4.6 SPSS File Specifications

When you select the SPSS format for an exported dataset, 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), will automatically load 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, and Measure. OpenClinica currently supports automated definition of Name, Type, Width, Decimals, Label, and Values using the SPSS Syntax file format (.sps).

The following topics describe the structure and syntax of the OpenClinica .sps dataset file and corresponding .dat file.

Approved for publication by Ben Baumann. Signed on 2014-03-24 8:24AM

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

4.6.1 SPSS Conceptual Mapping

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

SPSS Data Definition Metadata	**OpenClinica CRF Metadata**
Name	ITEM_NAME
Type	Mapped to DATA_TYPES
Width	Calculated from widest value in field
Decimals	If DATA_TYPES = Real, then calculated from most precise value in field. Else 0.
Label	DESCRIPTION_LABEL
Values	Generated from RESPONSE_OPTIONS_TEXT and RESPONSE_OPTIONS_VALUES
Missing	N/A
Columns	N/A
Align	N/A
Measure	N/A

4.6.2 Creation of SPSS Data Definitions from OpenClinica CRF Item Properties

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

SPSS Data Definition Property	**OpenClinica CRF Item Property**
Name	ITEM_NAME_[EVENT HANDLE]
Type	Mapped to DATA_TYPES
Width	If DATA_TYPE = ST, INT, REAL, or FILE, set to the width value of WIDTH_DECIMAL()
Decimals	If DATA_TYPE = REAL, then set to the decimal value of WIDTH_DECIMAL(). Else 0.
Label	DESCRIPTION_LABEL
Values	Generated from RESPONSE_OPTIONS_TEXT and RESPONSE_OPTIONS_VALUES
Missing	N/A
Columns	N/A
Align	N/A
Measure	N/A

4.6.3 Use of [EVENT HANDLE] and [CRF HANDLE] Appended to Variable Names

The [EVENT HANDLE] and [CRF HANDLE] refer to identifiers appended to each variable name to avoid duplication and confusion amongst the repeating data points collected in a study. See

https://docs.openclinica.com/3.1/technical-documents/openclinica-dataset-transformations/non-cdisc-data-export-formats for more detail.

4.6.4 Mapping between SPSS types and OpenClinica CRF ITEM Data Types

The table below describes the mapping of OpenClinica CRF ITEM data types [https://docs.openclinica.com/3.1/technical-documents/openclinica-item-data-specifications/canonical-datatypes] to SPSS types.

CRF data type	CRF Width(decimal)	CDISC ODM xml data type	SPSS variable type	SPSS Syntax for type Format
ST	n	text	String	An
INT	n	integer	Numeric	Fn.0

REAL	n(d)	float	Numeric	Fn.d
FILE	n	text	String	An
DATE	N/A	date	Date	ADATE10
PDATE	N/A	partialDate	String	A10.0

Notes:

1. 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.

2. 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.

4.6.5 Handling of OpenClinica Null values

When creating an Event Definition, the user can choose to allow certain codes to represent null values in the entered data. Examples are 'NI', 'NA' etc.

If a non-string item has one of the allowed OpenClinica null values as item data, SPSS will treat it as a system missing values, and an empty data is cell is displayed in the Data View of the SPSS tool. In case of an item of data type string (ST), the null value string is displayed as is.

4.6.6 Mapping Between SPSS Values and OpenClinica RESPONSE_OPTIONS

VALUE LABELS in the SPSS Syntax file map OpenClinica RESPONSE_OPTIONS to discrete value sets in SPSS. Only variables that are of RESPONSE_TYPE single select, or radio and that have a defined response set will be listed in the VALUE LABELS section.

