

Designing ADaM Datasets: KISS the Complexity Goodbye

David Jordan, SAS Institute, Cary, NC USA

ABSTRACT

The CDISC SDTM standard provides many predefined domains, each providing a model for a specific purpose with an appropriate set of columns. In the case of [CDISC ADaM](#), a large number of possible variables are defined and it is up to you to determine which variables are appropriate for your particular analysis modeling situation. Even though the ADaM ADSL and ADAE dataset templates serve a particular purpose, you still need to select a subset of the variables appropriate for your analysis needs. The BDS dataset template is a very general template that defines over 140 variables you can use, which can be overwhelming. Defining an ADaM analysis dataset is far more complex than using a pre-defined SDTM domain.

Metadata describing both the dataset templates in the standard and existing datasets in your data environment can be used when defining an ADaM dataset. When the user includes variables from existing datasets, this metadata can automatically provide traceability information. Some ADaM variables require the management of *xx*, *y*, and *zz* parameters within the ADaM variables. Software can assist in the management of this tedious and error-prone task.

The use of an underlying metadata framework and ADaM-specific algorithms to drive the dataset definition process in a graphical environment eases and accelerates the use of the ADaM standard. The ADaM dataset creation tool that has been incorporated into the SAS Clinical Data Integration® product is described in detail.

INTRODUCTION

This paper describes the process for defining a CDISC ADaM dataset using the ADaM dataset wizard implemented in SAS Clinical Data Integration release 2.4. Prior releases of SAS Clinical Data Integration supported the CDISC SDTM standard. Defining an ADaM dataset involves some issues that do not arise when defining an SDTM domain. The interface for defining an ADaM dataset in SAS Clinical Data Integration will be fully described, along with aspects of ADaM dataset definition that do not exist in the creation of SDTM domains.

SAS Clinical Data Integration is an extension of the SAS Data Integration® product, providing management of metadata and clinical trial data based on the CDISC standards. SAS Data Integration provides a nice user interface environment for graphically defining extract, transform, and load (ETL) procedures for mapping and storing data. Embedding metadata and CDISC-specific algorithms into this data integration platform eases the effort of mapping one's data from an existing data representation into a CDISC-compliant form.

INVOKING THE WIZARD

When setting up your SAS Clinical Data Integration environment, a standards administrator imports metadata associated with the clinical standards defined within the SAS Clinical Standards Toolkit®. The SDTM, ADaM and SEND standards are supported. This metadata provides a complete description of all the datasets and associated attributes and properties, as well as data and code used to validate the compliance of your datasets relative to the standards. When the metadata is imported, it is placed within the SAS Metadata Server®.

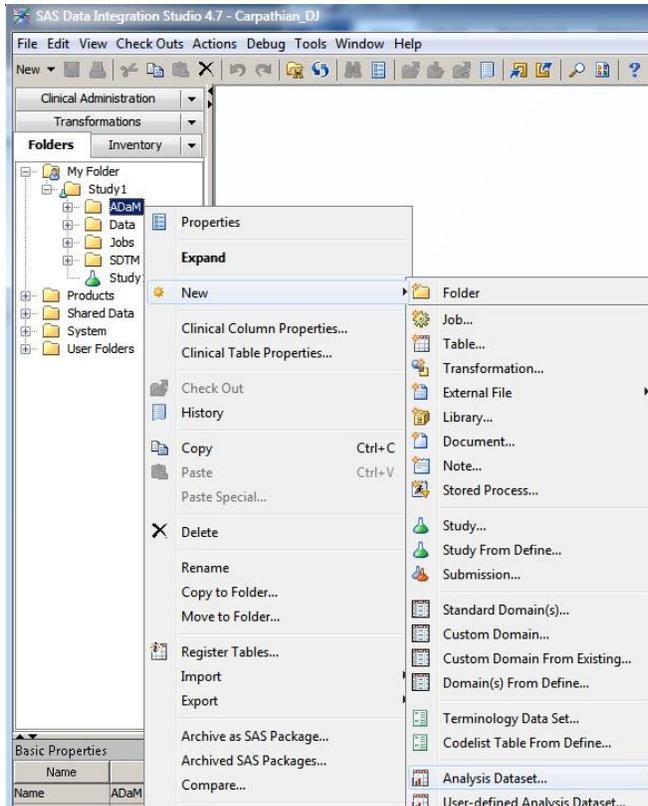
A study administrator is responsible for defining characteristics of a study. This includes the selection of the relevant standards, libraries in which to place datasets, and controlled terminologies to use with the study. The study administrator can also select a folder template which initializes a folder hierarchy in which to place the data associated with the study. A wizard walks you through the steps necessary to initialize the study.

