# OpenClinica - Data Import & Export

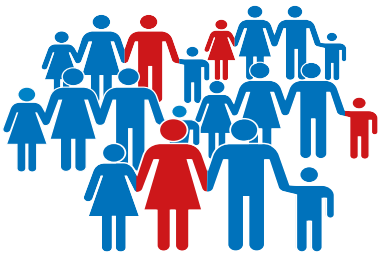
**Purpose of this document:** To provide instruction and best practice guidelines' for the import and export of study data into and out of OpenClinica.

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

## 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.

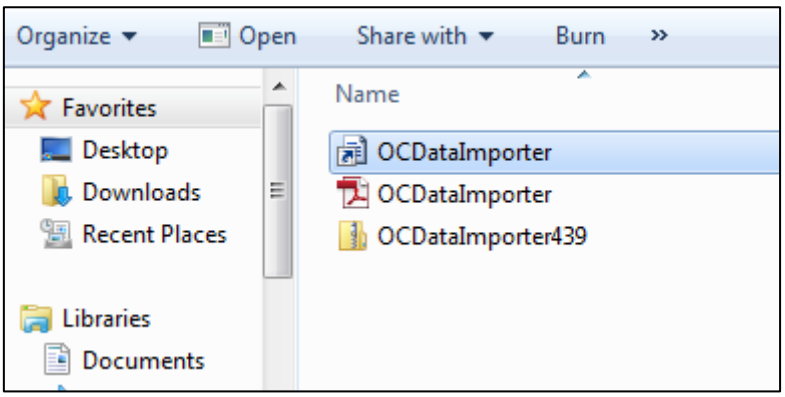
- TDS Importer is useful for those comfortable working in Access and primarily used for loading subjects and scheduling events, not for loading study data.

# 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.

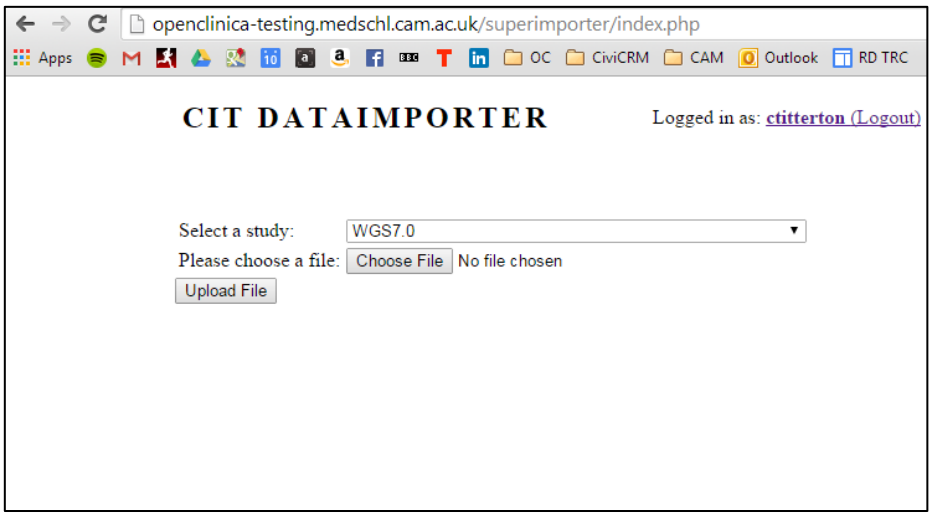
1.) Download OCdataimporter from https://community.openclinica.com/extension/ocdataimporter. Unzip the files and save to a new folder …/Desktop/OC Import.




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

## 2.2 CIT Importer

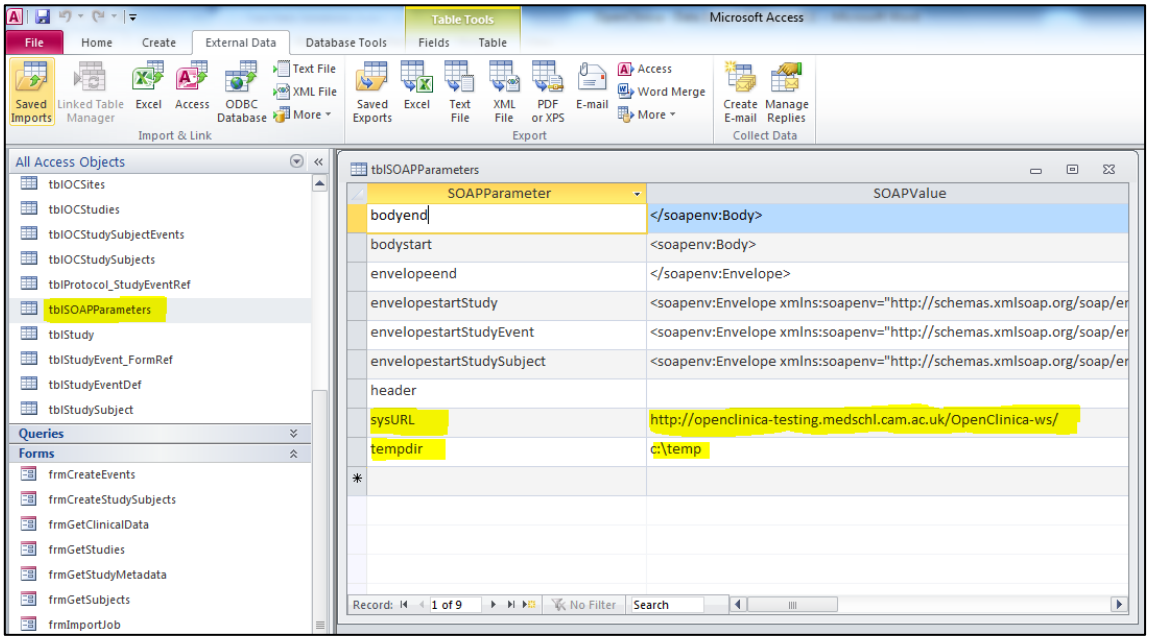
The CIT Importer does not require download. It is installed on your OpenClinica instance and available directly through a url http://openclinica-testing.medschl.cam.ac.uk/superimporter/index.php (the highlighted section will be specific to your instance). If you are using a locally managed instance of 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 **tblSOAPParameters**.




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.

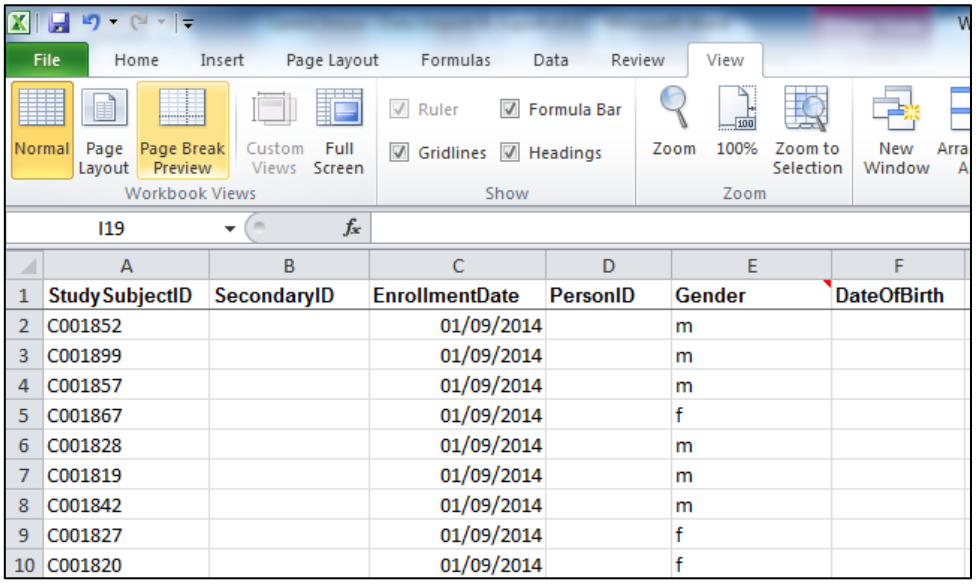
To use the TDS importer it is assumed you have your web-services running on OpenClinica and that your OpenClinica user account has been set up as an authorized user root to use the SOAP-services.

# 3 Import Subjects & Events

## 3.1 Prepare your subject and event data

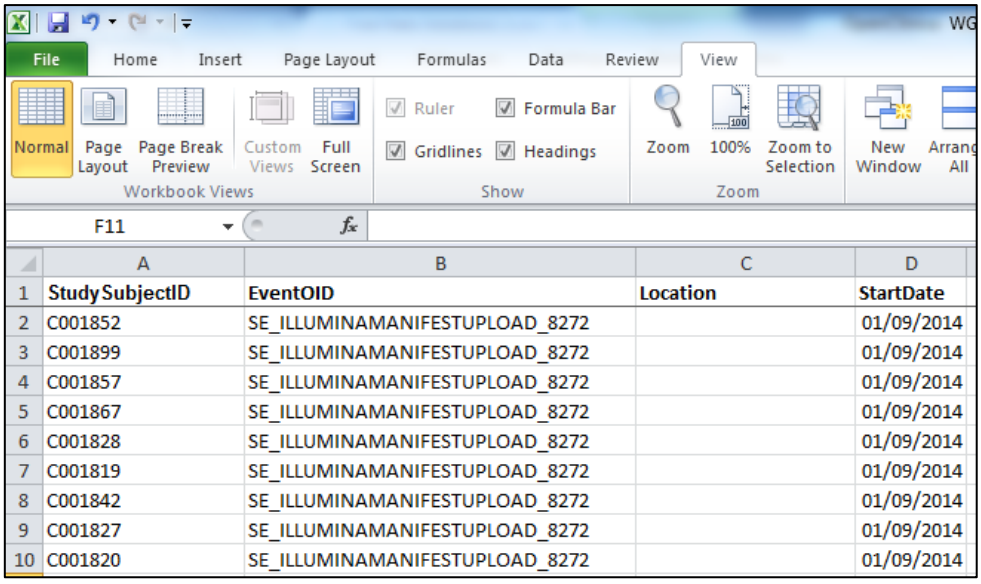
It is assumed your data is in an Excel format. The example below shows only three fields being populated. Depending on your study *Secondary ID*, *Person ID* and *Date of Birth* may also be populated.

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

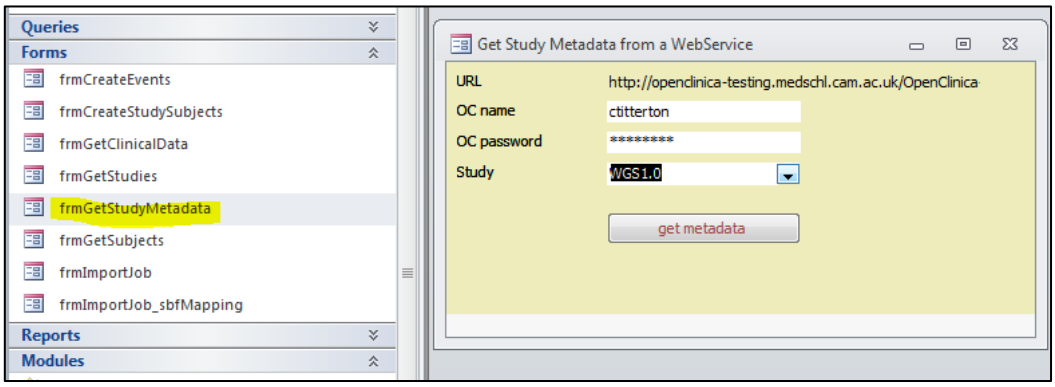



| | A | B | C | D |
|---|---|---|---|---|
| 1 | StudySubjectID | EventOID | Location | StartDate |
| 2 | C001852 | SE_ILLUMINAMANIFESTUPLOAD_8272 | | 01/09/2014 |
| 3 | C001899 | SE_ILLUMINAMANIFESTUPLOAD_8272 | | 01/09/2014 |
| 4 | C001857 | SE_ILLUMINAMANIFESTUPLOAD_8272 | | 01/09/2014 |
| 5 | C001867 | 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

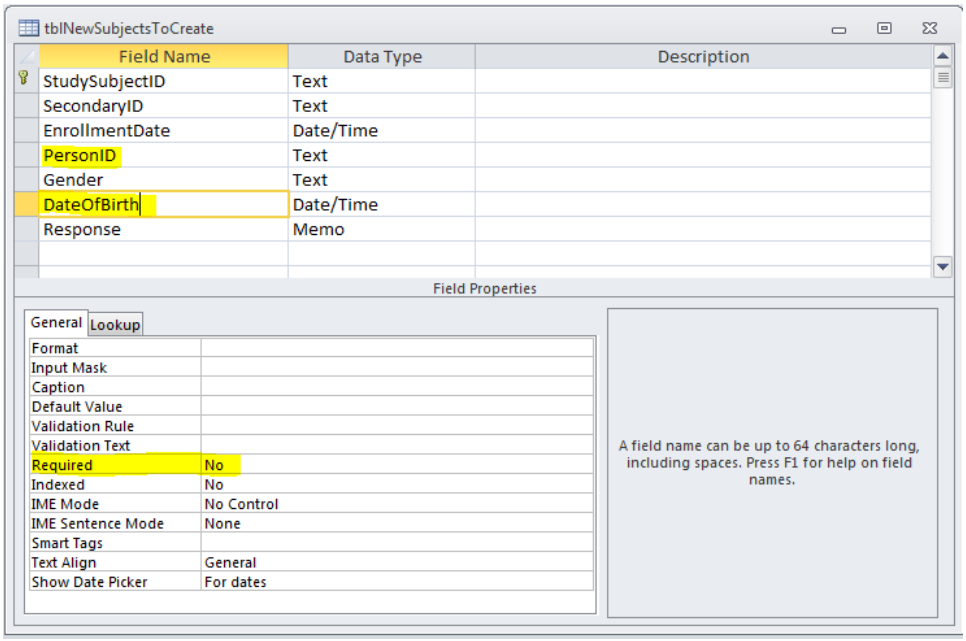
Study subjects can be added using the TDS Access database. (See section 2.3 TDS access database for installation instructions

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**.