4.6.6.1 Syntax for VALUE LABELS

VALUE LABELS

VARNAME1

RESPONSE_OPTIONS_VALUE[0] "RESPONSE_OPTIONS_TEXT[0]"

RESPONSE_VALUES[1] "RESPONSE_OPTIONS_TEXT[1]"

RESPONSE_VALUES[2] "RESPONSE_OPTIONS_TEXT[2]" /

VARNAME2

RESPONSE_OPTIONS_VALUE[0] RESPONSE_OPTIONS_TEXT[0]

RESPONSE_VALUES[1] RESPONSE_OPTIONS_TEXT[1]

RESPONSE_VALUES[2] RESPONSE_OPTIONS_TEXT[2] /

4.6.6.2 SPSS Data Definitions for Built-in System Fields

Subject Attribute: Date of Birth

SPSS Data Definition Property	**Value**	**Encoding**
Name	DateofBirth	DateofBirth
Type	Date	ADATE10
Width	N/A	
Decimals	N/A	
Label	Date of Birth	Date of Birth
Values	None	
Missing	None	
Columns	10	
Align	Right	
Measure	Unknown	

Subject Attribute: Sex

SPSS Data Definition Property	**Value**	**Encoding**
Name	Sex	Sex
Type	String	A
Width	1	1
Decimals	N/A	
Label	Date of Birth	Date of Birth
Values		Sex
	M, F	M Male F Female
Missing	None	
Columns	1	
Align	Left	
Measure	Unknown	

Subject Attribute: Subject Status

SPSS Data Definition Property	**Value**	**Encoding**
Name	SubjectStatus	SubjectStatus

Type	String	A
Width	[maximum length of subject status string across all the subjects]	[maximum length of subject status string across all the subjects]
Decimals	N/A	
Label	Subject Status	Subject Status
Values	None	
Missing	None	
Columns	[maximum length of subject status string across all the subjects]	[maximum length of subject status string across all the subjects]
Align	Left	
Measure	Unknown	

Subject Attribute: Person ID

SPSS Data Definition Property	Value	Encoding
Name	PersonID	PersonID
Type	String	A
Width	[maximum length of subject Unique Identifier string across all the subjects]	[maximum length of subject Unique Identifier string across all the subjects]
Decimals	N/A	
Label	Person ID	Person ID
Values	None	
Missing	None	
Columns	[maximum length of subject Unique Identifier string across all the subjects]	[maximum length of subject Unique Identifier string across all the subjects]
Align	Left	
Measure	Unknown	

Subject Attribute: Secondary ID

SPSS Data Definition Property	Value	Encoding
Name	SecondaryID	SecondaryID
Type	String	A
Width	[maximum length of subject Secondary Identifier string across all the subjects]	[maximum length of subject Secondary Identifier string across all the subjects]
Decimals	N/A	
Label	Secondary ID	Secondary ID
Values	None	

Missing	None	
Columns	[maximum length of subject Secondary Identifier string across all the subjects]	[maximum length of subject Secondary Identifier string across all the subjects]
Align	Left	
Measure	Unknown	

Event Attribute: Event Location

SPSS Data Definition Property	**Value**	**Encoding**
Name	LOCATION_[EVENT HANDLE]	LOCATION_[EVENT HANDLE]
Type	String	A
Width	[maximum length of event location string across all the subjects]	[maximum length of event location string across all the subjects]
Decimals	0	0
Label	Location for [EVENT NAME] (EVENT HANDLE)	Location for Event [EVENT NAME] (EVENT HANDLE)
Values	None	
Missing	None	
Columns	[maximum length of event location string across all the subjects]	[maximum length of event location string across all the subjects]
Align		
Measure		

Event Attribute: Start Date

SPSS Data Definition Property	**Value**	**Encoding**
Name	STARTDATE_[EVENT HANDLE]	STARTDATE_[EVENT HANDLE]
Type	Date	ADATE10
Width	N/A	
Decimals	N/A	
Label	Start Date for [EVENT NAME] (EVENT HANDLE)	Start Date for [EVENT NAME] (EVENT HANDLE)
Values	None	
Missing	None	
Columns	10	
Align	Right	
Measure	Unknown	