Once you have completed these steps, you can begin creating datasets for the study. You will likely place the datasets for SDTM and ADaM in separate folders within the study's folder hierarchy. The Folders tab within the user interface provides a graphical rendering of the folder hierarchy. The creation of an ADaM dataset involves the invocation of a wizard that guides you through the necessary steps. You can select the folder within the study in which you want to place your ADaM dataset, use the right-mouse-button context menu to select the New menu item, and choose the Analysis Dataset menu entry. This brings up the wizard to create an ADaM dataset. Another means

PhUSE 2013

of selecting the New menu is a drop-down menu that resides directly below the File button in the upper left corner of the SAS Data Integration Studio window.

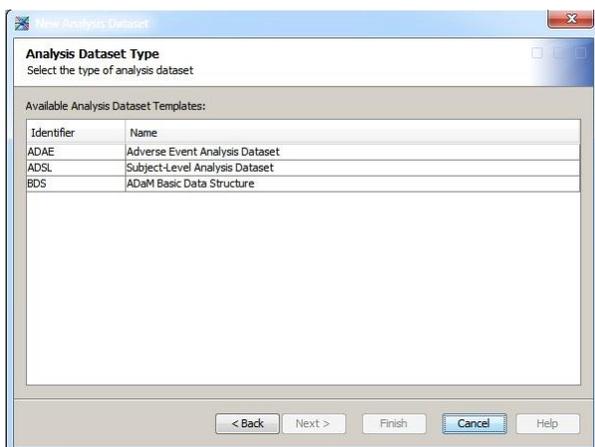
In the following screen shot, we have selected the folder named ADaM within Study1 and selected the New menu entry. There is an entry near the bottom of the new menu in the screen shot used to create an analysis dataset.



If you selected the New menu relative to the folder in which you want to place the ADaM dataset, the first page of the wizard identifies the location and study in which you invoked the wizard. You do have the option of browsing to select an alternative location, but normally you would press the Next button to proceed to the next page of the wizard. If the study happens to be associated with more than one ADaM standard, the next wizard page lists the ADaM standards and requires you to select one. Select the appropriate standard and press the Next button.

SELECT THE DATASET TYPE AND PROVIDE IT WITH A NAME AND IDENTIFIER

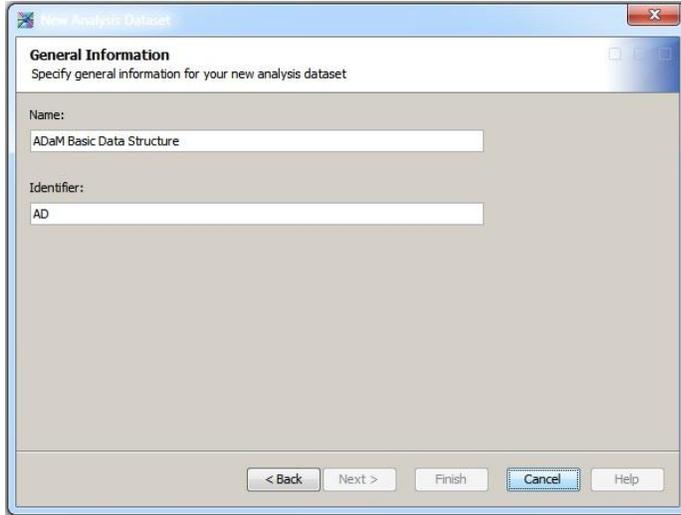
You must then select which ADaM dataset type you wish to create. Each type is presented in a list and you must select one of the entries.



PhUSE 2013

One, and only one, ADSL dataset can be defined for a given study and its identifier must be “ADSL”. If you have already created an ADSL dataset for the study, its entry in the list is disabled.

Press the Next button and you will be taken to the next wizard page where you can provide a name and identifier for the dataset.