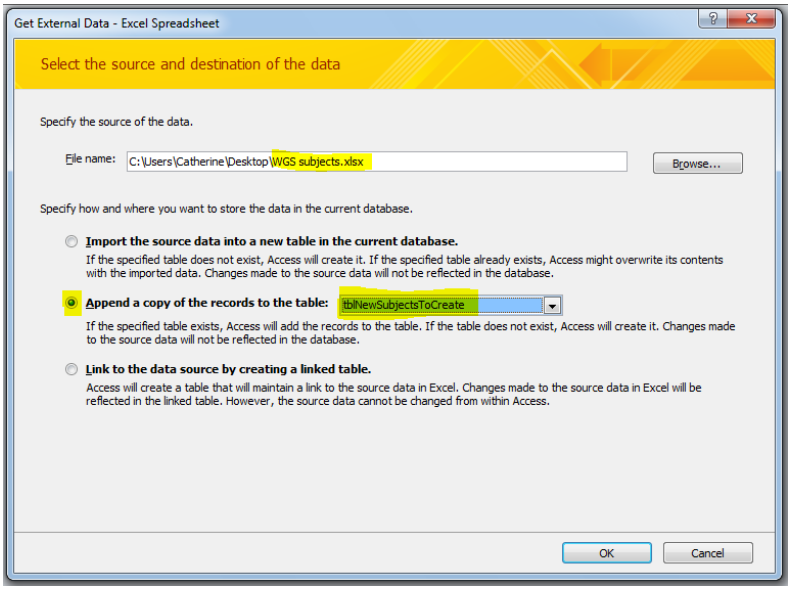

## 3.3 Import your subject records to TDS

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

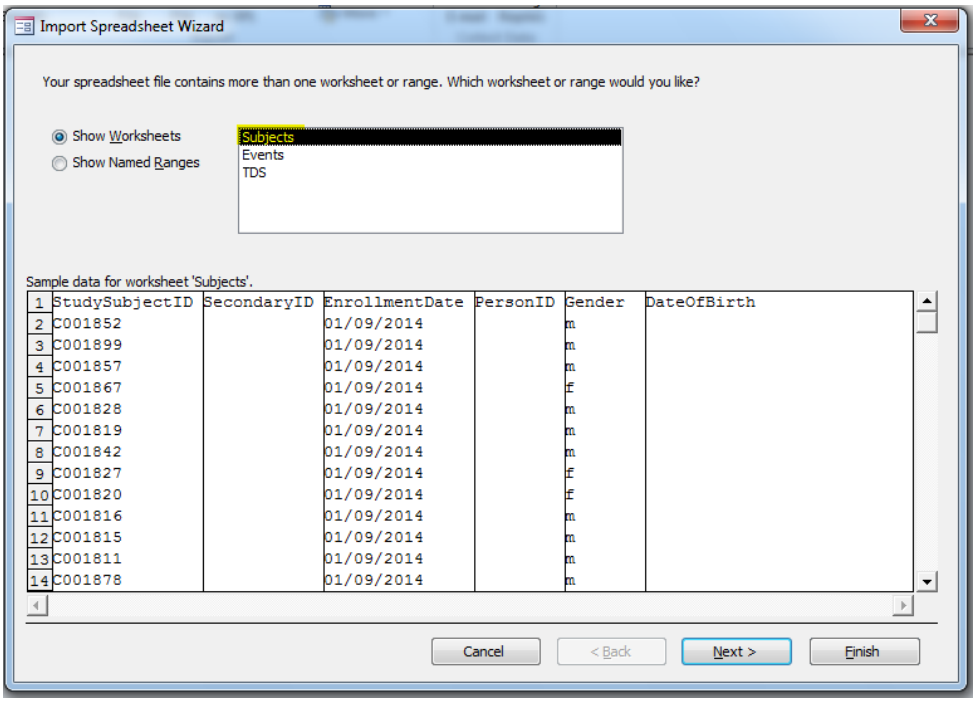



2. On the External data tab of the Access toolbar, click Excel (to import)

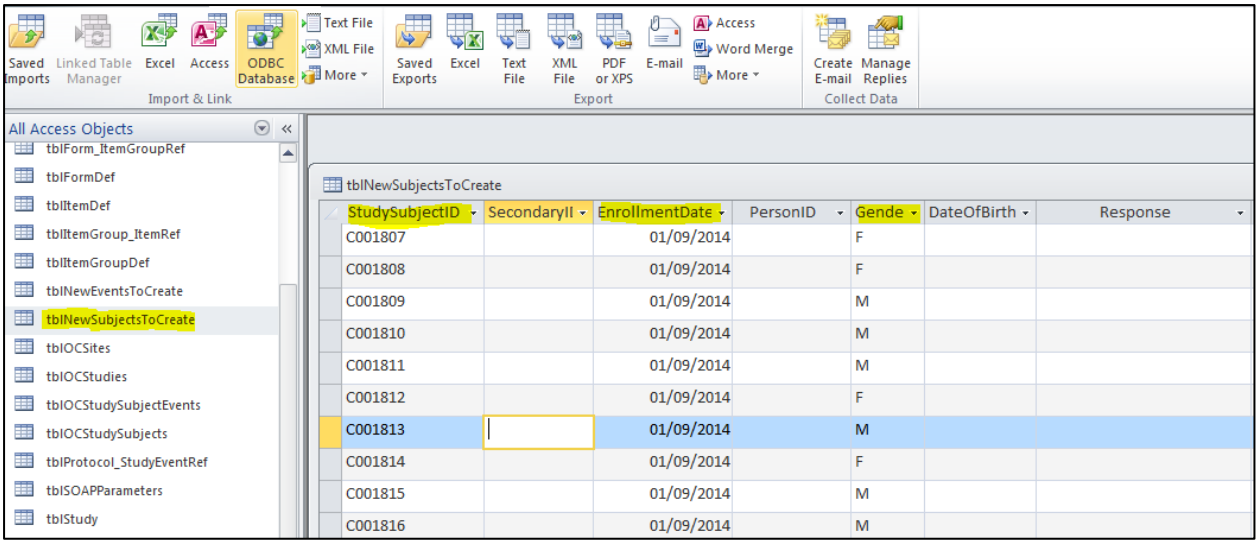
3. which opens the Get External Data dialog. Browse to your excel file containing your subject data.
4. select the option to Append a copy of records to the table and select from the dropdown **tblNewSubjectsToCreate**



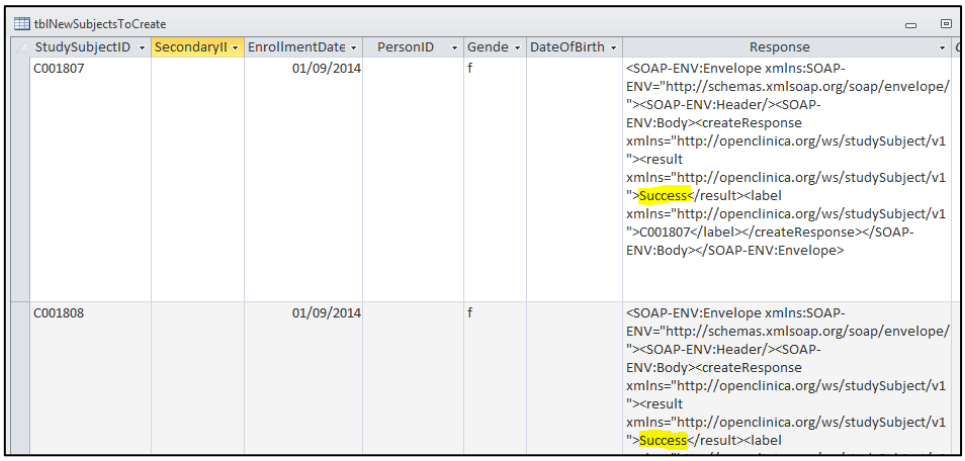

5. Click OK
6. Select the worksheet which contains subject data. (If the column headers in your spread sheet are named the same as the Access table columns you will not need to match the fields)



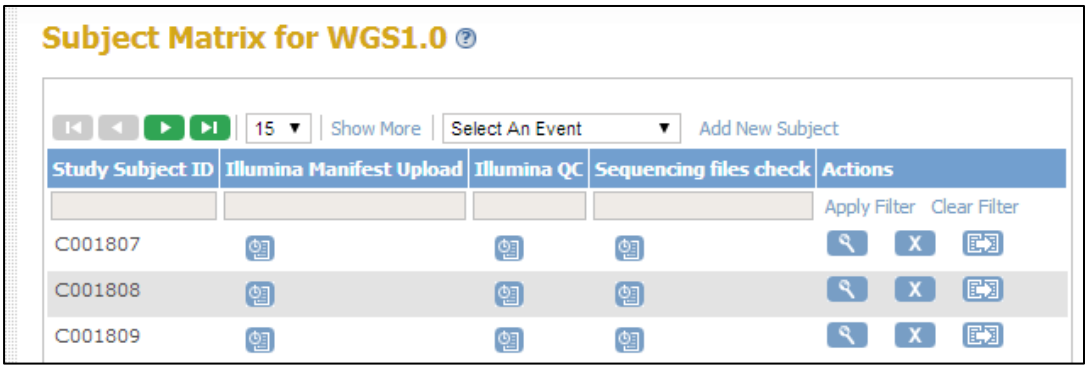

7. Click Next
8. Click Finish
9. Open the table **tblNewSubjectsToCreate** to show the subject records imported into Access.




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.

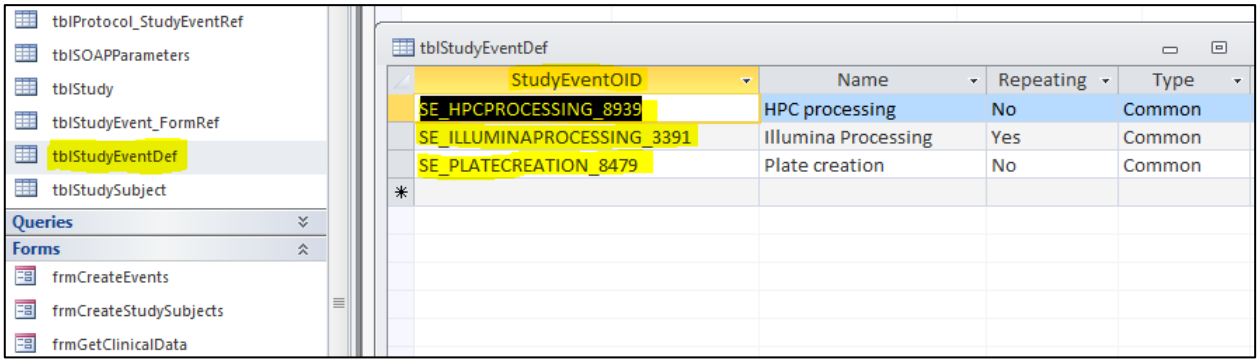



13. View the imported subject records in OpenClinica in the subject matrix



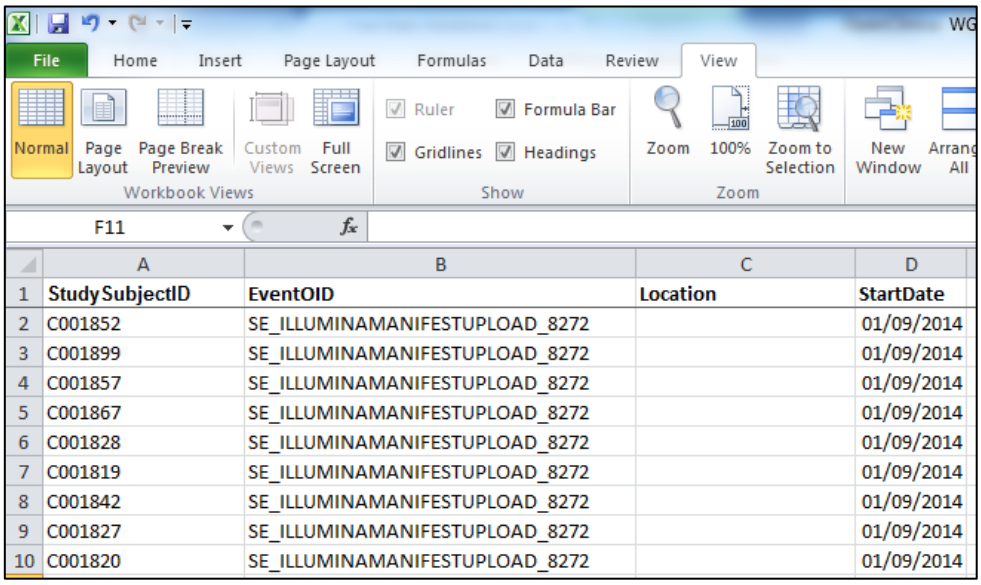
## 3.4 Import and Schedule Events with TDS

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

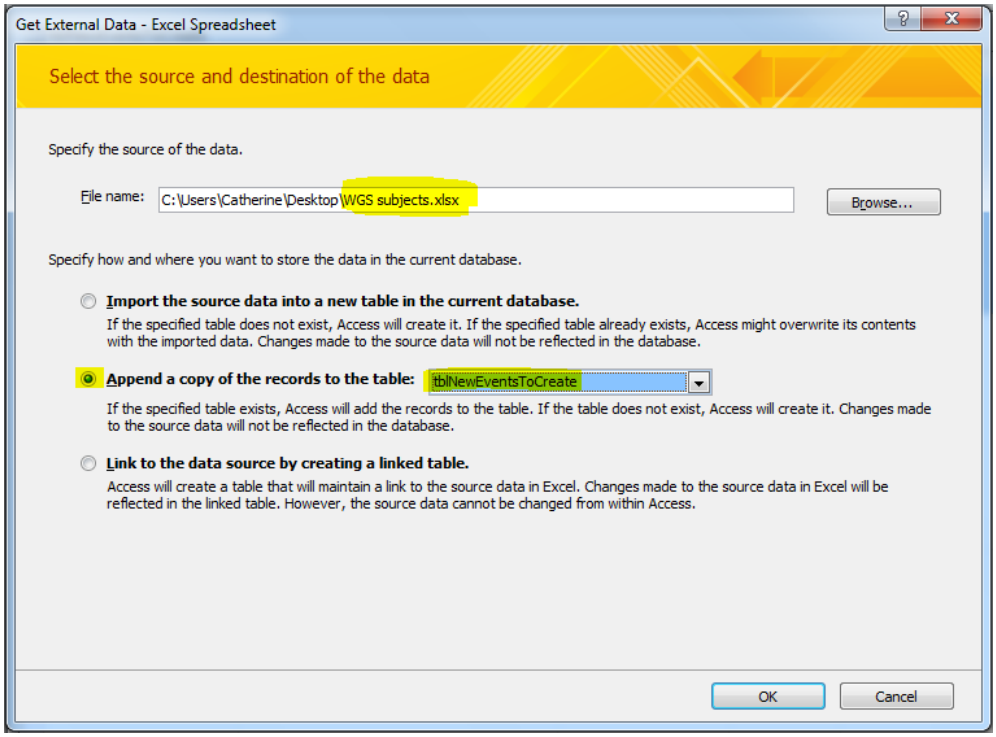



1. Copy and paste the event OID against your subject records in Excel. Note in the above example there are three events. For multiple events you need to create a row for each event against each subject (so three rows for each subject)

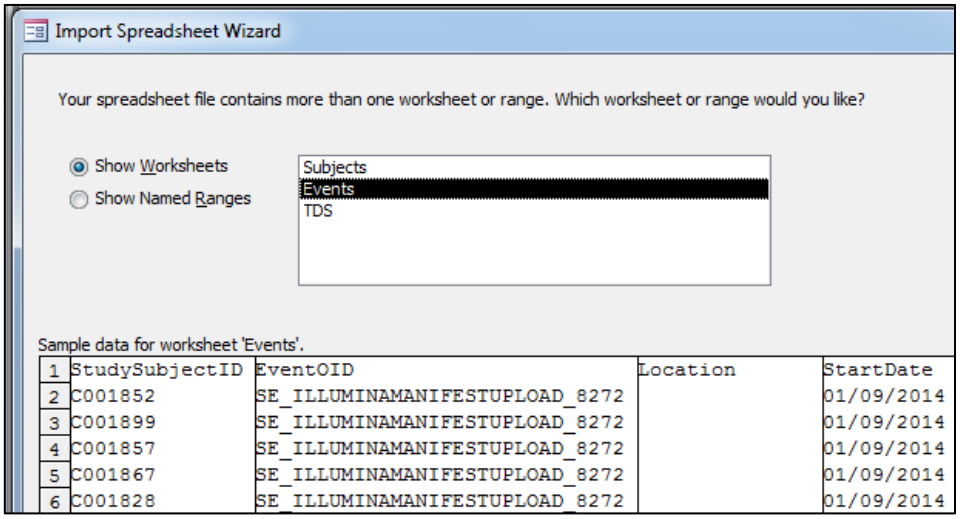
| StudySubjectID | EventOID | Location | StartDate |
|---|---|---|---|
| C001807 | SE_ILLUMINAMANIFESTUPLOAD_8272 | | 01/09/2014 |
| C001807 | SE_ILLUMINAQC_8933 | | 07/09/2014 |
| C001807 | SE_SEQUENCINGFILESCHECK_2213 | | 18/09/2014 |



