

Importing Data into OpenClinica3 (OC3)

Data importing is an alternative method of data entry in OpenClinica3 (OC3). This document provides guidance on how to import study data into OC3.

1. Precursor

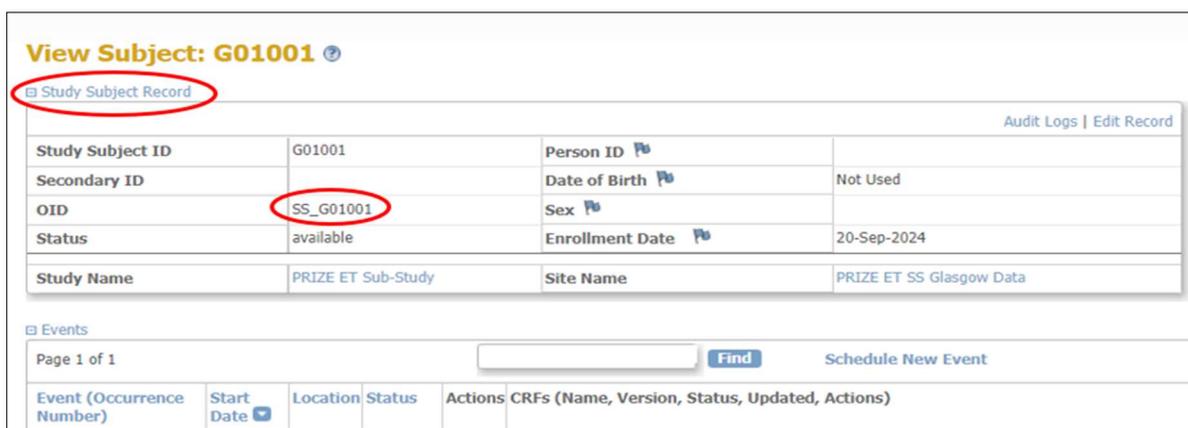
Complete FRM071 – Importing data into OpenClinica: Prior to importing data into OC, the import request should be signed off by the Clinical Project Manager (CPM) using FRM071 on OpenClinica eForms.

2. Importing Data

See Appendix 1 for an instructional video on how to import data into OC.

Step 1 – Preparing the import file

- a. Make a copy of the original data extract before replacing the patient's SSIDs with SSOIDs, usually in an Excel sheet. You can download the SSOID listing from DataMart or Insight, then use VLOOKUP to match each SSOID with the patient SSIDs in the import file. Alternatively, SSOIDs can be obtained directly from OpenClinica (see the screenshot below). Ensure there are no extra spaces in the SSOID fields.



View Subject: G01001			
Study Subject Record			
Study Subject ID	G01001	Person ID	
Secondary ID		Date of Birth	Not Used
OID	SS_G01001	Sex	
Status	available	Enrollment Date	20-Sep-2024
Study Name	PRIZE ET Sub-Study	Site Name	PRIZE ET SS Glasgow Data

Event (Occurrence Number)	Start Date	Location	Status	Actions	CRFs (Name, Version, Status, Updated, Actions)
Page 1 of 1					
Find Schedule New Event					

- b. The import file also requires the ItemOID of each field, which can be obtained in different ways from OpenClinica. One option is to check the CRFs page in the Task tab. Alternatively, ItemOIDs can be derived from the OC3 study metadata (See screenshots below): *Tasks > View Study (or Build Study) > Create Event Definitions.*

	Task	Status	Count	Mark Complete	Actions
1	Create Study	In Progress	NA	<input type="checkbox"/>	 
2	Create CRF	In Progress	9	<input type="checkbox"/>	 
3	Create Event Definitions	In Progress	15	<input type="checkbox"/>	 
4	Create Subject Group Classes	Not Started	0	<input type="checkbox"/>	
5	Create Rules	In Progress	46	<input type="checkbox"/>	 
6	Create Sites	In Progress	7	<input type="checkbox"/>	 
7	Assign Users	In Progress	Total : 54	<input type="checkbox"/>	 

GD045 Importing Data into OpenClinica3 (OC3)

