event (SE_SUBMITEGA) for subject (C02ATL): **Fail**
Cannot schedule an event for this Subject.

[Upload a mapping file](#) or [Continue to CRF definitions](#)

13. Upload an existing mapping file (see Section 6) or “Continue to CRF definitions”

Events	Default CRF for the Event	Check Version of CRF to use for import	Items to import
Receipt	A_WGS1 Sample Receipt - 1.2 <input checked="" type="checkbox"/>	A_WGS1 Sample Receipt - 1.2 ▼	<ul style="list-style-type: none"> <input checked="" type="checkbox"/> STUDY <input checked="" type="checkbox"/> BRIDGE_ID <input checked="" type="checkbox"/> SAMPLE_ID <input checked="" type="checkbox"/> SAMPLE_DATE_REC <input checked="" type="checkbox"/> SAMP_REC_COMMENT <input checked="" type="checkbox"/> GENDER
CATGO QC	A_WGS2 CATGO QC - 1.2 <input checked="" type="checkbox"/>	A_WGS2 CATGO QC - 1.2 ▼	<ul style="list-style-type: none"> <input checked="" type="checkbox"/> OD_QC <input checked="" type="checkbox"/> CONC_QC <input checked="" type="checkbox"/> QUANTITY_QC <input checked="" type="checkbox"/> CATGO_COMMENT <input checked="" type="checkbox"/> AGEL_QC
GEL OK	A_WGS2 GEL PASS - 1.0 <input checked="" type="checkbox"/>	A_WGS2 GEL PASS - 1.0 ▼	<ul style="list-style-type: none"> <input checked="" type="checkbox"/> GEL_OK <input checked="" type="checkbox"/> DATE_GEL_OK <input checked="" type="checkbox"/> GEL_COMMENT

- Select the events you wish to schedule and the related CRF version. By default the latest CRF version is selected.
- Click "Continue to mapping". Only the fields from the events ticked on the previous screen show on the left under CRF items.

CRF items	Associated item	
Receipt A_WGS1 Sample Receipt - 1.2 STUDY		OD 260/280
Receipt A_WGS1 Sample Receipt - 1.2 BRIDGE_ID		Tissue Source
Receipt A_WGS1 Sample Receipt - 1.2 SAMPLE_ID		Extraction Method
Receipt A_WGS1 Sample Receipt - 1.2 SAMPLE_DATE_REC		Ethnicity
Receipt A_WGS1 Sample Receipt - 1.2 SAMP_REC_COMMENT		Parent 1 ID
Receipt A_WGS1 Sample Receipt - 1.2 GENDER		Parent 2 ID
CATGO QC A_WGS2 CATGO QC - 1.2 OD_QC		Replicate(s) ID
CATGO QC A_WGS2 CATGO QC - 1.2 CONC_QC		Comment: plating
CATGO QC A_WGS2 CATGO QC - 1.2 QUANTITY_QC		GEL OK
CATGO QC A_WGS2 CATGO QC - 1.2 CATGO_COMMENT		GEL Ok Date
CATGO QC A_WGS2 CATGO QC - 1.2 AGEL_QC		GEL OK Comments
		ILMN GT
		ILMN quant
		ILMN QC comments
		Date files sent (ILMN)

Skip empty cells in datafile

- The field list on the left in blue displays the fields defined in your CRFs. The field list to the right in red are the column headings from your csv import file. Drag and drop the appropriate fields from the csv import file against the CRF fields.

CRF items	Associated item	
Receipt A_WGS1 Sample Receipt - 1.2 STUDY	Study	OD 260/280
Receipt A_WGS1 Sample Receipt - 1.2 BRIDGE_ID		Tissue Source
Receipt A_WGS1 Sample Receipt - 1.2 SAMPLE_ID	Sample_ID	Extraction Method
Receipt A_WGS1 Sample Receipt - 1.2 SAMPLE_DATE_REC		Ethnicity
Receipt A_WGS1 Sample Receipt - 1.2 SAMP_REC_COMMENT		Parent 1 ID
Receipt A_WGS1 Sample Receipt - 1.2 GENDER	Gender_MFU	Parent 2 ID
CATGO QC A_WGS2 CATGO QC - 1.2 OD_QC	OD_QC	Replicate(s) ID
CATGO QC A_WGS2 CATGO QC - 1.2 CONC_QC		Comment: plating
CATGO QC A_WGS2 CATGO QC - 1.2 QUANTITY_QC		GEL OK
CATGO QC A_WGS2 CATGO QC - 1.2 CATGO_COMMENT		GEL Ok Date
CATGO QC A_WGS2 CATGO QC - 1.2 AGEL_QC		GEL OK Comments

- Any fields not mapped, will not have data imported against them. If you have a mix of subjects who do and don't have data for a particular event, tick the bottom box "Skip empty cells in data file"
- Once you have mapped all fields, click "Start Import!"

