

SAS Enterprise Guide: Introduction to stored processes and custom tasks

Alex Dmitrienko, Ph.D., Eli Lilly and Company, Indianapolis, IN

ABSTRACT

SAS® Enterprise Guide® provides a powerful framework for creating new tools known as stored processes and custom tasks. These tools are conceptually similar to SAS macros but, unlike regular macros, they feature attractive graphical user interfaces and can significantly shorten the user's learning curve and greatly increase his/her productivity.

This paper provides an overview of main features of stored processes and custom tasks and discusses development/installation procedures. It also provides useful information for custom task developers, for example, describes the underlying data model, how custom tasks interact with the SAS System and how to build custom task using Microsoft Visual Basic or other languages.

ENTERPRISE GUIDE CUSTOM TASKS

SAS Business Intelligence products such as Enterprise Guide and Add-in for Microsoft Office ship with a large number of standard tasks for performing common data manipulations and generating data sets or output. The tasks feature graphical user interfaces and are conceptually similar to wizards in Microsoft Office programs such as Word and Excel.

In addition, Enterprise Guide provides a powerful framework to enable the users to create their own tools with graphical interfaces: *custom tasks* and *stored processes*. The tools can be thought of as SAS macros with a point-and-click interface. They are similar to SAS macros in that they input certain parameters, generate SAS code and create output or data sets. However, unlike regular SAS macros, custom tasks exhibit a number of useful features summarized below.

ENTERPRISE GUIDE CUSTOM TOOLS AND LEARNING CURVE

Custom tools (custom tasks and stored processes) are more "intelligent" and user-friendly than SAS macros. It is well known that powerful SAS macros are complex and create a steep learning curve even for experienced SAS users. This happens because SAS macros force the user to deal with each one of their numerous options or arguments. In contrast, custom tools offer a user-friendly interface that updates itself as the user makes his or her choices and reduces the number of arguments the user needs to specify. This feature helps eliminate the learning curve and enables less experienced users to quickly understand the process of performing a complex analysis or setting up a complex report.

DISTRIBUTION OF ENTERPRISE GUIDE CUSTOM TOOLS

Consider, for example, custom Enterprise Guide tasks. They are easy to distribute and upgrade because each task is stored in a single file. If custom tasks are saved in a central location, individual users can copy them to their hard drives and install them using the Add-in Manager in Enterprise Guide. Once installed, custom tasks can be set up to search for their own updates and, if an update is found, inform the user and help him or her upgrade itself.

VALIDATION ISSUES OF ENTERPRISE GUIDE CUSTOM TOOLS

Finally, since custom tools are similar to SAS macros, they require the same level of validation; for example, they can be validated using validation procedures for multiple-use SAS macros. The advantage of using custom tools is that they are tamper-proof (unlike SAS macros) and do not require any revalidation work once they have been distributed to users.

OVERVIEW

This paper focuses on custom Enterprise Guide tasks, provides an overview of their most important features and shows how the readers can boost their productivity by using custom tasks developed by other users or by building their own tasks.

Using a custom task created by the author (Data Review task), the paper explains how to install and execute custom tasks in the Enterprise Guide environment. It also describes the inner workings of custom tasks, interactions between tasks and Enterprise Guide and the role of the SAS System.

The reader interested in building tasks will find information about the technologies that provides a foundation for custom task development and a brief summary of useful information available on the SAS web site.