12 months post 2nd vaccine (PAP only)	SE_12MONTHSPOST2NDVACCINEPAPONL	No	Scheduled	Yes	02-Sep-2021 (tdevine)	02-Sep-2021 (tdevine)	  
Unscheduled visit (PAP only)	SE_UNSCHEDULEDVISITPAPONLY	Yes	Scheduled	Yes	16-Nov-2021 (tdevine)	18-May-2022 (tdevine)	  
End of study	SE_ENDOFSTUDY_5442	No	Scheduled	Yes	10-Nov-2020 (tdevine)	16-Nov-2021 (tdevine)	  
Admin (PAP only)	SE_ADMINPAPONLY	No	Scheduled	Yes	02-Sep-2021 (tdevine)	16-Nov-2021 (tdevine)	  
Vaccine data	SE_VACCINADATA	No	Scheduled	Yes	25-Aug-2022 (tdevine)	25-Aug-2022 (tdevine)	  

Order	Name	Required	Double Data Entry	Password Required	Default Version	Hide CRF	Participant Form	Allow Anonymous Submission	Submission URL	Offline	Null Values	Source Data Verification	Status	Actions
	HICC End of study	Yes	No	No	v1.0	No	No					Not Required	available	

View CRF Details

Name: HICC End of study
Description: v1.0
OID: F_HICCENDOFSTU

Version(s)

Version Name	oid	Description	Status	Revision Notes	Action
v1.0	F_HICCENDOFSTU_V10	v1.0	available	Thomas Devine 10/11/2020	  

Items

Name	Item_OID	Description	Data Type	Version(s)	Integrity Check
date_trial_completed	I_HICCE_DATE_TRIAL_COMPLETED	Date trial completed	date	v1.0	OK
other_reason	I_HICCE_OTHER_REASON	Please specify (Other)	character string	v1.0	OK
reason	I_HICCE_REASON	Reason trial completed	integer	v1.0	OK
withdrawal_consent_reason	I_HICCE_WITHDRAWAL_CONSENT_REASON	Please specify (Withdrawal of consent)	character string	v1.0	OK

Also, care should be taken to ensure the ItemOIDs are correctly matched to the columns. It is recommended to add the ItemOIDs underneath the initial headings provided. Once these have all been added and checked, the first row can be removed. Ensure there are no extra spaces in the Item OID names.

Note: This is a sample dataset, modified for illustration purposes.

SSID	reason	other_reason	date_trial_completed
SSOID	I_HICCE_REASON	I_HICCE_OTHER_REASON	I_HICCE_DATE_TRIAL_COMPLETED
SS_PAP1S	Other	Lost to follow-up	2022-07-19
SS_PAP9	Other	Lost to follow-up	2022-06-26
SS_PAP11S	Other	Lost to follow-up	2022-08-07
SS_PAP13S	Other	Lost to follow-up	2022-08-19
SS_PAP15	Completed Study		2022-07-28
SS_PAP17S	Completed Study		2022-08-07
SS_PAP2S	Screen failure		2022-08-07
SS_PAP21S	Other	Lost to follow-up	2022-08-19
SS_PAP22	Other	Lost to follow-up	2022-07-28

GD045 Importing Data into OpenClinica3 (OC3)

- c. For any radio/checkbox fields, the RESPONSE_VALUES need to be imported, as opposed to the RESPONSE_OPTIONS_TEXT. These can be found in the CRF Design spreadsheet or in the OpenClinica metadata. To find the Response Values in the OpenClinica metadata, follow the steps above to get to the following page and then select the icon <..>.

View CRF Details

Name:	HICC End of study
Description:	v1.0
OID:	F_HICCENDOFSTU

Version(s)

Version Name	oid	Description	Status	Revision Notes	Action
v1.0	F_HICCENDOFSTU_V10	v1.0	available	Thomas Devine 10/11/2020	  

Items

Name	Item_OID	Description	Data Type	Version(s)	Integrity Check
date_trial_completed	I_HICCE_DATE_TRIAL_COMPLETED	Date trial completed	date	v1.0	OK
other_reason	I_HICCE_OTHER_REASON	Please specify (Other)	character string	v1.0	OK
reason	I_HICCE_REASON	Reason trial completed	integer	v1.0	OK
withdrawal_consent_reason	I_HICCE_WITHDRAWAL_CONSENT_REASON	Please specify (Withdrawal of consent)	character string	v1.0	OK