C02ATL	SE_CATGOQC_6200	F_A_WGS2CATGOQ_12	I_A_WGS_OD_QC_7365	No	OK!
--------	-----------------	-------------------	--------------------	----	-----

IMPORT FINISHED, SESSION CLOSED.
[Start a new import](#)

- For every subject, against each event and each field, you will get a result of "OK" if the data has imported successfully. Any items that have not been imported will give an error message.
- Check if your records have imported as expected through the subject matrix in OpenClinica

Subject Matrix for WGS7.0							
Study Subject ID	Receipt	CATGO QC	GEL OK	Plating	Illumina QC	Receive Files	Delete Files
C029APO	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C02ADGA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C02AFCA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C02AGAA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

6 Advanced Data Import Mapping

The import steps described in the above sections for TDS import database, CIT Data Importer and the OC Data Importer can all be used without reference to the section on advanced data import mapping.

The advanced data import mapping process is useful where you importing large amounts of data between OpenClinica instances. In OpenClinica, for auditing purposes there is no possibility of deleting data once it has been entered into OpenClinica, therefore it is important to ensure data being loaded into OpenClinica is 'clean' and accurate. To ensure accuracy it is envisaged there is likely, for most studies, that a staging service will do the first check on data import and once verified a second import will take place to the final live service.

The steps and files described below facilitate this process.

6.1 Files required for advanced mapping import

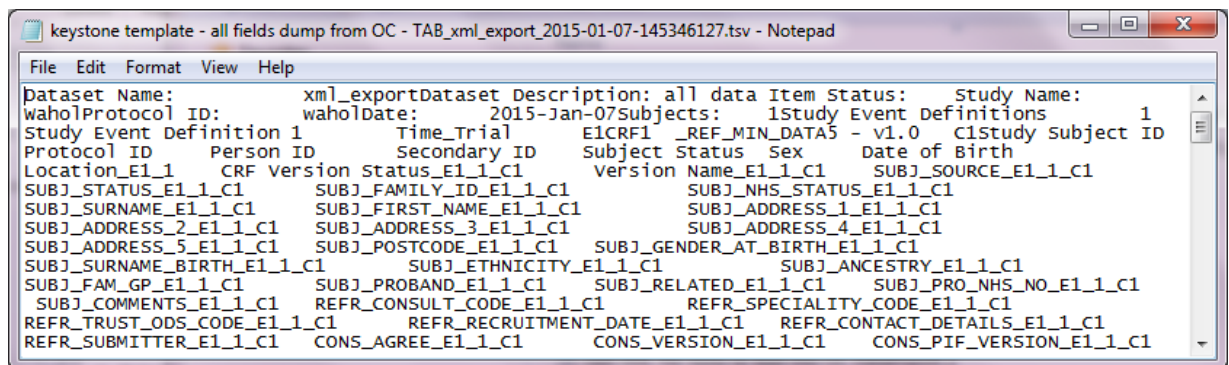
File	file type	file name	Used for
Study Metadata file	xml	DownloadStudyMetadata.xml	Provides an output of all study metadata
Keystone template	tsv	TAB_exportname_2015-01-21-114325879.tsv	Extracts
CRF	xls	as named	Needed to generate OpenClinica input screens and can be used for dummy data generation
Mapping Definitions file	xls	matching and substitutions between sample and study XML - wahl_keystone_grid_edits.xlsx	Needed to create the grid file from the keystone template and study metadata files
ODI Grid file	dmp	ODI mapping file generated and stored by ODI import1_grid.dmp	ODI mapping file generated and stored by ODI
ODI Parameters file	dmp	ODI parameters file generated by ODI import1_parameters.dmp	ODI parameters file generated by ODI.
Study Data Import file	xml	ODI XML file ready to lead into OC DataImport_1	XML import file generated by ODI of data to be imported into OpenClinica
			Subject creation
			Event Scheduling

6.2 Extract source files for creation of mapping definitions file

1. In OpenClinica create your study, setting all study parameters according to you study protocols.
2. Create CRFs and events against expected study data (and GEL min data set).

(Note: consider merging multiple CRFs into single reference CRF for study - will only work for studies collecting data outside of OpenClinica, if collecting within OpenClinica stay with study structure)

3. Create and populate a single subject record for every event and field
4. Export the single subject record with all events and fields from OpenClinica as tab delimited text file (tsv) to create the Keystone Template.



5. Download study metadata - go to Tasks > View Study > click download study metadata at top of screen (link highlighted). File will open, do a right-click and 'save as'.

Primary Immune Disorder ?

Download the study metadata [here](#). Click to open in your browser, or right click (option click for Mac users) to your computer. (Please note, you will still need to get the Study Subject OIDs from the Subject Matrix by selection the table.)

