User manual OpenClinica Data Importer (OCDI 2.0)

Created by
Ilona Pinter
Jolanda Strubel

Version:
1.3.3 (update Rene Breet)
june 2017

User Manual OpenClinica data importer Version_1_3-3.docx

Contents:

Abreviations:

OC   OpenClinica
OCWS   OpenClinica Webservices
OCDI   OpenClinica Data Importer
TraIT   Translational Research IT

This manual was created with great care, yet errors are possible. In no event does The Hyve or TraIT accept liability for consequential damages resulting from the use, misuse or inability to use the OCDI. If you have remarks on this manual, contact us via servicedesk@ctmm-trait.nl.

OCDI 2.0 supports recent versions of Firefox and Chrome, IE version 11 and higher

1. Introduction

The OpenClinica Data Importer (OCDI) is an open source web application. It is created by TraIT in collaboration with The Hyve. It enables the upload of clinical data to an OpenClinica (OC) study via OpenClinica Web Services (OCWS).
New subjects and unscheduled events can be created in OC with the OCDI. All data files need to be in a predefined structure (see Appendix 2). The data is validated before uploading to OC database.

Location of the web application OCDI 2.0: https://www.openclinica.nl/ocdi/views/login?logout

The OCDI user interface will guide the user through the upload in several steps (see Figure 1).

Figure 1: Steps within the OCDI tool:
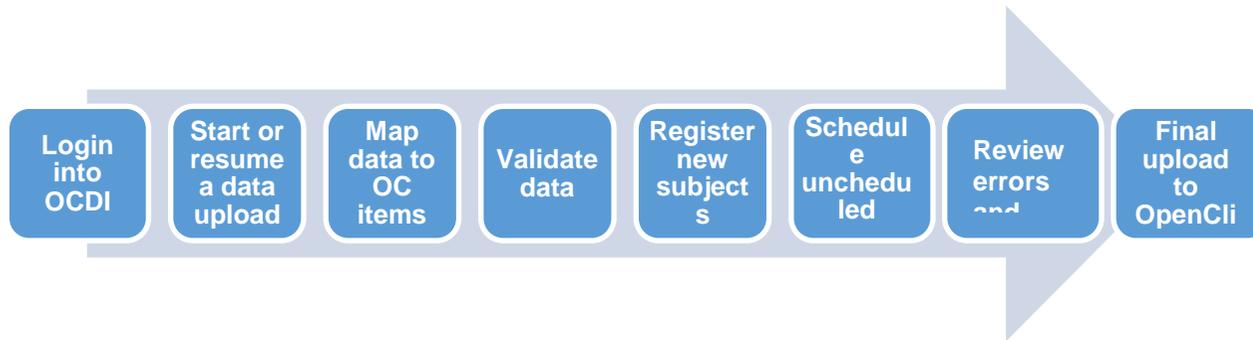
The data file that can be uploaded with OCDI needs to meet the following requirements:

- The format of the file is tab-delimited text file (encoded in UTF-8). The column names are defined as indicated in Appendix 2 of this manual.
- Only one event, one CRF and one CRF version are allowed per upload.
- Cells in the dataset may be left empty, unless the item is required in OC. For non-required items the empty cells are ignored.
- Partial dates (Pdates) are not accommodated by OCDI.
- The dataset should <u>only</u> contain items that are present in the target CRF version. If additional items are present in the dataset the entire upload will be blocked by OCDI (also see 'known issues' section, OCDUAT-193 and -51).

Need to knows

❏ Data upload is a complex procedure. Therefore it is mandatory to perform a complete data upload on the OpenClinica Sandbox (test) environment before a data upload on the OpenClinca production environment is allowed.
❏ User must have authorization for SOAP web services in the OpenClinica environment. Contact the TraIT servicedesk if not.
❏ OCDI accommodates partial uploads. So, if you ignore error messages and proceed with the upload the invalid data will be removed from the upload and only the valid data will be uploaded.
❏ If a required field is not present in the dataset, the entire upload will be blocked for that subject.
❏ If the Study Parameter Configuration is changed in OC from Date of Birth required to Year of Birth (YOB) required, it will not be possible to create subjects using the YOB. This is a general issue with OC, so not specific for subject creation with OCDI. Please contact the TraIT Servicedesk if this is the case in your situation.
❏ Subject attributes that are not set to required in OC (study parameter configuration) can't be uploaded using the OCDI tool. For example, gender is not required, but you would like to fill it. The workaround for now is to put the setting temporarily on required, upload the data and then change it back.

