

4 Getting Data for Your Study: Reports and Extracts

For Information on Getting Data for Your Study, See the Following Sections:

Approved for publication by Kerry Tamm. Signed on 2020-11-19 1:22PM

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

4.1 Data Review Tables

Definition:

•

Data Review Tables: Custom tables of form data used to review data and perform bulk actions in Study Runner.

Potential Uses:

- **Remove erroneous form data after an import:** Quickly review form data in bulk after importing data (such as data from devices). If issues are noticed, the form data can be removed without having to go into each form individually. The issue can then be addressed, and the data can be imported correctly.
- **Bulk-close automatic queries:** If automatic queries are set to appear incorrectly and the issue has been resolved, these automatic queries can be resolved in bulk.
- **Review information to catch potential errors:** Create a data review table to track a participant's weight from visit to visit and check for any potential errors. If an error is suspected, a query can be added or the value can be updated.

Users:

- Investigators, Clinical Research Coordinators, Data Managers, Data Entry Persons, and Data Specialists.
- Monitors can access Data Review Tables where they can view and close queries, but they cannot remove or restore forms.
- Site/Study Viewer users are not currently able to access Data Review Tables.

Data Review Tables are first created and configured within Study Designer by a Data Manager. After the table is configured and the study is published, the table can be accessed from the **Tasks** menu in Study Runner.

Table Features:

- Filter and sort data by any column
- Bulk-close open Queries on any items in the table
- Bulk remove/restore primary forms
- Drill into each data point in the source form
- Download the entire report to review or print

Each custom table includes items from a primary form. The primary form can either be in exactly one repeating event or in one or more non-repeating events. The table can include items from one or more additional forms in non-repeating events. If any forms within the table have permission tags, the entire table will only be available to users with those permission tags.

Accessing Data Review Tables:

1. From the **Tasks** menu in Study Runner, select the table name in the **Data Review Tables** section.
 - a. The **Data Review Tables** section will only appear if a table has been created for the study and the user has access to the Data Review Table.

Sort, Filter, and Review Tables:

1. Click **Active Records** to display the dropdown menu and select **Removed Records** to display *Removed* and *Archived* primary forms.
 - a. To toggle back to the *Active Records* view, click **Removed Records** and select **Active Records** from the dropdown menu.
2. Click the **arrows** or the heading name for any column to sort the column ascending or descending. Click the red **X** to clear the filter.
3. Click the **Filter** icon to filter the table to include only records with specific values. After selecting the filter criteria, click **Apply** to activate the filter.
 - a. Click the **Filter** icon again to update the filter, or click **Clear** to remove it.
4. The column headings will display a label for the required information columns or will display the item name.
5. Hover over the column heading to see which form the information is coming from. The hover text will display "primary form" after the form name if it is the primary form of the Data Review Table.
6. Click the **value** in a cell to open the form in *Edit* mode for the item that was clicked. Columns with a *Select One* item type will display the label and then have the coded value in parentheses.
7. Click the **Participant ID** to open the Participant Record on the Participant Details Page.
8. Click the **Download** icon to download a copy of the Data Review Table to review outside of OpenClinica and/or print a hard copy.

Participant ID	Site	Event Start Date	Form Status	Event Name	Event Status	Form Last ModifiedBy	Form Last Modified Date	Were any medical conditions or events re...	What was the medical condition or condition...	What is the Medical History - primary form
123	San Francisco	01-Mar-2024	data entry complete	Baseline Visit	data entry started	klambert@openclinica.com	08-May-2024	yes (1)	06-May-2024	headache
101	San Francisco	19-Apr-2024	data entry complete	Baseline Visit	data entry started	klambert@openclinica.com	08-May-2024	no (2)		
500	San Francisco	07-May-2024	data entry complete	Baseline Visit	data entry started	klambert+crc	08-May-2024	yes (1)	22-Apr-2024	heart attack
159	San Francisco	15-Mar-2024	initial data entry	Baseline Visit	data entry started	klambert+crc	07-May-2024	yes (1)		
006	The Best Hospital	30-Apr-2024	initial data entry	Baseline Visit	data entry started	klambert+crc	30-Apr-2024	no (2)		

Note: Forms with a status of "Not Started" will not appear anywhere on the Data Review Table.

Select Form Rows vs. Queries:

1. Select the checkbox in the checkbox column header to select all form rows and queries in the table.
2. Select an individual form row by clicking the checkbox in the first column of the table. This will also select the checkbox for any queries on the form row.
3. Select the checkbox next to any query to individually select one or more queries.
4. Select all open queries for a particular item by clicking the checkbox in the item column header.

<input type="checkbox"/>	Participant ID	Site	Event Start Date	Form Status	Event Name	Event Status	Form Last ModifiedBy	Form Last Modified Date	Were any medical conditions or events re...	What was the medical event or condition ...	What is the medical condition or event t...
<input type="checkbox"/>	500	San Francisco	07-May-2024	data entry complete	Baseline Visit	data entry started	klambert@ope nclinica.com	08-May-2024	yes (1)	22-Apr-2024	heart attack
<input type="checkbox"/>	123	San Francisco	01-Mar-2024	data entry complete	Baseline Visit	data entry started	klambert@ope nclinica.com	08-May-2024	yes (1)	06-May-2024	headache
<input type="checkbox"/>	101	San Francisco	19-Apr-2024	data entry complete	Baseline Visit	data entry started	klambert@ope nclinica.com	08-May-2024	no (2)		
<input type="checkbox"/>	159	San Francisco	15-Mar-2024	initial data entry	Baseline Visit	data entry started	klambert+crc	07-May-2024	yes (1)		

Note: Locked primary forms will not have checkboxes available because the form cannot be removed/restored when it is locked. Similarly, there will not be a checkbox for the queries which have a locked primary form because these queries cannot be closed in bulk.

Bulk Remove and Restore Primary Forms:

1. From the **Active Records** view, select the checkbox next to each form row you would like to remove.
2. Click **Remove Forms** to remove the primary form for the rows selected.
3. Add a reason for removal and click **Yes** to confirm.

