2.5 Import Data

OpenClinica allows you to import Event CRF data from a file as an alternative to performing data entry using the OpenClinica web interface. The import data feature is useful when the data you want for a Study in OpenClinica is captured using a system other than OpenClinica. When data is imported, the Audit Log records the userid of the person performing the import as the data entry person.

Before importing data into OpenClinica, the following tasks must be completed:

- Create the Study in OpenClinica
  Tasks>Create Study and then Tasks>Build Study>Create Study
- Create the CRFs for the data being imported
  Tasks>Build Study>Create CRF
- Add the Event Definitions for the data being imported, and associate CRFs for the data being imported with the Event(s)
  Tasks>Build Study>Create Event Definitions
- Enroll the Study Subjects for the data being imported
  Subject Matrix>Add New Subject
- For each of the Subjects, schedule the Events for which data is being imported
  Subject Matrix>Click on an Event for a Subject, then click Schedule and schedule the Event.

In addition, the following data characteristics are enforced when data is imported:

- Data cannot be imported into a CRF that has a status of DoubleDataEntryStarted or DoubleDataEntryComplete.
- Dates must be imported in the format YYYY-MM-DD
- Data types in the import file must match the item data type in OpenClinica (i.e., you cannot import a character string into an item that was defined as integer).
- Data values must match the response values specified in the eCRF. For example you cannot import "3" if the only response options provided were Yes (1) and No (2).
- Required fields cannot be empty.

The information in the import file can be for:

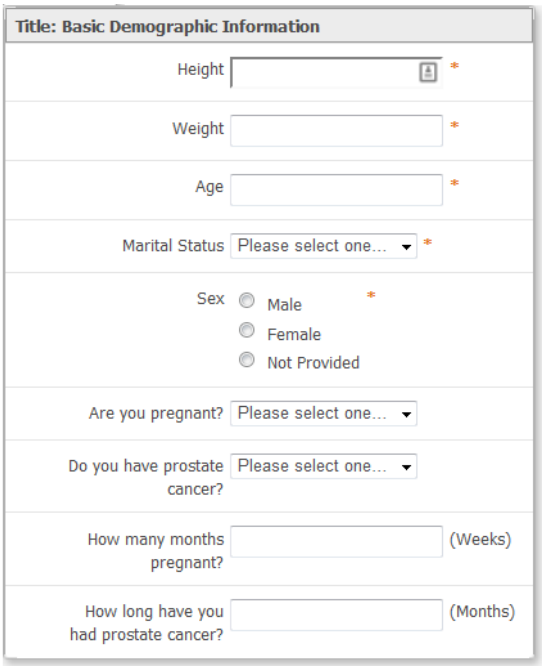
    - one or more Subject*s*
    - one or more Events for each Subject
    - one or more CRFs for each Event

The type or amount of information for each Subject does not have to be the same. For example, the file can include one Subject with two Events - each with one CRF, and another Subject with one Event that contains three CRFs.

Sample eCRF and Import File

This section presents a sample eCRF and the related XML file that would import data into that CRF. Prior to performing the import, all of the steps listed above were performed.

Sample eCRF:



Sample XML File:

The following XML file was used to import data into the above eCRF. In the sample below, all highlighted values are values that must be specified for each import as follows:

- Values in red are where your various OIDs must be provided
- Values in **bold** are where you specify the description and status conditions for the import
- Values highlighed in blue should be replaced with the actual data values that you want to import into the associated Item OID.

```
<?xml version="1.0" encoding="UTF-8"?><ODM xmlns="http://www.cdisc.org/ns/odm/v1.3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://www.cdisc.org/ns/odm/v1.3 ODM1-3.xsd" ODMVersion="1.3" FileOID="1D20080412202420"
FileType="Snapshot" Description="Demographics Import" CreationDateTime="2008-04-12T20:24:20" >
<ClinicalData StudyOID="S_THEJUNOD" MetaDataVersionOID="v1.0.0">
        <UpsertOn NotStarted="true" DataEntryStarted="false" DataEntryComplete="false"/>
    <SubjectData SubjectKey="SS_0001">
      <StudyEventData StudyEventOID="SE_FIRSTVISIT" StudyEventRepeatKey="1">
        <FormData FormOID="F_DEMOG_1" OpenClinica:Status="initial data entry">
              <ItemGroupData ItemGroupOID="IG_DEMOG_BASICS" ItemGroupRepeatKey="1" TransactionType="Insert" >
                        <ItemData ItemOID="I_DEMOG_PGDEMOG1" Value="58"/>
                        <ItemData ItemOID="I_DEMOG_PGDEMOG2" Value="116"/>
                        <ItemData ItemOID="I_DEMOG_PGDEMOG3" Value="53"/>
                        <ItemData ItemOID="I_DEMOG_PGDEMOG4" Value="2"/>
                        <ItemData ItemOID="I_DEMOG_PGDEMOG5" Value="2"/>
                        <ItemData ItemOID="I_DEMOG_PGDEMOG6" Value="2"/>
                        <ItemData ItemOID="I_DEMOG_PGDEMOG7" Value="2"/>
                        <ItemData ItemOID="I_DEMOG_PGDEMOG8" Value="0"/>
                        <ItemData ItemOID="I_DEMOG_PGDEMOG9" Value="0"/>
              </ItemGroupData>
        </FormData>
      </StudyEventData>
    </SubjectData>
</ClinicalData>
</ODM>
```