The RESPONSE_VALUES can be found here

Section Name	Title	Subtitle	Instructions	Page Number Label
End of study	End of study		<style type="text/css">.aka_text_block{width:150px;text-align:left;}</style>	

Group Name	OID	Header	Repeat Number	Repeat Max	Is shown?	Group Layout
end	IG_HICCE_END		1	1	Yes	Non-repeating

Name	Item_OID	GROUP_OID	Description	Group Name	Units	PHI Status	Item Data Type	Left Item Text	Right Item Text	Default Value	Response Layout	Response Label	Response Options	Response Values	Required	Is shown?
reason	I_HICCE_REASON	IG_HICCE_END	Reason trial completed	end		false	integer	Reason trial completed				Reason	(select), Completed study, Screen failure, Withdrawal of consent, Death, Other	, 1, 2, 3, 4, 5	Yes	Yes

SSID	reason	other_reason	date_trial_completed
SSOID	I_HICCE_REASON	I_HICCE_OTHER_REASON	I_HICCE_DATE_TRIAL_COMPLETED
SS_PAP1S	5	Lost to follow-up	2022-07-19
SS_PAP9	5	Lost to follow-up	2022-06-26
SS_PAP11S	5	Lost to follow-up	2022-08-07
SS_PAP13S	5	Lost to follow-up	2022-08-19
SS_PAP15	1		2022-07-28
SS_PAP17S	1		2022-08-07
SS_PAP2S	2		2022-08-07
SS_PAP21S	5	Lost to follow-up	2022-08-19
SS_PAP22	5	Lost to follow-up	2022-07-28

- d. Date entries in the import file needs to be converted to ISO 8601 format: YYYY-MM-DD

SSID	reason	other_reason	date_trial_completed
SSOID	I_HICCE_REASON	I_HICCE_OTHER_REASON	I_HICCE_DATE_TRIAL_COMPLETED
SS_PAP1S	5	Lost to follow-up	2022-07-19
SS_PAP9	5	Lost to follow-up	2022-06-26
SS_PAP11S	5	Lost to follow-up	2022-08-07
SS_PAP13S	5	Lost to follow-up	2022-08-19
SS_PAP15	1		2022-07-28
SS_PAP17S	1		2022-08-07
SS_PAP2S	2		2022-08-07
SS_PAP21S	5	Lost to follow-up	2022-08-19
SS_PAP22	5	Lost to follow-up	2022-07-28

- e. If the field is limited to a specified number of decimal places on Openclinica, this format needs to be the same in the import file. Please note, if the number of decimal places does not match, and OpenClinica needs to be amended, this needs to be added as a new field. If the existing field is modified to allow the correct number of decimal places, during importing, OpenClinica will still check the decimal places against the original version.
Also, agreeing on all acceptable formats for all data types is important during study build. However, since most data imports happens later in the study, CRFs for import may be built whenever the data extract is available. Otherwise, the formats may need to be reviewed and adjusted per study team's approval.
- f. Save the import data file under the type .CSV (Comma delimited).

Step 2 – Create the stylesheet: The stylesheet is used by Kernow to create the XML file from the CSV file. The template can be found here: S:\shared\OpenClinica\Importing_data_into_OC\Stylesheet template

Make a copy of the stylesheet for each new import. Open this in NotePad++ and edit the following lines:

LINE 10:

```
<xsl:param name="pathToCSV"
select="file:///S:/shared/OpenClinica/Importing_data_into_OC/hicc_eos_import.csv"/>
```

Update LINE 10 with the file path to your import data. Note that forward slashes should be used in the file path in the stylesheet. It is advised to put the import file in the following folder so only the end of the file path needs to be updated in the stylesheet:

S:\shared\OpenClinica\Importing_data_into_OC\file sub-folder\file name.csv

LINE 30:

```
<ClinicalData StudyOID="S_P01685" MetadataVersionOID="v1.0.0">
```

Update LINE 30 with the StudyOID. This is found under *Tasks > Build Study > Create Study (view)*:

GD045 Importing Data into OpenClinica3 (OC3)

Overview	
Name:	HICC
Unique Protocol ID:	P01685
OID:	S_P01685
Principal Investigator:	Dr Helen Baxendale
Brief Summary:	HICC: Humoral Immune Correlates for COVID19: Defining protective responses and critical readouts for Clinical Trials of Vaccines and Therapeutics
Owner:	tdevine
Date Created:	15-Sep-2020