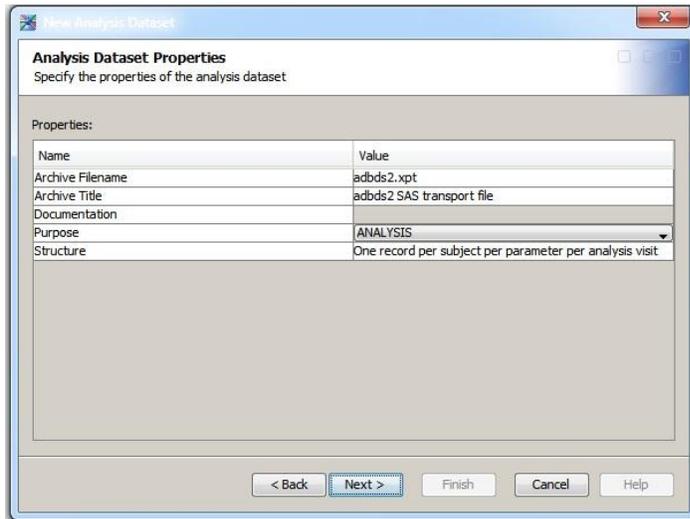


The screenshot shows the 'New Analysis Dataset' wizard window. The title bar reads 'New Analysis Dataset'. The main heading is 'General Information' with the subtitle 'Specify general information for your new analysis dataset'. There are two text input fields: 'Name:' containing 'ADaM Basic Data Structure' and 'Identifier:' containing 'AD'. At the bottom, there are five buttons: '< Back', 'Next >', 'Finish', 'Cancel', and 'Help'.

The identifier for all ADaM datasets must begin with the prefix “AD” and the wizard automatically provides this prefix in the identifier field. It also makes sure you provide a name that is compliant with the ADaM standard. If you are creating an ADSL dataset, the identifier is automatically provided and cannot be changed.

SPECIFY THE PROPERTIES AND THE LIBRARY

The next page of the wizard allows you to specify values for properties of the dataset. Default values are provided for these properties, but you can enter alternative values.



The screenshot shows the 'New Analysis Dataset' wizard window, 'Analysis Dataset Properties' page. The title bar reads 'New Analysis Dataset'. The main heading is 'Analysis Dataset Properties' with the subtitle 'Specify the properties of the analysis dataset'. Below the heading is a table with the following data:

Name	Value
Archive Filename	adbds2.xpt
Archive Title	adbds2 SAS transport file
Documentation	
Purpose	ANALYSIS
Structure	One record per subject per parameter per analysis visit

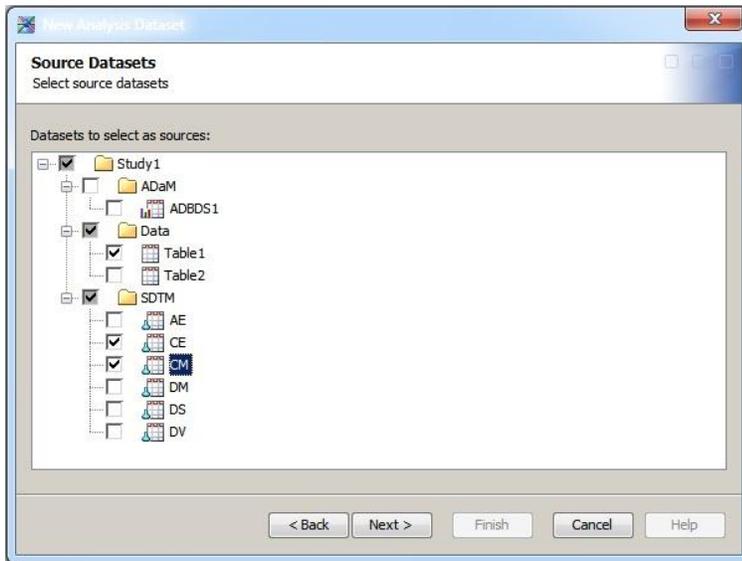
At the bottom, there are five buttons: '< Back', 'Next >', 'Finish', 'Cancel', and 'Help'.

The next wizard page allows you to select a library to assign to the dataset. If any default libraries have been specified at the study level, they are listed in the table. There is also a means to select a different library.