☰ Active Records Remove Forms Close Queries						
<h2>Medical History</h2> <p>Medical History</p>						
<input type="checkbox"/>	Participant ID	Site	Event Start Date	Form Status	Event Name	Event Status
	↑↓ ▾	↑↓ ▾	↑↓ ▾	↑↓ ▾	↑↓ ▾	↑↓ ▾
<input checked="" type="checkbox"/>	500	San Francisco	07-May-2024	data entry complete	Baseline Visit	data entry started
<input checked="" type="checkbox"/>	123	San Francisco	01-Mar-2024	data entry complete	Baseline Visit	data entry started
<input type="checkbox"/>	101	San Francisco	19-Apr-2024	data entry complete	Baseline Visit	data entry started

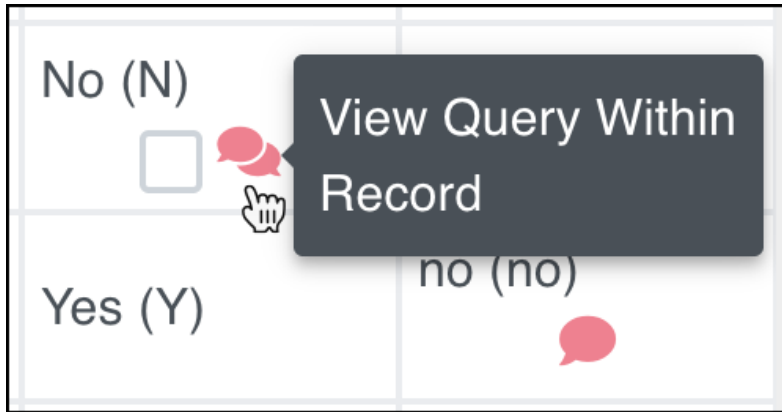
1. From the **Removed Records** view, select the checkbox for each form you would like to restore.
2. Click **Restore Forms** to restore the primary form back to an active record.
3. Add a reason for restoration and click **Yes** to confirm.

☰ Removed Records Restore Forms								
<h2>Medical History</h2> <p>Medical History</p>								
<input type="checkbox"/>	Participant ID	Site	Event Start Date	Form Status	Event Name	Event Status	Form Last ModifiedBy	Form Last Modified Date
	↑↓ ▾	↑↓ ▾	↑↓ ▾	↑↓ ▾	↑↓ ▾	↑↓ ▾	↑↓ ▾	↑↓ ▾
<input checked="" type="checkbox"/>	006	The Best Hospital	30-Apr-2024	initial data entry	Baseline Visit	data entry started	klambert@openclinica.com	08-May-2024
<input checked="" type="checkbox"/>	159	San Francisco	15-Mar-2024	initial data entry	Baseline Visit	data entry started	klambert@openclinica.com	08-May-2024

Note: When removing forms, you will remove the entire primary form even if all items are not shown on the custom table. Any other data from non-repeating events included beyond the primary form will not be removed.

View Queries:

1. Click the **Queries** bubble to individually view and respond to a query within the record.
 - a. Queries within Locked forms will open in read-only mode and cannot be responded to.



Bulk Close Queries:

1. Select the checkbox for each query you want to close. Alternately,
 - a. Select the checkbox for the form row to select all queries in the row; or
 - b. Select the checkbox for the item column to select all queries for the item.
2. Click **Close Queries**
3. Add a comment and click **Yes** to confirm.

Medical History											
Medical History											
<input type="checkbox"/>	Participant ID	Site	Event Start Date	Form Status	Event Name	Event Status	Form Last ModifiedBy	Form Last Modified Date	Were any medical conditions or events re...	What was the medical event or condition ...	What is the medical condition or
	↑ ↓ ▾	↑ ↓ ▾	↑ ↓ ▾	↑ ↓ ▾	↑ ↓ ▾	↑ ↓ ▾	↑ ↓ ▾	↑ ↓ ▾	↑ ↓ ▾ Ⓞ	↑ ↓ ▾ Ⓞ	↑ ↓ ▾ Ⓞ
<input type="checkbox"/>	005	The Best Hospital	30-Apr-2024	data entry complete	Baseline Visit	data entry started	klambert@openclinica.com	09-May-2024	no (2)		
<input type="checkbox"/>	123	San Francisco	01-Mar-2024	data entry complete	Baseline Visit	data entry started	klambert@openclinica.com	08-May-2024	yes (1)	06-May-2024	headache
<input checked="" type="checkbox"/>	101	San Francisco	19-Apr-2024	data entry complete	Baseline Visit	data entry started	klambert@openclinica.com	08-May-2024	no (2)		
<input type="checkbox"/>	004	The Best Hospital	30-Apr-2024	initial data entry	Baseline Visit	data entry started	klambert+crc	30-Apr-2024	yes (1)		
<input type="checkbox"/>	003	The Best Hospital	30-Apr-2024	data entry complete	Baseline Visit	data entry started	klambert+crc	30-Apr-2024	yes (1)	15-Jun-2022	headache
<input type="checkbox"/>	002	The Best Hospital	30-Apr-2024	data entry complete	Baseline Visit	data entry started	klambert+crc	30-Apr-2024	yes (1)		

Showing 1 to 10 of 10 entries << < 1 > >> 20 ▾ Ⓞ

Note: Monitors will not have the **Remove Forms** or **Restore Forms** buttons.

Use the [Create and Configure Data Review Tables](#) page for information on creating Data Review Tables as a Data Manager.

Functional approval by Kate Lambert. Signed on 2024-05-10 3:31PM

Approved for publication by Paul Bowen. Signed on 2024-07-01 11:52PM

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

4.2 Back-End Access Via Insight

You may have existing reporting/visualization/statistical analysis tools you want to use to track, analyze, or report on your OpenClinica data. The most common method for doing this is using downloadable data extracts from the OpenClinica user interface. However several other options exist, including using OpenClinica's API, using the OpenClinica Insight reporting tools, using Insight's API, or by establishing a direct read-only connection to the Insight data warehouse. This guide provides an overview of accessing data through a direct read-only connection to the Insight data warehouse. For existing institutional toolchains, such as SAS, Qlik, Tableau, SSRS, Shiny, Jupyter, Stata, R, Python, Power BI, etc, the customer can establish a direct, live, readonly connection with the Insight database. This method can also be used to incorporate your OpenClinica data into other databases via Postgres Foreign Data Wrappers (FDWs, or linked tables). Since the Insight back-end is a PostgreSQL database, allowing connections is a matter of setting up a secure, read-only database connection.

Pros:

- Works where SSH does: Linux, Windows (PuTTY), etc.
- Use the tool of your choice, so long as it supports remote connections
- Live access to data, not snapshots as in data extracts (can be pro & con)

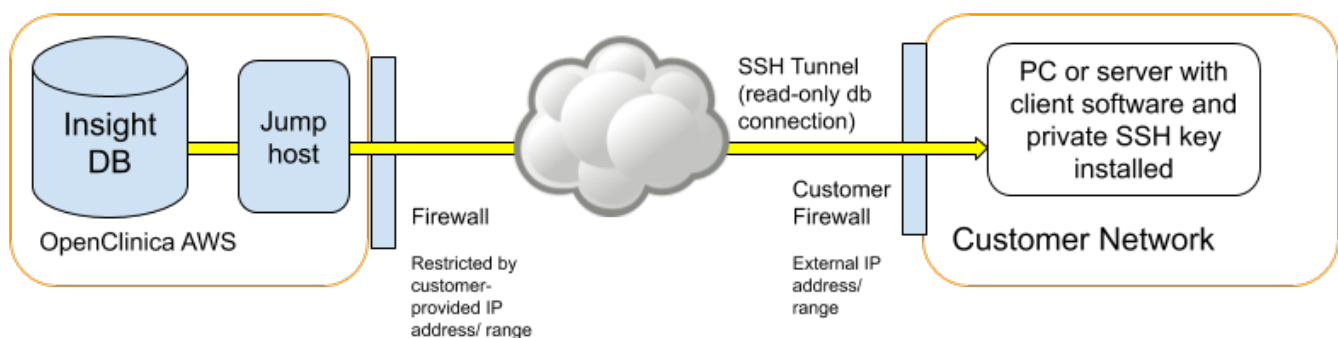
Cons:

- Mostly static data permissions (schema-level)
- User management is separate from the OpenClinica and Insight web interfaces
- OpenClinica team will support setup and maintenance of your connection, however, assistance with connectivity and use of your particular tool (such as SAS or Tableau) is best-effort. We are not able to support the wide variety of possible clients (SAS, Tableau, Power BI, etc) and cannot guarantee end-to-end connectivity, functionality, or level of service.