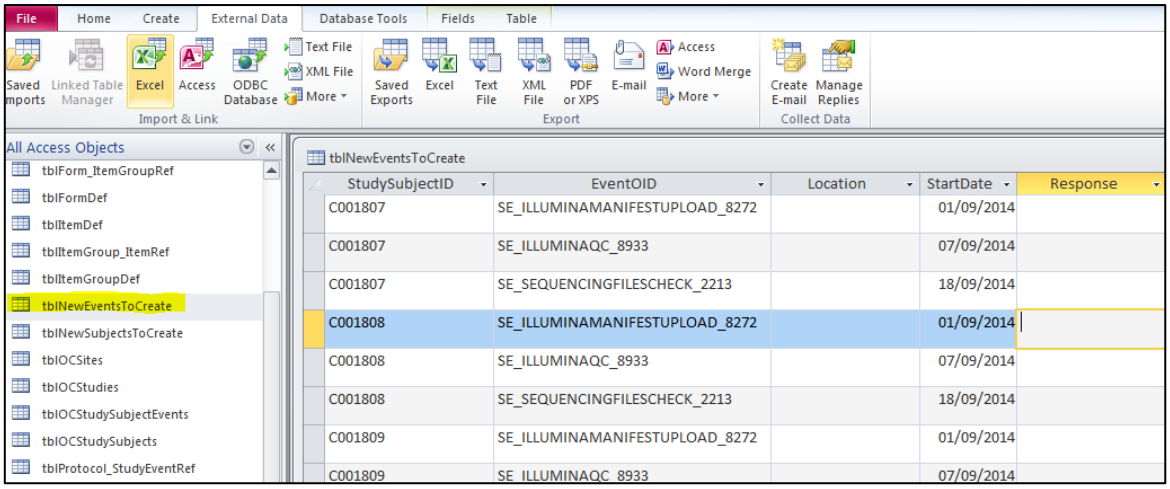

2. Follow the steps to import to Excel as per subject data > open the excel import option
3. Browse to the Excel file holding event data
4. Choose the append option and select the dropdown table **tblNewEventsToCreate**



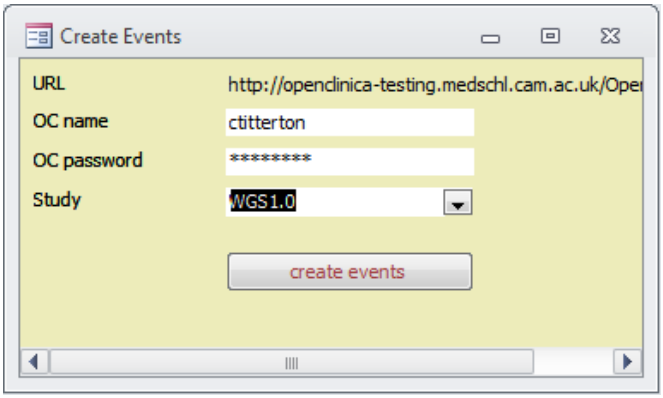

5. Ensure you select the worksheet Events



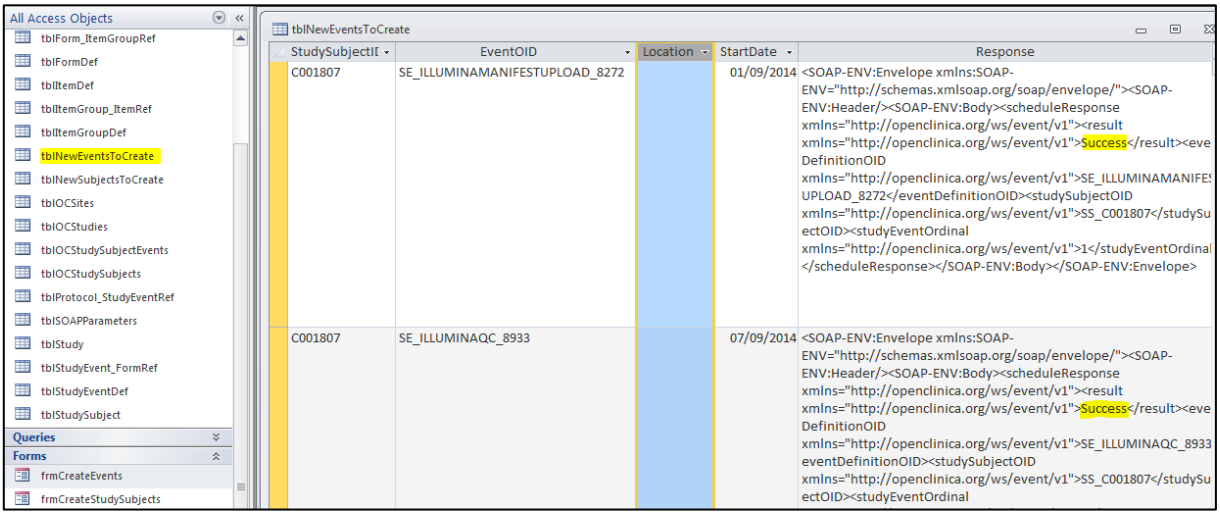

6. Click Next
7. Click Continue
8. Check the table **tblNewEventsToCreate** to see your imported records



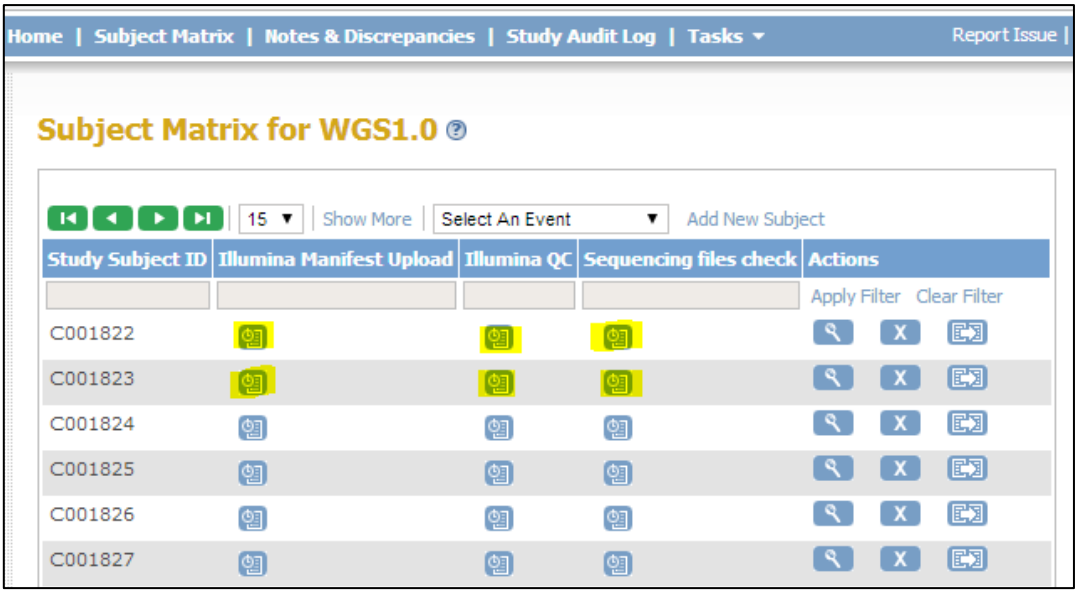

9. Open the form **frmCreateEvents** type in your *username* and *password* and select the appropriate study




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




11. Review your scheduled events inOpenClinica

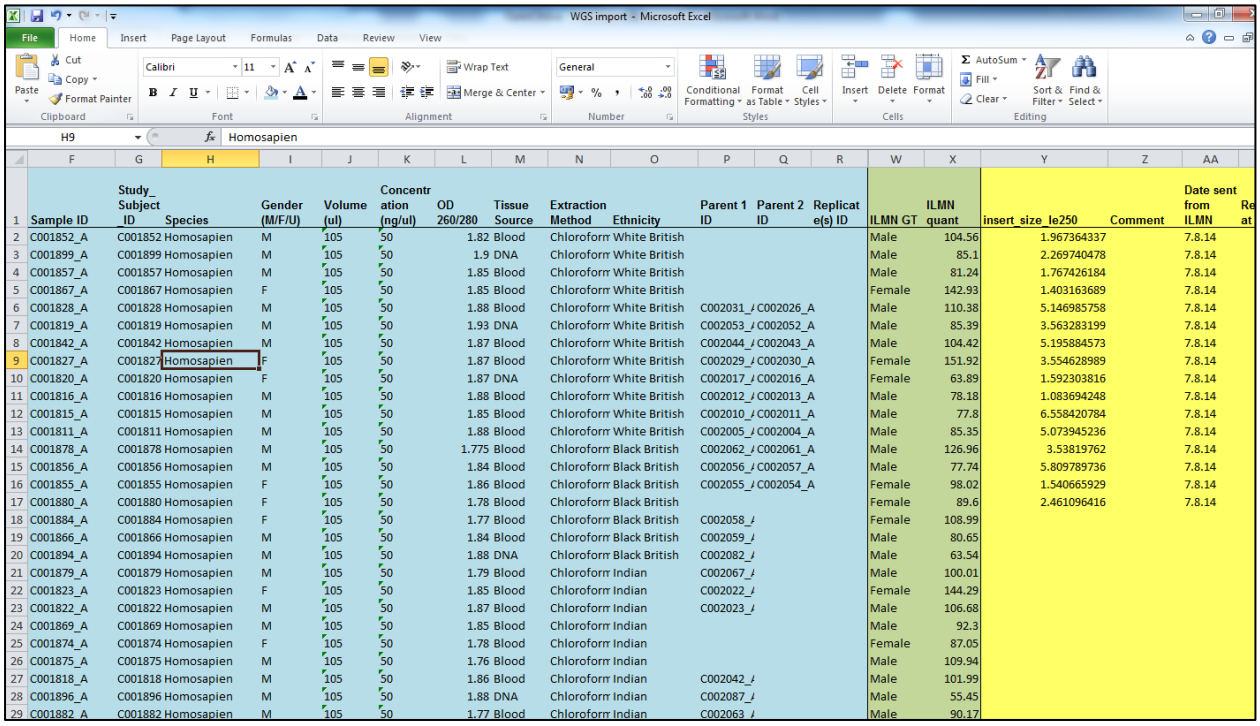

Home | Subject Matrix | Notes & Discrepancies | Study Audit Log | Tasks
Report Issue |
Subject Matrix for WGS1.0
15
Show More
Select An Event
Add New Subject
Study Subject ID
Illumina Manifest Upload
Illumina QC
Sequencing files check
Actions
Apply Filter
Clear Filter
C001822
C001823
C001824
C001825
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.

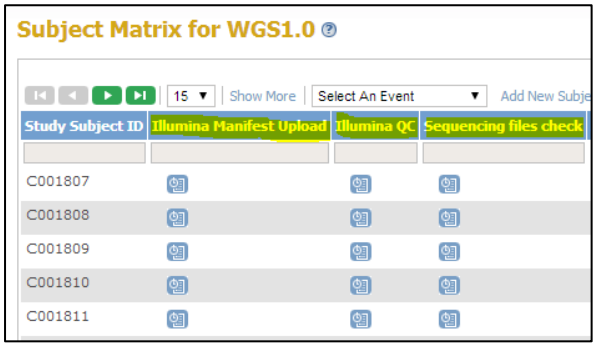
For simplicity the example below shows all data to be imported held in a single spreadsheet, although spread over three events/CRFs to be imported against. (Blue, green and yellow each represent a separate CRF)



WGS import - Microsoft Excel

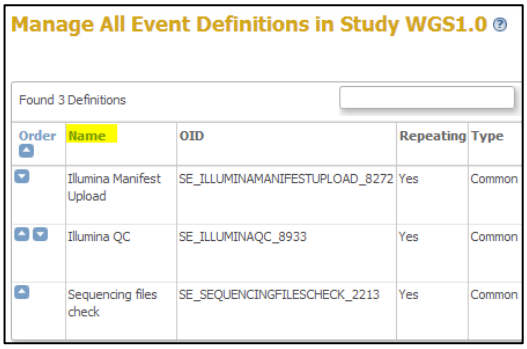
| | Sample ID | Study_ Subject _ID | Species | Gender (M/F/U) | Volume (ul) | Concentr ation (ng/ul) | OD 260/280 | Tissue Source | Extraction Method | Ethnicity | Parent 1 ID | Parent 2 ID | Replicat e(s) ID | ILMN GT | ILMN quant | insert_size_le250 | Comment | Date sent from ILMN | Re at |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | C001852_A | C001852 | Homosapien | M | 105 | 50 | 1.82 | Blood | Chloroform | White British | | | | Male | 104.56 | 1.967364337 | | 7.8.14 | |
| 3 | C001899_A | C001899 | Homosapien | M | 105 | 50 | 1.9 | DNA | Chloroform | White British | | | | Male | 85.1 | 2.269740478 | | 7.8.14 | |
| 4 | C001857_A | C001857 | Homosapien | M | 105 | 50 | 1.85 | Blood | Chloroform | White British | | | | Male | 81.24 | 1.767426184 | | 7.8.14 | |
| 5 | C001867_A | C001867 | Homosapien | F | 105 | 50 | 1.85 | Blood | Chloroform | White British | | | | Female | 142.93 | 1.403163689 | | 7.8.14 | |
| 6 | C001828_A | C001828 | Homosapien | M | 105 | 50 | 1.88 | Blood | Chloroform | White British | C002031_A | C002026_A | | Male | 110.38 | 5.146985758 | | 7.8.14 | |
| 7 | C001819_A | C001819 | Homosapien | M | 105 | 50 | 1.93 | DNA | Chloroform | White British | C002053_A | C002052_A | | Male | 85.39 | 3.563283199 | | 7.8.14 | |
| 8 | C001842_A | C001842 | Homosapien | M | 105 | 50 | 1.87 | Blood | Chloroform | White British | C002044_A | C002043_A | | Male | 104.42 | 5.195884573 | | 7.8.14 | |
| 9 | C001827_A | C001827 | Homosapien | F | 105 | 50 | 1.87 | Blood | Chloroform | White British | C002029_A | C002030_A | | Female | 151.92 | 3.554628989 | | 7.8.14 | |
| 10 | C001820_A | C001820 | Homosapien | F | 105 | 50 | 1.87 | DNA | Chloroform | White British | C002017_A | C002016_A | | Female | 63.89 | 1.592303816 | | 7.8.14 | |
| 11 | C001816_A | C001816 | Homosapien | M | 105 | 50 | 1.88 | Blood | Chloroform | White British | C002012_A | C002013_A | | Male | 78.18 | 1.083694248 | | 7.8.14 | |
| 12 | C001815_A | C001815 | Homosapien | M | 105 | 50 | 1.85 | Blood | Chloroform | White British | C002010_A | C002011_A | | Male | 77.8 | 6.558420784 | | 7.8.14 | |
| 13 | C001811_A | C001811 | Homosapien | M | 105 | 50 | 1.88 | Blood | Chloroform | White British | C002005_A | C002004_A | | Male | 85.35 | 5.073945236 | | 7.8.14 | |
| 14 | C001878_A | C001878 | Homosapien | M | 105 | 50 | 1.775 | Blood | Chloroform | Black British | C002062_A | C002061_A | | Male | 126.96 | 3.53819762 | | 7.8.14 | |
| 15 | C001856_A | C001856 | Homosapien | M | 105 | 50 | 1.84 | Blood | Chloroform | Black British | C002056_A | C002057_A | | Male | 77.74 | 5.809789736 | | 7.8.14 | |
| 16 | C001855_A | C001855 | Homosapien | F | 105 | 50 | 1.86 | Blood | Chloroform | Black British | C002055_A | C002054_A | | Female | 98.02 | 1.540665929 | | 7.8.14 | |
| 17 | C001880_A | C001880 | Homosapien | F | 105 | 50 | 1.78 | Blood | Chloroform | Black British | | | | Female | 89.6 | 2.461096416 | | 7.8.14 | |
| 18 | C001884_A | C001884 | Homosapien | F | 105 | 50 | 1.77 | Blood | Chloroform | Black British | C002058_A | | | Female | 108.99 | | | | |
| 19 | C001866_A | C001866 | Homosapien | M | 105 | 50 | 1.84 | Blood | Chloroform | Black British | C002059_A | | | Male | 80.65 | | | | |
| 20 | C001894_A | C001894 | Homosapien | M | 105 | 50 | 1.88 | DNA | Chloroform | Black British | C002082_A | | | Male | 63.54 | | | | |
| 21 | C001879_A | C001879 | Homosapien | M | 105 | 50 | 1.79 | Blood | Chloroform | Indian | C002067_A | | | Male | 100.01 | | | | |
| 22 | C001823_A | C001823 | Homosapien | F | 105 | 50 | 1.85 | Blood | Chloroform | Indian | C002022_A | | | Female | 144.29 | | | | |
| 23 | C001822_A | C001822 | Homosapien | M | 105 | 50 | 1.87 | Blood | Chloroform | Indian | C002023_A | | | Male | 106.68 | | | | |
| 24 | C001869_A | C001869 | Homosapien | M | 105 | 50 | 1.85 | Blood | Chloroform | Indian | | | | Male | 92.3 | | | | |
| 25 | C001874_A | C001874 | Homosapien | F | 105 | 50 | 1.78 | Blood | Chloroform | Indian | | | | Female | 87.05 | | | | |
| 26 | C001875_A | C001875 | Homosapien | M | 105 | 50 | 1.76 | Blood | Chloroform | Indian | | | | Male | 109.94 | | | | |
| 27 | C001818_A | C001818 | Homosapien | M | 105 | 50 | 1.86 | Blood | Chloroform | Indian | C002042_A | | | Male | 101.99 | | | | |
| 28 | C001896_A | C001896 | Homosapien | M | 105 | 50 | 1.88 | DNA | Chloroform | Indian | C002087_A | | | Male | 55.45 | | | | |
| 29 | C001882_A | C001882 | Homosapien | M | 105 | 50 | 1.77 | Blood | Chloroform | Indian | C002063_A | | | Male | 90.17 | | | | |