Overview

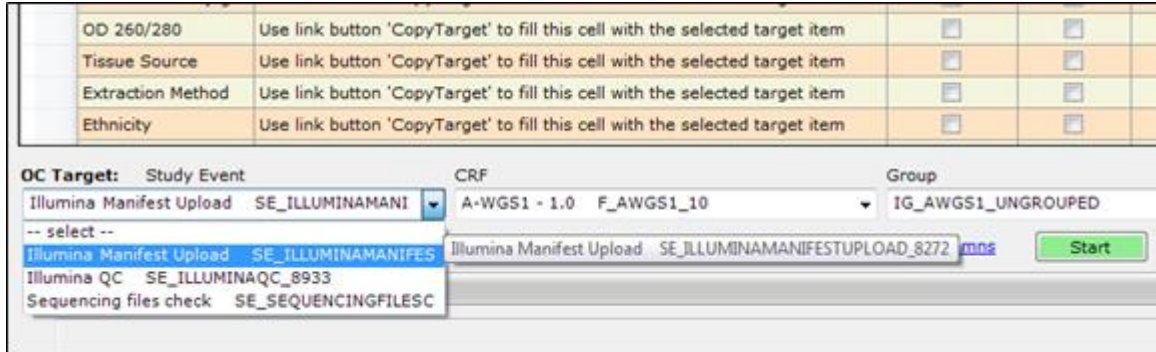
Name:	Primary Immune Disorder
Unique Protocol ID:	PID
OID:	S_PID

The end result you should have two files to be used in Excel to create the Mapping Definitions file:

- Study Metadata file: [DownloadStudyMetadata.xml](#)
- Keystone template: [TAB_\[exportname\]_2015-01-21-114325879.tsv](#)

6.3 Create Mapping Definitions file

When using OCDataImporter, the standard process is to load the Study Metadata file ([DownloadStudyMetadata.xml](#)) and the data import file and then manually map each column of data in your data import file against events and items, as shown below.



The process described below bypasses the manual mapping and generates the parameters and grid file through Excel.

6.3.1 In Excel - create the mapping definitions file

To create Mapping Definitions file from the column headings of the Keystone Template (tsv) file and the study metadata definitions - do the following:

1. Create two sheets first called *Grid* and second called *field conversion*
2. Import study metadata into *field conversion* sheet as using the excel import wizard, as a txt file, fixed width which will bring in every line of xml as a single excel row/cell.

Data on *Field conversion sheet* will look like below:

	A	B	C
1	<?xml version="1.0" encoding="UTF-8"?>		
2	<ODM FileOID="Study-MetaD20150107135726+0000" Description="Study Metadata" CreationDateTime="2015-01-07 13:57:26">		
3	<Study OID="S_DEFAULTS1">		
4	<GlobalVariables>		
5	<StudyName>Wahol</StudyName>		
6	<StudyDescription>		
7	test of input		
8	</StudyDescription>		
9	<ProtocolName>wahol</ProtocolName>		
10	</GlobalVariables>		
11	<BasicDefinitions>		
12	<MeasurementUnit OID="MU_HGNC" Name="HGNC">		
13	<Symbol>		
14	<TranslatedText>HGNC</TranslatedText>		
15	</Symbol>		

3. Import ODI Grid file into *grid* sheet as using the excel import wizard, again as a txt file. Go to Data tab > From Text > change file import dialog to 'All files' and select the ODI Grid file.

- Select 'delimited' > click next > tick Other on delimiters options and add a ~ to separate data. Click next and finish.

Data on *GRID sheet* will look like below:

	A	B	C	D
1	Study Subject ID	none	TRUE	FALSE
2	Protocol ID	none	FALSE	FALSE
3	Person ID	none	FALSE	FALSE
4	Secondary ID	none	FALSE	FALSE
5	Subject Status	none	FALSE	FALSE
6	Sex	none	FALSE	FALSE
7	Date of Birth	none	FALSE	FALSE
8	Location_E1_1	none	FALSE	FALSE
9	CRF Version Status_E1_1_C1	none	FALSE	FALSE
10	Version Name_E1_1_C1	none	FALSE	FALSE
11	SUBJ_SOURCE_E1_1_C1	SE_TIME_TRIAL.F_REF_MIN_DAT_9575_V10.IG_REF	FALSE	FALSE
12	SUBJ_STATUS_E1_1_C1	SE_TIME_TRIAL.F_REF_MIN_DAT_9575_V10.IG_REF	FALSE	FALSE
13	SUBJ_FAMILY_ID_E1_1_C1	SE_TIME_TRIAL.F_REF_MIN_DAT_9575_V10.IG_REF	FALSE	FALSE
14	SUBJ_NHS_STATUS_E1_1_C1	SE_TIME_TRIAL.F_REF_MIN_DAT_9575_V10.IG_REF	FALSE	FALSE
15	SUBJ_SURNAME_E1_1_C1	SE_TIME_TRIAL.F_REF_MIN_DAT_9575_V10.IG_REF	FALSE	FALSE

- On the *Field conversion sheet* filter on <ItemRef ItemOID to list all the items