There are Three Main Parts of the Set Up:

1. Securing the connection itself
2. Authenticating the user that is connecting
3. Defining what the user is authorized to do

We do this with a secure, read-only, direct interface to the Insight Data Warehouse, via an SSH tunnel. It will require the involvement of your internal IT team



Technical Information

Connections to the Insight back-end (PostgreSQL) database are secured by using a SSH tunnel via a

jump host. A "jump host" means that your Insight server is not directly reachable from the Internet, so instead connections are made via an intermediate server that is reachable from the Internet but from which connections are allowed to the Insight server. A "SSH tunnel" means that the nature of the connection to Insight will be via a SSH (secure shell) session between your computer (or server) and the jump host. This SSH session is then configured to forward database connection requests to a certain local port (such as 65432) to the Insight server's and database port (such as 5432). Database traffic forwarded in this way is secured. The exact configuration of back-end connections on your side depends on the intended setup. A SSH client will be required on your computer or server. On Linux, SSH is installed by default (ssh). On Windows, there is a SSH tool called "PuTTY" which has a user friendly interface. Alternatively on Windows, "Git for Windows" (Git is a change tracking tool) comes with a Windows version of Linux's SSH tool.

To Set Up Back-End Access Via Insight:

1. Consider the pros and cons and decide if a back-end connection is the right solution for you.
2. Submit a support ticket to the OpenClinica team to get the process started. Include:
 - The IP addresses from which you will connect. This could be your computer IP, or a server IP. This is used to restrict access to the jump host to only those nominated IP addresses, for additional security.
 - The user names and a "PEM" format SSH public key (see below) for the user(s) that will connect. If you are setting this up for a server then there could be just one "service account" type user, or there could be multiple individual users.
 - For each user, which study environment(s) / schemata that they should have access to. The Insight "databases" shown in the Metabase front-end are actually database schemata. The default is access to all schemata, but it is possible to assign read-only access to one schema or many.
 - The public key. These are files that end with ".pub" Please do not share the private key(s).
 - Steps to set up the SSH with your preferred tool.
3. Once the OpenClinica team has set up the connection, test it with your preferred tool.

The Following Links Describe How to Create SSH Keys:

- Linux/Mac: <https://www.digitalocean.com/docs/droplets/how-to/add-ssh-keys/create-with-openssh/>
- Windows: <https://www.digitalocean.com/docs/droplets/how-to/add-ssh-keys/create-with-putty/>

Please be sure to use a passphrase for your keys so that they cannot be used if they fall out of your possession. Once the SSH tunnel is in place, it will be possible to connect to the Insight back-end PostgreSQL server. If using ODBC, the "psqlODBC" driver will need to be installed on the connecting computer/server (<https://odbc.postgresql.org/>). This driver allows applications to talk to PostgreSQL in the ODBC protocol. Then, either a DSN or connection string can be set up in the desired application. We can provide examples of these with your settings.

Approved for publication by Kerry Tamm. Signed on 2020-12-01 9:17AM

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

4.3 Extract Data

OpenClinica allows the user to select specific Events, Forms, and/or items (or all data) to be included in a dataset. The dataset can then be exported on demand or on a regular schedule.

You can create a dataset with any data that has been entered into OpenClinica. You can also combine data from different Events and Forms as needed when creating a dataset.

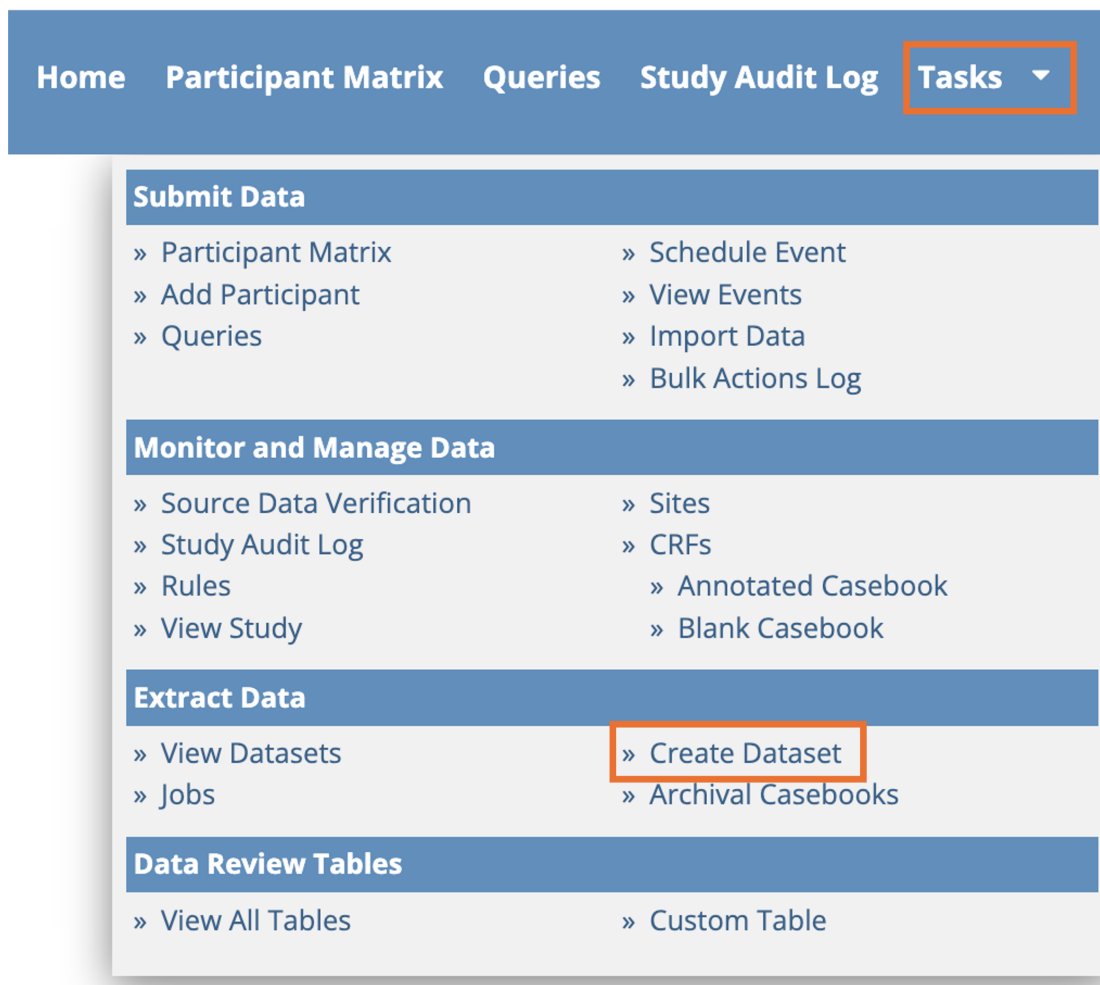
Once you define the data that you want to extract, you can choose from a number of different formats in which to view and work with that data. Once the extract is run, datasets include data from all sites and all participants within those sites.