❑ Person id: if a person id that you upload via subject registration file already exists in the OpenClinica environment (i.e. used for a subject in another study) then the subject attributes -like DoB and sex- of that other subject will be linked to your subject. You won't notice this, unless you check the subject attributes after uploading in OC is completed. To avoid problems, we provide a warning in OCDI to contact the TraIT servicedesk in case you upload Person ids via the subject registration file. See also OCDUAT-224 in the Bugs / known issues section.

❑ If data are uploaded to existing subjects in OC, site specification is not required for subjects that have been created on site level. If a site is indicated for subjects that have been created on study level a mismatching site error message will be thrown.

❑ OCDI uploads data to 'signed' Events. In the process OCDI gives a warning "for subject x the Event has status data entry complete", but does not indicate when the Event is signed (see 'known issues' OCDUAT-172)

❑ The current version of OCDI does not support Event start time. If Event start time is indicated in the Event registration file, it is not uploaded to OpenClinica. (this is a bug)

❑ It shouldn't be possible to schedule an event if it does not match the event indicated in the dataset. If this is the case then OCDI provides an error message "event in the event scheduling file does not match the event in the data file". However, if you ignore this error and click "next" then OCDI will schedule the event that is indicated in the event scheduling file. Data are not uploaded because of the mismatch. (this is a bug)

❑ In case of a repeating event: OCDI only checks whether required fields are present in the dataset at the first (or one of the) event occurrences for a subject. If required fields are left empty for one/more occurrences while present in at least one other occurrence OCDI will not throw an error and upload will be performed for that subject. Consequently, OC CRF(s) contain empty required items. A validation error will occur in OC when you edit the CRF ("required item left empty") (see bugs / known issues OCDUAT-192, -130-, 131)

❑ In case of a repeating event: if an occurrence for a certain subject is left empty, then this occurrence will be ignored during the upload. So, it is not possible to schedule an occurrence if no data are uploaded to it.

❑ CRF versions:
- OCDI cannot check if CRF version that is indicated in the dataset is the same as selected as default version on site or study level. Data will be uploaded to the CRF version that is indicated in the dataset, even if that version is not in use in the OC study. In fact, this is not an issue but expected behavior: via the UI it is also possible to enter data in other CRF version than the one selected as default.
- OCDI cannot handle changes in the CRF regarding simple conditional display settings. Changes in a new CRF version are not taken into account.

- OCDI cannot handle changes in the CRF regarding single/multiple-selects. Changes in a new CRF version are not taken into account. E.g. in version 0.1 the allowed values are 'Aap, Noot, Mies'. Subsequently a new CRF-version is released with a new value: 'Aap, Noot, Mies, Wim'. OCDI cannot detect this change and therefore can't upload value 'Wim' (this is a bug)
❑ In the mapping step <u>all dataset items</u> (on right side of the screen) should be mapped to CRF items. If you leave items unmapped, the entire upload will be blocked (this is a bug)
❑ The status of the CRF can be changed via OCDI by indicating the desired CRF status after upload. Be aware that it is not possible to change the status of the CRF when it has the status 'marked complete' in OpenClinica. It is possible to upload data to a CRF that is marked complete, but the status cannot be changed.
❑ In case default value is defined in the CRF, this value must be provided in the dataset. OCDI does not throw an error message if default value is missing in the dataset, but the entire record (entire upload for that subject) will be blocked. This is a bug.
❑ In the dataset underscores in item names are used to indicate repeating groups. Please remove underscores from the item names (column headers) in the dataset. You can manually map the items to the appropriate CRF items in the OCDI mapping step.
❑ OCDI ignores validation rules that are defined in OC for the particular CRF.

1. Login into OCDI

On the OCDI login page select the appropriate OC environment from the dropdownlist. Use the OC login details for that environment. The credentials are only used by OCDI to connect to the OCWS and are not stored.
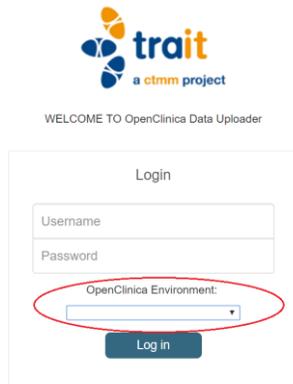
Changing your password can only be done in the appropriate OC instance.