SELECTING VARIABLES FROM EXISTING DATA SOURCES

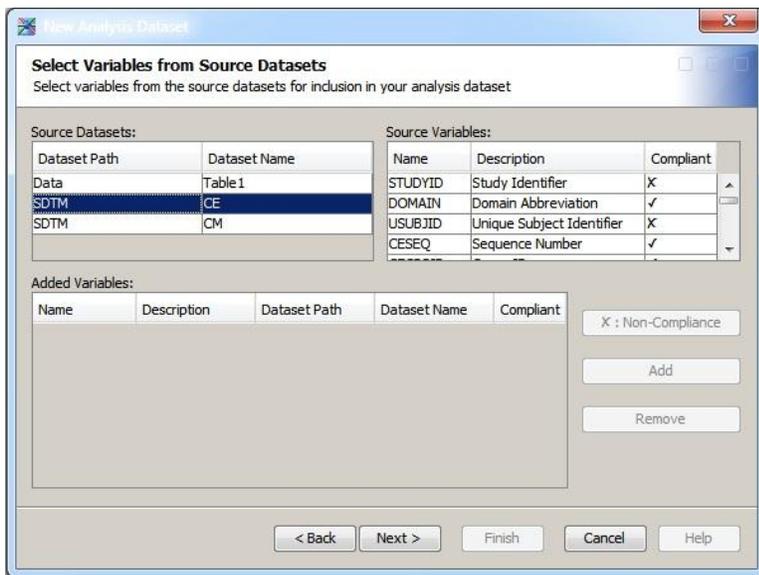
ADaM allows you to select variables from existing data sources within your study to use in your dataset. This can include other ADaM datasets, SDTM domains, and ordinary SAS data sets. The next wizard page presents all of these data sets in a manner that reflects the folder hierarchy within the study.

PhUSE 2013



Set the checkboxes on any datasets that have variables you want to include. Then press the Next button to advance to the next wizard page. If you did not select any source datasets, the next wizard page which allows you to select variables is skipped.

You can then select variables from each of the source datasets.

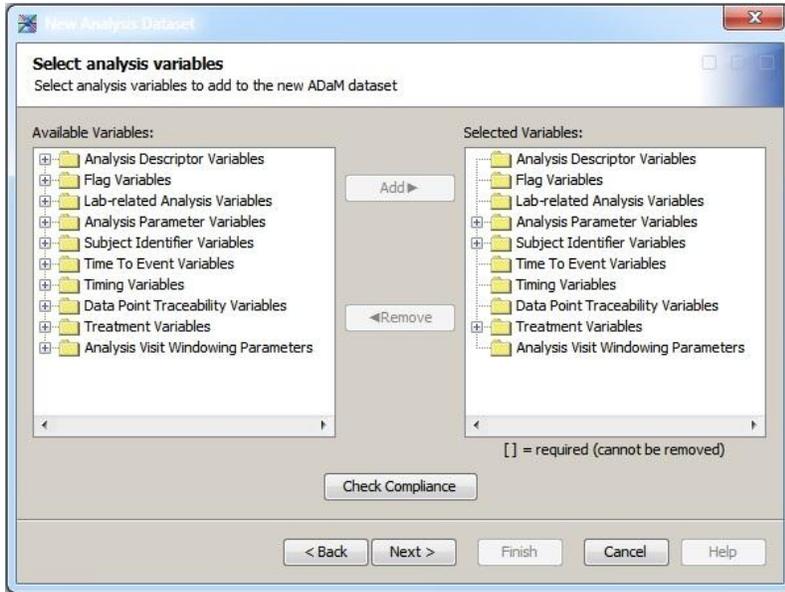


The source datasets are listed in the table positioned at the top left of the wizard page. When you select a specific dataset, its variables are listed in the table on the top right-hand side of the wizard page, as is shown for the CE SDTM domain. The name and description of each variable is provided. If a particular variable is not compliant with the ADaM standard for some reason, the third column labeled "Compliant" will have an X mark in it. Select a non-compliant variable and press the button labeled "X : Non-Compliance" and a popup window will explain why the variable is not compliant.

Once you have selected one or more variables in the top right-hand table, you can press the Add button and they will be added to the table in the bottom left part of the window. It was decided that the user should be allowed to add non-compliant variables, so you can still add them. If you added one or more variables and decide that you would rather not include them in the dataset, you can select them and press the Remove button. Once you have selected all the variables that you want to include from the source datasets, press the Next button to advance to the next wizard page.