You can also download Participant Casebooks to extract all data on a single participant. Details about this feature can be found under [Generating Participant Casebooks](#).

Archived and removed Forms are not included in any extracts by default (although archived Form Versions are included). This setting can be changed for Participant Casebooks or when using the API.

To Create a Dataset:

1. Click **Tasks** in the header bar, and select **Create Dataset**.



2. In the left panel, expand the Event that contains the Form data that you want to extract. If the Form is in multiple Events, select one Event now, and add more Events and Forms later.

OpenClinica Training Study : Site A (A) Test Environment | Change Need help?

Participant ID View → Home

Alerts & Messages ▶ **Create Dataset**

Instructions ▶ Select a CRF to choose individual (or all) items for export. You can also export participant, event, and CRF attributes.

Info

Study Events:

← Use the Info panel on the left to select CRFs

- Baseline
 - Demographics
 - Medical History
 - Physical Exam
 - Vital Signs
- Week 2
- Week 3
- Monthly Followup
- Adverse Events
- ConMeds
- Termination

Event Attributes

Participant Attributes

CRF Attributes

View Selected Items

Select All Items in Study

3. Select the Form that contains the data you want to extract.

OpenClinica Training Study : Site A (A) Test Environment | Change Need help? trainingdev (Data Manager)

Participant ID View → Home Participant Matrix Queries Study Audit Log Tasks

Alerts & Messages ▶ **Create Dataset**

Instructions ▶ Select a CRF to choose individual (or all) items for export. You can also export participant, event, and CRF attributes.

Info

Study Events:

← Use the Info panel on the left to select CRFs

Event Name: Baseline

CRF Name: Physical Exam

Description:

Select All Items

Include the selected items in this dataset:

Name	Description	Version(s)	Section(s)	Group(s)	Data Type	Units	Response Type	Response Label	PHI	Required?	Double Data Entry	Default Value	Max Repeats
<input type="checkbox"/> RG_NUM_INIT		2		R_Group1	st		calculation	RG_NUM_INIT	No	No			40
<input type="checkbox"/> RG_NUM		2		R_Group1	st		calculation	RG_NUM	No	No			40
<input type="checkbox"/> ROW_NUMBER		2		R_Group1	st		calculation	ROW_NUMBER	No	No			40
<input type="checkbox"/> BODYSYSTEM		2		R_Group1	st		text	BODYSYSTEM	No	No			40
<input type="checkbox"/> BODYSYSTEM_STATUS		2		R_Group1	st		radio	BODYSYSTEM_STATUS	No	Yes			40
<input type="checkbox"/> BODYSYSTEM_COMMENTS		2		R_Group1	st		text	BODYSYSTEM_COMMENTS	No	Yes			40

4. Select items individually, or to select all items in that Form, check the **Select All Items**

checkbox above the item list.

- (Optional) To select additional items, either from the same Form in other Events, or from a different Form, click **Save and Add More Items**. Then repeat steps 2 through 4 until you have selected all of the items you want to include in the dataset.

Note: In addition to selecting data from Forms, you can also select data from Events, Participants, and CRFs from the **Event/Participant/CRF Attributes screens**. You can also click **Select All Items in the Study** to include all items in the study.

- Once you have selected all of the items you want to include in the dataset, click **Save and Define Scope**. The **Name and Description** screen appears.

Create Dataset: Name and Description

* indicates required field.

Name *

Description *

Item status

- Data from CRFs marked complete
- Data from CRFs not marked complete
- Data from all available CRFs

- Enter a name and description for the dataset.

Note: Names must be alphanumeric characters, but underscores are permitted.

- In the **Item Status** field, select the completion status of the data you want to include in the extract. You can choose to include data from: CRFs marked Complete, CRFs not marked Complete, and all CRFs.
- Ignore the message and fields on the lower portion of the screen, and click **Confirm and Save**.

Create Dataset: Select Format

Dataset name: Physical Exam Data
Dataset description: Physical Exam Data from Baseline Visit
Item status: Data from CRFs marked complete

Note: Large extracts may take some time to run. Once complete, you will receive an email notification and you can download your extract in the table below.

- CDISC ODM XML 1.3 Full with OpenClinica extensions [Run Now](#)
- CDISC ODM XML 1.3 Clinical Data with OpenClinica extensions [Run Now](#)
- CDISC ODM XML 1.3 Clinical Data [Run Now](#)
- CDISC ODM XML 1.2 Clinical Data with OpenClinica extensions [Run Now](#)
- CDISC ODM XML 1.2 Clinical Data [Run Now](#)
- View as HTML [Run Now](#)
- Excel Spreadsheet [Run Now](#)
- Excel Spreadsheet (no headers) [Run Now](#)
- Tab-delimited Text [Run Now](#)
- Tab-delimited Text (no headers) [Run Now](#)
- SPSS Data and Syntax [Run Now](#)
- SAS Data and Syntax [Run Now](#)

This will save your dataset definition, and it can now be extracted on demand or as part of a scheduled job. As more data is added to OpenClinica, the contents of this dataset will be updated to include any new data that meet the definition you have saved.

10. On the **Select Format** screen, select the output format for the extracted dataset, and click **Run Now** to extract the data.

Note: The first option is the most complete extract format, and it is the only option that includes the audit log data as well as all of the clinical data and metadata. If using Participate forms with data entered by the participant, it is included in study data. It can also be included in extracts if those forms are added.

Dataset Format	File Name	Run Time (Seconds)	File Size (Bytes)	Created Date	Created By	Status	Action
ODM XML 1.3 Full with OC extensions	odm1.3_fullTest_SAS_Dataset_2020-01-22-200318020.xml.zip	1.0	9316	22-Jan-2020 20:03:19	sjones	COMPLETED	Download Delete
SAS	SAS_Test_SAS_Dataset_2020-01-13-195644047.sas.zip	1.0	1543	13-Jan-2020 19:56:45	sjones	COMPLETED	Download Delete
ODM XML 1.3 with OC extensions	odm1.3_clinical_ext_Test_SAS_Dataset_2020-01-13-195636246.xml.zip	1.0	5061	13-Jan-2020 19:56:37	sjones	COMPLETED	Download Delete
SAS		0.0	0	09-Dec-2019 20:47:40	sjones	IN_PROGRESS	

OpenClinica displays a screen that indicates your extract is running.