For each Item OID, specify the value that should be imported.

Import Status and Form Status

Beginning with version 3.6 of OpenClinica, the UpsertOn tag in the XML code for importing data allows you to specify whether to perform the import or not based on the current CRF status. This provides the option to update data that has already been imported or entered. For example, the following specification would only import the data if the current status of the form is "not started":

<UpsertOn NotStarted="true" DataEntryStarted="false" DataEntryComplete="false"/>

If you want to use the import to update data that is already in the database, but has not yet been marked complete, then use the following settings:

<UpsertOn NotStarted="false" DataEntryStarted="true" DataEntryComplete="false"/>

If all three statuses were set to true, then the import would occur regardless of the CRF status. If all three statuses were set to false, data would not be imported.

If you choose to remove the UpsertOn option from the XML file, then OpenClinica performs the import as if all UpsertOn statuses were set to "true."

Note:

- For any data that is being imported and does not match the criteria for the UpsertOn settings, a list of "skipped" records is displayed for your review. Based on the data that is displayed, you may need to modify the UpsertOn settings, or you may want to continue with the import if the data was skipped appropriately.

- You cannot import data into a CRF that has a status of DoubleDataEntryStarted or DoubleDataEntryComplete. OpenClinica does not include that data in the import, but does provide information about the skipped data in the Summary Statistics and Import CRF Data details window.

Assign a Status

By default, when an import is complete, OpenClinica sets the CRF status to "complete." OpenClinica version 3.6 provides the option to instead set the status of the CRF to "data entry started" after the import is complete. To do this, set the "OpenClinica:Status=" to "inital data entry" (case sensitive). The following tag includes the Form OID (in this example F_DEMOG_1) as well as the status setting for that Form once the import is complete:

<FormData FormOID="F_DEMOG_1" OpenClinica:Status="initial data entry">

In this example, once the data is imported, the form will have a status of initial data entry.

If you remove the OpenClinica:Status clause from the XML file, or provide any value other than "initial data entry", OpenClinica implements the default behavior and sets the status of the form to "complete". Although you may omit the OpenClinica:Status="initial data entry" clause, the rest of the FormData statement must remain in place in order for the import to function.

The data import file must be a properly-formatted XML file. For details, see Data File for Import.

Automatic Edit Checks

For Rules to run automatically upon import, the **Run** parameter **ImportDataEntry** must be set to **"true"**. For more information, see Creating Rules.

Edit checks are processed as follows:

- If Rules are set to run during import, failed Rules automatically create Failed Validation Discrepancy Notes.
- If there are errors with the import definition (for example you tried to import text into an item defined as a date), then an error message displays in the the Alerts and Messages section of the screen.
- If there are no errors, the CRF is automatically marked with the status provided in the "OpenClinica:Status=" clause.
- Any subsequent changes can be done by editing the data via the Subject Matrix or by re-importing using the appropriate UpsertOn value. If you had set the OpenClinica:Status to "complete" on the initial import, the data can only be edited using Administrative Editing.
- If there are any calculated value fields in the CRF, you must either supply the correct data in the import file or open and save the CRFs manually to generate the calculations.

You can automate the importing of data, which is useful when you import the data files on a recurring basis: see Scheduled Import Jobs. Another alternative for getting data into OpenClincia is with SOAP Web Services: see the import Data Web Service API in the *OpenClinica Technical Documentation*.

Functional approval by Laura Keita. Signed on 2016-06-06 11:56AM

Approved for publication by Ben Baumann. Signed on 2016-06-07 11:14AM

Not valid unless obtained from the OpenClinica document management system on the day of use.