LINE 36:

<StudyEventData StudyEventOID="**SE_ENDOFSTUDY_5442**">

Update LINE 36 with the StudyEventOID. This can be found under *Tasks > Build Study > Create Event Definitions (view)*:

Unscheduled visit (PAP only)	SE_UNSCHEDULEDVISITPAPONLY	Yes	Scheduled	Yes	16-Nov-2021 (tdevine)	18-May-2022 (tdevine)	
End of study	SE_ENDOFSTUDY_5442	No	Scheduled	Yes	10-Nov-2020 (tdevine)	16-Nov-2021 (tdevine)	
Admin (PAP only)	SE_ADMINPAPONLY	No	Scheduled	Yes	02-Sep-2021 (tdevine)	16-Nov-2021 (tdevine)	
Vaccine data	SE_VACCINADATA	No	Scheduled	Yes	25-Aug-2022 (tdevine)	25-Aug-2022 (tdevine)	

LINE 37:

<FormData FormOID="**F_PRODOSERANDO_V10**">

Update LINE 37 with the FormOID. This FormOID must include the version number. This can be found under the same page that contains the ItemOIDs:

View CRF Details	
Name:	HICC End of study
Description:	v1.0
OID:	F_HICCENDOFSTU

Do NOT use this FormOID because it does not contain the version number

Version(s)					
Version Name	oid	Description	Status	Revision Notes	Action
v1.0	F_HICCENDOFSTU_V10	v1.0	available	Thomas Devine 10/11/2020	

Items					
Name	Item_OID	Description	Data Type	Version(s)	Integrity Check
date_trial_completed	I_HICCE_DATE_TRIAL_COMPLETED	Date trial completed	date	v1.0	OK
other_reason	I_HICCE_OTHER_REASON	Please specify (Other)	character string	v1.0	OK
reason	I_HICCE_REASON	Reason trial completed	integer	v1.0	OK

LINE 38:

<ItemGroupData ItemGroupOID="**IG_HICCE_END**" TransactionType="Insert" >

GD045 Importing Data into OpenClinica3 (OC3)

Update LINE 38 with the ItemGroupOID, this can be found under the same page that contains the Response Values:

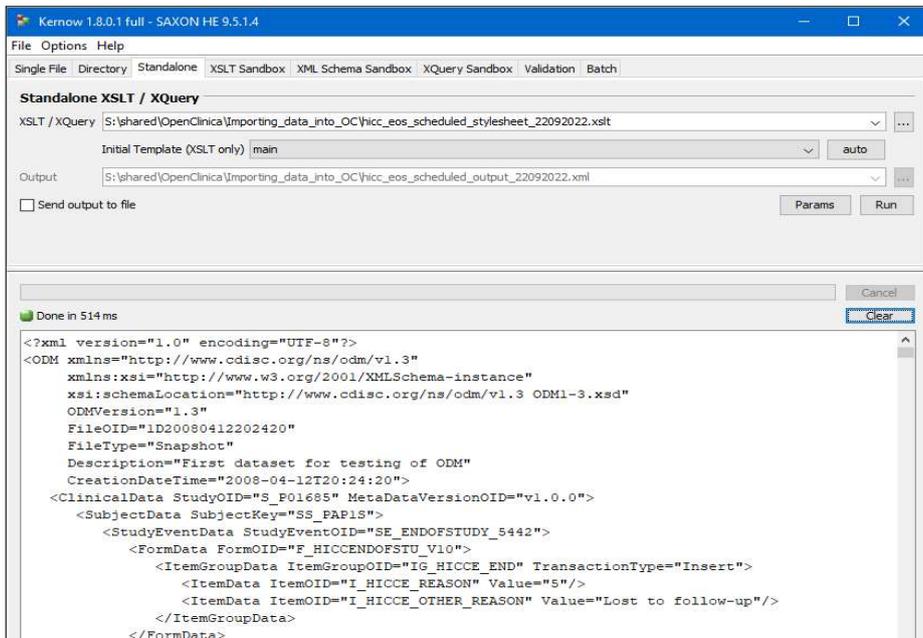
SECTION			
Section Name	Title	Subtitle	Instructions
End of study	End of study		<style type="text/css">.aka_text_block{width:150px;text-align:left;}</style>