	A	B	C	D
1	<?xml version="1.0" encoding="UTF-8"?>	Unique item id	Core Item ID	Table
56	<ItemRef ItemOID="I_REF_SUBJ_SOURCE_118" OrderNumber="1" Mandatory="No"/>	I_REF_SUBJ_SOURCE_118	I_REF_SUBJ_SOURCE	SUBJ_SOURCE
57	<ItemRef ItemOID="I_REF_SUBJ_STATUS_5168" OrderNumber="2" Mandatory="No"/>	I_REF_SUBJ_STATUS_5168	I_REF_SUBJ_STATUS	SUBJ_STATUS
58	<ItemRef ItemOID="I_REF_SUBJ_FAMILY_ID_6122" OrderNumber="3" Mandatory="No"/>	I_REF_SUBJ_FAMILY_ID_6122	I_REF_SUBJ_FAMILY_ID	SUBJ_FAMILY
59	<ItemRef ItemOID="I_REF_SUBJ_NHS_NUMBER_5559" OrderNumber="4" Mandatory="No"/>	I_REF_SUBJ_NHS_NUMBER_5559	I_REF_SUBJ_NHS_NUMBER	SUBJ_NHS_NU
60	<ItemRef ItemOID="I_REF_SUBJ_NHS_STATUS_3217" OrderNumber="5" Mandatory="No"/>	I_REF_SUBJ_NHS_STATUS_3217	I_REF_SUBJ_NHS_STATUS	SUBJ_NHS_STA
61	<ItemRef ItemOID="I_REF_SUBJ_SURNAME_3761" OrderNumber="6" Mandatory="No"/>	I_REF_SUBJ_SURNAME_3761	I_REF_SUBJ_SURNAME	SUBJ_SURNAM
62	<ItemRef ItemOID="I_REF_SUBJ_FIRST_NAME_8350" OrderNumber="7" Mandatory="No"/>	I_REF_SUBJ_FIRST_NAME_8350	I_REF_SUBJ_FIRST_NAME	SUBJ_FIRST_NA
63	<ItemRef ItemOID="I_REF_SUBJ_ADDRESS_1_2527" OrderNumber="8" Mandatory="No"/>	I_REF_SUBJ_ADDRESS_1_2527	I_REF_SUBJ_ADDRESS_1	SUBJ_ADDRESS
64	<ItemRef ItemOID="I_REF_SUBJ_ADDRESS_2_9588" OrderNumber="9" Mandatory="No"/>	I_REF_SUBJ_ADDRESS_2_9588	I_REF_SUBJ_ADDRESS_2	SUBJ_ADDRESS
65	<ItemRef ItemOID="I_REF_SUBJ_ADDRESS_3_2857" OrderNumber="10" Mandatory="No"/>	I_REF_SUBJ_ADDRESS_3_2857	I_REF_SUBJ_ADDRESS_3	SUBJ_ADDRESS
66	<ItemRef ItemOID="I_REF_SUBJ_ADDRESS_4_7773" OrderNumber="11" Mandatory="No"/>	I_REF_SUBJ_ADDRESS_4_7773	I_REF_SUBJ_ADDRESS_4	SUBJ_ADDRESS
67	<ItemRef ItemOID="I_REF_SUBJ_ADDRESS_5_5064" OrderNumber="12" Mandatory="No"/>	I_REF_SUBJ_ADDRESS_5_5064	I_REF_SUBJ_ADDRESS_5	SUBJ_ADDRESS
68	<ItemRef ItemOID="I_REF_SUBJ_POSTCODE_7902" OrderNumber="13" Mandatory="No"/>	I_REF_SUBJ_POSTCODE_7902	I_REF_SUBJ_POSTCODE	SUBJ_POSTCO
69	<ItemRef ItemOID="I_REF_SUBJ_GENDER_AT_BIRTH_974" OrderNumber="14" Mandatory="No"/>	I_REF_SUBJ_GENDER_AT_BIRTH_974	I_REF_SUBJ_GENDER_AT_BIRTH	SUBJ_GENDER

6.3.2 Creating the ODI grid.dmp file

In order to create the ODI grid.dmp file some cleanup must be done on the *grid* and *file conversion* sheets. The mapping definitions spreadsheet will be provided in a format, with the correct cleanup formulas, so that all that is needed is a paste of the keystone template and study metadata is needed. Explanation of formulas is provided as a reference only.

With the below formulas applied copy/paste the resulting column L into Wordpad and save as ODI [study_name]_grid.dmp to the ODI folder.