OCDI will open in the data upload screen after logging in.

2. Start a data upload

In the data upload page you can:
- Start a new upload session by uploading a new data file.
- Resume a previous, interrupted upload (bottom of the screen).

The data will be validated at different steps during the OCDI process. To prevent errors at that stage, make sure your data file is setup correctly before uploading it.

2.1. Preparing the data file

The data file should contain data for only one study, one event and one CRF version. Data are imported into the indicated CRF version even if this is another CRF version than the one indicated in your OC study.
If the event is a repeating event, data for more than one occurrence can be uploaded. The details of the requirements for this file can be found in Appendix 2.1.

The following general conventions are in place for the data:
- If no value is provided in the data file for a required item in the OC CRF , the data for that CRF (row) of that subject will not be uploaded in OC. For repeating item groups required fields can be left empty , if all items for the repeating item group are left empty.
- Data are not uploaded to items which are hidden in another way than by "Simple Conditional Display".
- Hidden items (using "Simple Conditional Display") are only validated when the display condition is met (e.g. "Did any AE occur" is answered yes, then values for the entered AEs are validated by OCDI).
- The data types "File", "Calculation" and "Group-calculation" are not allowed in OCDI.
- Fields that are left empty which are not required in OC will be ignored during the upload. So, existing data in OC CRFs will not be 'erased' by empty fields in the dataset.

Next to checking whether the value is conform the format indicated in OC (e.g. string, integer etc) additional validation is performed for the following data types:

Numeric value:
values are validated according to the ranges specified in OC. Regular expression validations are not performed by OCDI. The width (=number of characters) should be smaller or equal to the setting in OC (including the decimal sign for a Real number).

Real number (REAL):
if a decimal separator used (is not mandatory) it should be a dot.

Converting data with a comma to data with a dot as decimal separator can also affect other fields, in which a comma is used (eg. for items for which more than 1 value is allowed, the values are separated by a comma). A workaround is to open the data in Excel, select the only the columns with Real data types and replace commas for dots in the selection.

Integer (INT):
no decimals are allowed (e.g. 1.0)

String value (ST):
the number of characters should be smaller or equal to the setting in OC. In most cases no width will be defined for this data type in OC, which may give problems. The workaround for now is to set the "width_decimal" setting in OC for all free text items to 255 (the maximal value allowed in OC).

Date:
the format should be dd-mm-yyyy. "Leading zeros" are removed in Excel files, but are needed in OC. You can circumvent this by putting an apostrophe in front of the date value (e.g. '01-04-2002) or by opening your file in a text editor.

Partial Date (PDate):
can't be imported using OCDI. Either use a normal date field or create separate items for (month and) year.

Radio buttons and single select:
OCDI checks whether the value is present in the response values indicated in OC.

2.2.   *Start a new data upload*
To start a new upload
- Give a unique and meaningful name to this upload. The upload is saved under this name and can be resumed if for some reason the upload was not finished
- Click on the '**Choose File**' button and select a file

● Click on the '**Next**' button to continue
The application will check if the data file has the right format and structure. It will return an error message if this is not the case. Otherwise it will bring you to the next step: the mapping screen.

2.3.   Resume a previous upload

All uploads previously started by the user are shown at the bottom of the screen. Click on the name to open it. The upload will open in the same stage as it was in, when it was interrupted.
Please note that an upload session is linked to one specific data file. If you change data in that file (e.g. because errors were present), you need to start a new session with the corrected file. The subject file and event file can be replaced with updated versions within an upload session.

3. Mapping data file to OC items

In this step the columns in the data file are mapped to the corresponding items in OC.

3.1.   Mapping

All items present in OC in the indicated CRF in the data file are shown on the left side. Columns with identical names are automatically mapped. The data file columns that could not be matched are at the right side.
You can scroll through both lists independently and filter on (part of the) item name using the field at the top of the screen to shorten the list.

The columns from the data file that are not be mapped automatically, can be mapped manually by dragging and dropping them to the left hand side. The following additional actions are available:

● '**Clear mapping**' deletes the complete mapping (including automatically mapped items).