The three events are:



Subject Matrix for WGS1.0

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



Manage All Event Definitions in Study WGS1.0

Found 3 Definitions

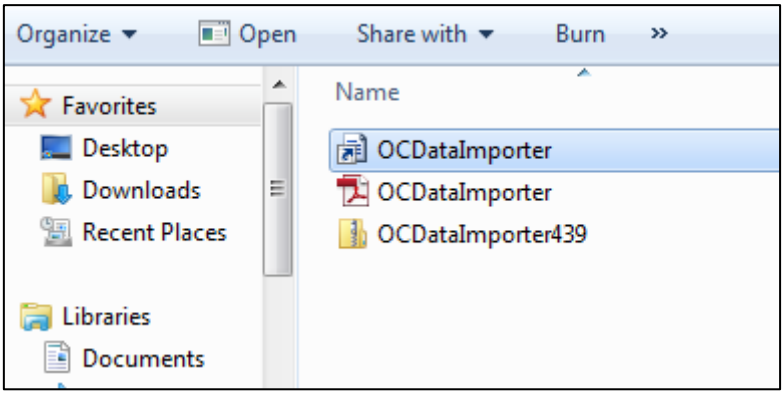
| Order | Name | OID | Repeating | Type |
|---|---|---|---|---|
| | Illumina Manifest Upload | SE_ILLUMINAMANIFESTUPLOAD_8272 | Yes | Common |
| | Illumina QC | SE_ILLUMINAQC_8933 | Yes | Common |
| | Sequencing files check | SE_SEQUENCINGFILESCHECK_2213 | Yes | Common |

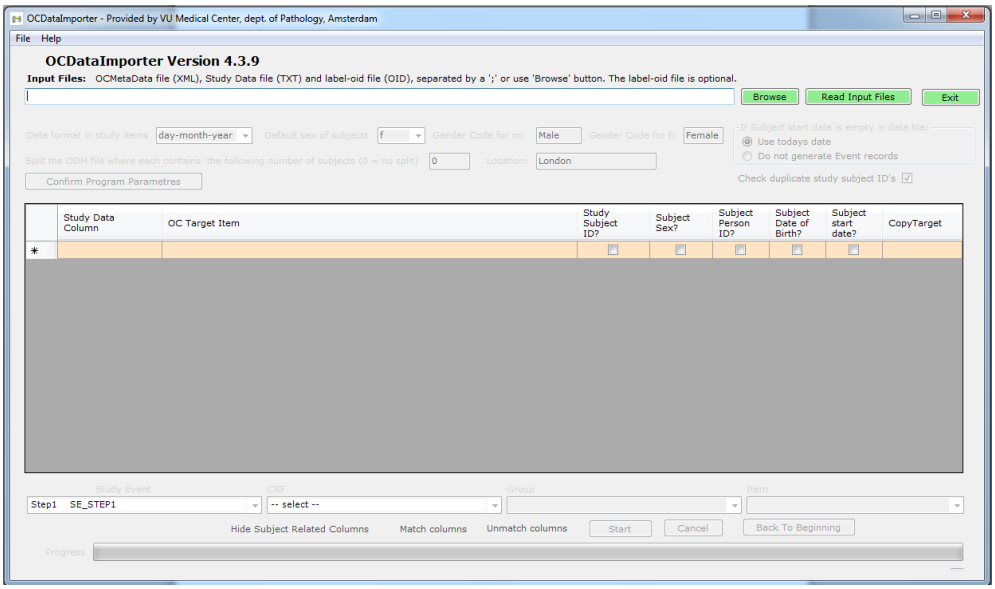
And the three associated CRFs are:

| 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

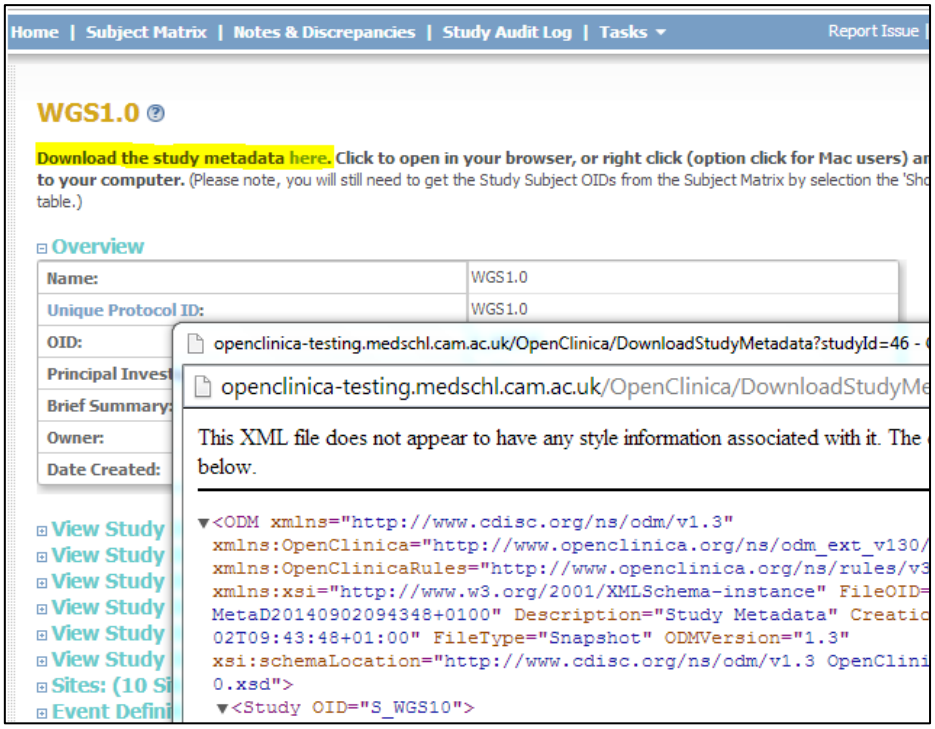
Open OCDataImporter from the shortcut created when you downloaded OCDataImporter (See above Section 2 for instructions to download)



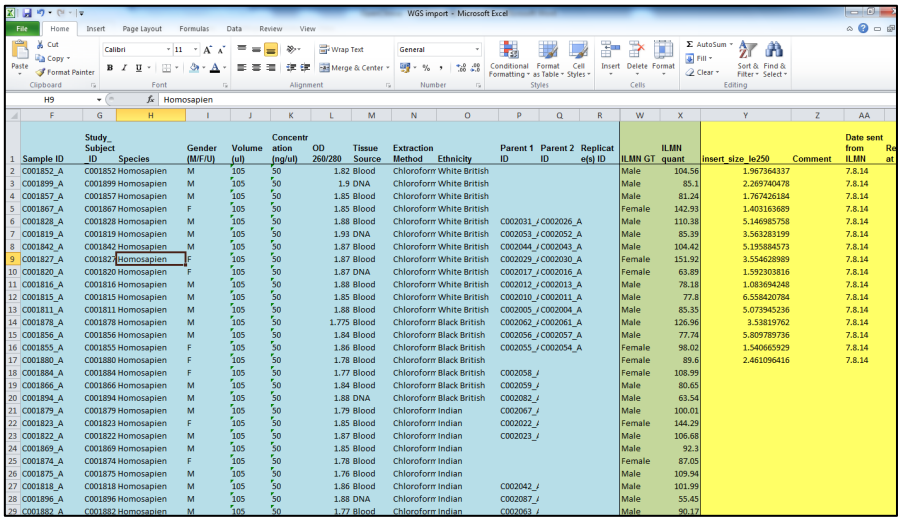
## 4.2 Download your study metadata file

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.



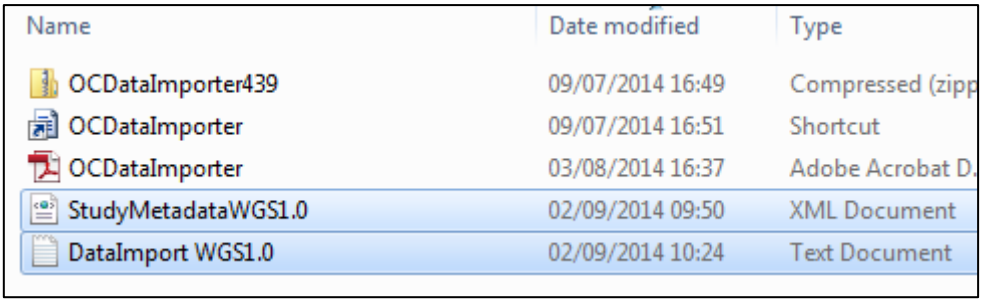

2.) Right mouse click the XML file and do a 'Save As' to the folder …/Desktop/OC Import. You may want to rename the file to reflect your study (here it is called StudyMetadataWGS1.0)

3.) Save your data file as a tab delimited txt file to the same location.



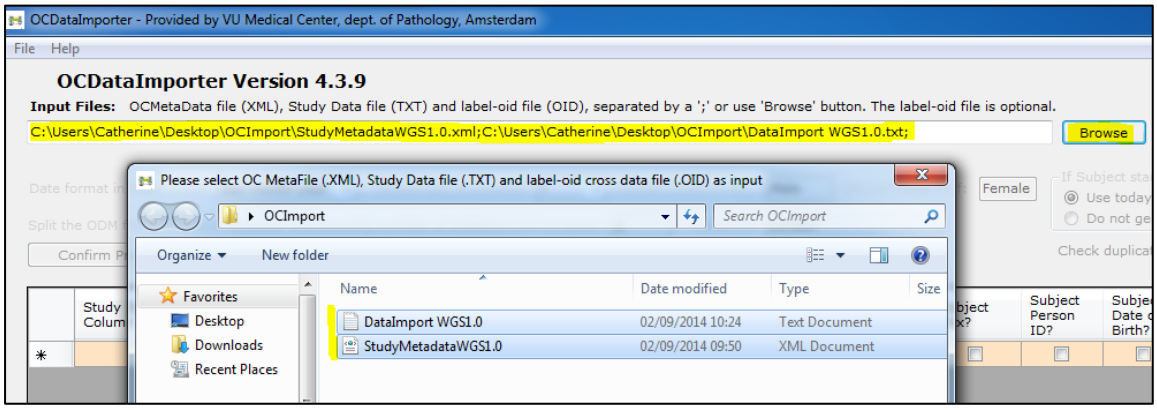
| Sample ID | Study_ Subject ID | Species | Gender (M/F/U) | Volume (ul) | Concentration (ng/ul) | OD 260/280 | Tissue Source | Extraction Method | Ethnicity | Parent 1 ID | Parent 2 ID | Replicate(s) ID | ILMN GT | ILMN quant | insert_size_le250 | Comment | Date sent from ILMN |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C001852_A | C001852 | Homosapien | M | 105 | 50 | 1.82 | Blood | Chloroform | White British | | | | Male | 104.56 | 1.967364337 | | 7.8.14 |
| C001899_A | C001899 | Homosapien | M | 105 | 50 | 1.9 | DNA | Chloroform | White British | | | | Male | 85.1 | 2.269740478 | | 7.8.14 |
| C001857_A | C001857 | Homosapien | M | 105 | 50 | 1.85 | Blood | Chloroform | White British | | | | Male | 81.24 | 1.767426184 | | 7.8.14 |
| C001867_A | C001867 | Homosapien | F | 105 | 50 | 1.85 | Blood | Chloroform | White British | | | | Female | 142.93 | 1.403163689 | | 7.8.14 |
| C001828_A | C001828 | Homosapien | M | 105 | 50 | 1.88 | Blood | Chloroform | White British | C002031_A | C002026_A | | Male | 110.38 | 5.146985758 | | 7.8.14 |
| C001819_A | C001819 | Homosapien | M | 105 | 50 | 1.93 | DNA | Chloroform | White British | C002053_A | C002052_A | | Male | 85.39 | 3.563283199 | | 7.8.14 |
| C001842_A | C001842 | Homosapien | M | 105 | 50 | 1.87 | Blood | Chloroform | White British | C002044_A | C002043_A | | Male | 104.42 | 5.195884573 | | 7.8.14 |
| C001827_A | C001827 | Homosapien | F | 105 | 50 | 1.87 | Blood | Chloroform | White British | C002029_A | C002030_A | | Female | 151.92 | 3.554628989 | | 7.8.14 |
| C001820_A | C001820 | Homosapien | F | 105 | 50 | 1.87 | DNA | Chloroform | White British | C002017_A | C002016_A | | Female | 63.89 | 1.592303816 | | 7.8.14 |
| C001816_A | C001816 | Homosapien | M | 105 | 50 | 1.88 | Blood | Chloroform | White British | C002012_A | C002013_A | | Male | 78.18 | 1.083694248 | | 7.8.14 |
| C001815_A | C001815 | Homosapien | M | 105 | 50 | 1.85 | Blood | Chloroform | White British | C002010_A | C002011_A | | Male | 77.8 | 6.558420784 | | 7.8.14 |
| C001811_A | C001811 | Homosapien | M | 105 | 50 | 1.88 | Blood | Chloroform | White British | C002005_A | C002004_A | | Male | 85.35 | 5.073945236 | | 7.8.14 |
| C001878_A | C001878 | Homosapien | M | 105 | 50 | 1.775 | Blood | Chloroform | Black British | C002062_A | C002061_A | | Male | 126.96 | 3.53819762 | | 7.8.14 |
| C001856_A | C001856 | Homosapien | M | 105 | 50 | 1.84 | Blood | Chloroform | Black British | C002056_A | C002057_A | | Male | 77.74 | 5.809789736 | | 7.8.14 |
| C001855_A | C001855 | Homosapien | F | 105 | 50 | 1.86 | Blood | Chloroform | Black British | C002055_A | C002054_A | | Female | 98.02 | 1.540665929 | | 7.8.14 |
| C001880_A | C001880 | Homosapien | F | 105 | 50 | 1.78 | Blood | Chloroform | Black British | | | | Female | 89.6 | 2.461096416 | | 7.8.14 |
| C001884_A | C001884 | Homosapien | F | 105 | 50 | 1.77 | Blood | Chloroform | Black British | C002058_A | | | Female | 108.99 | | | |
| C001866_A | C001866 | Homosapien | M | 105 | 50 | 1.84 | Blood | Chloroform | Black British | C002059_A | | | Male | 80.65 | | | |
| C001894_A | C001894 | Homosapien | M | 105 | 50 | 1.88 | DNA | Chloroform | Black British | C002082_A | | | Male | 63.54 | | | |
| C001879_A | C001879 | Homosapien | M | 105 | 50 | 1.79 | Blood | Chloroform | Indian | C002067_A | | | Male | 100.01 | | | |
| C001823_A | C001823 | Homosapien | F | 105 | 50 | 1.85 | Blood | Chloroform | Indian | C002022_A | | | Female | 144.29 | | | |
| C001822_A | C001822 | Homosapien | M | 105 | 50 | 1.87 | Blood | Chloroform | Indian | C002023_A | | | Male | 106.68 | | | |
| C001869_A | C001869 | Homosapien | M | 105 | 50 | 1.85 | Blood | Chloroform | Indian | | | | Male | 92.3 | | | |
| C001874_A | C001874 | Homosapien | F | 105 | 50 | 1.78 | Blood | Chloroform | Indian | | | | Female | 87.05 | | | |
| C001875_A | C001875 | Homosapien | M | 105 | 50 | 1.76 | Blood | Chloroform | Indian | | | | Male | 109.94 | | | |
| C001818_A | C001818 | Homosapien | M | 105 | 50 | 1.86 | Blood | Chloroform | Indian | C002042_A | | | Male | 101.99 | | | |
| C001896_A | C001896 | Homosapien | M | 105 | 50 | 1.88 | DNA | Chloroform | Indian | C002087_A | | | Male | 55.45 | | | |
| C001882_A | C001882 | Homosapien | M | 105 | 50 | 1.77 | Blood | Chloroform | Indian | C002063_A | | | Male | 90.17 | | | |