6.3.2.1 File conversion sheet formulas

	A	B	C	
1	<?xml version="1.0" encoding="UTF-8"?>	Unique item id	Core item ID	Table
56	<itemRef ItemOID="I_REF_SUBJ_SOURCE_118" OrderNumber="1" Mandatory="No"/>	I_REF_SUBJ_SOURCE_118	I_REF_SUBJ_SOURCE	SUBJ_SOURCE
57	<itemRef ItemOID="I_REF_SUBJ_STATUS_5168" OrderNumber="2" Mandatory="No"/>	I_REF_SUBJ_STATUS_5168	I_REF_SUBJ_STATUS	SUBJ_STATUS
58	<itemRef ItemOID="I_REF_SUBJ_FAMILY_ID_6122" OrderNumber="3" Mandatory="No"/>	I_REF_SUBJ_FAMILY_ID_6122	I_REF_SUBJ_FAMILY_ID	SUBJ_FAMILY
59	<itemRef ItemOID="I_REF_SUBJ_NHS_NUMBER_5559" OrderNumber="4" Mandatory="No"/>	I_REF_SUBJ_NHS_NUMBER_5559	I_REF_SUBJ_NHS_NUMBER	SUBJ_NHS_NUM
60	<itemRef ItemOID="I_REF_SUBJ_NHS_STATUS_3217" OrderNumber="5" Mandatory="No"/>	I_REF_SUBJ_NHS_STATUS_3217	I_REF_SUBJ_NHS_STATUS	SUBJ_NHS_STA
61	<itemRef ItemOID="I_REF_SUBJ_SURNAME_3761" OrderNumber="6" Mandatory="No"/>	I_REF_SUBJ_SURNAME_3761	I_REF_SUBJ_SURNAME	SUBJ_SURNAM
62	<itemRef ItemOID="I_REF_SUBJ_FIRST_NAME_8350" OrderNumber="7" Mandatory="No"/>	I_REF_SUBJ_FIRST_NAME_8350	I_REF_SUBJ_FIRST_NAME	SUBJ_FIRST_NA
63	<itemRef ItemOID="I_REF_SUBJ_ADDRESS_1_2527" OrderNumber="8" Mandatory="No"/>	I_REF_SUBJ_ADDRESS_1_2527	I_REF_SUBJ_ADDRESS_1	SUBJ_ADDRESS
64	<itemRef ItemOID="I_REF_SUBJ_ADDRESS_2_9588" OrderNumber="9" Mandatory="No"/>	I_REF_SUBJ_ADDRESS_2_9588	I_REF_SUBJ_ADDRESS_2	SUBJ_ADDRESS
65	<itemRef ItemOID="I_REF_SUBJ_ADDRESS_3_2857" OrderNumber="10" Mandatory="No"/>	I_REF_SUBJ_ADDRESS_3_2857	I_REF_SUBJ_ADDRESS_3	SUBJ_ADDRESS
66	<itemRef ItemOID="I_REF_SUBJ_ADDRESS_4_7773" OrderNumber="11" Mandatory="No"/>	I_REF_SUBJ_ADDRESS_4_7773	I_REF_SUBJ_ADDRESS_4	SUBJ_ADDRESS
67	<itemRef ItemOID="I_REF_SUBJ_ADDRESS_5_5064" OrderNumber="12" Mandatory="No"/>	I_REF_SUBJ_ADDRESS_5_5064	I_REF_SUBJ_ADDRESS_5	SUBJ_ADDRESS
68	<itemRef ItemOID="I_REF_SUBJ_POSTCODE_7902" OrderNumber="13" Mandatory="No"/>	I_REF_SUBJ_POSTCODE_7902	I_REF_SUBJ_POSTCODE	SUBJ_POSTCOA
69	<itemRef ItemOID="I_REF_SUBJ_GENDER_AT_BIRTH_974" OrderNumber="14" Mandatory="No"/>	I_REF_SUBJ_GENDER_AT_BIRTH_974	I_REF_SUBJ_GENDER_AT_BIRTH	SUBJ_GENDER

Column B - strips out the Unique item ID with

=MID(A156,(FIND(CHAR(34),A156)+1),(-1+FIND(CHAR(34),A156,(FIND(CHAR(34),A156,1)+1))-(FIND(CHAR(34),A156))))

Column C - creates the Core item ID (without suffix automatically added by OC)

=LEFT(B156,(-1+FIND("@",SUBSTITUTE(B156,"_","@",LEN(B156)-LEN(SUBSTITUTE(B156,"_",""))))))

Column D - generates table name (with hardcoded event info - E1__1__C1 this is only possible if all data is held in a single CRF and a dingle event)

=CONCATENATE(SUBSTITUTE(B156,"I_REF_",""),"_E1__1__C1",")

Column E - generates values for XML (includes hardcoding which will need modifying for each study)

=CONCATENATE("SE_TIME_TRIAL.F_REF_MIN_DAT_9575_V10.IG_REF_PEDIGREE_431.",B156)

A **named range** is created called 'autolink' and is used on vlookup between data on the *Grid sheet* and *Field Conversion sheet*