- '**Match item names**' automatically maps the items from the data to the items in OC if the item names are identical. This will overrule any manual mapping made. Therefore, make sure that you first match item names before continuing with a manual mapping.

Confirm the mapping by clicking on the '**Next**' button. The validation of the data is started if no mapping errors are present.
Please note: fields in the data file that could not be mapped, because there are items present in OC, will block the complete import of the data. Remove any fields from the original data file that are not importer in OC.

3.2. Data Validation

The data file is validated to prevent an unsuccessful upload to OC or data inconsistencies.
Error and warning messages are reported on the '**Feedback of Data Upload**' screen, grouped by error type. For each error the Item name and study subject ID is indicated.

4. Register new subjects

The application detects if the data file contains subjects that are not registered in OC. If your file contains unregistered subjects, the subject registration screen appears. In this step you can:
1. Download a template for the subject registration file
2. Upload a subject registration file
If your data file only contains existing subjects, this will be indicated and the step will be skipped.

The registration of subjects with "Person IDs" in OC is not supported by the OCDI. Please make sure that the field Person ID is not required in your study before uploading subjects.

4.1.　Download and complete the subject registration file

Download the template file by clicking the "Download subject template" button. This template will indicate which subjects need to be registered and which columns are required (e.g. "Gender", "Year of birth") depending on the settings in OC. Don't remove or add any columns, only complete the data in this template.

The "Study subject ID" and "Study" columns are prefilled with the appropriate values from the data file.

The file always contains a column "site", even if no sites are defined in OC. Entering data in this column is optional, even when sites are defined. When providing a site for a subject this subject will be registered at site level. If no site is provided the subject will be registered at study level.

The subject registration file is a tab separated file. Use Excel or another spreadsheet editor to get an overview of which value belongs to which column. Excel removes "leading zero's" in date fields, but these are needed in OC. You can prevent this by putting an apostrophe in front of the date (e.g. '02-12-2014), which will convert it into text. You can check in a text editor whether your date values are correct.

4.2.　Upload a subject registration file

Complete and save the template subject registration file and upload it in OCDI:
- Click on "bestand kiezen" and select file.
- Click on the "next" button to continue.

The file will be validated and any validation errors will be shown on the "Feedback of subject registration" screen.

The actual subject registration in OC will occur in the last step in the process ("upload data to OC").

5. Schedule events

The application detects if the uploaded data file contains events in the data file that is not yet scheduled in OC for that subject (unscheduled events). If so, the "event scheduling" screen appears.

In this step you can:

1. Download a template for the "event scheduling file".
2. Upload an event scheduling file.

5.1.   Download and complete the event scheduling file

Download the template file by clicking the "Download event template" button.

In this file the "Study subject ID", "Study name", "Event name" and "repeat number" are prefilled. The "event repeat" column is mandatory and the value should be "1" for non-repeating events. Leave the prefilled values as provided in the template, as this represents the data in the data file. Changing these values can give rise to errors in the uploading process (for instance changing the event to another unscheduled event will give a warning in OCDI, but the event will be scheduled in OC after all. No data will be uploaded however).

If column "Site" is prefilled by the system with value "null" this value must be deleted.

This is a tab separated file. Use Excel or another spreadsheet editor to get an overview of which value belongs to which column. Excel removes "leading zero's" in date fields, but these are needed in OC. You can prevent this by putting an apostrophe in front of the date (e.g. '02-12-2014), which will convert it into text. You can check in a text editor whether your date values are correct.

The "Event Start Date" is mandatory. If not provided, the event will not be uploaded to OC.
The "Event Start *time"* can be provided but will not be uploaded to OC in the current version of the OCDI.

5.2.   Upload an event scheduling file

Upload the completed event scheduling file. Click on the "next" button to continue. The file will be validated and any validation errors are shown in the "Feedback event scheduling" screen. The actual event scheduling in OC will occur in the last step in the process ("upload data to OC").

6. Final upload to OpenClinica

The "final overview" screen is shown. This contains a summary of all errors and warnings. You may ignore the errors and click "next", but in that case the invalid data will not be uploaded to OC. The other option is to click "back" and correct the erroneous data, or abort the process, correct the data and restart the process again.