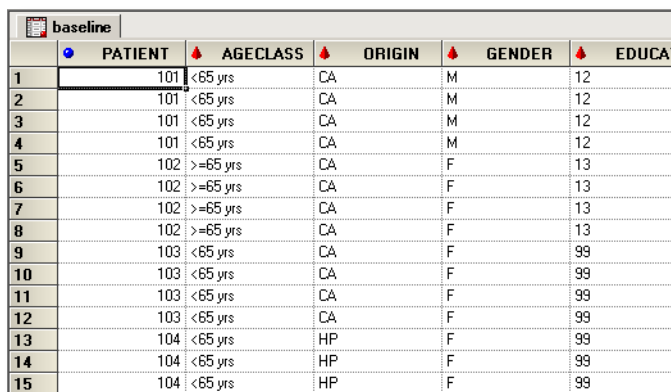
DATA REVIEW TASK

To describe key features of custom tasks in the Enterprise Guide environment, consider a custom task for reviewing the quality of data sets. SAS and Enterprise Guide users frequently deal with data files that come from multiple sources (both reliable and unreliable) and often employ informal data inspections aimed at assessing the overall quality of the data and identify spurious data values or data collection/data entry errors.

An efficient way of identifying data problems/inconsistencies is based on a review of frequency counts for important character variables and extreme (smallest and largest) values for important numeric, date and time variables. By examining frequency counts, SAS and Enterprise Guide users can assess the prevalence of missing values or data entry errors (e.g., misspelled words). Likewise, a review of extreme values helps identify data points that are not consistent with the rest of the data (outliers) and most likely represent data collection or entry errors.

The Data Review task introduced below provides a useful tool that facilitates a review of the distribution of important variables. This task relies on a wizard-type interface with three screens:

1. Specification of the ID variable that uniquely identifies each record in the data set.
2. Selection of review variables for which a frequency count or extreme values need to be generated.
3. Selection of secondary variables that help examine the distribution of the review variables.



	PATIENT	AGECLASS	ORIGIN	GENDER	EDUCAT
1	101	<65 yrs	CA	M	12
2	101	<65 yrs	CA	M	12
3	101	<65 yrs	CA	M	12
4	101	<65 yrs	CA	M	12
5	102	>=65 yrs	CA	F	13
6	102	>=65 yrs	CA	F	13
7	102	>=65 yrs	CA	F	13
8	102	>=65 yrs	CA	F	13
9	103	<65 yrs	CA	F	99
10	103	<65 yrs	CA	F	99
11	103	<65 yrs	CA	F	99
12	103	<65 yrs	CA	F	99
13	104	<65 yrs	HP	F	99
14	104	<65 yrs	HP	F	99
15	104	<65 yrs	HP	F	99

REVIEW OF DEMOGRAPHIC DATA

As an illustration, consider a data set containing patient demographic data (patient ID, age class, racial origin, gender, education, body weight, height, etc). The data set contains one record per patient. Suppose that you are interested in examining the consistency of the body weights in the data set and identifying outliers.

The five steps below illustrate the process of creating a summary of the extreme body weight values using the Data Review task. It will be assumed that you have Enterprise Guide 3.0 installed on your personal computer along with SAS Release 8.2 or later.

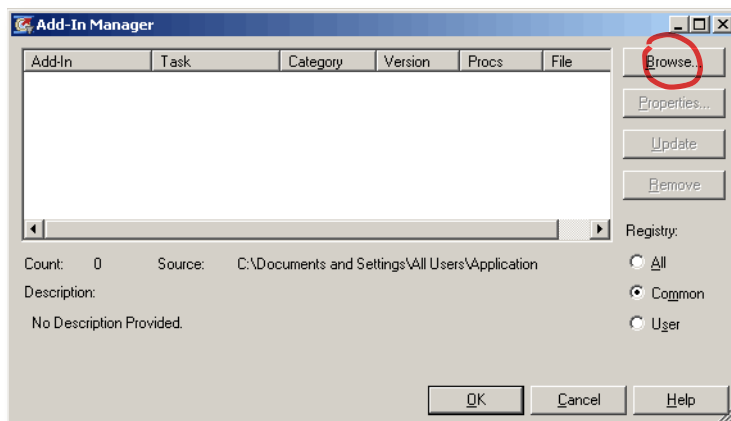
Figure 1. Demographic data

STEP 1: INSTALL THE DATA REVIEW TASK

Before you can run the Data Review task, you need to install it on your computer. The task can be found in the Business Intelligence Exchange on the BISUG web site:

1. Point your Internet browser to the Business Intelligence Exchange page (<http://www.bisug.org/bix/general.html>) and select the Data Review task (bixge001).
2. Save the task file (bixge001.dll) anywhere on your hard drive.