6.3.2.2 GRID sheet formulas

	A	C	D	E	F	G	H	I	J	K	L
1										Matched Ref	dmp file
2	Study Subject ID	none	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	CopyTarget		Study Subject ID~TRUE~FAL
3	Protocol ID	none	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	CopyTarget		Protocol ID~FALSE~FALSE~f
4	Person ID	none	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	CopyTarget		Person ID~FALSE~FALSE~FA
5	Secondary ID	none	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	CopyTarget		Secondary ID~FALSE~FALSE
6	Subject Status	none	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	CopyTarget		Subject Status~FALSE~FALS
7	Sex	none	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	CopyTarget		Sex~FALSE~FALSE~FALSE~F
8	Date of Birth	none	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	CopyTarget		Date of Birth~FALSE~FALSE
9	Location_E1_1	none	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	CopyTarget		Location_E1_1~FALSE~FALS
10	CRF Version Status_E1_1_C1	none	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	CopyTarget		CRF Version Status_E1_1_C
11	Version Name_E1_1_C1	none	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	CopyTarget		Version Name_E1_1_C1~FA
12	SUBJ_SOURCE_E1_1_C1	none	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	CopyTarget	SE_TIME_TRIAL.F_REF_MIN_DAT_957	SUBJ_SOURCE_E1_1_C1~SE
13	SUBJ_STATUS_E1_1_C1	none	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	CopyTarget	SE_TIME_TRIAL.F_REF_MIN_DAT_957	SUBJ_STATUS_E1_1_C1~SE
14	SUBJ_FAMILY_ID_E1_1_C1	none	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	CopyTarget	SE_TIME_TRIAL.F_REF_MIN_DAT_957	SUBJ_FAMILY_ID_E1_1_C1
15	SUBJ_NHS_STATUS_E1_1_C1	none	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	CopyTarget	SE_TIME_TRIAL.F_REF_MIN_DAT_957	SUBJ_NHS_STATUS_E1_1_C
16	SUBJ_SURNAME_E1_1_C1	none	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	CopyTarget	SE TIME TRIAL.F_REF MIN DAT 957	SUBJ_SURNAME_E1_1_C1~

Columns A to J have been imported as described above

Column K - Matched ref - does a vlookup against *Field Conversion sheet* data

```
=VLOOKUP(CONCATENATE("I__REF__",SUBSTITUTE(B14,"_E1_1_C1","",1)),autolink,3,FALSE)
```

Column L - Dmp file - is final output to be pasted into ODI Grid (dmp) file

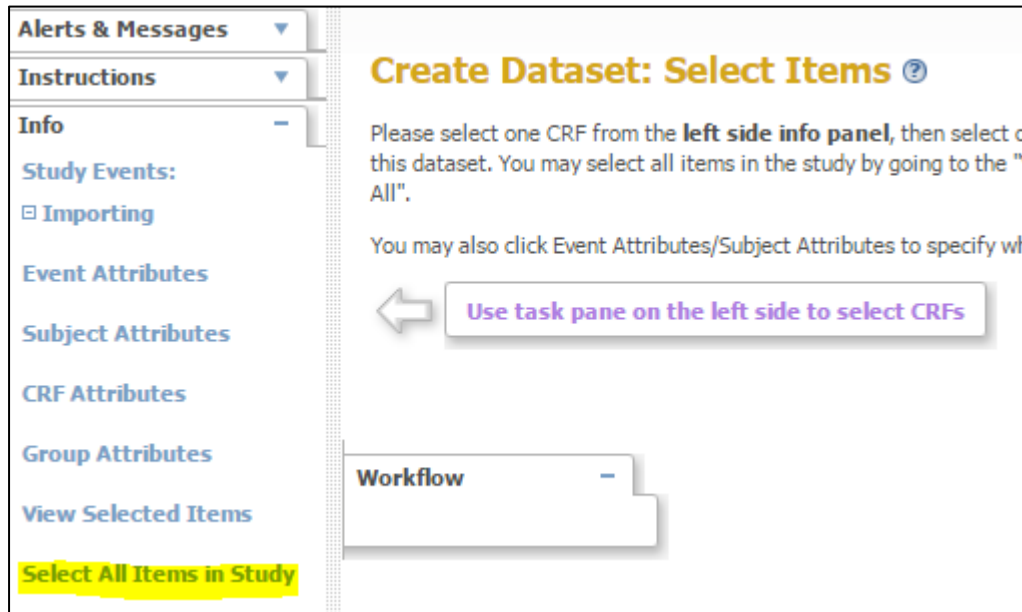
```
=IF((ISERROR(K13)),CONCATENATE(A13,CONCATENATE("~",C13),"~",D13,"~",E13,"~",F13,"~",G13,"~",H13,"~",I13,"~",J13),CONCATENATE(A13,CONCATENATE("~",K13),"~",D13,"~",E13,"~",F13,"~",G13,"~",H13,"~",I13,"~",J13))
```

6.3.3 Generate XML data import file

Having generated your ODI Grid.dmp file you can now run through the standard OCDataImprter steps outlined in section 4.3 to generate the study data XML file for import into OpenClinica.

7 Data Export from OpenClinica

1. In OpenClinica go to Tasks > Extract Data > Create dataset > click “select all items in study”



2. Save and define scope > select earliest to latest dates possible 01/01/1980 – 31/12/2015

Create Dataset: Define Temporal Scope

Filter your data by subject enrollment date by selecting a date range below.

