Easier and More Accurate Importing of Data from Excel® into SAS®

Arnold R. Cowan

Cowan Research LC and Iowa State University

This article provides a tutorial on using SAS Enterprise Guide, included with SAS for Windows (PC SAS), to import data from Excel workbook files into SAS data sets. If you use Eventus® for Windows or use PC SAS/Connect to submit Eventus code to run on WRDS, you might need to create a SAS data set from an Excel file while preparing your data. However, the content below is not specific to Eventus. A beginner's knowledge of SAS for Windows is helpful but not essential for fully understanding this article. Some researchers choose to use Enterprise Guide (EG) for all their SAS and Eventus computing, while others prefer regular PC SAS. PC SAS and EG each has advantages and disadvantages in general; importing Excel files is an area where EG shines.

Older SAS methods of directly importing Excel data, such as the File Import wizard in regular PC SAS or manually coding a Proc Import step, can be useful but are not always fully suitable. Dates sometimes import incorrectly, and numeric data sometimes import as text that cannot be used for the intended purpose without first coding a conversion. For some Excel files, coding a Proc Import step can involve a substantial learning curve. Some users routinely re-save their Excel files in comma separated (CSV) format in preparation for importing into SAS. While this can improve the consistency of importation, Excel sometimes drops needed information when saving to CSV. An example is the leading zeros of numeric values in a mixed numeric-alphanumeric column, such as a CUSIP column, that should be imported as text with the zeroes intact. Enterprise Guide's Excel importation feature largely overcomes the limitations of the older methods. Most importantly, it allows the user to interactively manage the importation at the column level.

To use Enterprise Guide, you need to have SAS for Windows. Enterprise Guide is an included but separately run application that normally appears in the SAS folder of the

Windows Start menu or Start screen. If you have SAS but can't find EG, it may have been omitted during installation, in which case you need to re-install SAS with EG.

Before importing, check your Excel file for a couple of common situations where EG needs a little help. If a column that should import as text has all-numeric values in some cells, select the column in Excel, right-click, select Format Cells, and select Text format. If a numeric column has values that are displayed with excessive rounding, similarly use Format Cells (in Excel) to select Number format with plenty of decimal places. (Display rounding does not affect calculations within either Excel or SAS, but EG sometimes "sees" the displayed values instead of the underlying values when importing an Excel worksheet.) Save the Excel workbook file.

The step-by-step process for importing from Excel follows. If your Excel file is relatively large, one or two steps could take a minute for EG to process.

1. Launch SAS Enterprise Guide.
2. If the welcome screen appears, choose New Project.
3. On the menu bar just under the words "SAS Enterprise Guide", click File, then select Import Data.
4. Browse for and select the Excel file to import.
5. The Specify Data screen comes up. Click Browse, and edit the contents of the File name box if necessary. Choose a 1-32 character name that consists only of letters, digits and the underscore character (_), has no blanks, and does not start with a digit. (EG is actually a little more flexible than this, but following these rules will simplify some things for most research users.) Click Save, then Next.
6. On the Select Data Source screen:
   a. Click on the worksheet to import. If the Excel workbook file contains only one worksheet, it should already be selected. If you have more than one worksheet to import from the same workbook, do each one separately.
   b. In case you need to limit the importation to a specific range of cells, there is a checkbox to enable entering the range. For the typical worksheet with column names in a single row at the top (blank rows before or after the column names don't matter), there is no need to specify a range. To import only some columns and omit others, don't specify a range here just for that purpose; there will be an easier way in a moment.

c. First row of range contains field names may be checked already; if not, select it if the Excel worksheet has column headings. Blank rows at the very top are ignored, i.e. the "first row" doesn't have to be literally row 1 as shown in Excel. Blank rows after the heading row are okay too.

d. Select Rename columns to comply with SAS naming conventions.

e. Click Next.

7. The Define Field Attributes box, which now appears, allows corrections of EG's mistaken guesses about the nature of the data and changes to the proposed variable names and labels. For example, in the screen shot below, Enterprise Guide identifies some variables as character that should be numeric variables. The Excel columns will still be columns, also called variables, in the SAS data set, but are shown one to a row in the Define Field Attributes box. (The box does not have a maximize widget. If you prefer to see more rows at once or see roomier columns, expand the box by clicking the right or bottom edge or lower-right corner and dragging.)
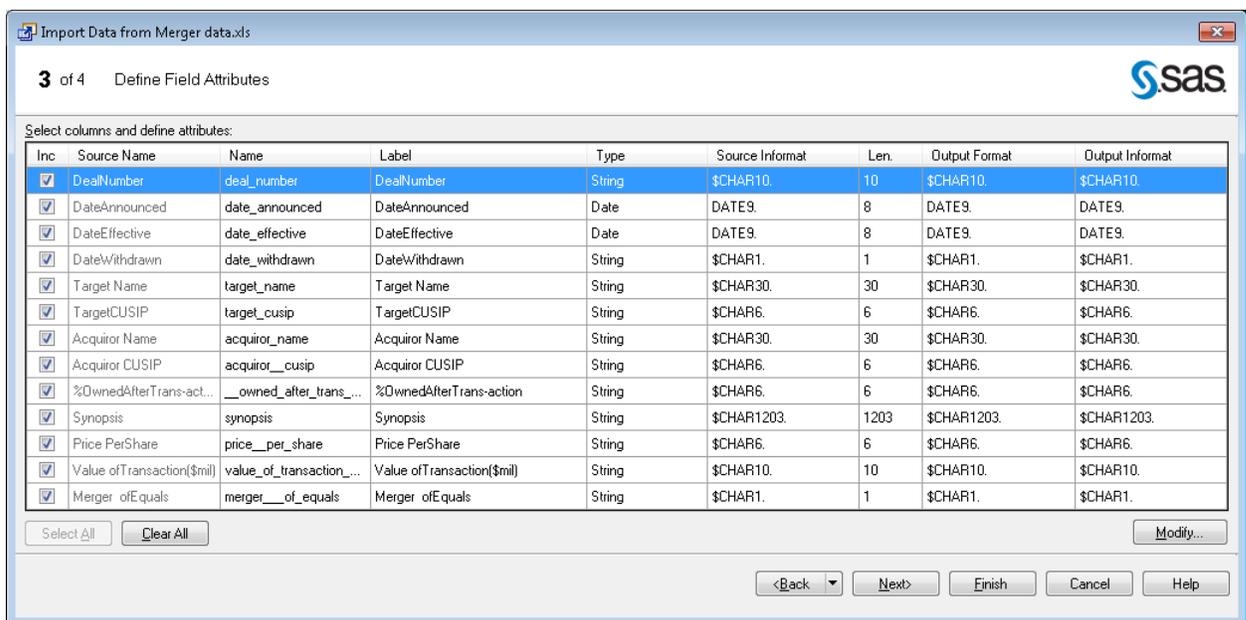
a. *Correcting a data type.* The deal number is all numeric but is importing as text. This can be changed if desired. On the DealNumber line, click on the cell in the Type column (which currently shows String),