2.5.1 Import the Data File

Users whose role has appropriate permission can import data from the OpenClinica web interface. The Study or Site in the import file must be the same as the current Study or Site.

To import the data:

1. From the navigation bar, select **Tasks > Import Data**.
   The Import CRF Data page opens.

2. Browse for the XML file that contains the import data, select the file, and click **Open**.
   The filename you selected is listed on the Import CRF Data page.

3. On the Import CRF Data page, click **Continue**.
   OpenClinica validates the XML file. If there are any validation errors, the Alerts and Messages sidebar panel reports the problems. Correct the errors in the file, then in the Import CRF Data page, browse to select the filename, and click Continue.

   If there are no validation errors, OpenClinica displays a page summarizing the CRF data being imported, and presents the results of the Hard Validation Error Checks, including any errors. For example, it reports an error if a date format is invalid.

Correct any problems in the file, then, in the Import CRF Data page, browse to select the filename, and click Continue.

4. If there are no Hard Validation Errors, to continue to import the data, click **Continue**.

5. In the confirmation dialog box, to complete the data import, click **OK**. To cancel the import, click **Cancel**.
   If you completed the import, the CRF(s) for the Subject(s) specified in the import file now contain the imported data, and the status of the CRF(s) are set to the status specified in the import file.

6. If the CRF contains empty fields for any calculated values, manually open the CRF and save each page that contains the calculated values.
   OpenClinica only calculates the values when you save the page manually. For details, see Complete a CRF.

7. If there are Rules associated with the CRFs where the data was imported, check to see if you need to manually run the Rules. As of OpenClinica 3.1.3, the Rule option **<Run ImportDataEntry="true">** ensures that Rules are run when importing data. If this option was set to "false" in the Rule, go to **Tasks>Monitor and Manage Data>CRFs**. Then, on the top row for the CRF (original), click the view icon. Scroll to the bottom of the page and click **Run All Rules for this CRF**.

2.5.2 Data File for Import

To import Event CRF data, generate a properly-formatted XML file that contains the data to import. The XML file includes the Object Identifiers (OIDs) that OpenClinica uses to uniquely identify the entities in the Study. The XML file is compliant with the Clinical Data Interchange Standards Consortium Object Data Module (CDISC ODM), version 1.2 or 1.3.

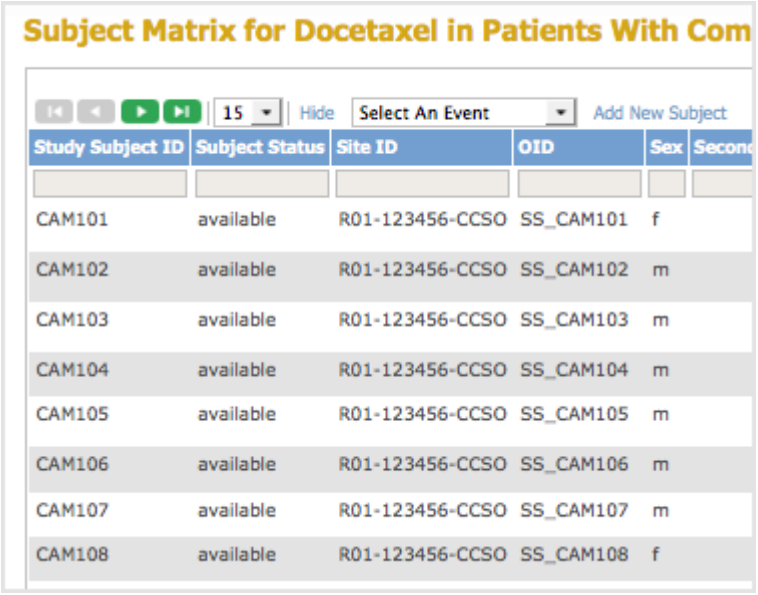
To generate the XML file for importing data:

1. Determine OIDs for the Subjects, Study, Event Definitions, Forms, Item Groups, and Items that you want to import data for. See details at Determine OIDs.
2. Generate the import file using the OIDs and the data. For details, see Generate Data Import File.

2.5.3 Determine OIDs

Use the following methods to determine Object Identifiers (OIDs). Record the OIDs for use in the data import file. The OIDs for Study, Study Event Definitions, Forms (CRFs), Item Groups, and Items were created as part of Study Setup. The OIDs for Study Subjects were created when the Subject was added to the Study.