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




## 4.3 Match you study data and study metadata

1.) Click Browse and select both files

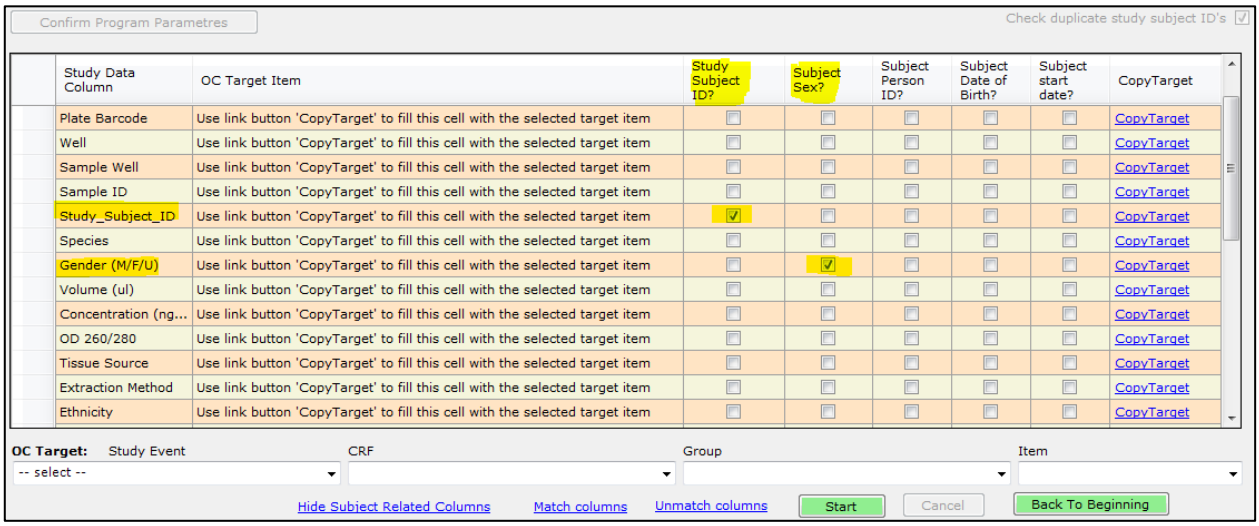



2.) Click 'Read Input files'

3.) Set the program parameters for date format, gender (OpenClinica expects lowercase **m** and **f**), subject start date, duplicate subject IDs, location.

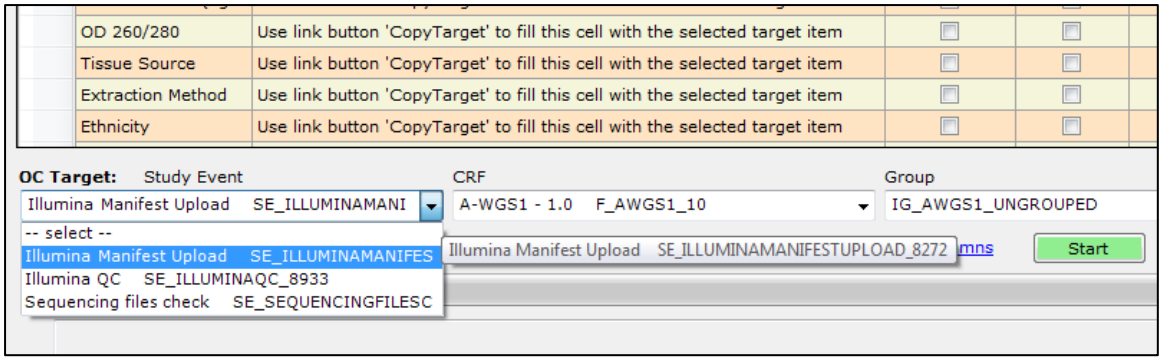
4.) Leave ODM file split as '0'

5.) Click 'Confirm Program Parameters'



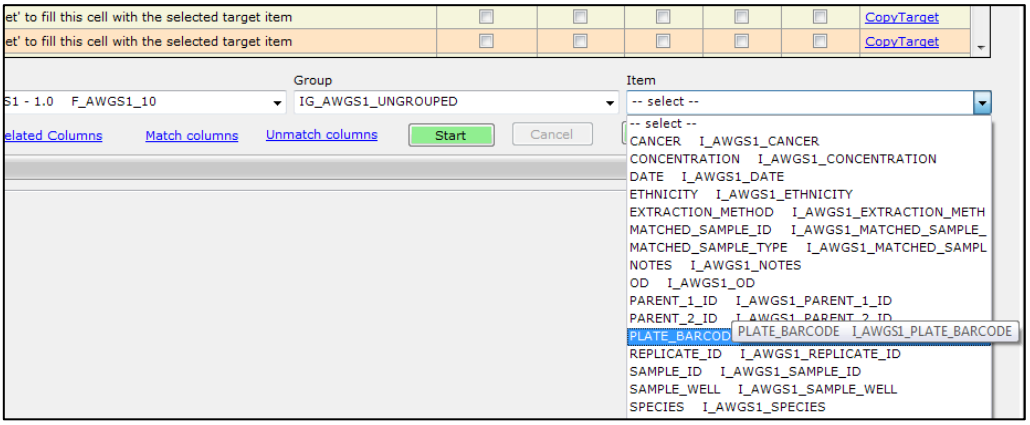

6.) Your excel data columns will be loaded into the first column 'StudyData Column'. OCdataimporter will attempt to match again the core subject fields as show above.

7.) Select the **OC targe: Study Event** (in the example below there are three study events)

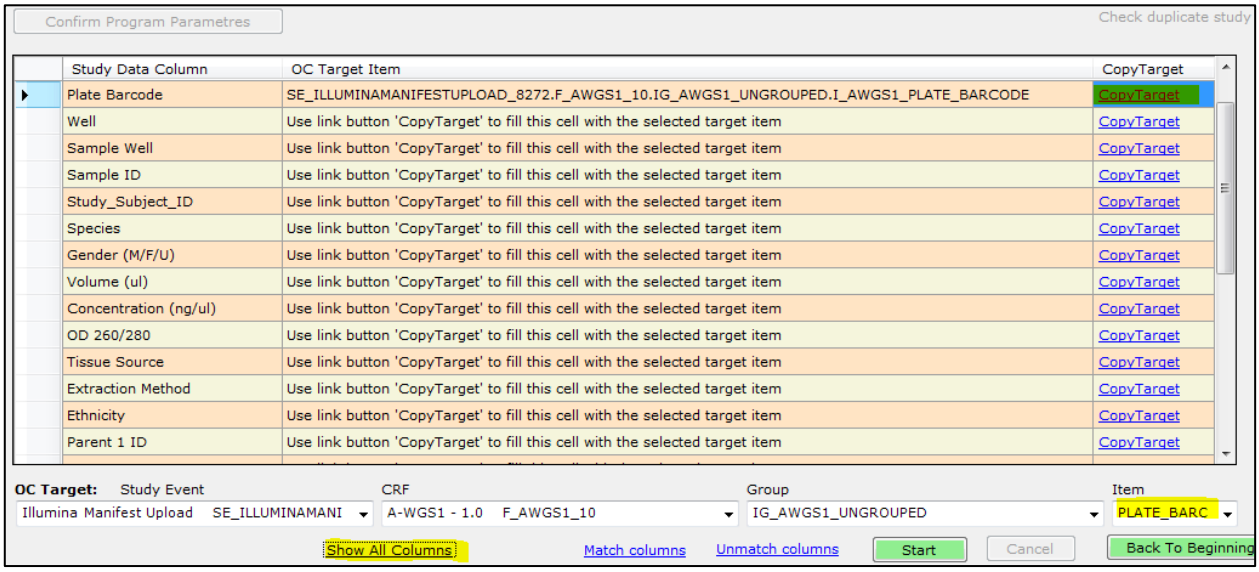



8.) Select the correct CRF and Group (if groups exist in your CRF)

9.) The last drop down will then present you with all the fields present in the CRF



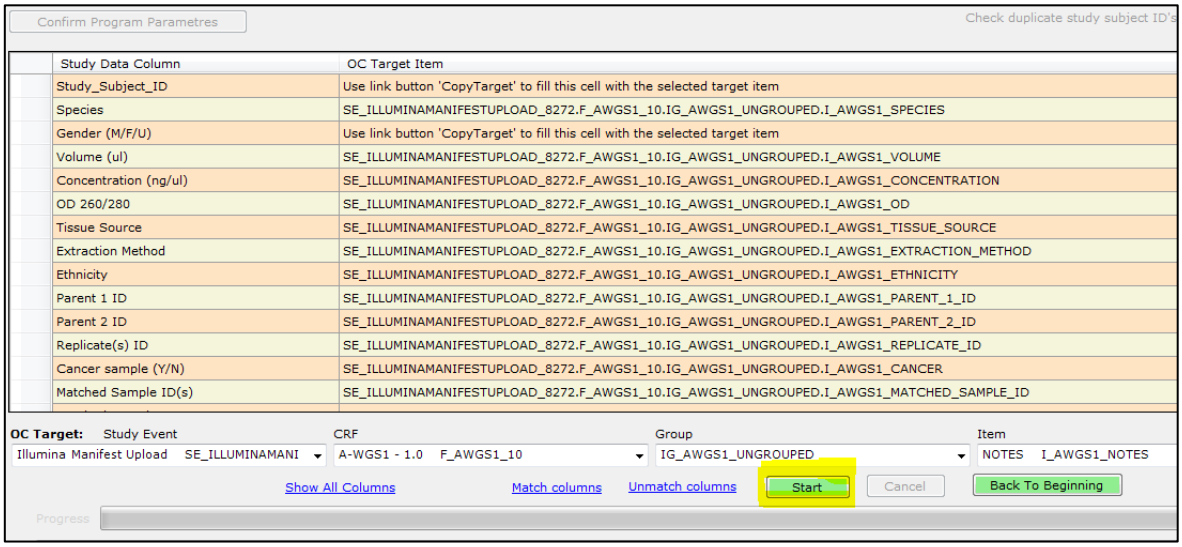

10.) Select the appropriate CRF field and then click the 'CopyTarget' link against the related row. (Note: to clear the screen you can 'show/hide the subject related fields' by clicking the link at the bottom of the screen)



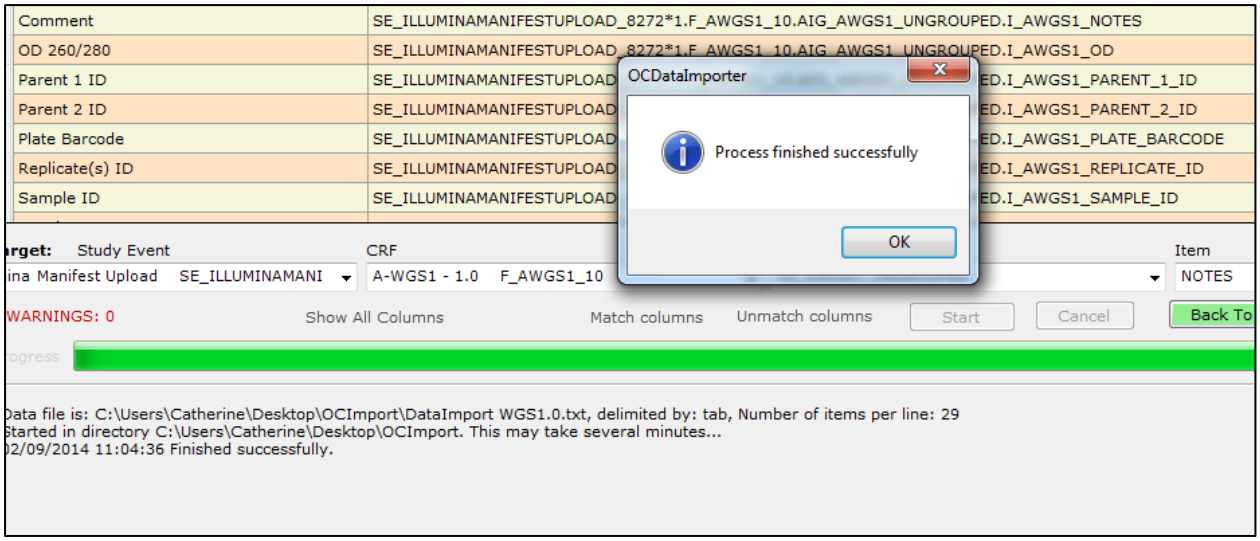

11.) Continue matching fields until all dropdown CRF items have been associated with a row in the main part of the screen.