SELECTING ADAM VARIABLES

The next wizard page allows you to select the ADaM variables for your dataset. You are presented with the variables associated with the particular type of dataset that you are creating. They are organized into the same categories that are provided in the ADaM specification. Any variables that are required are preselected and their names will be contained in square brackets. It is not possible to remove any of these required variables. The following screen dump displays this wizard page.



The set of available variables are displayed in a list on the left and those that have been added to the dataset are placed in the list on the right. To add one or more variables, select them in the list on the left and then press the Add button. This causes them to be moved to the list of selected variables on the right. If you had selected a variable to include but decide that you do not want it in the dataset, select it in the list of selected variables on the right and press the Remove button, which will place it back in the list of available variables.

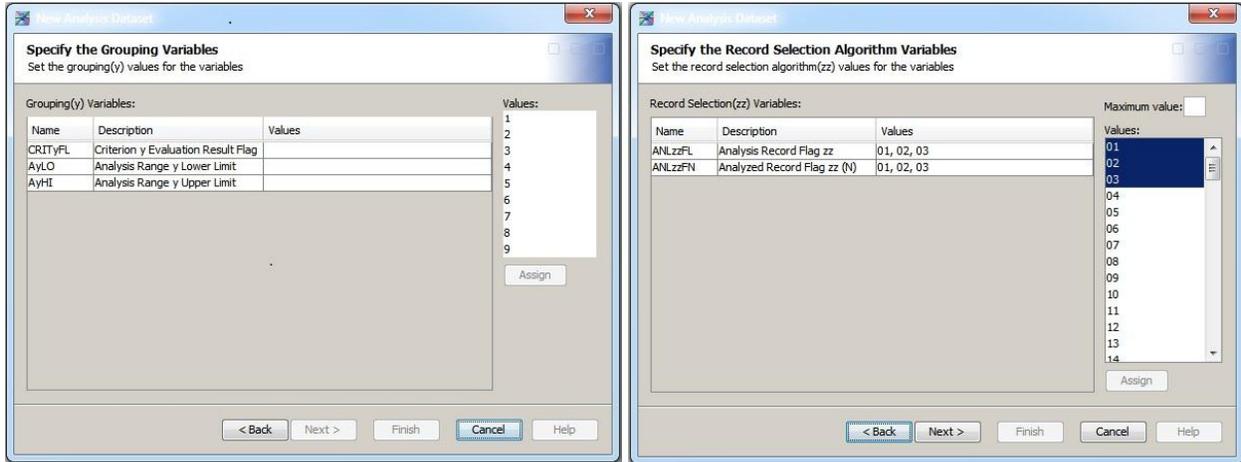
There are cases where inclusion of one variable requires the inclusion or exclusion of another variable based on constraints in the ADaM specification. These variable selection dependencies can be evaluated by pressing the button labeled "Check Compliance". The set of variables that you have selected will be examined and you will be told whether any variable dependency constraints have been violated. Once you complete the process of selecting the variables for your dataset, press the Next button to proceed to the next wizard page. It is also always possible in the wizard to use the Back button to return to a previous wizard page if you need to make changes to any prior selections you have made.

PROVIDING VALUES FOR VARIABLES WITH XX, Y, ZZ PARAMETERS

You may have selected variables that contain parameters denoted by a lower-case xx, y, or zz. These are placeholders to be replaced by numeric digits. The xx refers to a specific period of the study and must be replaced with a zero-padded two digit integer (01-99) denoting a specific period. A single digit value for the y parameter is used to refer to a grouping or other categorization, analysis criterion, or analysis range. A zz parameter is an index for the zzth record selection algorithm. It should be replaced with a zero-padded two digit integer (01-99). You would typically have multiple actual variables derived from the same variable template that has a parameter, with the integer substitution for the parameter providing a unique name for each variable.

A separate wizard page is provided for the setting of each parameter type (xx, y, zz). A wizard page is only shown for the parameter types included in the variables that you chose. In the following screen dumps, variables with y and zz parameters were selected for inclusion in the dataset.