Click **Back to Dataset** to view the status of extracted data. The **Select Format** screen appears again, and your dataset is listed at the bottom of the screen.

For large datasets it might take a while to extract data. If not complete, you will see that the dataset is still **IN PROGRESS**.

To Download the Dataset:

Click the **Download** button in the **Actions** column.

To Delete the Dataset:

(This only deletes the data definition; it does NOT delete data from the database)

Click the **Delete** button in the **Actions** column.

Note: The data in the dataset reflects the OpenClinica database at the time the dataset file was generated, not at the time when you downloaded the file. The dataset file name includes the date and time that the file was generated.

You can run the same extract in a number of different formats. OpenClinica retains one dataset file for each format for the dataset definition. If you generate a dataset and the dataset format already exists, the file you create overwrites the existing file.

For example, if the same extract was run and you selected **Excel**, that extract would overwrite the original Excel file. However, if you selected **HTML** for the second extract of this dataset, both the Excel file and the HTML file would be available.

If there is a permission tag on the Form that prevents you from accessing the Form data, you also cannot extract the data.

If you do not have access to any of the selected Forms, an error message appears to inform you that you do not have permission to access that data. If you do not have access to one or more Forms in the extracted dataset file, you cannot access, download, or delete the dataset.

Scheduled Export Jobs

To View Scheduled Jobs:

Click **Tasks** in the header bar of Study Runner and select **Jobs** under **Extract Data** to go to the **Scheduled Export Data Jobs** screen.

*Certain users can also access the Scheduled Jobs API on the **Web Services Information screen**.

*Users who can access scheduled job files via the API:

- User Type of Admin or User
- **Study-level roles:** *Data Managers, Data Specialists, and Data Monitors* can call the job execution API, call the job file retrieval API and get a file.
- **Site-level roles with site-specific job privileges:** *Site Monitors and Investigators* can call the job execution API and receive a list of job UUIDs consistent with the jobs scheduled for their site, call the job file retrieval API and get a file for a job from their site.
 - *Site Monitors and Investigators cannot* call the job file retrieval API and get a file for a job from a different site or the study level.
 - *Site Monitors and Investigators* can only access scheduled job lists for the studies/sites they have valid access to and can only access dataset files that they have access to (study/site/role permission and permission tags).
- *CRCs, Data Entry Persons, and Site and Study Viewers cannot* access either API.

Job execution API endpoint for getting list of job execution UUIDs for studies/sites the user has access to, and file retrieval API for getting a file for a given execution UID:

GET /auth/api/extractJobs/{jobUuid}/jobExecutions

GET /auth/api/extractJobs/jobExecutions/{jobExecutionUuid}/dataset

To View Details of Scheduled Jobs:

1. Click **Tasks** in the header bar of Study Runner and select **Jobs** under **Extract Data** to go to the **Scheduled Export Data Jobs** screen.
2. Click the **View** button in the **Actions** column of the table.

To Create a Scheduled Job:

Scheduled jobs can only be created by users with the Admin User Type.

1. Click **Tasks** in the header bar of Study Runner and select **Jobs** under **Extract Data** to go to the **Scheduled Export Data Jobs** screen.
2. Click the **Create New Scheduled Extract** link at the top of the screen.
3. Enter data into each field.
4. Click **Confirm and Save** to save the scheduled job. Click **Cancel** to cancel the scheduled job. You will receive an email when the job is complete.


The **number of files to save** field can be used to configure the job to preserve up to 10 extract files of the job. The scheduled job will always email the recipients indicated the most recent file, but the API can be used to retrieve the older files.

Note: *This field is set to the current server time by default. Any date/time after the current server time is valid.*

Create Scheduled Job: Export Dataset

Note that the job is set to run on the server time. The current server time is 21-Jan-2023 08:52:01.

* indicates required field.

Job Name: *	<input type="text"/>
Description: *	<input type="text"/>
Please Pick a Dataset to Export: *	test dataset ▾
Period to Run: *	Daily <input type="radio"/> Weekly <input type="radio"/> Every 4 Weeks <input type="radio"/>
Start Date/Time: *	21-Jan-2023  09 ▾ : 00 ▾ (DD-MMM-YYYY HH:MM)
File Format: *	CDISC ODM XML 1.3 Full with OpenClinica extensions <input type="radio"/> CDISC ODM XML 1.3 Clinical Data with OpenClinica extensions <input type="radio"/> CDISC ODM XML 1.3 Clinical Data <input type="radio"/> CDISC ODM XML 1.2 Clinical Data with OpenClinica extensions <input type="radio"/> CDISC ODM XML 1.2 Clinical Data <input type="radio"/> View as HTML <input type="radio"/> Excel Spreadsheet <input type="radio"/> Excel Spreadsheet (no headers) <input type="radio"/> Tab-delimited Text <input type="radio"/> Tab-delimited Text (no headers) <input type="radio"/> SPSS Data and Syntax <input type="radio"/> SAS Data and Syntax <input type="radio"/>
Number of Files to Save:	1 ▾
Contact Email: *	<input type="text"/>
<input type="button" value="Confirm and Save"/> <input type="button" value="Cancel"/>	

To Edit a Scheduled Job:

1. Click **Tasks** in the header bar of Study Runner and select **Jobs** under **Extract Data** to go to the **Scheduled Export Data Jobs** screen.
2. Click the **Edit** button in the **Actions** column of the table.
3. Change the answers to one or multiple fields.
4. Click **Confirm and Save** to save the scheduled job. Click **Cancel** to cancel the scheduled job.

To Remove a Job:

1. Click **Tasks** in the header bar of Study Runner and select **Jobs** under **Extract Data** to go to the **Scheduled Export Data Jobs** screen.
2. Click the **Remove** button in the **Actions** column of the table. Removing a job will cause it to stop running, but it can be restored later if you want it to resume.
3. Confirm that you want to remove the job.

To Restore a Job:

1. Click **Tasks** in the header bar of Study Runner and select **Jobs** under **Extract Data** to go to the **Scheduled Export Data Jobs** screen.
2. Click the **Restore** button in the **Actions** column of the table.

To Delete a Job:

1. Click **Tasks** in the header bar of Runtime and select **Jobs** under **Extract Data** to go to the **Scheduled Export Data Jobs** screen.
2. Click the **Delete** button in the **Actions** column of the table.
3. Confirm that you want to delete the job.

Note: Unlike the **remove** button, the **Delete** button completely deletes the job.

Formats for DataSets and Scheduled Export Jobs

You can download datasets in several formats based on how you want to view the data. Tabular formats (Tab Delimited Text, HTML, and Excel) are the easiest to read.

The table below summarizes the different formats:

[table id=23 /]

For more information, see [OC Data Extracts and Reporting Types](#).

