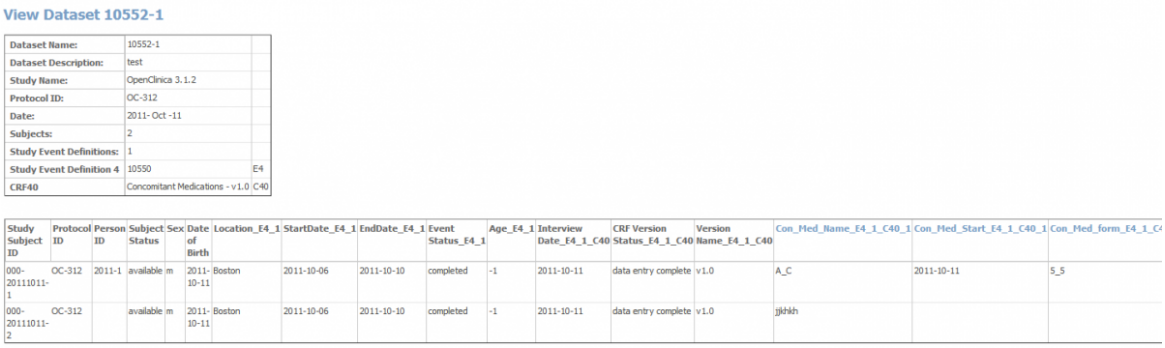
4.5.1 Tab-Delimited Text Format

OpenClinicas standard tabular (non-CDISC XML) data export formats are HTML, tab-delimited, Excel, and SPSS. The HTML, tab-delimited, and Excel formats each contain 2 tables a header table that contains reference information about the dataset contents, and the data table. The SPSS data export format has a data table similar in structure and format to the others, but does not have a header table. Instead it includes a separate .sps syntax file that describes the dataset. The following image shows a sample output for HTML format.



View Dataset 10552-1

Dataset Name:	10552-1	
Dataset Description:	test	
Study Name:	OpenClinica 3.1.2	
Protocol ID:	OC-312	
Date:	2011- Oct -11	
Subjects:	2	
Study Event Definitions:	1	
Study Event Definition 4	10550	E4
CRF40	Concomitant Medications - v1.0	C40

Study Subject ID	Protocol ID	Person ID	Subject Status	Sex	Date of Birth	Location_E4_1	StartDate_E4_1	EndDate_E4_1	Event Status_E4_1	Age_E4_1	Interview Date_E4_1_C40	CRF Version Status_E4_1_C40	Version Name_E4_1_C40	Con_Med_Name_E4_1_C40_1	Con_Med_Start_E4_1_C40_1	Con_Med_form_E4_1_C4
000-20111011-1	OC-312	2011-1	available	m	2011-10-11	Boston	2011-10-06	2011-10-10	completed	-1	2011-10-11	data entry complete	v1.0	A_C	2011-10-11	5_5
000-20111011-2	OC-312		available	m	2011-10-11	Boston	2011-10-06	2011-10-10	completed	-1	2011-10-11	data entry complete	v1.0	jjkhkh		

This page is not approved for publication.

4.5.1.1 Header Table Format

The header table includes the following information:

- Dataset name
- Dataset description
- Study name
- Protocol ID the study protocol ID
- Date the date the data set was created
- Subjects the number of subject records in the dataset
- Study Event Definitions the number of study event definitions included in the dataset
- For each of the included study event definitions, the name of the event defintion 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

4.5.1.2 Data Table Format

To avoid duplication and confusion amongst the data points 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 appendages will help identify the event, CRF and item the value was collected in. The

identifiers are defined in the header table for tab, HTML, and Excel formats. The identifiers are defined in a separate syntax (.sps) file for SPSS. The following scheme will be implemented:

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 detail 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. As an 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 in the following way.

DEMO_E1_3_C1_5

For an item in a repeating event, but not part of a repeating group, the variable would be identified in the following way:

DEMO_E1_3_C1

4.5.1.3 Variable naming convention

To avoid duplication and confusion amongst the data points 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 appendages will help identify the event, CRF and item the value was collected in. The identifiers are defined in the header table for tab, HTML, and Excel formats. The identifiers are defined in a separate syntax (.sps) file for SPSS. The following scheme will be implemented:

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 detail 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. As an 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 in the following way.

DEMO_E1_3_C1_5

For an item in a repeating event, but not part of a repeating group, the variable would be identified in the following way:

DEMO_E1_3_C1