## 4.4 Generate the XML CRF data import file

1.) Click 'Start'

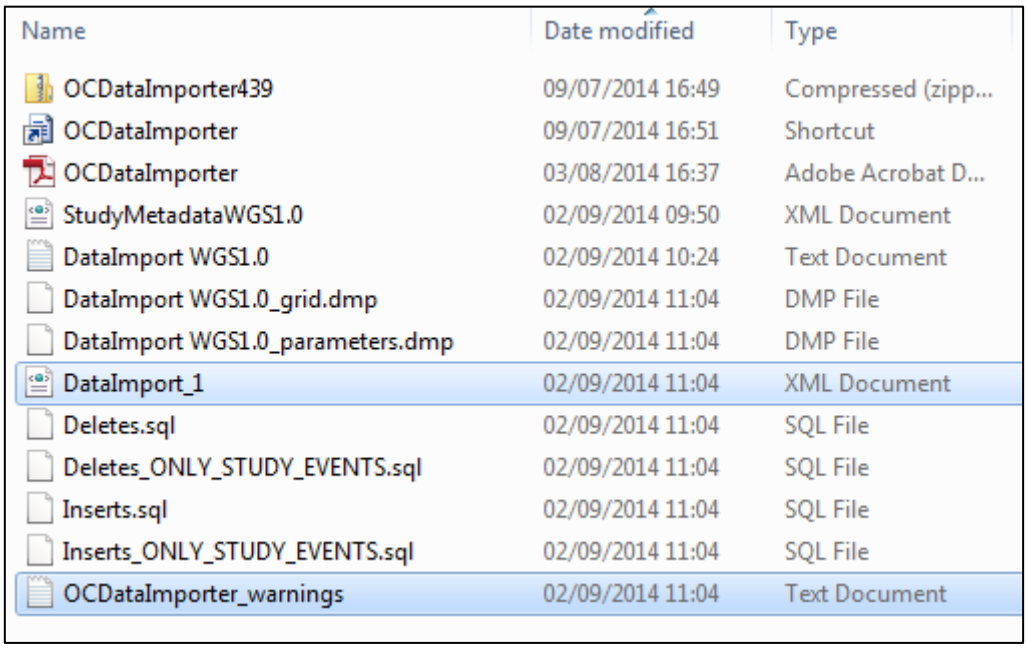



2.) If all records were successfully processed the below screen will show. If not review the errors which will be displayed in the lower pane as well as in the file OCdataImporter_warnings in …/Desktop/OC Import.



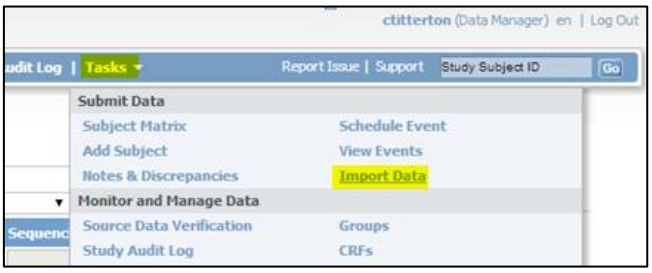

3.) OCdataimporter generates a set of files and places them in the …/Desktop/OC Import folder. The SQL files can be ignored (this is an alternate method for importing data for which there are detailed notes in the original OCDataImporter.pdf help file found in the folder below ).

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.

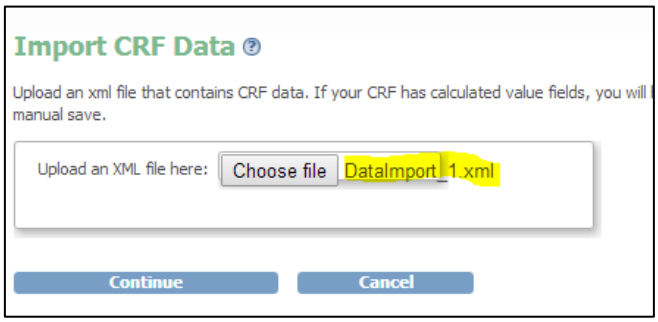



## 4.5 Import your CRF data XML file into OpenClinica

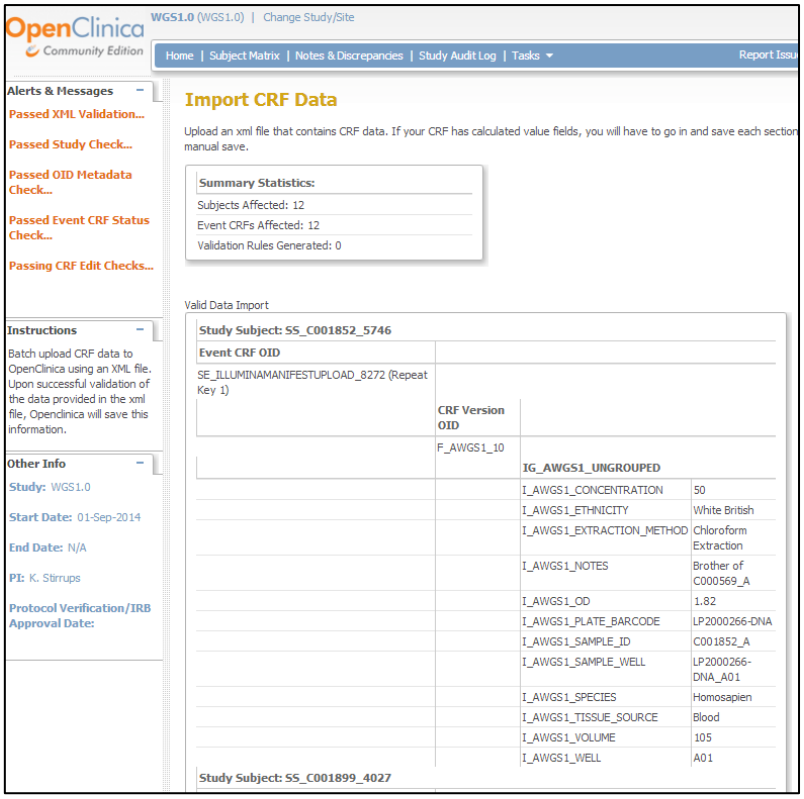
1.) In OpenClinica click Tasks > Import Data




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

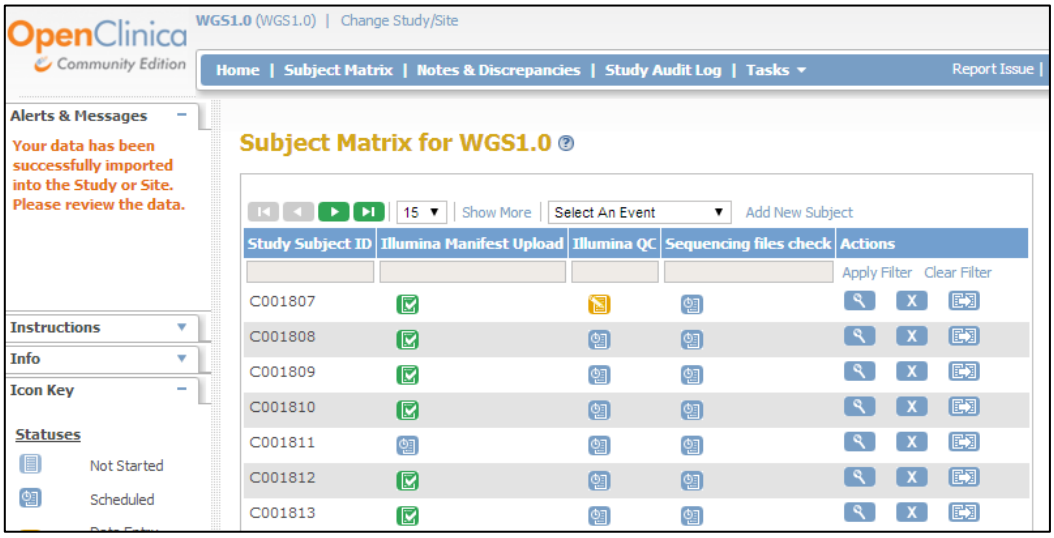



3.) OpenClinica checks XML validation, study checks,OID metada, event CRF status and CRF edit checks and present the following screen (or a screen showing errors and which points it failed on)




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:

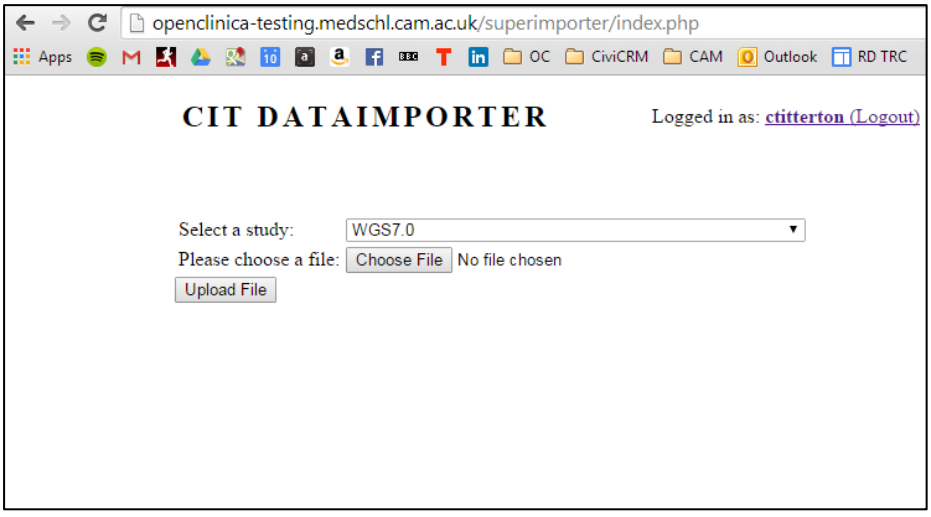
- Subject creation
- Event scheduling
- Study data import

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

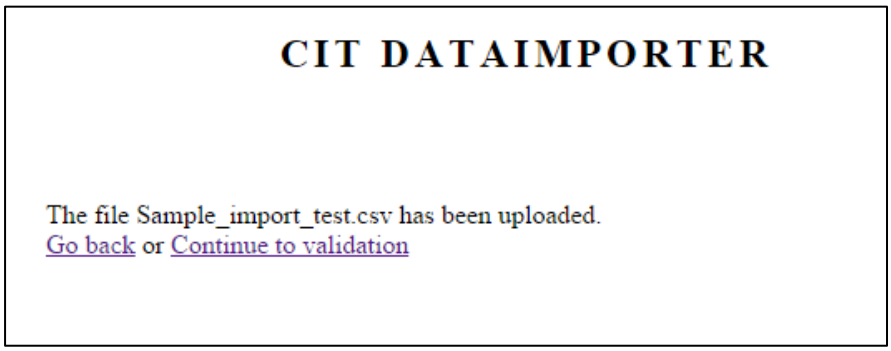
1. Save your data file as [studyname].csv
2. Columns A to F must contain the column headings *SubjectID*, *SecondaryID, EnrollmentDate, PersonID, Gender* and *DOB* as shown below. The import is case sensitive so ensure column headings are exactly as show.
3. If you study is not collecting date of birth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

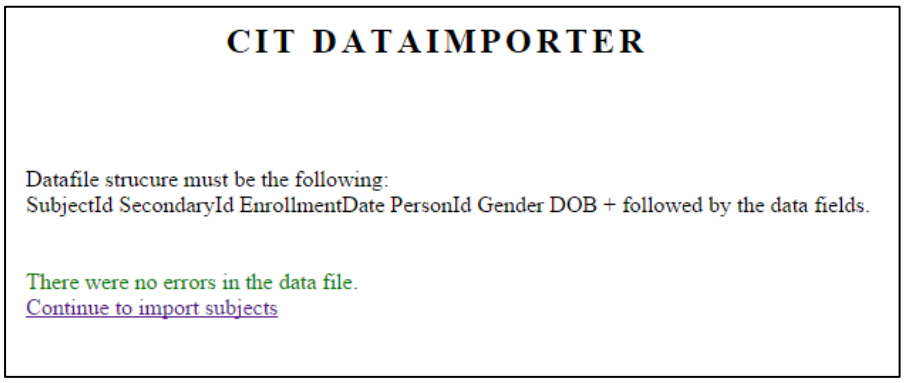
Open the CIT data importer at the url provided to you by the CIT team e.g. [http://openclinica-testing.medschl.cam.ac.uk/superimporter/index.php](http://openclinica-testing.medschl.cam.ac.uk/superimporter/index.php) . You will be presented with the screen shown below



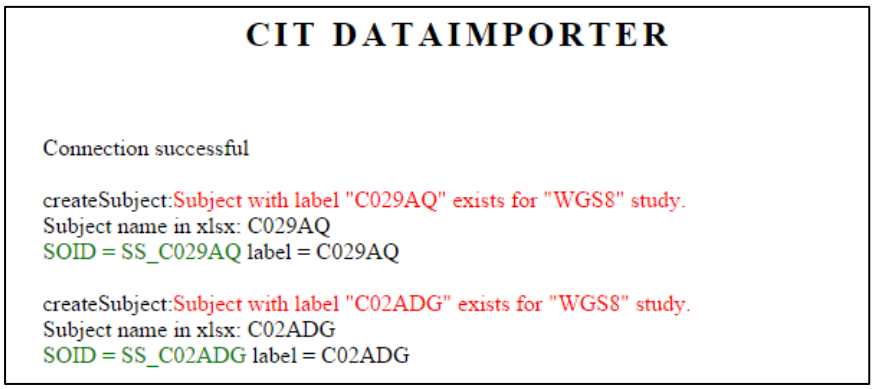
4. Select the study you are importing to from the dropdown list.
5. Click"Choose file" and navigate to the file [studyname].csv
6. Click "Upload file"




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




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



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

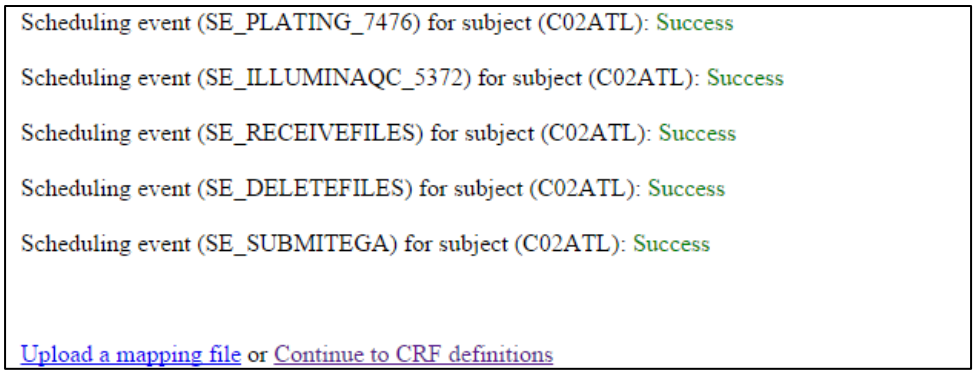
Continue to scheduling

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)