Below Are Some Images of Extract Formats:

Tab-Delimited

Dataset Name:	Demographics
Dataset Description:	Demographics
Item Status:	
Study Name:	New Cancer Drug
Protocol ID:	NewCancerDrug
Date:	2020-Mar-10
Participants:	1
Study Event Definitions	1
Study Event Definition 1	Screening Visit E1
CRF4	Demographics C4

Participant ID	Protocol ID	DATE_E1_C4	AGE_E1_C4	GENDER_E1_C4	ETHNICITY_E1_C4	RACE_E1_C4
003	NewCancerDrug	2020-03-10	23	1	7 Hispanic	1 2 1 1

HTML Format

View Dataset Demographics

Dataset Name:	Demographics	
Dataset Description:	Demographics	
Study Name:	New Cancer Drug	
Protocol ID:	NewCancerDrug	
Date:	2020- Mar -10	
Participants:	1	
Study Event Definitions:	1	
Study Event Definition 1	Screening Visit	E1
CRF4	Demographics	C4

Participant ID	Protocol ID	DATE_E1_C4	AGE_E1_C4	GENDER_E1_C4	ETHNICITY_E1_C4	RACE_E1_C4
003	NewCancerDrug	2020-03-10	23	1	1	7

When viewing the HTML file, you can view the metadata for an Item by clicking its column header.

Excel Format

	A	B	C	D	E	F	G
1	Dataset Name:	Demographics					
2	Dataset Description: Demographics						
3	Item Status:						
4	Study Name	New Cancer Drug					
5	Protocol ID	NewCancerDrug					
6	Date:	2020-Mar-10					
7	Participant	1					
8	Study Event	1					
9	Study Event	Screening \ E1					
10	CRF4	Demograph C4					
11							
12							
13							
14							
15							
16	Participant	Protocol ID	DATE_E1_C	AGE_E1_C	GENDER_E	ETHNICITY	RACE_E1_C
17	3	NewCancer	3/10/2020	23	1	1	7
18							

CDISC ODM XML Format

```

<?xml version="1.0" encoding="US-ASCII"?>
- <ODM xsi:schemaLocation="http://www.cdisc.org/ns/odm/v1.3 OpenClinica-ODM1-3-0-OC3-0.xsd" ODMVersion="1.3" FileType="Snapshot" CreationDateTime="2020-03-10T19:43:42+00:00" Description="Demographics" FileOID="DemographicsD20200310194342+0000" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:OpenClinicaRules="http://www.openclinica.org/ns/rules/v3.1" xmlns:OpenClinica="http://www.openclinica.org/ns/odm_ext_v130/v3.1" xmlns="http://www.cdisc.org/ns/odm/v1.3">
  - <Study OID="S_NEWCAN(TEST)">
    - <GlobalVariables>
      <StudyName>New Cancer Drug</StudyName>
      <StudyDescription> </StudyDescription>
      <ProtocolName>NewCancerDrug</ProtocolName>
    </GlobalVariables>
    - <MetaDataVersion OID="v1.0.0" Name="MetaDataVersion_v1.0.0">
      - <Protocol>
        <StudyEventRef Mandatory="Yes" OrderNumber="0" StudyEventOID="SE_SCREENINGVISIT"/>
        <StudyEventRef Mandatory="Yes" OrderNumber="1" StudyEventOID="SE_BASELINEVISIT"/>
        <StudyEventRef Mandatory="Yes" OrderNumber="2"

```

When you select one of the CDISC ODM formats for the dataset, OpenClinica exports the dataset to an .xml file that complies with the Operational Data Model (ODM) of the Clinical Data Interchange Standards Consortium (CDISC) standard. These are the different parameters for the available ODM formats:

1.3 or 1.2: refers to the version of the ODM specification.

With extensions: Includes OpenClinica entities that are not part of the ODM specification, such as OpenClinica:SdvStatus.

Full: Includes **Discrepancy Notes (Queries)** and the **Audit Log**.

SAS Format

Obs	SubjectID	StudyEvent	StudyEventRepeatKey	Form	FormVersion	ItemGroupRepeatKey	_VISIT	_TEMP	_HR	_MAP	_SABP	_DABP	_CVP
1	SDASA001	Event 1	.	Form 1	1	1	Baseline
2	SDASA002	Event 1	.	Form 1	2	1	Cycle 1/Day 1	0	34
3	SDASA003	Event 1	.	Form 1	3	1	Cycle 3/Day 1	35	44	51	.	.	.
4	SDASA004	Event 1	.	Form 1	4	1	Cycle 4/Day 1	40	100	100	100	.	.
5	SDASA005	Event 1	.	Form 1	5	1	Cycle 2/Day 1	40	100	100	100	100	.
6	SDASA006	Event 1	.	Form 1	6	1	Baseline	40	85	82	210	45	32

SPSS Format

You can open the **.DAT** file in a Text Editor.

The SPSS Format appears in a table similar to Excel.

Tab-Delimited, HTML, and Excel:

To avoid duplication and confusion about the data collected in a study, certain identifiers and ordinal numbers must be appended to each variable name.

These variable names can be used in multiple CRFs across multiple Events. These appended numbers will help identify the event, CRF and item the value was collected in.

The identifiers are defined in the header table for Tab Delimited, HTML, and Excel formats. The identifiers are defined in a separate syntax (.sps) file for SPSS.

The following conventions apply to Tab Delimited, HTML, and Excel formats:

- **E1:** **E** specifies that the appendage represents the event. 1 specifies which event the variable is from, as defined in the header table. If the event is repeating, it would be represented as **E1_1**, **E1_2**, **E1_3**, etc.
- **C1:** **C** specifies that the appendage represents a CRF. 1 specifies which CRF the variable is from, as defined in the header table. For repeating events and repeating groups, additional information must be provided to specify which occurrence of the event and/or which repeat of the group the item value comes from. This is done by appending **_X** where **X** is the ordinal or repeat number.

For example:

An item called DEMO appearing in the 3rd occurrence of a repeating event, and the 5th repeat of the group called Example would be identified as: **DEMO_E1_3_C1_5**

For an item in a repeating event, but not part of a repeating group, the variable would be identified as: **DEMO_E1_3_C1**

The **[EVENT HANDLE]** and **[CRF HANDLE]** refer to identifiers appended to each item name to

avoid duplication and confusion amongst the repeating data points collected in a study.

Tabular formats (Tab Delimited Text, HTML, and Excel) contain both a header table and a data table.

The header table includes the following information:

- Dataset name
- Dataset description
- Study name
- Protocol ID
- Date
- Subjects
- Study Event Definitions
- CRFs

For each of the included study event definitions, the name of the event definition plus an identifier which is used to reference the event definition in the data table

For each of the included case report forms (CRFs), the name of the CRF plus an identifier which is used to reference the CRF in the data table

The data table includes the data you selected for the dataset.

CDISC:

CDISC ODM is a vendor neutral, platform independent format for interchange and archive of data collected in clinical trials. The model represents study metadata, data, and administrative data associated with a clinical trial. The ODM has been designed to be compliant with guidance and regulations published by the FDA for computer systems used in clinical trials.

The ODM model categorizes a clinical study's data into several kinds of entities including subjects, study events, forms, item groups, items, and annotations. The metadata of a study describes the types of study events, forms, item groups, and items that are allowed in the study. The clinical data of a study will typically have many actual entities corresponding to their definitions described in the metadata.