Event Attribute: End Date

SPSS Data Definition Property	Value	Encoding
Name	EndDate_[EVENT HANDLE]	EndDate_[EVENT HANDLE]
Type	Date	ADATE10
Width	N/A	
Decimals	N/A	
Label	End Date for [EVENT NAME] (EVENT HANDLE)	End Date for [EVENT NAME] (EVENT HANDLE)
Values	None	
Missing	None	
Columns	10	
Align	Right	
Measure	Unknown	

Event Attribute: Status

SPSS Data Definition Property	Value	Encoding
Name	EventStatus_ [EVENT HANDLE]	EndDate_[EVENT HANDLE]
Type	String	A
Width	[maximum length of event status string across all the subjects]	[maximum length of event status string across all the subjects]
Decimals	N/A	
Label	Event Status For [EVENT NAME] (EVENT HANDLE)	End Date for [EVENT NAME] (EVENT HANDLE)
Values	None	
Missing	None	
Columns	[maximum length of event status string across all the subjects]	[maximum length of event status string across all the subjects]
Align	Right	
Measure	Unknown	

CRF Attribute: Interview Date

SPSS Data Definition Property	Value	Encoding
Name	InterviewDate_[EVENT HANDLE]_[CRF HANDLE]	InterviewDate_[EVENT HANDLE]_[CRF HANDLE]
Type	Date	ADATE10
Width	N/A	

Decimals	N/A	
Label	Interviewer Date For [EVENT NAME]	Interviewer Date For [EVENT NAME]
Values	None	
Missing	None	
Columns	10	
Align	Right	
Measure	Unknown	

CRF Attribute: Interviewer Name

SPSS Data Definition Property	**Value**	**Encoding**
Name	Interviewer_[EVENT HANDLE]_[CRF HANDLE]	Interviewer_[EVENT HANDLE]_[CRF HANDLE]
Type	String	A
Width	[maximum length of interviewer name string across all the event CRFs]	[maximum length of interviewer name string across all the event CRFs]
Decimals	N/A	
Label	Interviewer Name for [EVENT NAME]	Interviewer Name for [EVENT NAME]
Values	None	
Missing	None	
Columns	[maximum length of interviewer name string across all the event CRFs]	[maximum length of interviewer name string across all the event CRFs]
Align	Left	
Measure	Unknown	

CRF Attribute: CRF Version Status

SPSS Data Definition Property	**Value**	**Encoding**
Name	CRFVersionStatus_[EVENT HANDLE]_[CRF HANDLE]	CRFVersionStatus_[EVENT HANDLE]_[CRF HANDLE]
Type	String	A
Width	[maximum length of CRF version status string across all the event CRFs]	[maximum length of CRF version status string across all the event CRFs]
Decimals	N/A	
Label	CRF Version Status For [EVENT NAME]	CRF Version Status For [EVENT NAME]
Values	None	
Missing	None	
Columns	[maximum length of CRF version status string across all the event CRFs]	[maximum length of CRF version status string across all the event CRFs]
Align	Left	
Measure	Unknown	

CRF Attribute: CRF Version Name

SPSS Data Definition Property	Value	Encoding
Name	VersionName_ [EVENT HANDLE]_[CRF HANDLE]	VersionName_ [EVENT HANDLE]_[CRF HANDLE]
Type	String	A
Width	[maximum length of CRF version name string across all the event CRFs]	[maximum length of CRF version name string across all the event CRFs]
Decimals	N/A	
Label	Version Name For [EVENT NAME]	Version Name For [EVENT NAME]
Values	None	
Missing	None	
Columns	[maximum length of CRF version name string across all the event CRFs]	[maximum length of CRF version name string across all the event CRFs]
Align	Left	
Measure	Unknown	

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. Sixty-four 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, 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 (implemented in OpenClinica 3.1.3)

OpenClinica does not correct for other SPSS variable name validity constraints.

A future OpenClinica release may automatically correct for additional SPSS validity constraints. See https://issuetracker.openclinica.com/view.php?id=13686:

- 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.