To determine OIDs for Study Subjects, view the Subject Matrix, then click Show More. The Subject OIDs are in the OID column. For the example Subject Matrix shown here, the OID for Subject CAM107 is SS_CAM107:

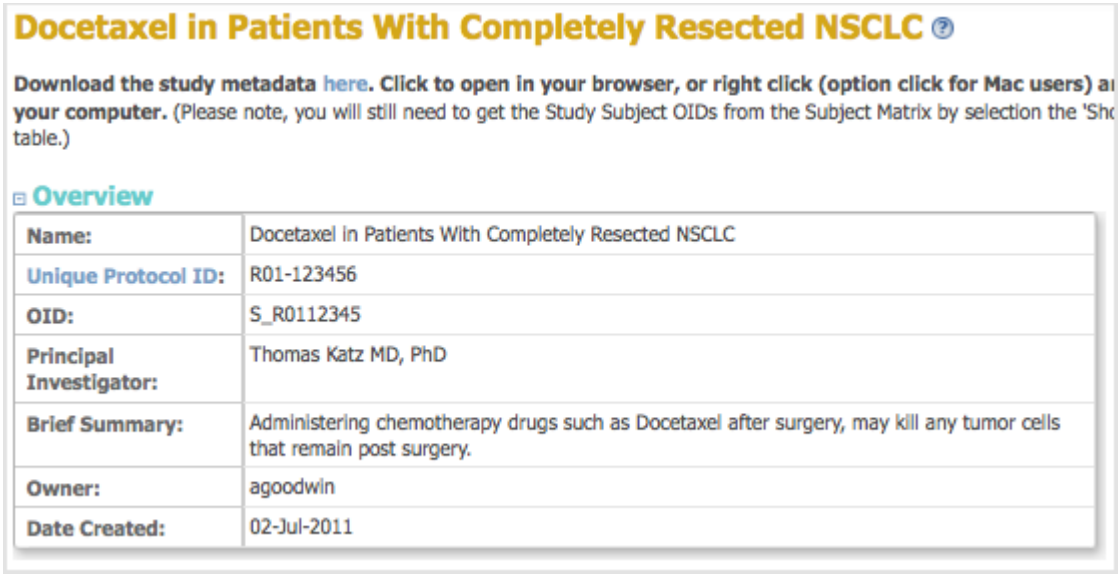


To determine OIDs for the Study, Study Event Definitions, CRF, Item Groups, and Items, use one of these methods:

- View study details in the OpenClinica web interface.
- View the Study Metadata file. You should be familiar with XML file structure to use this method.

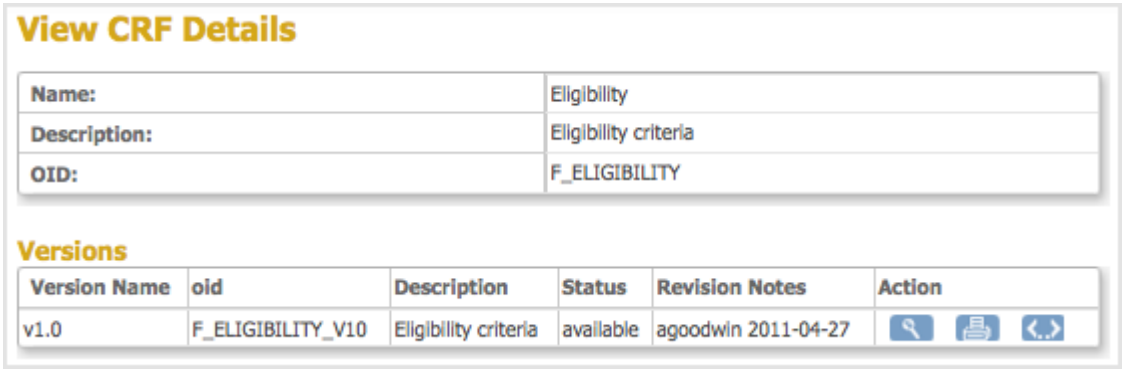
2.5.3.1 View Study Details to Determine OIDs

1. Select Tasks > View Study.
   A page with details about the current study opens.
2. From the Oveview section, record the Study OID.



3. From the Sites section, record the Site OID.
4. From the Event Definitions section, record the Event OIDs.
5. For the Event Definition whose CRF OIDs you want, click the View icon.
   The View Event Definition page opens, displaying a table of CRFs for that Event.
6. In the CRFs table, click the View icon for the CRF whose OIDs you want.
   The View CRF Details page opens and displays a table of Versions.