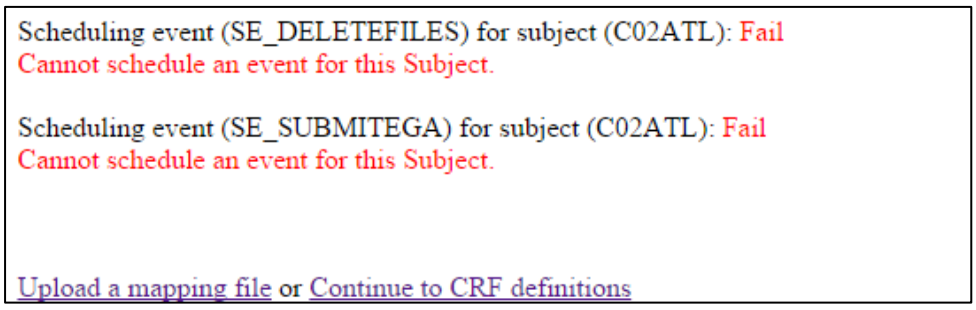
| Event name | Schedule? | Date | Time | Forms |
|---|---|---|---|---|
| Receipt | ☑ | 2015-01-26 | 14:11 | • A_WGS1 Sample Receipt - 1.2<br>• A_WGS1 Sample Receipt - 1.1<br>• A_WGS1 Sample Receipt - 1.0 |
| CATGO QC | ☑ | 2015-01-26 | 14:11 | • A_WGS2 CATGO QC - 1.2<br>• A_WGS2 CATGO QC - 1.1<br>• A_WGS2 CATGO QC - 1.0 |
| GEL OK | ☑ | 2015-01-26 | 14:11 | • A_WGS2 GEL PASS - 1.0 |
| Plating | ☑ | 2015-01-26 | 14:11 | • A_WGS3 Plating - 1.4<br>• A_WGS3 Plating - 1.3<br>• A_WGS3 Plating - 1.2 |
| Illumina QC | ☑ | 2015-01-26 | 14:11 | • A_WGS4 Illumina QC - 1.2<br>• A_WGS4 Illumina QC - 1.1 |
| Receive Files | ☑ | 2015-01-26 | 14:11 | • A_WGS5 Files Receipt - 1.2<br>• A_WGS5 Files Receipt - 1.1 |
| Delete Files | ☑ | 2015-01-26 | 14:11 | • A_WGS6 Files Destroy - 1.0 |
| Submit EGA | ☑ | 2015-01-26 | 14:11 | • A_WGS8 EGA - 1.0 |

Schedule selected events!

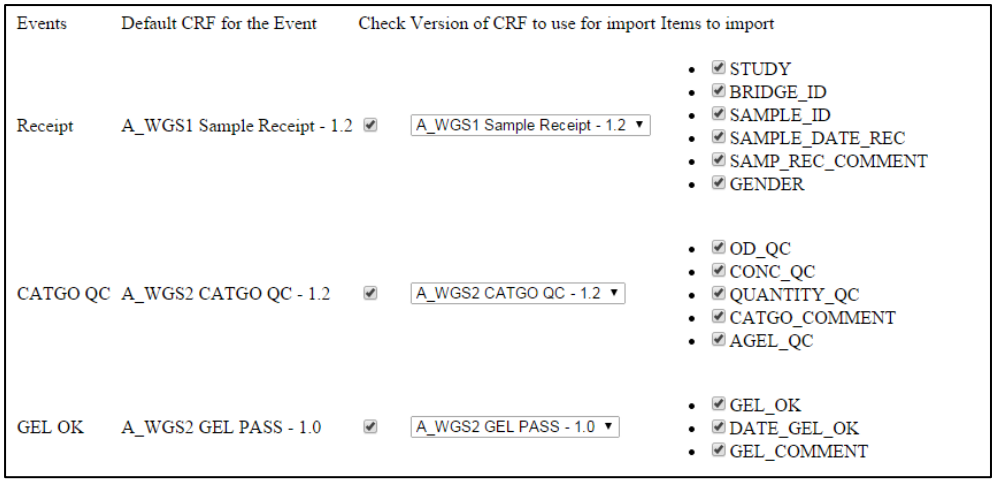
12. Click “Schedule selected events”. The next screen will show a line for every event per subject as scheduled with “success”. If the event has already been scheduled for a particular subject, a warning will show. (You cannot reschedule dates within CIT data importer, if they have already been scheduled.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



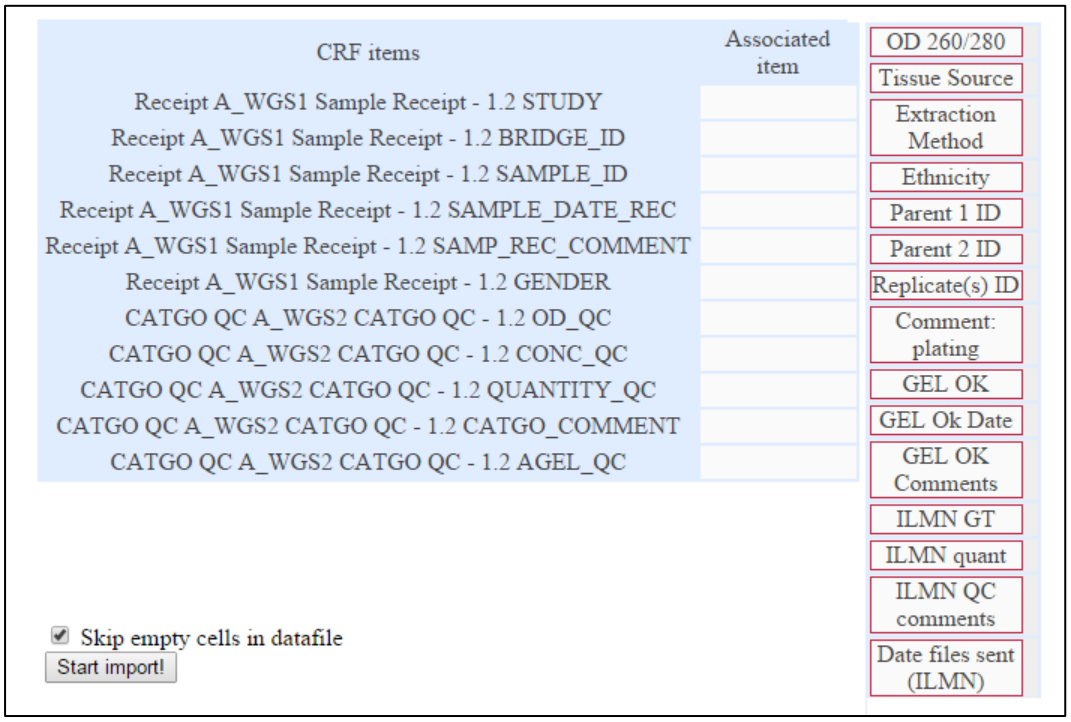

If the event has already been scheduled for a particular subject, a warning will show. (You cannot reschedule dates within CIT data importer, if they have already been scheduled. You can manually reschedule events through the interface, but this needs to be done on an event / by subject record, prior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




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




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




16. The field list on the left in blue dispalys 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 form the csv import file against the CRF fields.

| CRF items | Associated item |
|---|---|
| Receipt A_WGS1 Sample Receipt - 1.2 STUDY | Study |
| Receipt A_WGS1 Sample Receipt - 1.2 BRIDGE_ID | |
| Receipt A_WGS1 Sample Receipt - 1.2 SAMPLE_ID | Sample_ID |
| Receipt A_WGS1 Sample Receipt - 1.2 SAMPLE_DATE_REC | |
| Receipt A_WGS1 Sample Receipt - 1.2 SAMP_REC_COMMENT | |
| Receipt A_WGS1 Sample Receipt - 1.2 GENDER | Gender_MFU |
| CATGO QC A_WGS2 CATGO QC - 1.2 OD_QC | OD_QC |
| CATGO QC A_WGS2 CATGO QC - 1.2 CONC_QC | |
| CATGO QC A_WGS2 CATGO QC - 1.2 QUANTITY_QC | |
| CATGO QC A_WGS2 CATGO QC - 1.2 CATGO_COMMENT | |
| CATGO QC A_WGS2 CATGO QC - 1.2 AGEL_QC | |

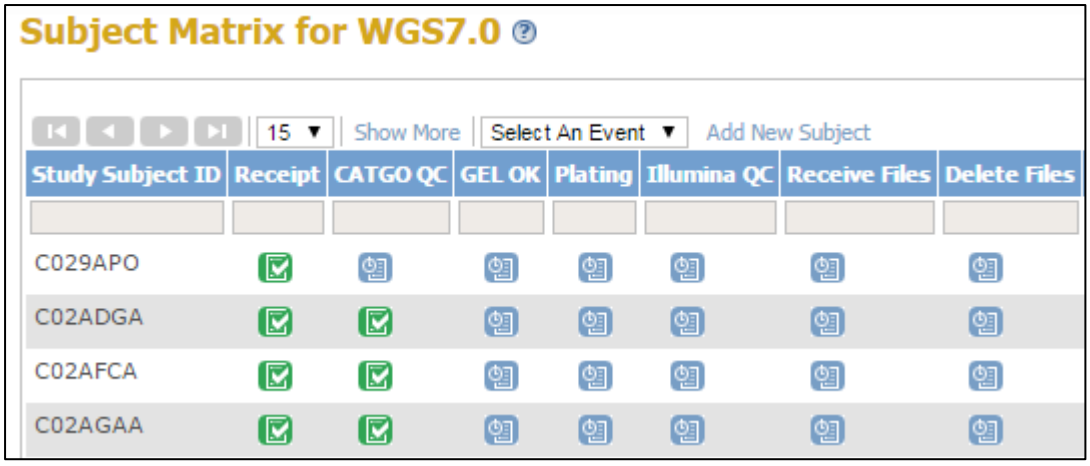
OD 260/280
Tissue Source
Extraction Method
Ethnicity
Parent 1 ID
Parent 2 ID
Replicate(s) ID
Comment: plating
GEL OK
GEL Ok Date
GEL OK Comments

17. 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 particulat event, tick the bottom box "Skip empty cells in data file"
18. 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

19. 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.
20. Check if you records have imported as expected through the subject matrix in OpenClinica



# 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

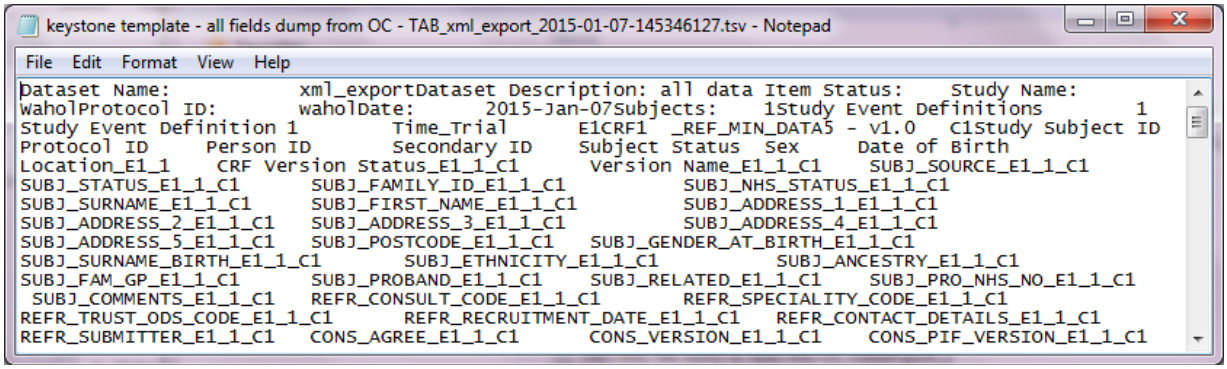
| File | file type | file name | Used for |
|---|---|---|---|
| Study Metadata file | **xml** | DownloadStudyMetadata.xml | Provides an output of all study metadata |
| Keystone template | **tsv** | TAB_*exportname*_2015-01-21-114325879.tsv | Extracts |
| CRF | **xls** | as named | Needed to generate OpenClinica input screens and can be used for dummy data generation |
| Mapping Definitions file | **xls** | matching and substitutions between sample and study XML - wahol_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

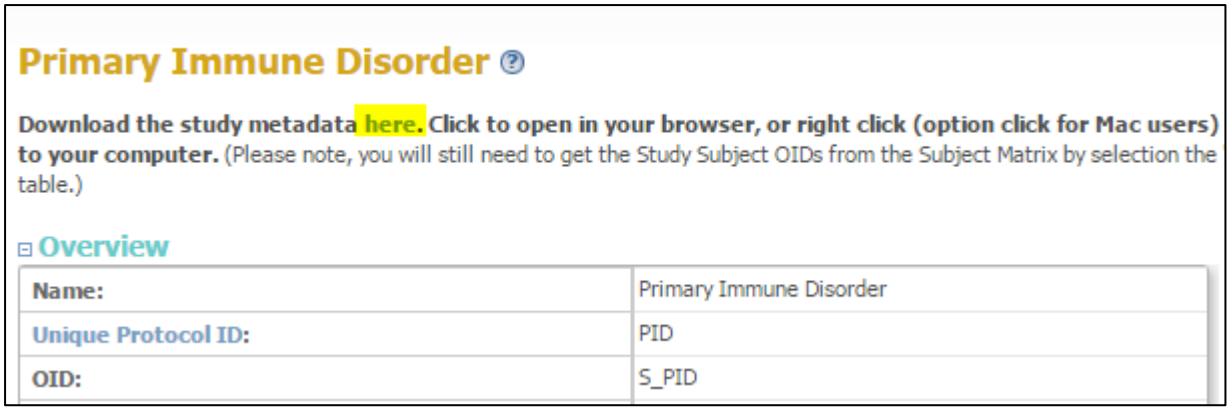
1. In OpenClinica create your study, setting all study parameters according to you study protocols.
2. Create CRFs and events against expected study data (and GEL min data set).

   (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'.

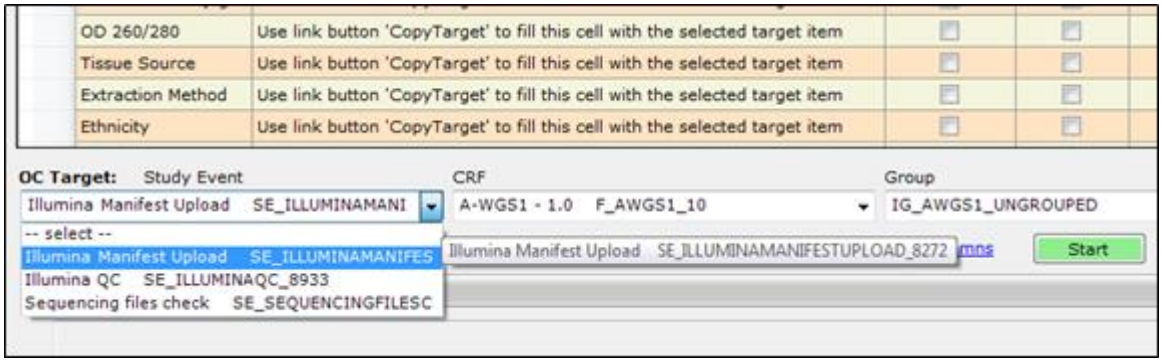



The end result you should have two files to be used in Excel to create the Mapping Definitions file:

- 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="2015-01-0 | | |
| 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.

4. Select ‘delimited’ > click next > tick Other on delimiters options and add a ~ to seperate data. Click next and finish.