Like any XML file, an ODM file consists of a tree of elements that correspond to entities. Each element consists of required attributes and optional attributes. An ODM file type must be either Snapshot or Transactional.

A Snapshot file shows the current state of the included data.

A Transactional file shows both the latest state and (optionally) some prior states of an included entity.

An ODM file has a Granularity attribute which describes the coverage information of the ODM file.

The ODM file consists of two parts: metadata followed by Subject data. The metadata provides OIDs for the Study units (as defined when the CRFs were created), Event information, CRF information including Item Groups and Items with information about validations, and user account information. The Subject data provides Subject information, Event information, CRF information, and then the values.

SAS:

The Output from OpenClinica Includes Three Files:

- **SAS_DATA.xml:** The extracted data.
- **SAS_Format.sas:** For items defined as select_one or select_multiple, OpenClinica creates the library and maps response values to the appropriate response text

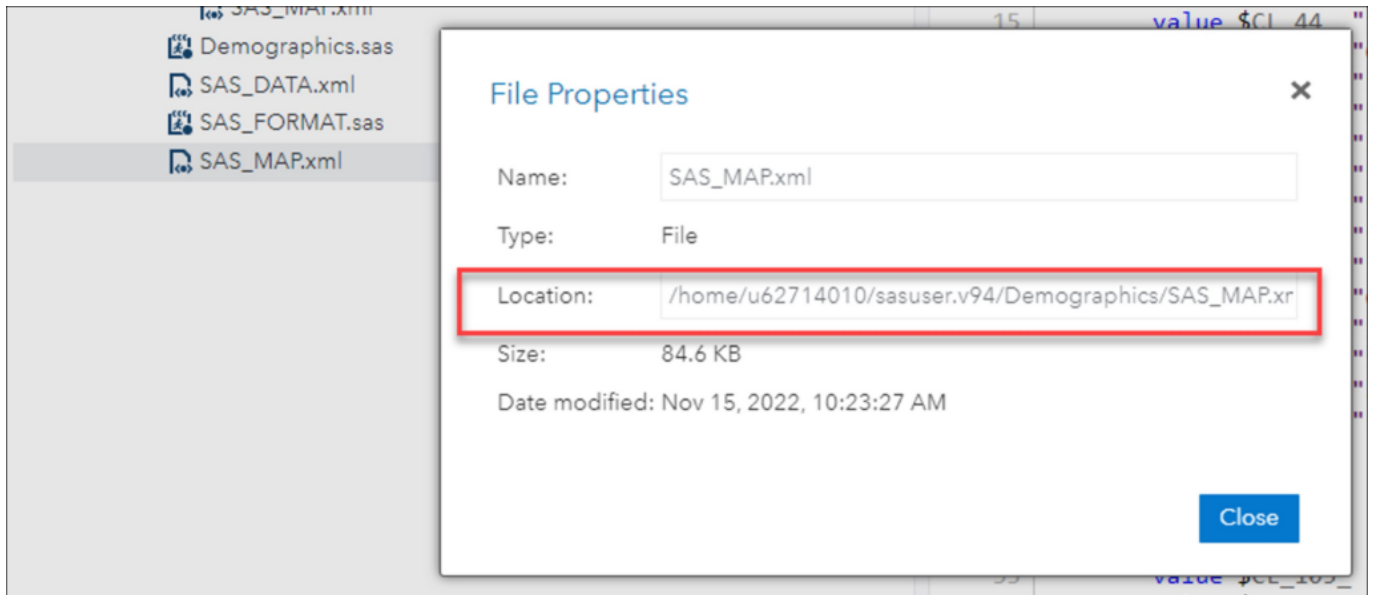
Note: Because select multiple and checkbox items include multiple values in a string format in OpenClinica (e.g., 1,2,7), these cannot be mapped to individual response text options.

- **SAS_MAP.xml:** A mapping file that maps the data to the appropriate structures

If using SAS Studio, please follow the directions below. Other versions of SAS will retain the basic instructions related to what needs to be uploaded and the code that needs to be run.

After Creating a Dataset in OpenClinica and Downloading it in SAS Format:

1. Create an account for SAS Studio by going to <https://odamid.oda.sas.com/SASODAControlCenter/>
 - a. When logging in after creating an account, click **SAS® Studio (Launch)**.
2. Once in "SAS® Studio", right-click **Files**.
 - a. Select **New > Folder**.
 - b. Enter a name for the folder.
 - c. Click **Save**.
3. To upload the data file (xml) and the map file (xml), click **Upload Files** at the top of the sidebar or right-click the folder and select **Upload Files**.
 - a. Click **Choose Files** after confirming the folder.
 - b. Select the **SAS_MAP** and **SAS_DATA** xml files to upload and click **Open**.
 - c. Verify the information and click **Upload**.
4. Click **New** at the top of the sidebar or right-click on your folder and select **New > SAS Program (F4)** to open a new Program window.
5. Open the **SAS_FORMAT** file in an external text editor.
6. Before running this code, edit the first three lines of the code by replacing the ~ with the path of the files.
 - a. Find the paths by right-clicking the folder that contains these files and selecting **Properties**.



The First 3 Lines of Your Format File (Before):

```
FILENAME S100_155 "~/SAS_DATA.xml";
```

```
FILENAME map "~/SAS_MAP.xml";
```

```
LIBNAME S100_155 xml xmlmap=map access=readonly;
```

```

1
2 FILENAME S100_155 "~/SAS_DATA.xml";
3 FILENAME map "~/SAS_MAP.xml";
4 LIBNAME S100_155 xml xmlmap=map access=readonly;
5 proc datasets library=S100_155;
6 copy out=work;
7 run;

```

The First 3 Lines of Your Format File (After):

```
FILENAME S100_155 "/home/u62714010/sasuser.v94/Demographics/SAS_DATA.xml/";
```

```
FILENAME map "/home/u62714010/sasuser.v94/Demographics/SAS_MAP.xml";
```

```
LIBNAME S100_155 xml xmlmap=map access=readonly;
```

```

2
3 FILENAME S100_155 "/home/u62714010/sasuser.v94/Demographics/SAS_DATA.xml/";
4 FILENAME map "/home/u62714010/sasuser.v94/Demographics/SAS_MAP.xml";
5 LIBNAME S100_155 xml xmlmap=map access=readonly;
6 proc datasets library=S100_155;
7 copy out=work;
8 run;

```

13. Add the code from the text editor to the to the Click **Run** at the top of the screen.

14. Click the new **Output Data** tab that appears to see the results of your export.

SAS MAP.xml x SAS_FORMAT.sas x

CODE LOG RESULTS **OUTPUT DATA**

Table: WORK_DEMOG_DM_2620 View: Column names Filter: (none)

Columns Total rows: 2 Total columns: 16

SubjectID	SiteID
1 100-009	100
2 RTEST-017	100

Columns list (all checked):

- Select all
- SubjectID
- SiteID
- StudyEvent
- StudyEventRepeatKey
- Form
- FormVersion
- FormStatus
- ItemGroupRepeatKey
- _DM_BRTHDAT
- _DM_AGEALC
- _DM_AGE_5025
- _DM_AGEU
- _DM_SEX_4074
- _DM_ETHNIC_909

Property Value

Label

Name

Output:

- This generates all the data tables based on Item Groups.
- OpenClinica Items become SAS Column Names.
- Tables include the master set of items (i.e., Item Groups span CRF Versions, though the SAS file does not indicate which version of the CRF was the source for the item.)
- There are two resulting data types: Numeric or Char. All OpenClinica items that are Integer or Real are classified as Numeric. All other OpenClinica data types are classified as Char.
- The SAS datasets/tables are generated from the OpenClinica metadata. Tables are created for all Item Groups in the extract. If no data was entered for a specific item group, the SAS table is still created, but is empty.