PhUSE 2013



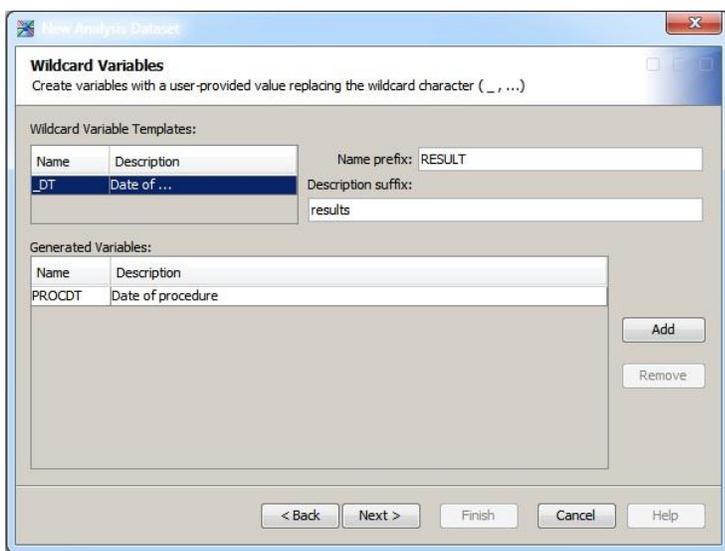
Each wizard page has a similar layout. A table is provided on the left that lists the variables that contain the particular parameter type. Each variable's name and description are provided and the third column lists the values that you have assigned for the parameter. On the right-side of the wizard page is a list of values that can be associated with the parameter. You can select one or multiple values from this list. Once you have selected one or more variables and one or more values for their parameters, you press the Assign button which places the values you selected into the third column of the table labeled Values.

For the y parameter type, the value list only contains the values 1-9. For the xx and zz parameter types, a list is provided of all 99 values from 1-99. You can enter a maximum value in the field above the values to reduce the number of entries in this list of values. Once you have assigned at least one value for each variable with an xx, y, or zz parameter, you can advance to the next wizard page.

For a given variable, if you provide N values, you will get N corresponding variables added to the dataset. There are a few variables that include both an xx and y parameter. In that case you will get N*M variables, where N is the number of x values and M is the number of y values you have assigned.

DEALING WITH WILDCARD CHARACTERS

Some ADaM variables include a wildcard character prefix (_) in the name and wildcard suffix (...) in the description, allowing you to provide your own substitution characters. If you selected any of these variables, a wizard page is presented so you can provide the text to substitute for the wildcard character. The wizard page presents a list of the selected wildcard variable templates.



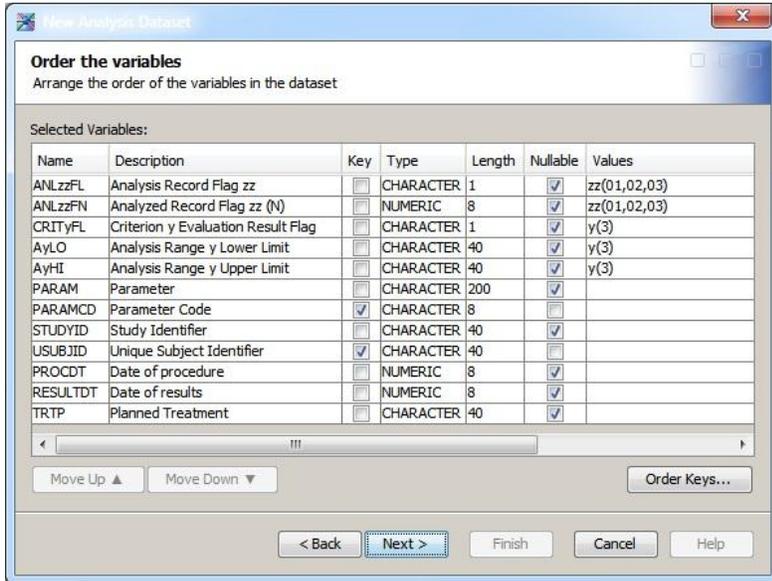
You should select a variable, specify a name prefix and description suffix and then press the Add button. This will add the variable to the list of generated variables. You can add additional prefix and suffix values for the same

PhUSE 2013

variable, resulting in additional variables in the dataset. If you add a variable and decide you no longer want to include it, select the variable in the list of generated variables and press the Remove button.