then click on the drop-down arrow that appears, then click Number.

b. *Fine tuning with the informat.* Enterprise Guide is excellent at figuring out the formats of many types of numeric data on its own. In other cases, you may need to fix how a variable imports by changing its Source Informat setting (SAS input format). For example, if a column contains numeric data displayed with commas when appropriate, but many rows contain absolute values too small to need a comma, EG may not select an informat that accepts commas. Change the source informat for such a column to COMMA$w.d$ where w is the maximum width of a displayed value and d is the number of decimal places. Usually EG automatically sets the w correctly and you just need to adjust d. Documentation of all SAS informats is [here](#) among other places. Each variable's SAS storage length appears under Len; for text columns, this follows from the informat, and for numeric columns it correctly defaults to eight. Don't change Len manually.

c. *Output formats.* The Output Format affects how the variable looks when you view the data set in SAS and how it comes out if you make viewable output from it. The output format is only a template that SAS uses to display the column; it doesn't affect the stored value or calculations. You can set this property now if you like, but it is relatively easy to change the format within SAS or override it at any point in a program. (Ignore the oxymoronically named Output Informat setting; it has no effect on anything one would do for academic research or similar uses of SAS.)

d. *Renaming a variable.* The Excel column that contains percent owned after the transaction is given an automatic SAS variable name that is technically valid but unwieldy: "__owned_after_trans__action". To change it, click on the cell containing the name, click again to activate editing mode, and type a new name. A similar method can be used to modify the variable label, which may not be worth the time. The variable label is for display and isn't used in coding. If an automatic label that EG shows for a particular column could be confusing or misleading when you or a co-author view the data in SAS later on, you might want to delete it and leave this property blank now. Variables without labels always show up under their variable names when SAS displays them.

e. *Omitting an Excel column.* To omit a column from the SAS data set, uncheck the box at the left end of the corresponding row of the Define Field Attributes screen.

f. *Finish.* When done with the Define Field Attributes screen, click Finish.

8. The imported data set appears. The data set is ready to use in the current EG session. To save the data set for use in regular PC SAS (as opposed to EG): Click the Data drop-down that appears just above the data-set column names and select Download
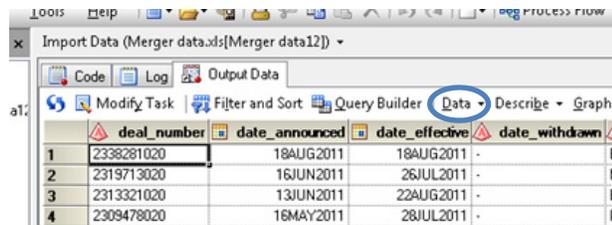
    Data Files to PC. On the Select the SAS data sets to copy to your PC screen, click Add and select the data set just created, then click the Open button, then the Next button. On the Select the folder for the SAS data sets on your PC screen, browse for or type in the folder in which to save the data set. You can use any folder. After specifying the folder, click Finish.

9. Enterprise Guide makes it easy to go back and make changes to a data importation task without losing prior work. To revise and resubmit the importation, click the Modify Task button, or find the Excel file listed under Process Flow, right-click on Import Data under or beside it and select Modify.

10. Optionally, before leaving EG, save the project (File, Save Project) so that if you later discover anything else that needs to be changed, you can come back and make modifications instead of starting over.

11. If you are ready to use the data set in regular PC SAS (with or without Eventus), close EG and launch or switch to PC SAS. In case you are unfamiliar with saving and re-using SAS data sets, here are the basics of using PC SAS to access the data set you just created in and "downloaded" from EG.

    A library name (also called a libref) is a SAS pointer to a folder. SAS documentation refers to a folder that contains SAS data sets, and has been assigned a library name, as a library. The same Windows folder can contain other types of files and SAS will generally ignore them. If there is not already a library name pointing to the folder where you saved the data set, create one now. Go to the Explorer tab in SAS, right-click Libraries and select New. In the Name box, type a one-to-eight character name starting with a letter or underscore and having the remaining characters be letters, underscores or digits. Click Enable at Startup if desired, to save the name for future use. In the Path box, browse for or type in the Windows folder path and name in which the data set was saved from Enterprise Guide. There is no need to do anything

with Engine or Options. Click OK. Use a two-part name to refer to the data set in SAS and Eventus programs. For example, if the data set name created in EG is merger_data and the library name just created is acq, refer to the data set in SAS and Eventus programs as acq.merger_data. A library may contain as many data sets as desired.