SPSSDataConsolidator (SDC) User Guide

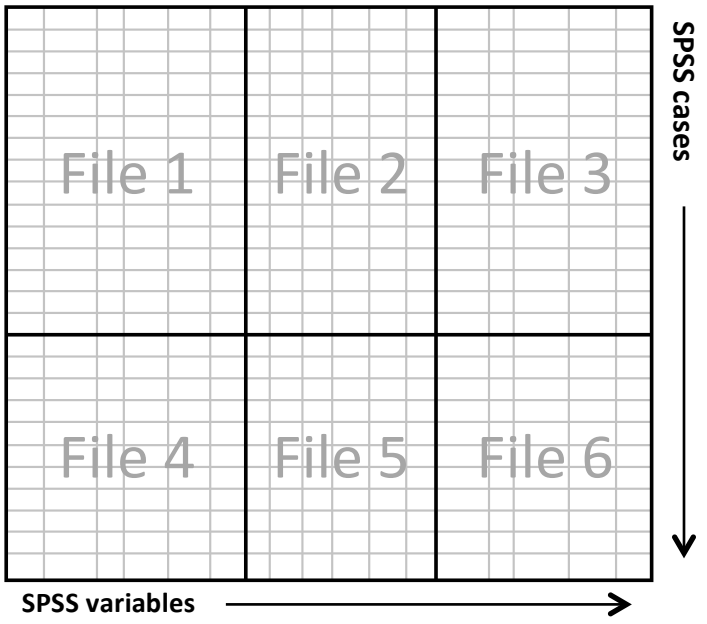
SDC is a program for consolidating FILMAN and comma separated variable (CSV) files into a SPSS-compatable file (.sav) that can be used as data input to many statistical packages, such as SYSTAT and SPSS. This format has many advantages, including the ability to have lengthy variable names of 60 characters or more (depending on the statistical package used).

Definitions:

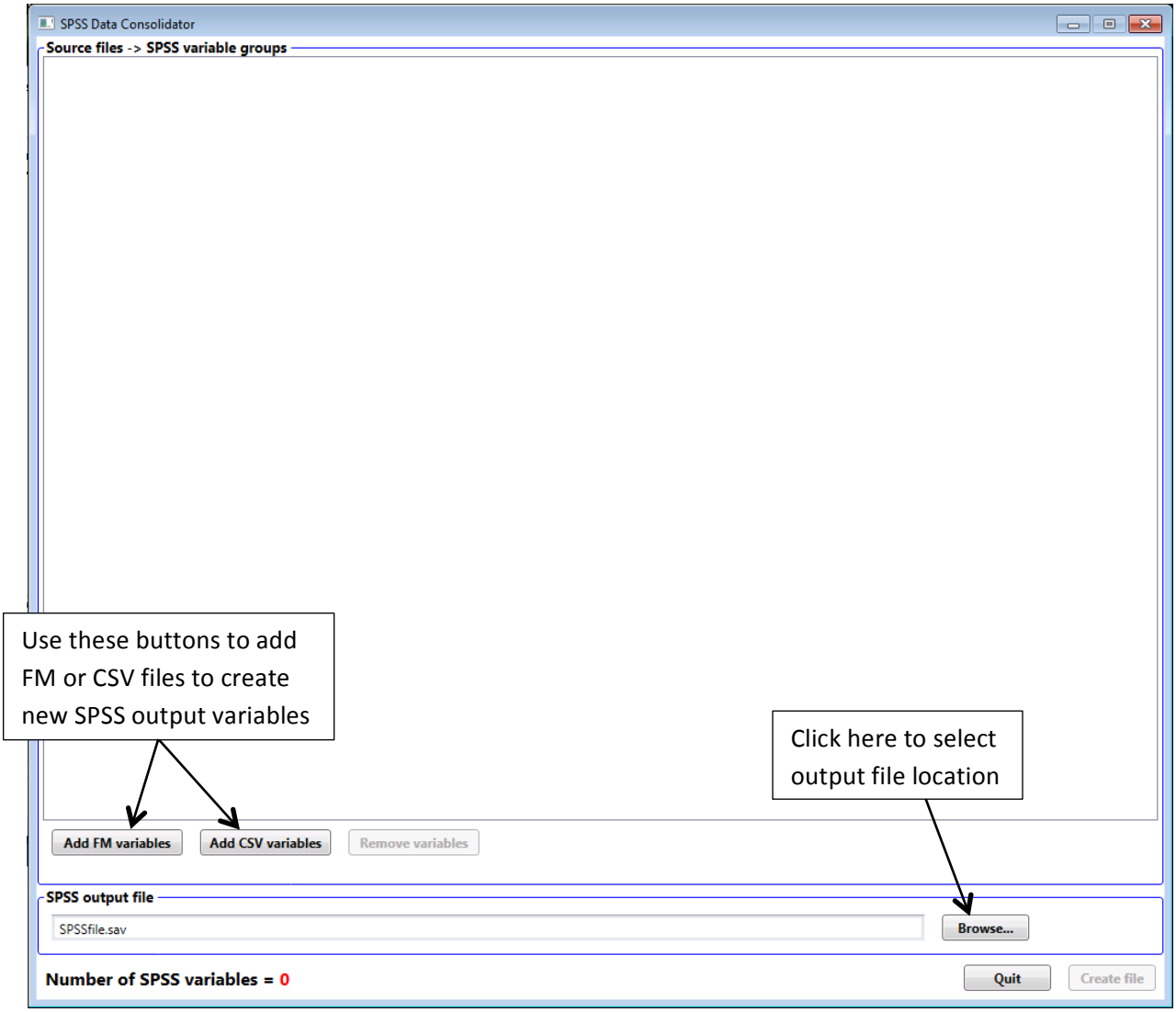
SPSS dataset	This is the ouput file generated by SDC; it is a (strictly) rectangular array of data values; each column contains the values for a particular Variable and each row, the values for each SPSS Case; for SDC purposes, the SPSS dataset is conceptually divided into sub-arrays, each from a constituent FILMAN or CSV file
Individual SPSS column	Contains the values for a single SPSS Variable
Individual SPSS row	Contains the values for each of the Variables in a single SYSTAT Case
Collected column	The values for a SPSS Variable Group; these values are selected from a set of files of a single type (FILMAN or CSV) that contain this information; the files that may be included in a Variable Group column must be of the same type; if they are CSV, they must have the same number of items in each row and the names of those items (in the first row) must be identical; if they are FILMAN files, they must have the same number of GVs, channels, and data points
Collected row	The Trials from a single experimental session; the files containing these the values may be of either FILMAN or CSV types (or both); all the files must have the same number of records (CSV) and/or recordsets (FILMAN), each representing a given SYSTAT Trial
CSV file	Comma Separated Value file; the rows in this file consist of a series of values (either numbers or strings in quotation marks) separated by commas; the first row must contain a list of the names of the Variables; these names must start with a capital letter and contain only letters, numbers and underscore; Variables which encode strings should end with a $ (which is added, if not already present); each row ends in a CR/LF; rows after the first contain a list of values for the corresponding Variable; Excel CSV files produce an acceptable format