ORDERING THE VARIABLES

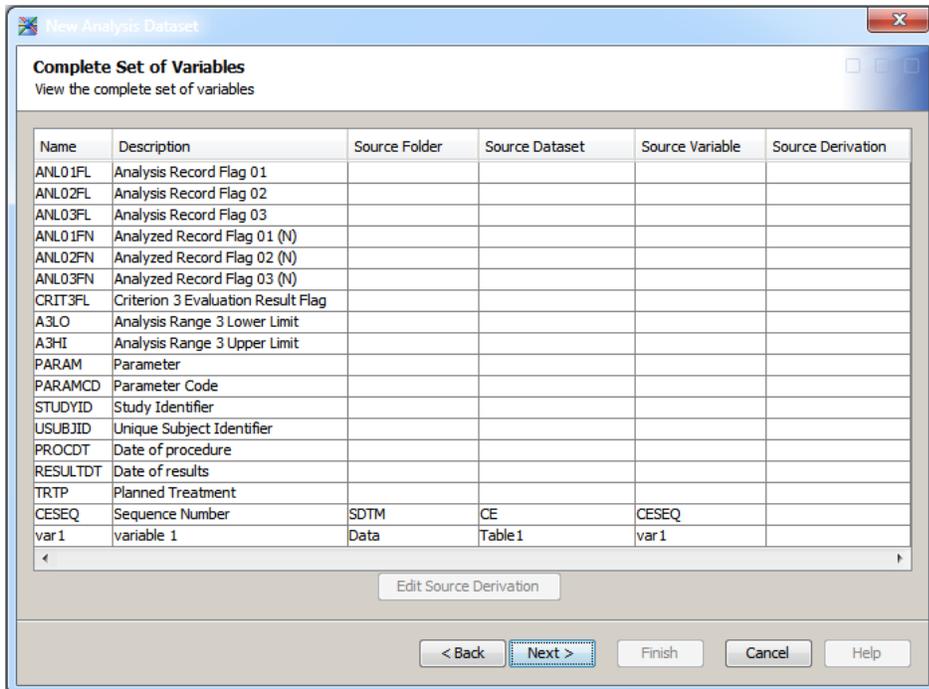
You can change the order of the variables in the dataset on the next wizard page. Select one or more contiguous entries in the table and press either the Move Up or Move Down buttons to rearrange the order of variables in the dataset.



There is also a button provided to order the keys for the dataset. Pressing the button labeled "Order Keys" brings up a separate dialog listing the key variables, allowing you to rearrange them in a similar manner.

ADDING TRACEABILITY INFORMATION

The next wizard page displays a complete list of all the variables in your dataset. This includes all the substitutions for wildcard characters as well as all values you provided for the xx, y, and zz parameters. For variables you directly included from a source dataset, the folder, dataset, and variable columns will have the appropriate values.



PhUSE 2013

You can enter values for each of these columns to provide traceability information about the derivation of your variables. The source derivation column on the far right has an associated multi-line editor you can use to provide a detailed description or pseudo-code describing the algorithm to be used in the derivation. This editor is invoked by pressing the button labeled “Edit Source Derivation.”

GENERATING THE DATASET

A final wizard page is displayed with some summary information about the dataset you are creating and a Finish button. When you press the Finish button, the ADaM dataset and associated metadata is created.

CREATING A USER-DEFINED TEMPLATE FOR DATASET CONSISTENCY

Users of SAS Clinical Data Integration may want to define a set of their own fully-defined custom templates for their ADaM datasets that can be shared across studies. This allows you to establish organization-wide consistency in your dataset definitions. This is possible in SAS Clinical Data Integration by promoting a dataset to be a user-defined template, which adds it to the list of datasets associated with the standard.

Once you have defined a dataset and promoted it to the standard, you can directly create an instance of one of these datasets without going through the process of defining all the variables to be included. Select the folder in which you want to place the new dataset and under the New menu, a menu entry labeled “User-defined Analysis Dataset” is used to create an instance of the dataset.

CONCLUSION

The new ADaM dataset creation wizard in SAS Clinical Data Integration allows you to easily generate an ADaM-compliant dataset. It is driven by metadata describing the standard’s dataset templates. The user interface allows you to quickly identify data from other datasets to be used as sources for variables in the ADaM dataset. It also provides an interface for defining the values for the xx, y, and zz parameters found in some ADaM variables. It also provides a means for generating variables that include wildcard characters. The wizard provides a quick and easy means of generating your ADaM datasets.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

David Jordan
SAS Institute
SAS Campus Drive
Cary, NC 27513
Work Phone: 919-531-1233
Email: david.jordan@sas.com

Brand and product names are trademarks of their respective companies.