Select the event start and end dates below. To specify data from the beginning of the year, you may leave the month field blank. However, if you specify a month, you must also specify the year.

If you do not wish to filter study event data by date of subject enrollment, you may leave the date fields blank and continue.

Beginning Date:

End Date:

Save and Export your dataset now or Apply a Filter to specify additional selection criteria based on values of parameters in CRFs.

3. Provide name for data set and description, select item status “Data from all available CRFs”
4. Click Continue > Confirm and save
5. Select your preferred data format

- [CDISC ODM XML 1.3 Full with OpenClinica extensions](#) [Run Now](#)
- [CDISC ODM XML 1.3 Clinical Data with OpenClinica extensions](#) [Run Now](#)
- [CDISC ODM XML 1.3 Clinical Data](#) [Run Now](#)
- [CDISC ODM XML 1.2 Clinical Data with OpenClinica extensions](#) [Run Now](#)
- [CDISC ODM XML 1.2 Clinical Data](#) [Run Now](#)
- [View as HTML](#) [Run Now](#)
- [Excel Spreadsheet](#) [Run Now](#)
- [Tab-delimited Text](#) [Run Now](#)
- [SPSS data and syntax](#) [Run Now](#)

6. Click “Run now”. An email will be sent once available for download and an alert notice will show in top left pane. Download and open files in preferred analysis application (SPSS, excel etc)

8 Data Cleaning

Section coming.

Excel, Google Refine, eyeballs...

9 Test Data Generation (optional)

Section coming.

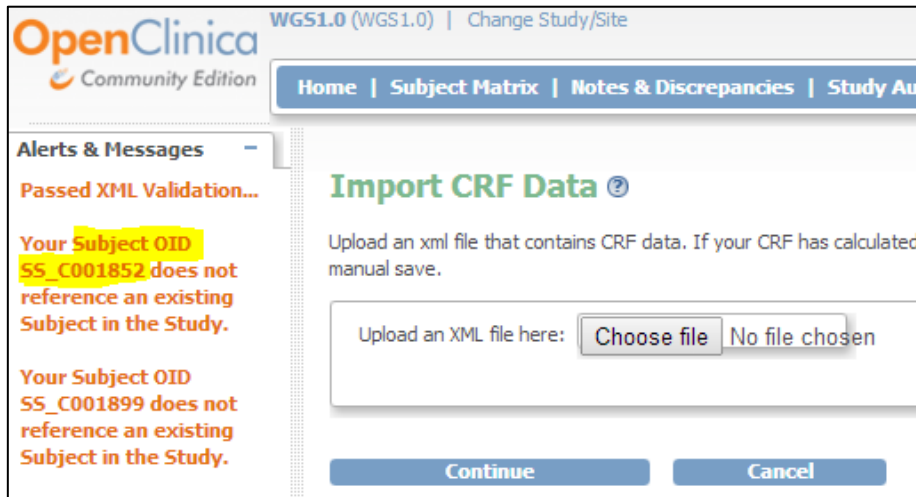
OPTIONAL Data Generation for testing / creating dummy data only, step not needed for regular study data upload)

- a. Open a copy of the original CRF used to loaded into the study in step 2 above
- b. Import into a new sheet in the CRF the file Keystone template [TAB_\[exportname\]_2015-01-21-114325879.tsv](#)

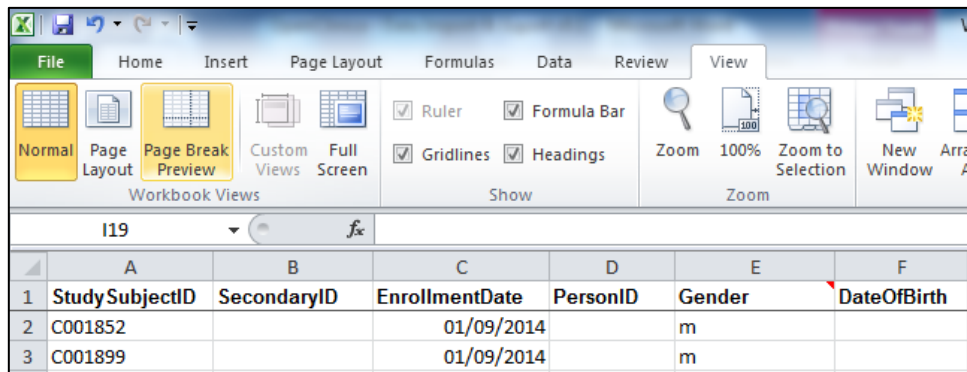
10 Common Import Errors

10.1 Study Subject IDs and OIDs

- 1.) A common error on importing CRF data is that the Subject OID is not found in the study, even though you have already added all your subject records.