Data on *GRID sheet* will look like below:

E6 TRUE

| | 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 |

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

| | 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_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_POSTCOI |
| 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_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_ |

**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

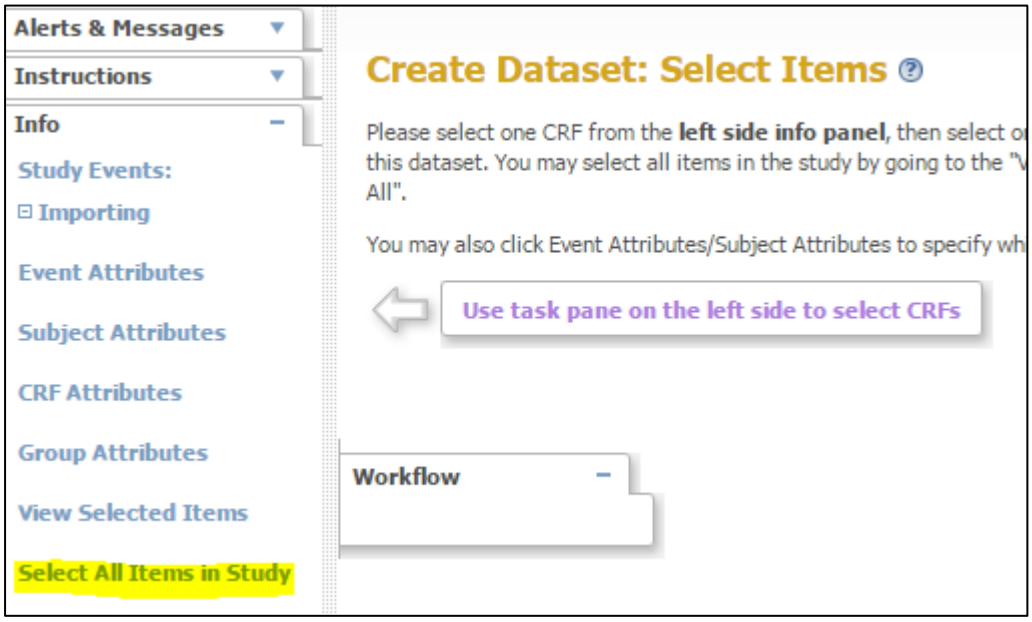
=IF((ISERROR(K13)),CONCATENATE(A13,CONCATENATE("~",C13),"~",D13,"~",E13,"~",F13,"~",G13,"~",H13,"~",I13,"~",J13),CONCATENATE(A13,CONCATENATE("~",K13),"~",D13,"~",E13,"~",F13,"~",G13,"~",H13,"~",I13,"~",J13))

### 6.3.3 Generate XML data import file

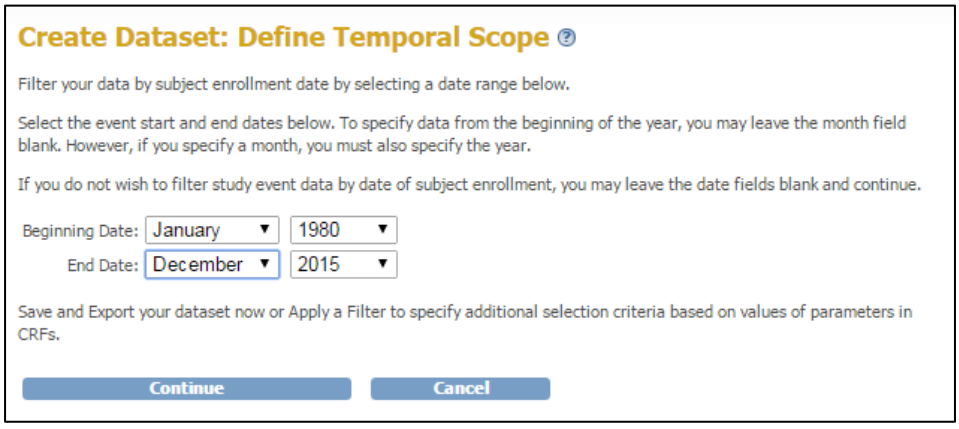
Having generated your ODI Grid.dmp file you can now run through the standard OCDataImprter steps outlined in section 4.3 to generate the study data XML file for import into OpenClinica.

# 7 Data Export from OpenClinica

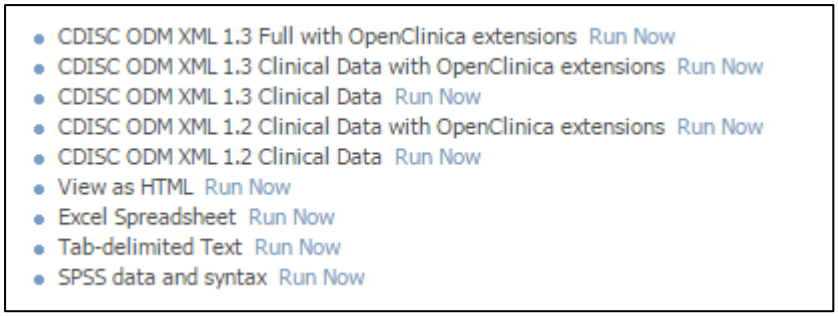
1. In OpenClinica go to Tasks > Extract Data > Create dataset > click “select all items in study”




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




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




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)

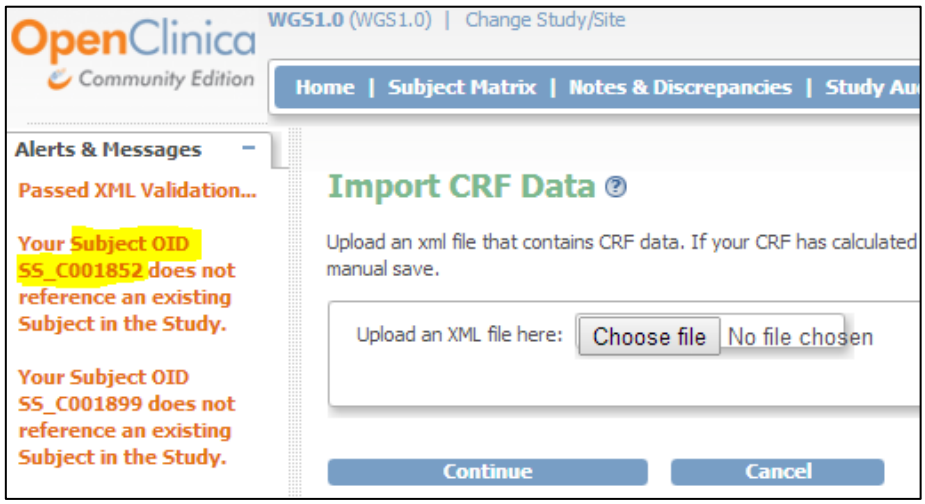
a. Open a copy of the original CRF used to loaded into the study in step 2 above

b. Import into a new sheet in the CRF the file Keystone template TAB_[exportname]_2015-01-21-114325879.tsv

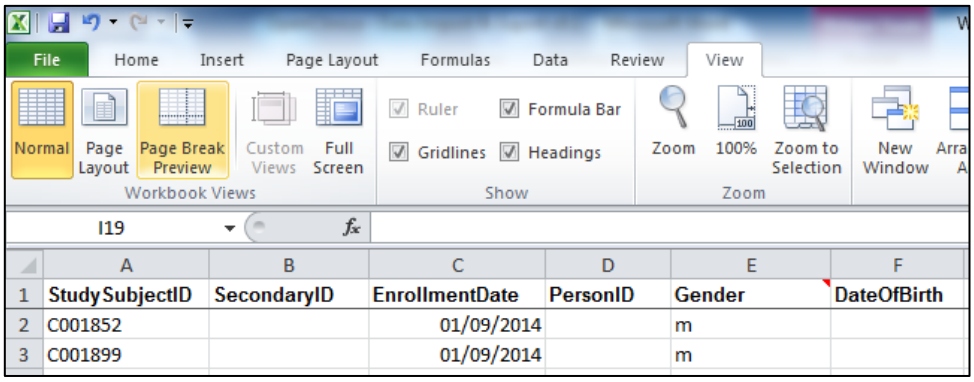
# 10 Common Import Errors

## 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 you 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**

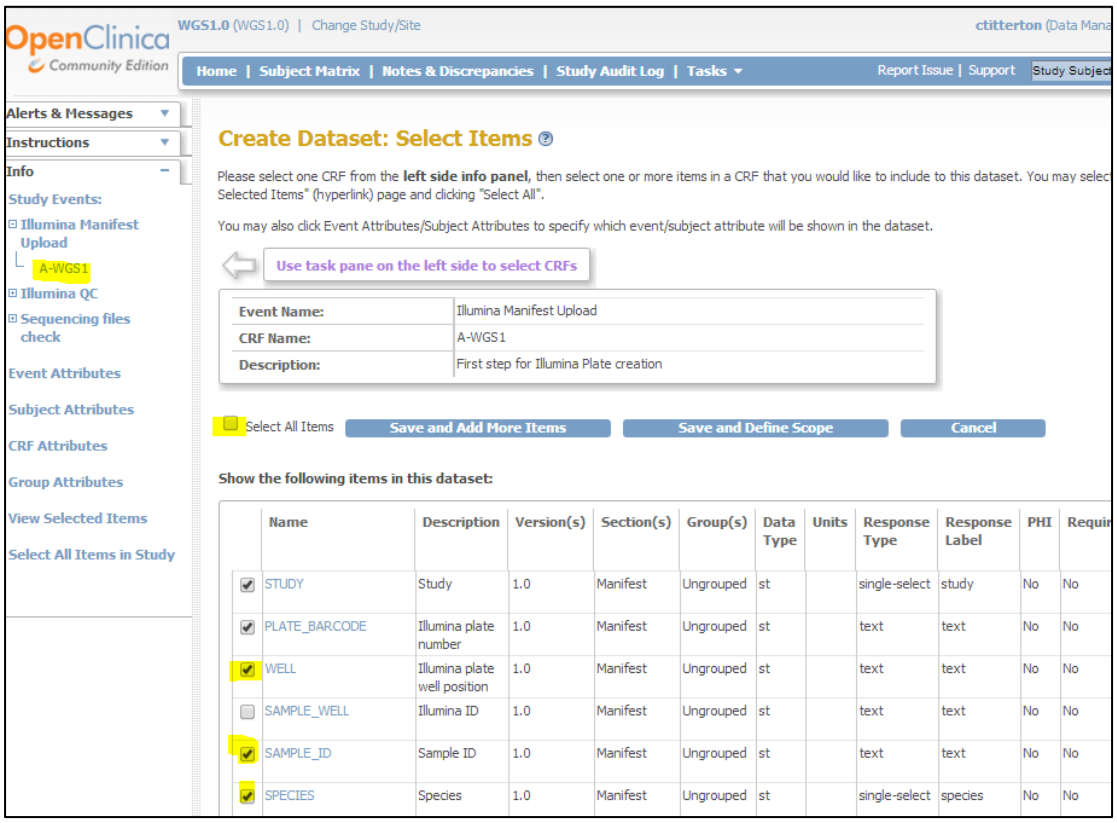



| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | StudySubjectID | SecondaryID | EnrollmentDate | PersonID | Gender | DateOfBirth |
| 2 | C001852 | | 01/09/2014 | | m | |
| 3 | C001899 | | 01/09/2014 | | m | |

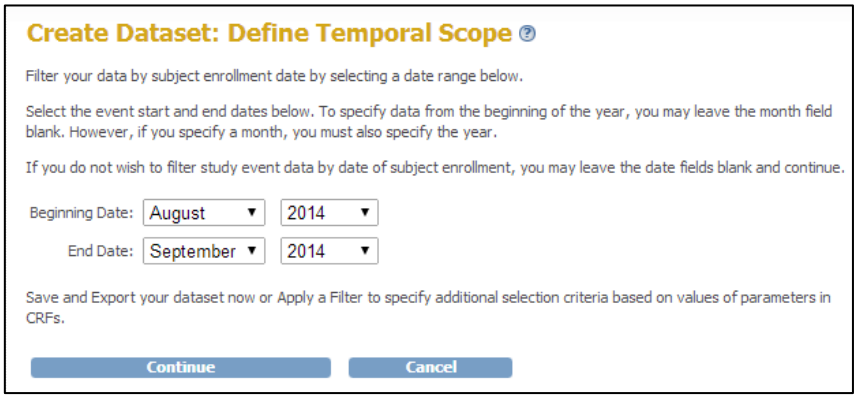
In your tab delimited txt file ensure your **Study Subject ID** includes any suffixes (the prefix will automatically beincluded)

# 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.

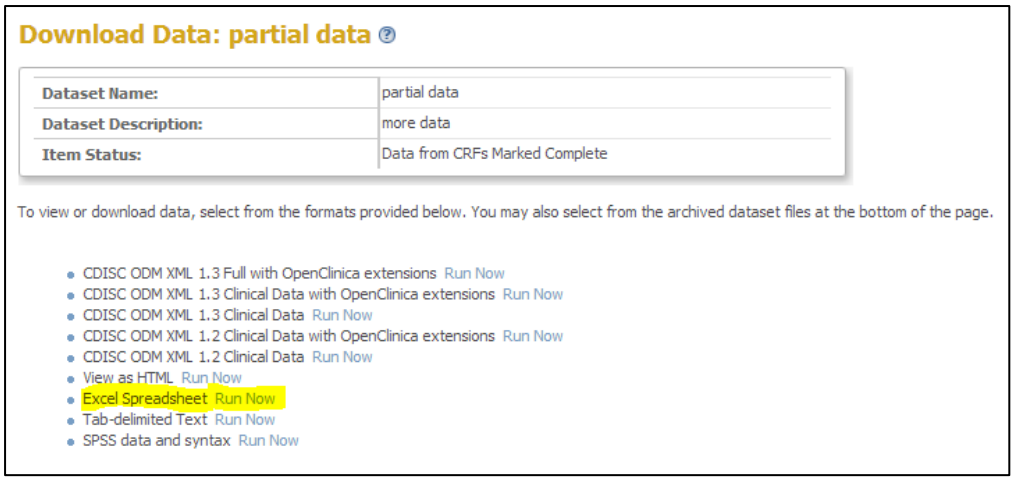



4.) Click 'Save and Add more Items' to navigate to another CRF, Event etc
5.) Click 'Save and Define Scope' to move to next screen to set you dates you wish to export against

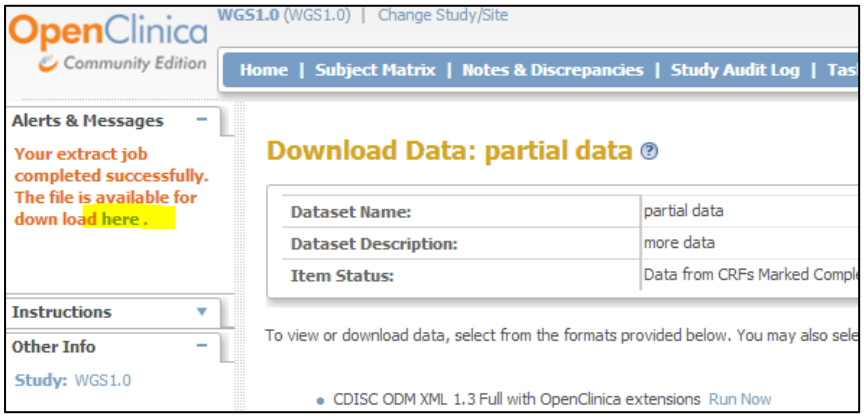



6.) Add the export set name, description

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)




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.




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