7. From the Versions table, record the OID for the version of the CRF you want to use.



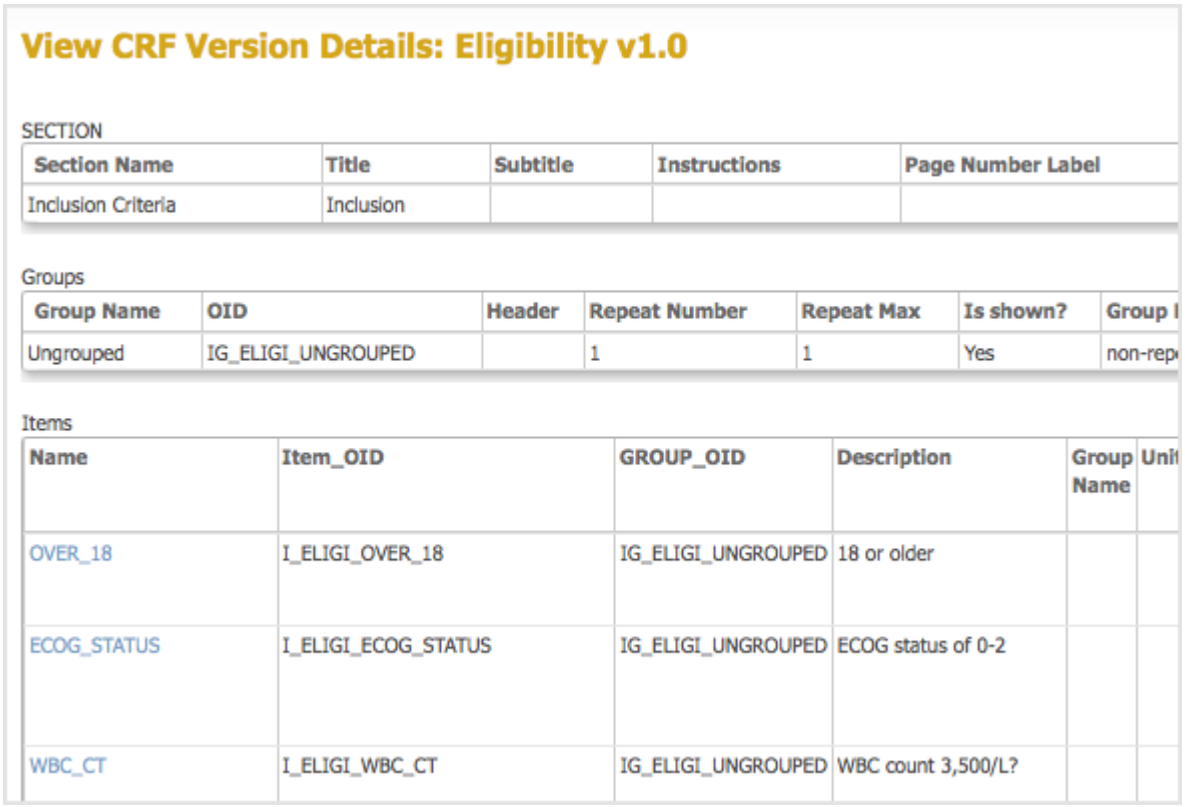
View CRF Details

Name:	Eligibility
Description:	Eligibility criteria
OID:	F_ELIGIBILITY

Versions

Version Name	oid	Description	Status	Revision Notes	Action
v1.0	F_ELIGIBILITY_V10	Eligibility criteria	available	agoodwin 2011-04-27	

8. For the CRF version you want, click the Metadata icon .
   The View CRF Version Details page opens.
9. From the View CRF Version Details page, record the OIDs for the Item Groups and Items.



View CRF Version Details: Eligibility v1.0

SECTION

Section Name	Title	Subtitle	Instructions	Page Number Label
Inclusion Criteria	Inclusion			

Groups

Group Name	OID	Header	Repeat Number	Repeat Max	Is shown?	Group I
Ungrouped	IG_ELIGI_UNGROUPED		1	1	Yes	non-rep

Items

Name	Item_OID	GROUP_OID	Description	Group Name	Unit
OVER_18	I_ELIGI_OVER_18	IG_ELIGI_UNGROUPED	18 or older		
ECOG_STATUS	I_ELIGI_ECOG_STATUS	IG_ELIGI_UNGROUPED	ECOG status of 0-2		
WBC_CT	I_ELIGI_WBC_CT	IG_ELIGI_UNGROUPED	WBC count 3,500/L?		

2.5.3.2 View the Study Metadata File to Determine OIDs

If you are familiar with XML file structure, you can determine OIDs using the Study metadata file:

1. Download the Study metadata file: Select Tasks > View Study, then follow the instructions at the top of the page.
2. Open the Study metadata file in a browser or text editor.