FILMAN file	Standard format FILMAN file created by FILMAN processing; a given FILMAN file is described by the number of GVs, number of channels, number of datapoints, and number of recordsets it contains. Sometimes FILMAN files entered into SPSSDataConsolidator will be processed into a single channel so that a single record(set) represents a single trial. If more than one channel is present in the FILMAN file, the data from the various channels will be entered as separate individual SPSS columns
Missing data	Indicated by the value NaN (Not a Number) in a file (SPSS, FILMAN or CSV); this value can be used to fill in "holes" in a data file in order to make files compatible

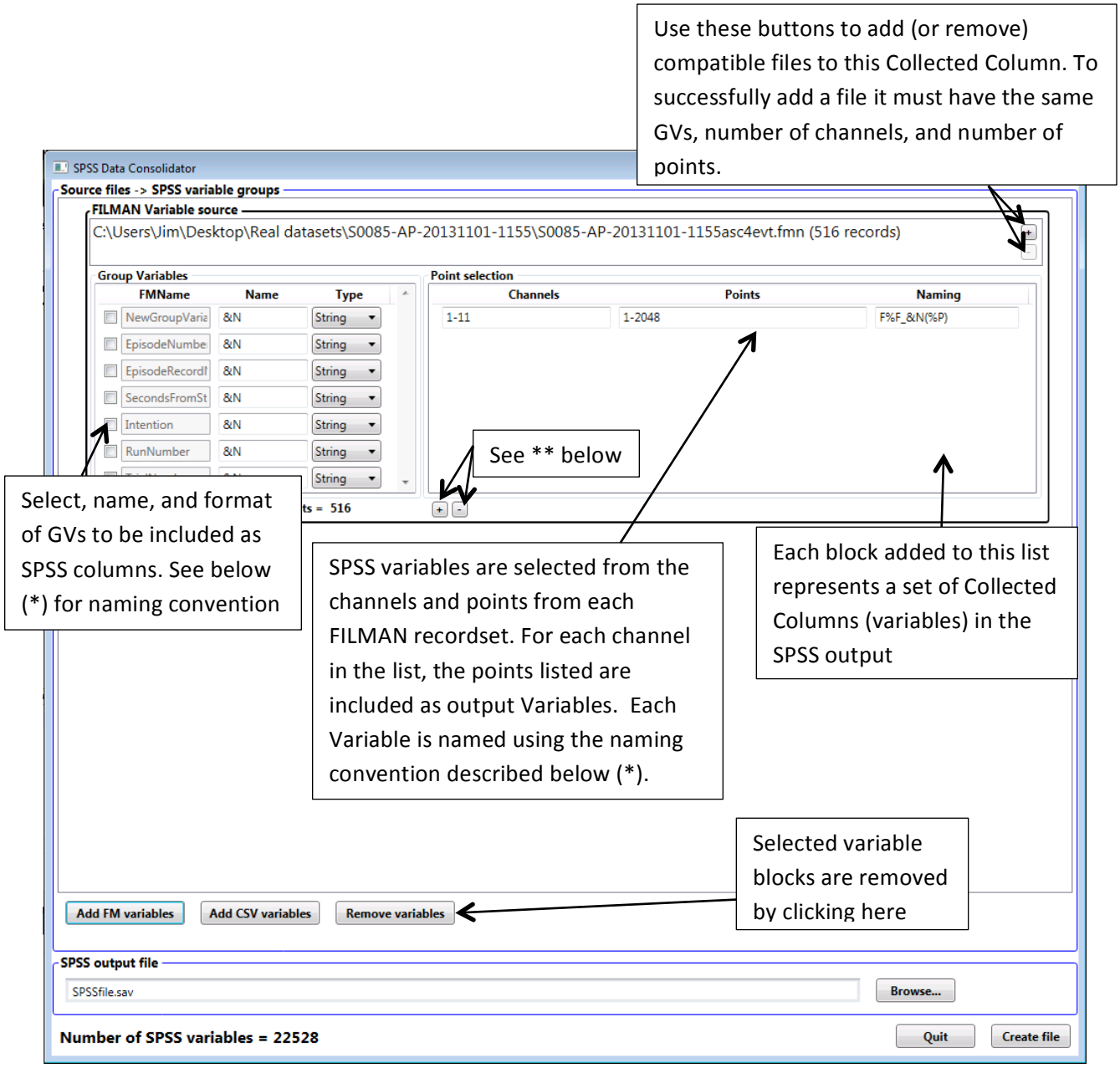
Below is a diagram of a SDC output file, a file that can be input into statistical packages such as SYSTAT for analysis. The individual small gray rectangles each represent a single datum; each gray row is a SPSS trial; and each gray column a SPSS variable. Each of the larger black rectangles represents an individual FILMAN or CSV input file. Notice that File 1 and File 4 must have the same number, ordering, and type of columns but may have different number of trials. This assures that the SPSS variables "line up" in the consolidated output file. The same requirements are true for File 2 and File 5, and File 3 and File 6. Also note that File 1, File 2, and File 3 must have the same number of trials, though they may (and almost certainly will) have different variables. This is analogously true for File 4, File 5, and File 6. SDC enforces these conditions and will proceed only if they are met. One mechanism for overcoming this "limitation" is the use of Missing data.



Now let's look at SDC itself. Here's the main window immediately after opening the program.



After we've added a FILMAN file, the window looks like this.



* Naming SPSS variables is somewhat problematic when there are many variables to be named. This system for naming points and GVs from FILMAN files attempts to simplify this problem by automatically generating names that "make sense". Think of this as a "format" or macro string for creating a name for each variable. The macros in the naming convention are two-symbol codes as follows:

%F Each input file is assigned a unique number from 1 to the number of files; this symbol represents this number and thus permits creating unique variable names even though several files are involved which may otherwise have identical variables

%f	Index assigned to each input file
%C	The original channel number from the FILMAN file
%c	Renumbered channel in the order selected in the channel selection field
%P	Original point number from the FILMAN file
%p	Renumbered point number in the order selected in the point selection field
&N	The channel name from the FILMAN file

Group variables may have macros as well using these codes:

%G	Original number of the GV in the FILMAN file
%g	Renumbered (based on the GVs selected) GV
%F and %f	As in the point naming convention above, referencing file that this GV is from
&N	The name of the GV in the FILMAN header
&n	The name of the GV in the HDR dataset header, if available; otherwise the FILMAN GV name

Any of the numeric codes (those beginning with a "%") can have an optional integer after the % to indicate a fixed field width, padded left with zeros. Thus the variable name for channel 16, point 123, would encode from "V%C(%P)" as "V16(123)", while "V%3C(%4P)" would result in "V016(0123)".

An alphabetic code (those beginning with a "&") can also have an optional number after the & that indicates the length to which the name will be trimmed if it is too long.

Valid SPSS variable names begin with a capital letter followed by letters, numbers, and "_". A set of parentheses may be used to enclose a number at the end of the name (as in the above example). Any blank characters generated in a name will be changed to "_".

** These buttons add (or remove) naming conventions for channel/point subsets.