Note that you must be logged in to be able to access the Business Intelligence Exchange.



To install the downloaded file on your system, launch Enterprise Guide, choose the Add-In menu on the Enterprise Guide menu bar and select Add-In Manager.

This will bring up the Add-In Manager window displayed in Figure 2. Select Browse to specify the location of the task file.

Figure 2. Add-In Manager window

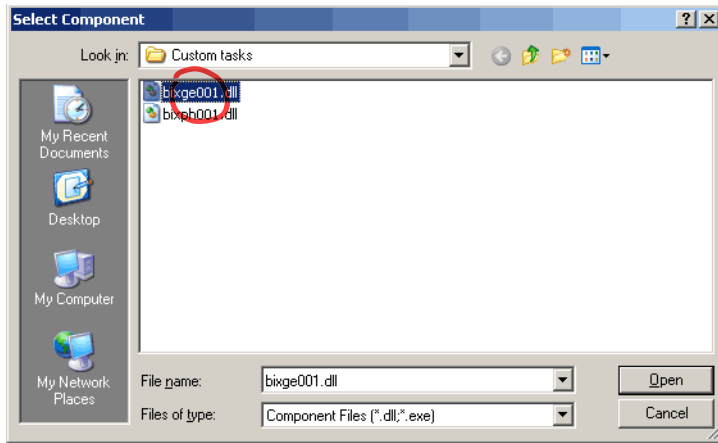


Figure 3. Specify the location of the task file

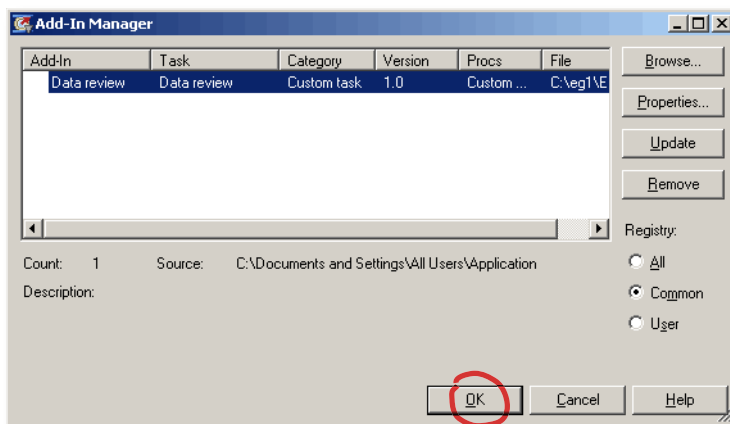


Figure 4. Data Review task displayed in the Add-in Manager window

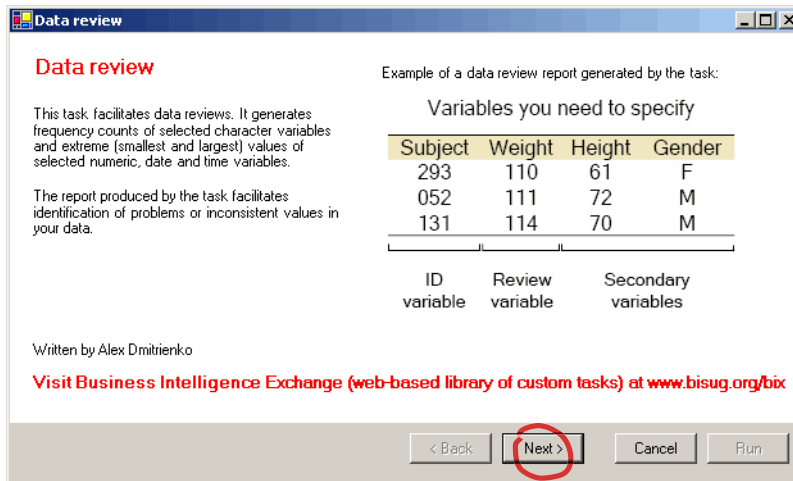


Figure 5. Data Review (Welcome screen)

Find bixge001.dll and click Open.