Next the "upload page" is shown. On this page you indicate:
- what status the CRFs should have after uploading
- to which CRF status data upload is allowed

The actual upload of the data, new subjects and unscheduled events starts after clicking on the "upload" button. This process can't be stopped.

The upload will be done for each subject separately. If data for one subject is not accepted by OC the upload will continue and an error message indicating which subject upload failed is shown at the end.

Please note that it is not possible to change the status of an EventCRF to 'data entry started' if the initial status is 'marked complete' in OC.

Appendix 1: Known issues / bugs

- OCDUAT-51: all dataset items should be mapped to CRF items. If not, the entire upload will be blocked
- OCDUAT-106a: there is a problem with the CRF-versions for single/multiple-selects. OpenClinica does not report back changes in the allowed values in a new CRF-version. E.g. in version 0.1 the allowed values are 'Aap, Noot, Mies'. Subsequently a new CRF-version is released with a new value: 'Aap, Noot, Mies, Wim'. OCDI cannot detect this change and therefore can't upload value 'Wim'..
- OCDUAT-106b: For the case where an integer (or real) is mapped to an item that is a codelist (single select) two error messages are provided instead of one. The messages are: 1) One or more of your items exceeds allowed length, and 2) One or more values in your data is not present in the multiselect or codelist for that item.
- OCDUAT-131 and -130: Correct error/warning message on missing mandatory field (non group item) - one event. No warning if dataset contains repeating events, where the concerning mandatory item is missing in first event, while present in the 2e event. It appears that the system only checks the presence of one (last?) mandatory item in case of repeating events.
- OCDUAT-176: When removing / changing a simple conditional display in OpenClinica CRF's then OCDI does not take this into account. So, OCI works with simple conditional display definition of the first CRF version.
- OCDUAT-184: when you upload to a repeating event with event status "stopped" OCDI does not throw an error message. However, data are not uploaded in OC.
- OCDUAT-191: if 'century indication' is missing in a date value (e.g. 01-01-62) in subject registration file, then OCDI does not report an error after data validation. However, the creation of that specific subject is blocked by the issue. Error message produced on the last screen: "creating subject_x failed, OC error: Unparsable date: 0062-01-01"
- OCDUAT-193: No error is reported by OCDI when uploading an item to a CRF-version where that item does not exist. At the last page OCDI returns the following error: "_Failed upload for subject EV-OCDUAT187-09. Cause: No Item could be found with the key . _". Steps to reproduce: create a new CRF-version and add a new item to it. Add that new item to the data file but specify the previous CRF-version. The error **blocks the entire upload**

- OCDUAT-205: in the subject validation step only one type of error on date values is indicated. Example: in subject registration file different date errors are present for different subjects. First only error reported "date can't be in the future". After correcting that error, the second error was reported "wrong date format". After correcting that error, the last warning was given "no date provided, todays date is used"

- OCDUAT-211 and -191: no error message is given by OCDI in case leading zero's are missing for date of birth in subject registration file. Missings zero's are added and the subject is created. E.g. if DoB 1-1-1960 is provided in the file, the subject is created in OpenClinica with DoB 01-Jan-1960. Note, missing leading zero's in dates are reported as error if provided in the datafile.
- OCDUAT-224 (person id): If a person id that you upload via subject registration file already exists in the OpenClinica environment (i.e. used for a subject in another study) then the subject attributes -like DoB and sex- of that other subject will be linked to your subject. You won't notice this, unless you check the subject attributes after uploading in OC is completed. To avoid problems, we provide a warning in OCDI to contact the TraIT servicedesk in case you upload Person ids via the subject registration file.

Appendix 2: Format data files

Three types of files can be uploaded in the OCDI: the Data File (data that should be imported in the OC CRF), Subject Registration file (with new Subjects, if applicable) and the Event Scheduling File (with subjects for which an event needs to be scheduled, if applicable). The column names for these files need to be exactly the same (including capitals) as defined in the tables below. In these tables it is also indicated when columns are mandatory in the file. Mandatory means that if no value is provided, the data for the concerning row will not be uploaded into OC.