When study subjects are added directly through the OpenClinica interface, a **Study Subject ID OID** is generated. This system prefixes your study subject Id with **SS_** so **C001852** becomes **SS_C001852**



	A	B	C	D	E	F
1	StudySubjectID	SecondaryID	EnrollmentDate	PersonID	Gender	DateOfBirth
2	C001852		01/09/2014		m	
3	C001899		01/09/2014		m	

In your tab delimited txt file ensure your **Study Subject ID** includes any suffixes (the prefix will automatically be included)

11 Exporting Data

- 1.) Click Tasks > Create Dataset
- 2.) Click the button 'Proceed to Create a Dataset'
- 3.) You are presented with a navigation pane on the left of you study events, subject, group and CRF attributes etc. Click on the appropriate item in the navigation pane and tick the items you wish to export.

Create Dataset: Select Items

Please select one CRF from the left side info panel, then select one or more items in a CRF that you would like to include to this dataset. You may select Selected Items" (hyperlink) page and clicking "Select All".

You may also click Event Attributes/Subject Attributes to specify which event/subject attribute will be shown in the dataset.

Use task pane on the left side to select CRFs

Event Name: Illumina Manifest Upload
 CRF Name: A-WGS1
 Description: First step for Illumina Plate creation

Select All Items Save and Add More Items Save and Define Scope Cancel

Show the following items in this dataset:

	Name	Description	Version(s)	Section(s)	Group(s)	Data Type	Units	Response Type	Response Label	PHI	Required
<input checked="" type="checkbox"/>	STUDY	Study	1.0	Manifest	Ungrouped	st		single-select	study	No	No
<input checked="" type="checkbox"/>	PLATE_BARCODE	Illumina plate number	1.0	Manifest	Ungrouped	st		text	text	No	No
<input checked="" type="checkbox"/>	WELL	Illumina plate well position	1.0	Manifest	Ungrouped	st		text	text	No	No
<input type="checkbox"/>	SAMPLE_WELL	Illumina ID	1.0	Manifest	Ungrouped	st		text	text	No	No
<input checked="" type="checkbox"/>	SAMPLE_ID	Sample ID	1.0	Manifest	Ungrouped	st		text	text	No	No
<input checked="" type="checkbox"/>	SPECIES	Species	1.0	Manifest	Ungrouped	st		single-select	species	No	No

- 4.) Click 'Save and Add more Items' to navigate to another CRF, Event etc
- 5.) Click 'Save and Define Scope' to move to next screen to set you dates you wish to export against

Create Dataset: Define Temporal Scope

Filter your data by subject enrollment date by selecting a date range below.

Select the event start and end dates below. To specify data from the beginning of the year, you may leave the month field blank. However, if you specify a month, you must also specify the year.

If you do not wish to filter study event data by date of subject enrollment, you may leave the date fields blank and continue.

Beginning Date: August 2014
 End Date: September 2014

Save and Export your dataset now or Apply a Filter to specify additional selection criteria based on values of parameters in CRFs.

Continue Cancel

- 6.) Add the export set name, description

- 7.) Select whether you wanted data from completed, incomplete or all CRFs.
- 8.) Click continue
- 9.) Click 'Confirm and Save'
- 10.) Select the format you wish to export to (in this case we are selected Excel)

Download Data: partial data [?](#)

Dataset Name:	partial data
Dataset Description:	more data
Item Status:	Data from CRFs Marked Complete

To view or download data, select from the formats provided below. You may also select from the archived dataset files at the bottom of the page.

- [CDISC ODM XML 1.3 Full with OpenClinica extensions](#) Run Now
- [CDISC ODM XML 1.3 Clinical Data with OpenClinica extensions](#) Run Now
- [CDISC ODM XML 1.3 Clinical Data](#) Run Now
- [CDISC ODM XML 1.2 Clinical Data with OpenClinica extensions](#) Run Now
- [CDISC ODM XML 1.2 Clinical Data](#) Run Now
- [View as HTML](#) Run Now
- [Excel Spreadsheet](#) Run Now
- [Tab-delimited Text](#) Run Now
- [SPSS data and syntax](#) Run Now

- 11.) Click 'Run Now'
- 12.) Navigate back to datasets – In alerts navigation pane a message will appear when your data extract job has run providing a link to download the file.

OpenClinica

WGS1.0 (WGS1.0) | [Change Study/Site](#)

[Home](#) | [Subject Matrix](#) | [Notes & Discrepancies](#) | [Study Audit Log](#) | [Tasks](#)

Alerts & Messages

Your extract job completed successfully. The file is available for download [here](#).

Instructions

Other Info

[Study: WGS1.0](#)

Download Data: partial data [?](#)

Dataset Name:	partial data
Dataset Description:	more data
Item Status:	Data from CRFs Marked Compl

To view or download data, select from the formats provided below. You may also sele

- [CDISC ODM XML 1.3 Full with OpenClinica extensions](#) Run Now

- 13.) You can also access the data set and any previous ones run by click Tasks > View Datasets > and clicking the Export Dataset icon.