Enterprise Guide will install the Data Review task and add it to the list of available custom tasks. Click OK to close the Add-In Manager.

To launch the Data Review task, you will need to choose the Add-In menu again. You will notice that the Data Review task has been added to the menu. Click on Data Review to open the task.

STEP 2: DATA REVIEW TASK'S WELCOME SCREEN

The first screen of the Data Review task describes the main features of this task and provides an example of a report it generates.

The screen also defines the variables you need to specify to produce a data review report:

- ID variable,
- Review variable,
- Secondary variables.

The Subject variable is as a good example of an ID variable since it uniquely identifies each record in this data set. The Weight variable serves as a review variable and the reports lists its three smallest values. Lastly, the Height and Gender variables are (optional) secondary variables that will help you determine whether or not the three weights are feasible.

Click Next to proceed to the next screen.

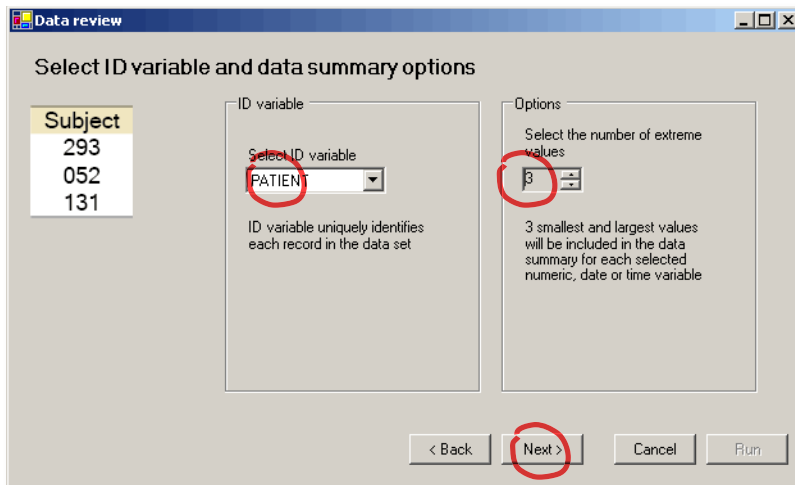


Figure 6. Data Review (Select the ID variable)

STEP 3: SELECT THE ID VARIABLE

This screen includes two panes. In the left-hand pane you need to select the ID variable from a dropdown list of all variables in the selected data set (in this case it is the Patient variable).

The right-hand pane shows the number of extreme values that will be listed for each numeric, date and time review variable. Change the default value (5 extreme values) to 3. Note that it generally makes sense to select less than 10 extreme values.

Click Next to proceed to the next screen.