Groups						
Group Name	OID	Header	Repeat Number	Repeat Max	Is shown?	Group Layout
end	IG_HICCE_END		1	1	Yes	Non-repeating

Items															
Name	Item_OID	GROUP_OID	Description	Group Name	Units	PHI Status	Item Data Type	Left Item Text	Right Item Text	Default Value	Response Layout	Response Label	Response Options	Response Values	Required
reason	I_HICCE_REASON	IG_HICCE_END	Reason trial completed	end		false	integer	Reason trial completed				Reason	(select), Completed study, Screen failure, Withdrawal of consent, Death, Other	, 1, 2, 3, 4, 5	Yes

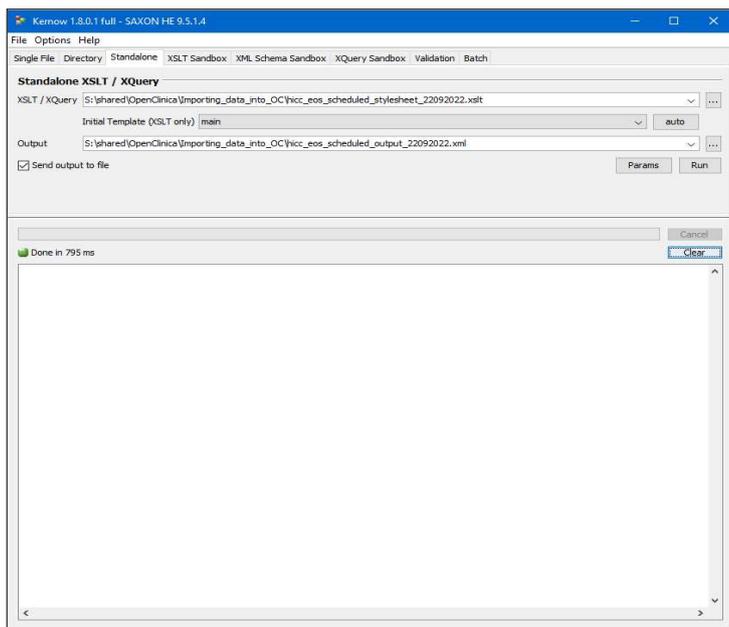
Once all of these lines have been updated the stylesheet is ready.

Step 3 – Generate the XML with Kernow: To generate the XML file open Kernow and select tab 'Standalone'. Under XSLT/XQuery, load the stylesheet you created. Set Initial Template (XSLT only) to 'main' by selecting 'auto'. Prior to writing the XML to a file, 'Send output to file' can be unselected, this will display the output in the console, and it will show if there any errors. If a message display that the file cannot be found, check the file path in the stylesheet is correct. See Troubleshooting at the end of this document for other potential issues.



If the console displays the XML correctly this can be written to an XML file. Select 'Send output to file' and specify the file location and name (ending in .xml). Select 'Run', this will create the import XML file in the location specified:

GD045 Importing Data into OpenClinica3 (OC3)



Note: If required, Xerrow licence can be obtained from a member of the Data team with authorised access.

Step 4 – Updating the XML file: Once the XML file has been created, open this in NotePad++ to review. The XML file may need to be updated, there are 3 things to consider: missing data, CRF status, and event occurrence.

Missing data

If there is missing data in your data set the XML will produce the line with a blank value:

```
10 <ClinicalData StudyOID="S_P01685" MetaDataVersionOID="v1.0.0">
11   <SubjectData SubjectKey="SS_PAP1S">
12     <StudyEventData StudyEventOID="SE_ENDOFSTUDY_5442">
13       <FormData FormOID="F_HICCENDOFSTU_V10">
14         <ItemGroupData ItemGroupOID="IG_HICCE_END" TransactionType="Insert">
15           <ItemData ItemOID="I_HICCE_REASON" Value=""/>
16           <ItemData ItemOID="I_HICCE_OTHER_REASON" Value="Lost to follow-up"/>
17         </ItemGroupData>
18       </FormData>
19     </StudyEventData>
20   </SubjectData>
```