Following is an excerpt of a Study metadata file that provides an example of the structure, showing only the tags that are relevant to finding and determining OIDs needed for the data import file.

Comments <!-- ... --> indicate where the OID values are. You can download and view the complete Study metadata file for the example Study.

Starting with the <Protocol> tag, the general structure is that the OID for an element is defined in a <...Ref> tag. After that, a <...Def> tag for the element contains tags that define the OIDs for the subelements, and so on, through the hierarchy. The heirarchy is: Study Event Definition > Form (CRF) > Item Group > Item.

<?xml version ...?>

<ODM ... >

<Study OID="*YourStudyOID*"> <!-- Get the Study OID here. -->

<GlobalVariables>

...

</GlobalVariables>

<BasicDefinitions>

...

</BasicDefinitions>

<MetaDataVersion OID="*YourMetaDataVersionOID*"... > <!-- Get the MetaDataVersion OID here. -->

<Protocol>

<StudyEventRef StudyEventOID="*YourFirstStudyEventOID*" .../> <!-- Get the OID for your first Study Event here -->

<StudyEventRef StudyEventOID="*YourSecondStudyEventOID*" .../> <!-- Get the OID for your second Study Event here -->
...

</Protocol>

<StudyEventDef OID="*YourFirstStudyEventOID*" ... > <!-- This section contains the OIDs for the Forms (CRFs) in the Study Event whose OID is *YourFirstStudyEventOID* -->
<FormRef FormOID="*YourFirstFormOID*" ... /> <!-- Get the OID for the first CRF used in this Study Event here -->
<FormRef FormOID="*YourSecondFormOID*" ... /> <!-- Get the OID for the second CRF used in this Study Event here -->
...

```
</StudyEventDef>

<StudyEventDef OID="YourSecondStudyEventOID" ... > <!-- This section contains the OIDs for the Forms (CRFs) in the Study Event whose OID is YourSecondStudyEventOID -->
</StudyEventDef>
...

<FormDef OID="YourFirstFormOID" ...> <!-- This section contains the OIDs for the Item Groups in the CRF whose OID is YourFirstFormOID -->
<ItemGroupRef ItemGroupOID="YourFirstItemGroupOID" ... /> <!-- Get the OID for the first Item Group in this CRF here -->
<ItemGroupRef ItemGroupOID="YourSecondItemGroupOID" .../> <!-- Get the OID for the second Item Group in this CRF here -->
...
<OpenClinica:FormDetails ... >
...
</OpenClinica:FormDetails>
</FormDef>

<FormDef OID="YourSecondFormOID" ...> <!-- This section contains the OIDs for the Item Groups in the CRF whose OID is YourSecondFormOID -->
...
</FormDef>

<ItemGroupDef OID="YourFirstItemGroupOID" ... > <!-- This section contains the OIDs for the Items in the Item Group whose OID is YourFirstItemGroupOID -->
<ItemRef ItemOID="YourFirstItemOIDInYourFirstItemGroup" /> <!-- Get the OID for the first Item in this Item Group here -->
<ItemRef ItemOID="YourSecondItemOIDInYourFirstItemGroup" /> <!-- Get the OID for the second Item in this Item Group here -->
...
<OpenClinica:ItemGroupDetails ... >
...
</OpenClinica:ItemGroupDetails ...>
</ItemGroupDef>


<ItemGroupDef OID="YourSecondItemGroupOID" ... > <!-- This section contains the OIDs for the Items in the Item Group whose OID is YourSecondItemGroupOID -->
...
</ItemGroupDef>
...

<ItemDef OID="YourFirstItemOIDInYourFirstItemGroup" ... > <!-- This section contains details for the Item whose OID is YourFirstItemOIDInYourFirstItemGroup -->
...
</ItemDef>
<ItemDef OID="YourSecondItemOIDInYourFirstItemGroup" ... > <!-- This section contains details for the Item whose OID is YourSecondItemOIDInYourFirstItemGroup -->
...
</ItemDef>
```

...

2.5.4 Create the Data Import File

Create the data import XML file using the OIDs for your Study, and the data for each Item. Use the example in this section as a template for the structure of your data import file. You can also download example files.

Starting with the <SubjectData> tag, the general structure is that the tag for an element contains the tags for its subelements, and so on, through the hierarchy. The heirarchy is: Study Subject > Study Event Definition > Form (CRF) > Item Group > Item. So within a <SubjectData> tag are the <EventData> tags for that Subject, and so on.