Troubleshooting:

OpenClinica allows **3,999** single-byte characters in a text field. When this size string is extracted to SAS, the full string is in the SAS_DATA.xml file.

SAS data set names must not exceed **32** characters and must start with either a letter (**A-Z**) or underscore.

As a result, Openclinica uses a modified Item Group OID for the data set name as follows:

- If group is Ungrouped use the CRF Name, otherwise:
 - To reduce the number of characters the remove the prepended IG (This means Group

- labels start with "_" + 5CHAR (of CRF Name) + _GROUPLABEL)
- If the resulting value exceeds 35 characters, OpenClinica appends the dataset name with the three- or four-digit number appended to the IG_OID
- SAS column names must not exceed 32 characters and must start with a letter (A-Z) or underscore. As a result, OpenClinica uses a modified Item OID for the column names as follows:
 - Truncate from the left to remove the I_5CHAR prefix to each Item Name.
 - Retain the portion of the OID starting with _ (underscore) followed by ITEMNAME (this ensures no Column Names start with a number.)
 - Retain appended three- or four-digit numbers to ensure item/column name uniqueness.

SPSS

When you select the SPSS format, the extracted .zip file contains two different files: a .dat file, which is a tab-delimited data file, and an .sps file, which is an SPSS data definition script.

To access the data, save the .dat and .sps files to the same location, then open the .sps file in the **IBM SPSS** program. If the **.sps** and **.dat** files are not in the same location, change the file location in the .sps file to point to the physical location of the **.dat** file. Then from SPSS, select **Run > All** to load the data into the application.

You can preview the **.dat** file by opening it in a text editor.

SPSS File Specifications

If you select the SPSS format, OpenClinica produces a package of files for use by the SPSS program. The files have been tested with the SPSS Windows software, version 20.

Although SPSS can read almost any ASCII file and deduce parameters for some of these variable attributes, any other attributes must be typed in by hand, which is tedious for large datasets.

Instead of generating an ASCII format dataset file from OpenClinica for use with SPSS, select the OpenClinica SPSS Syntax file format (.sps) which, in association with the data file (.dat), automatically loads the data with the correct variable definitions and attributes into SPSS.

SPSS Data Definitions Cover Ten Main Properties for Any Variable:

- Name
- Type
- Width
- Decimals
- Label
- Values
- Missing
- Columns
- Align
- Measure.

OpenClinica Currently Supports Automated Definition of:

- Name
- Type

- Width
- Decimals
- Label
- Values

SPSS Conceptual Mapping

This table presents the conceptual mapping of SPSS Data Definitions to OpenClinica data element metadata:

[table id=24 /]

Mapping between SPSS types and OpenClinica CRF Item Types

The table below describes the mapping of OpenClinica CRF ITEM data types to SPSS types.

[table id=25 /]

Notes:

- *Items of type ST, INT, and REAL are considered multi-select items when they are associated with a CRF response type of multi-select or checkbox. In this case, the item will be defined as a string (A) in SPSS and the selected values shown as a comma separated list in the field, even if the CRF data type is INT or REAL.*
- *SPSS can only handle up to 17 significant figures. If you use more than 17 significant figures you will lose accuracy in exporting to SPSS, but that is a limitation of SPSS not the OpenClinica export.*

Examples:

- *If you enter **12345678901234567890** (20 digits) into a numeric field the value **12345678901234567000** will be stored.*
- *If you enter **0.1234567890123456789** into a numeric field the value **0.123456789012345** will be stored*

Mapping Between SPSS Values and OpenClinica Choice Label

VALUE LABELS in the SPSS Syntax file map OpenClinica choice labels to discrete value sets in SPSS. Only variables that are of the type **select_one** or **select_multiple** are listed in the **VALUE LABELS** section.

Syntax for VALUE LABELS

Value labels appear for each variable as follows:

VARNAME1

Choice Name[0] "Choice Label[0]"

Choice Name [1] "Choice Label[1]"

Choice Name [2] "Choice Label[2]"

VARNAME2

Choice Name[0] "Choice Label[0]"Choice Name [1] "Choice Label[1]"

Choice Name [2] "Choice Label[2]"

SPSS Data Definitions for Built-in System Fields

Subject Attribute: Subject Status

[table id=26 /]

Event Attribute: Start Date

[table id=27 /]

Event Attribute: Status

[table id=29 /]

CRF Attribute: CRF Version Status

[table id=30 /]

CRF Attribute: CRF Version Name

[table id=31 /]

The Following Rules Apply to Variable Names in SPSS:

- Must begin with a letter. Remaining characters can be any letter, any digit, a period, or the symbols @, #, _, or \$.
- A \$ sign in the first position indicates that the variable is a system variable. The \$ sign is not allowed as the initial character of a user-defined variable.
- Avoid ending with a period, since the period may be interpreted as a command terminator.
- Avoid ending with an underscore to prevent conflict with variables automatically created by some procedures.
- Length of name cannot exceed 64 bytes. 64 bytes typically means 64 characters in single-byte languages (for example, English, French, German, Spanish, Italian, Hebrew, Russian, Greek, Arabic, Thai) and 32 characters in double-byte languages (for example, Japanese, Chinese, Korean).
- Cannot include spaces and special characters (for example, !, ?, ', and *).
- Must be unique.
- Cannot use reserved keywords: **ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, or WITH.**
- Can use any mixture of uppercase and lowercase characters; case is preserved for display purposes.
- When long variable names need to wrap onto multiple lines in output, SPSS attempts to break the lines at underscores, periods, and changes from lower case to upper case.
- OpenClinica follows certain rules for automatically converting an invalid dataset variable name to a valid SPSS variable name:
 - If the first character is not a letter, **V** is used as the first letter.
 - Any invalid characters are replaced with the symbol **#**.
 - If the last character is a period or an underscore, it is replaced by **#**.
 - If a name is longer than 64 characters, it is truncated to 64 characters.
 - If long variable names result in non-unique names in a data file, sequential numbers are used to replace its letters at the end. By default, the size of sequential numbers is **3**.

- If a reserved keyword has been used as a variable name, sequential numbers are appended to it.

Approved for publication by Paul Bowen. Signed on 2024-07-18 12:29AM

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