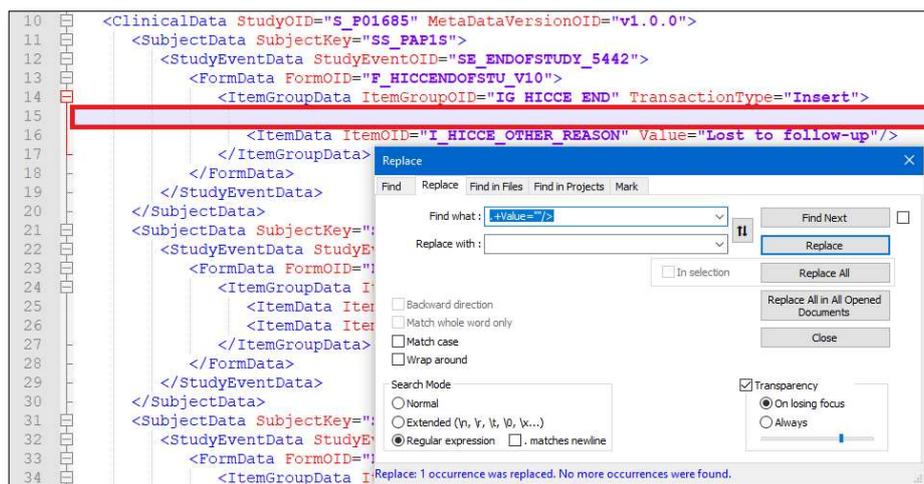
These lines should be removed with Find and Replace using Regular expression.

Find : `+.Value=""/>`

Replace : The 'Replace with' box should remain empty.

Select 'Regular expression' in the Search Mode section and 'Replace All' to remove all of these lines:

GD045 Importing Data into OpenClinica3 (OC3)



CRF status

By default, when importing, the CRF status will be set to Complete. If you need the status to be set as Initial Data Entry, the line containing the FormOID line needs to be edited in each case. The line needs to be edited from:

`<FormData FormOID="F_HICCENDOFSTU_V10">`

to

`<FormData FormOID="F_HICCENDOFSTU_V10" OpenClinica:Status="initial data entry">`

This can be done for all of lines by using Find and Replace. For this reason, if there some participants that need the CRF marked as Complete and some participants that need the CRF marked as Initial Data Entry started in the same data set, it is easier to do these as 2 separate imports. This change could also be made directly in the stylesheet before converting to XML with Kernow.

Event occurrence

If you are importing data into a repeatable event in OpenClinica, you need to specify which occurrence that data should be imported into within the XML file. In a similar way to the CRF status, this can be done with Find and Replace once the XML has been produced, or it could be done directly in the stylesheet before converting to XML. To specify the event occurrence, the line needs to be edited from.

The example below shows importing into occurrence 2

`<StudyEventData StudyEventOID="SE_SAES_263">`

to

`<StudyEventData StudyEventOID="SE_SAES_263" StudyEventRepeatKey="2">`

If the data is being imported into the first occurrence of a repeating event, you do not need to specify this.

If your import data has no missing data, the CRF status needs to be marked as Complete and there are no repeating events, Step 4 can be completely skipped.

GD045 Importing Data into OpenClinica3 (OC3)

Step 5 – Import the XML file: First, schedule the event in OpenClinica. The XML file can then be imported by Selecting Tasks > Import Data:



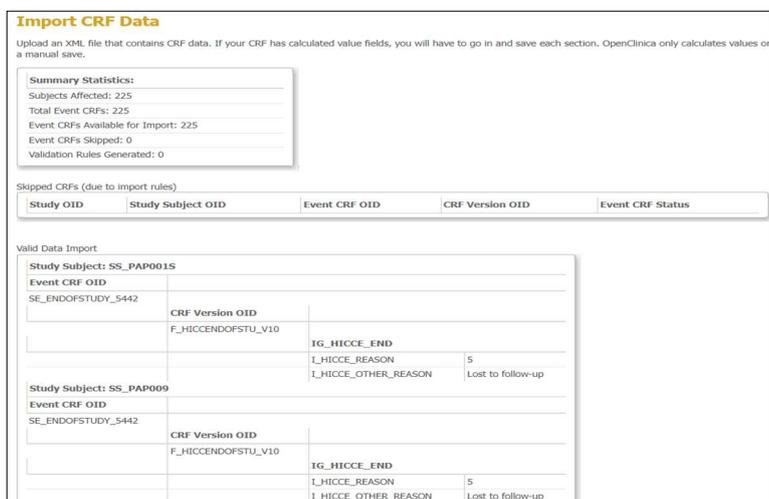
Import CRF Data

Upload an XML file that contains CRF data. If your CRF has calculated value fields, a manual save.