For Repeating Events and Repeating Item Groups, specify the value 1 for the first Repeat Key, 2 for the second Repeat Key, and so on.

Following is an XML import file that imports data into a demographics form in the sample Juno study:

```
<?xml version="1.0" encoding="UTF-8"?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="
http://www.cdisc.org/ns/odm/v1.3 ODM1-3.xsd" ODMVersion="1.3" FileOID="1D20080412202420" FileType="Snapshot" Description=
"OpenClinica Data Import" CreationDateTime="2015-04-12T20:24:20" >

<!-- Replace the S_OID with the Study OID of the Study to which you are importing data. -->
<ClinicalData StudyOID="S_OID" MetaDataVersionOID="v1.0.0">

        <!-- Specify whether or not to update existing data based on the current status of the data. May be set to "true"
        (update) or "false" (do not update) for each of the listed statuses.-->
        <UpsertOn NotStarted="true" DataEntryStarted="true" DataEntryComplete="true"/>

        <!-- Replace the SS_OID with the Study Subject OID for whom you're importing data. This Subject must already be
        enrolled and scheduled for the Event to which you are importing. -->
        <SubjectData SubjectKey="SS_OID">

            <!-- Replace the SE_OID with the OID of the Study Event. For non-repeating Events, the StudyEventRepeatKey
            should be set to 1. For repeating Events, specify the ordinal for the occurrence of the Event to which you are
            importing. For example, specify 2 to import the data into the second repeat occurrence of an Event. -->
            <StudyEventData StudyEventOID="SE_OID" StudyEventRepeatKey="1">

                <!-- Replace the F_OID with the OID of the Form that will be populated with the imported data. Also specify
                the status the Form should have after the import is complete. Options are initial data entry or complete.-->
                <FormData FormOID="F_OID" OpenClinica:Status="initial data entry">

                <!-- Replace the IG_OID with the Item Group OID. If the Group is a Grid, specify the row number for the
                insert in ItemGroupRepeatKey. TransactionType for an import should always be Insert.-->
                    <ItemGroupData ItemGroupOID="IG_OID" ItemGroupRepeatKey="1" TransactionType="Insert" >

                    <!-- For each Item in the Form, replace the I_OID with the Item OID and specify the value that should
                    be imported into that Item. Add or remove lines below as needed to match the number of Items in your
                    Form.-->
                        <ItemData ItemOID="I_OID" Value="68"/>
                        <ItemData ItemOID="I_OID2" Value="116"/>
                        <ItemData ItemOID="I_OID3" Value="53"/>
                        <ItemData ItemOID="I_OID4" Value="2"/>
                    </ItemGroupData>
                </FormData>
            </StudyEventData>
        </SubjectData>
        <!-- To import data for more than one Subject, copy the opening and closing SubjectData tags, and everything between
        those tags. Then update OIDs and values as needed.-->
    </ClinicalData>
</ODM>
```

To create an XML data import file:

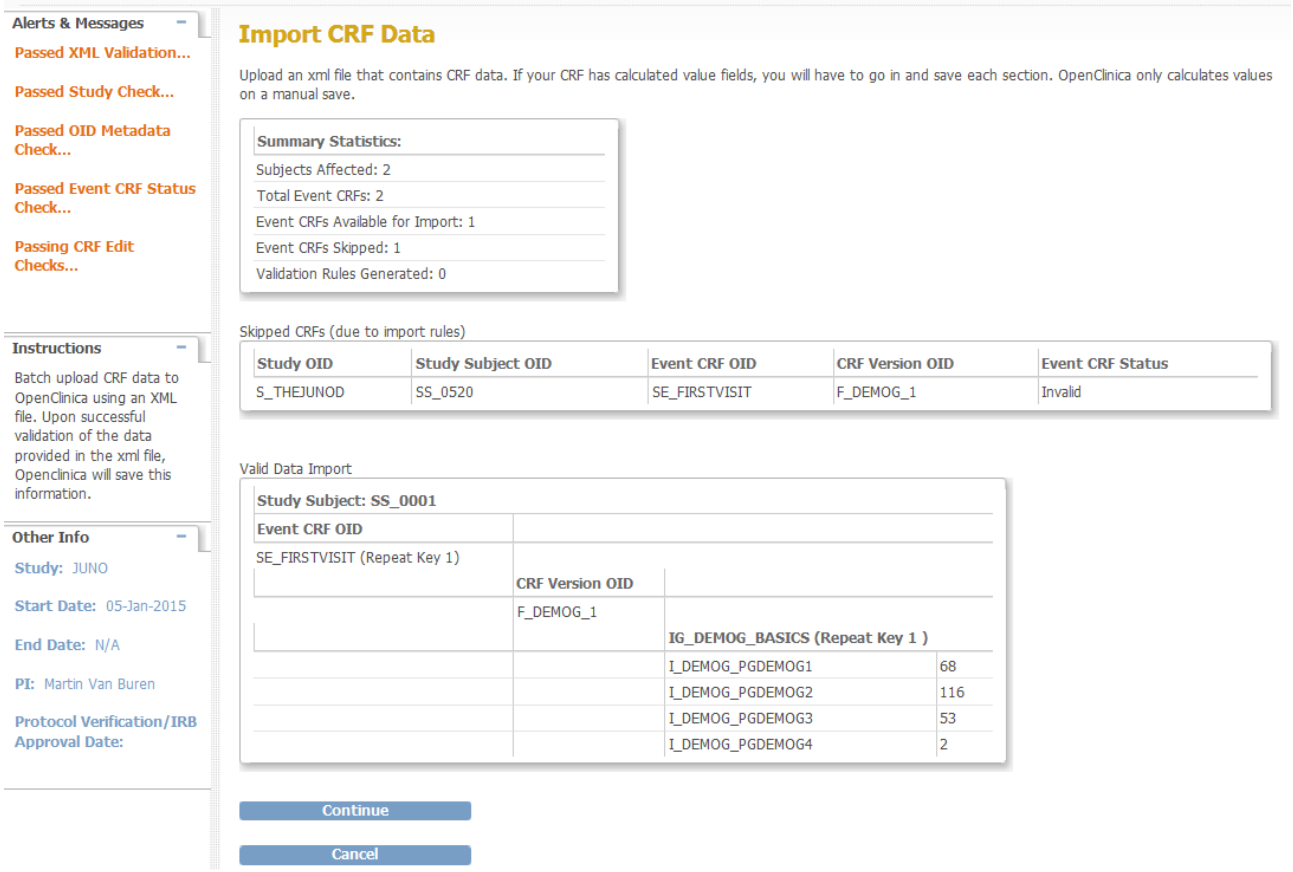
1. Download the import template
2. Open the import template in an XML editor
3. Save the template as an XML file.

4. Modify the template as needed to meet the requirements of the data you are importing. Instructions are provided throughout the template for guidance.

2.5.5 Validate and Check the Data Import File

When you import a data file, OpenClinica validates the file format and performs validity checking for the data, then displays any error messages in the Alerts and Messages sidebar panel. Correct any errors in the file and import it again.

OpenClinica also displays Summary Statistics of imported and skipped records and displays detailed information about the import. Details on the validation and rules applied during import are listed below the screen shot.



Validation that OpenClinica Performs on the Data Import File

- XML is well-formed against the ODM.
- The Study OID in the import file matches the current Study in the OpenClinica session you are performing the import from.
- Study Subject OIDs exist for the current Study.
- Study Event Definition OIDs exist for the current Study.
- OIDs for the CRF versions exist for the Study Event Definition OID.
- Study Subject OIDs are scheduled for the Study Events.
- Data values for the Items comply with the metadata. For example, if the metadata for an Item calls for a value of 1 or 2, the data value for that Item in your import file must be 1 or 2.
- Data values are in the correct format. For example, when an integer is required, a decimal

value generates an error.

Edit Checks that OpenClinica Performs on the Data Import File

If your study is configured to flag errors in Edit Checks, OpenClinica checks the following:

- The status of the CRF (prior to import) is not "double data entry started" or "data entry complete."
- Data is in the correct format. For example, a date that is not in the required YYYY-MM-DD format generates an error.
- Data values are within limits specified for CRF Items. For example, when the CRF accepts 90 to 110 degrees for temperature, a value of 200 is not accepted.
- Data values match acceptable criteria for CRF Items. For example, if a Yes answer is required for an Item and the value in the import file for that Item is No, a discrepancy note is required.
- Required fields are not empty.

2.5.6 Example Data Import File

This sample file is provided as an example only - your OIDs and item values will vary.

Click the file to open or download it, using the standard features in your browser for accessing linked files.

For best results, copy the content of the file to an XML editor and save as an XML file type. XML editors Notepad++ (for Windows) and TextWrangler (for Mac) are available for free download on the internet.

- Simple Data Import File with One Item Group: One Subject, one Event, one CRF, one Item Group, multiple Items.