2.1 Data file

Only one event, one CRF and one CRF version can be present in the data file. Use the column names indicated in table 2.1.1

Table 2.1.1: setup of the data file

Column name	Mandatory	Comment
StudySubjectID	YES	StudySubjects must be unique within the study/site
Study use the 'Unique Protocol ID'	YES	
Site Use the 'unique identifier'	NO	This column is always present but is optional. If sites are defined in OC and this field is left empty for a new subject, the subject will be created at study level.
EventName	YES	
EventRepeat	YES	1 for non-repeating. Gaps in EventRepeat numbers are not supported. The occurrence numbers are made consecutive after uploading in OC.
CRFName	YES	
CRFVersion	YES	
<Item Name>	YES	<Item Name> = the name used in the CRF.
<Item Name>_<Group Repeat>	YES (in case of repeating item)	<Item Name> = the name used in the CRF. <Group Repeat> = the repeat number to which the data belongs (required if item is repeating).

Table 2.1.2: example of a data file (Item1 is not repeated, Item2 is repeated)

StudySubjectID	Unique Protocol ID	EventName	EventRepeat	CRFName	CRFVersion	Item1	Item2_1	Item2_2
TR01	TRACER-ESRA	Repeating Visit	1	Lab	0.1	1234	basd	gffa
TR01	TRACER-ESRA	Repeating Visit	2	Lab	0.1	4325		

2.2 Subject Registration file

If subjects are present in the data file that are not present in OC, a subjects registration template is created by OCDI. Use this template to enter your subject data.The template contains the following columns:

Table 2.2.1: Setup of the subject registration file (created by OCDI)

Column name	Mandatory	Comment
StudySubjectID	YES	The downloaded template file will always contain this column
Gender	NO	Present and mandatory only if study requires it
DateOfBirth /	NO	Present and mandatory only if study requires it
DateOfEnrolment	YES	Date of enrolment is mandatory. If not provided, the date is set to today. Required format is dd-mm-yyyy.
SecondaryID	NO	Mandatory if study requires it
Unique Protocol ID	YES	The downloaded template file will always contain this column
Site	NO	The column is always present but optional (does not require values)

20

Table 2.2.2: example of subject registration template (as created by OCDI)

StudySubjectID	DateOfEnrolment	Gender	Unique Protocol ID	Site (optional)
TR01			TRACER-ESRA	
TR02			TRACER-ESRA	
TR03			TRACER-ESRA	

Table 2.2.3: example of completed subject registration file (see table 2.2.2 for the template), ready to be uploaded in OCDI

StudySubjectID	DateOfEnrolment	Gender	Unique Protocol ID	Site (optional)
TR01	15-12-2015	M	TRACER-ESRA	1
TR02		F	TRACER-ESRA	1
TR03	02-01-2016	M	TRACER-ESRA	

2.3 Events Registration

The columns mentioned below are present in the template that is created by OCDI if subjects are present in the data file with an unscheduled event in OC. Leave the column names as is and just enter the values for the events that need to be created in OC.

Table 2.3.1: Setup of the event scheduling file (created by OCDI)

Column name	Mandatory for User	Comment
Study Subject ID	YES	The template will always contain this column
Event Name	YES	The template will always contain this column
Unique Protocol ID	YES	The template will always contain this column
Site	Not Applicable	leave empty (delete "null" value, if present)
Start Date	YES	The template will always contain this column, currently mandatory
Start Time	NO	The template will always contain this column
End Date	NO	The template will always contain this column
End Time	NO	The template will always contain this column; mandatory if values are present for Start date and time and End date.
Repeat Number	YES	1 for non-repeating events. The template will always contain this column

Table 2.3.2: example of the event scheduling template (created by OCDI)

Study Subject ID	Event Name	Unique Protocol ID	Site	Start Date	Start Time	End Date	End Time	RepeatNumber
TR01	Repeating Visit	TRACER-ESRA	null					
TR02	Repeating Visit	TRACER-ESRA	null					

Table 2.3.3: example of the completed event scheduling file (see table 2.3.2 for the template), ready to be uploaded in OCDI

Study Subject ID	Event Name	Unique Protocol ID	Site (leave empty)	Start Date	Start Time	End Date	End Time	RepeatNumber
TR01	Repeating Visit	TRACER-ESRA		12-12-2016				
TR02	Repeating Visit	TRACER-ESRA		18-05-2016	12:00	25-12-2016	10:00	