Choose File No file chosen

Continue Cancel

Select the XML file. If the following page displays the data to be imported, then everything has worked correctly:



Import CRF Data

Upload an XML file that contains CRF data. If your CRF has calculated value fields, you will have to go in and save each section. OpenClinica only calculates values on a manual save.

Summary Statistics:

Subjects Affected:	225
Total Event CRFs:	225
Event CRFs Available for Import:	225
Event CRFs Skipped:	0
Validation Rules Generated:	0

Skipped CRFs (due to import rules)

Study OID	Study Subject OID	Event CRF OID	CRF Version OID	Event CRF Status

Valid Data Import

Study Subjects	Event CRF OID	CRF Version OID	IG_HICCE_END	L_HICCE_REASON	L_HICCE_OTHER_REASON
SS_PAP001S	SE_ENDOFSTUDY_5442	F_HICCENDOFSTU_V10			
				5	Lost to follow-up
Study Subject: SS_PAP009	SE_ENDOFSTUDY_5442	F_HICCENDOFSTU_V10			
				5	Lost to follow-up

If there are errors at this stage, review the steps completed. Usually this is due to the field not allowing the import data, for example, trying to import a value with 3 decimal places into an OpenClinica field that only allows 2 decimal places. Refer to the *Troubleshooting section* below for other potential issues.

Step 6 – Validation: The import should be validated by comparing the import csv file to a data export from OpenClinica, ensuring both entries are a MATCH. The two data sets should be compared in a new file (this can be done using the VLOOKUP function in Excel) and stored appropriately.

After completing the import, FRM071 should be updated with the location of the following files:

Location of csv import file

Location of stylesheet

Location of import xml file

Location of import validation

Step 7- TMF: All the files associated with the import should be saved in Section 11.6 Data Imports in the paper TMF (for a CTIMP) or in the eTMF (for other types of studies). For example, for NOTACS:

N:\Shared\R&D PROJECTS\P02590 NOTACS\Sponsor File\11.0 Data Management\11.6 Data Imports. Some previous imports have been saved in \Sponsor File\11.4 Data Amendment Form(s) – so this folder should be checked if you are looking for the files for a previous import.

3. Limitations

- The event must be scheduled in OpenClinica for the import to work.
- Data cannot be imported into a locked CRF.
- The transform to XML step only works correctly if there are no line breaks in the data (so no multi-line text fields).
- XML characters in the input like < and > will cause problems.
- The XML is easiest to construct if all the data is for the same event, event occurrence, CRF and group.

4. Troubleshooting

- Cannot find path in Kernow -> Check the file path in the stylesheet.
- If you are continuing to have an issue with Kernow, open the import file in Notepad++ to see if the csv contains any hidden special characters.
- Message 'Initial template not set!' appears in Kernow -> Under Initial Template (XSLT only) select auto (step 3).
- Cannot recognise OID when importing into OC -> Update all the OIDs in the stylesheet (step 1).
- XML passed with no errors, but no data is shown next to IDs -> Check the data type is right for each field (step 1). It could also mean the field options for that ItemOID were previously changed, causing the import to not recognise the new response values. In this case, a design change might be required to assign a new OID and response label to the field, or enter the affected field manually, depending on the study team's decision.

5. Best Practice

- If a significant import error occurs, notify the relevant study team to discuss the resolution plan before making any corrections. Also, document this event in the post-import comments section of FRM071.
- Avoid editing the original data extract directly. Always create a copy before making changes to prevent data loss or errors.

Appendix 1

- This method was developed using the following website:
https://en.wikibooks.org/wiki/OpenClinica_User_Manual/ImportingData#Converting_CSV_to_CDISC_ODM_XML_using_XSLT
- The following guidance document details importing data into OpenClinica3 by converting data in a CSV file to XML using XSLT transform (Kernow). A screen recording of this process can be found here:
<S:\shared\R&D\SOPs\FORMS, TEMPLATES & GUIDANCE DOCUMENTS\GUIDANCE DOCUMENTS\GD045 Importing data into OpenClinica\Importing data into OpenClinica video guide.mp4>