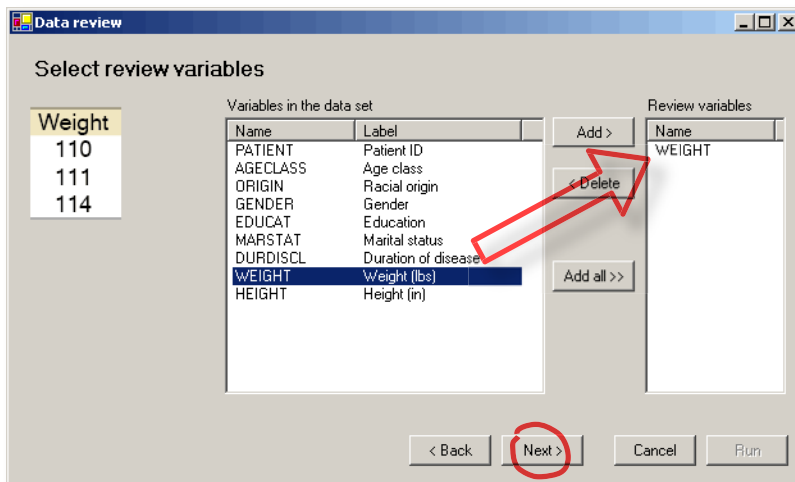


Figure 7. Data Review (Select the review variables)

STEP 4: SELECT REVIEW VARIABLES

This screen enables you to specify review variables that will be included in the data quality review report.

To choose the Weight variable, select it in the left-hand pane and click Add.

If you would like to select multiple variables, press the Shift or Ctrl key and use the mouse pointer to select variables in the left-hand pane.

To generate data quality review report containing all of the variables, click the Add all button.

Click Next to proceed to the next screen.

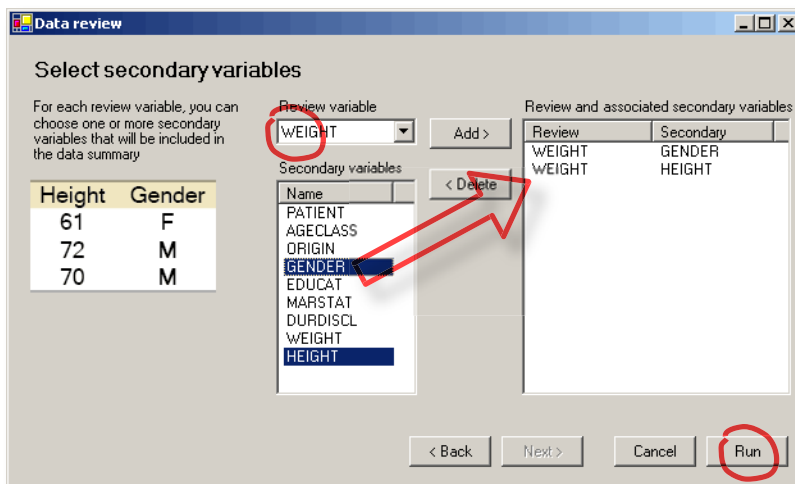


Figure 8. Data Review (Select the secondary variables)

STEP 5: SELECT SECONDARY VARIABLES

This is an optional step that deals with the so-called secondary variables.

As was explained earlier, secondary variables help identify feasible values of review variables. For example, it is prudent to display height and gender of patients with a very low weight.

To add the Height and Gender variables as secondary variables, select the Weight variable from the dropdown list and then choose Height and Gender in the list of secondary variables (press the Ctrl key to select the two variables simultaneously). Now click the Add button. The task will add two rows to the right-hand pane to indicate that the data review report for the Weight variable will include patients' height and gender.

Click Run to run the task.

patient	weight	height	gender
1703	104	63	Female
1004	109	51	Female
104	111	54	Female
2804	280	64	Female
2907	287	68	Male
1202	290	67	Male

After you click the Run button, the SAS code generated by the Data Review task will be submitted to the SAS System and the output will be displayed in the Workspace window.

The resulting data review report is shown in Figure 9. The report lists the three smallest and three largest values of the Weight variable. It also displays Patient IDs to make it easier to generate queries if the listed patient weights appear problematic.

A quick review of the report reveals that the Weight values are generally consistent with the height values except for Patient 1703 who is 63 inches tall yet weighs a little over a hundred pounds. The patient's weight or height could have been entered incorrectly and it is prudent to check the patient's records in order to determine where the problem lies.

Figure 9. Output generated by the Data Review task

CUSTOM TASKS IN ENTERPRISE GUIDE ENVIRONMENT

Now that you have become familiar with user-developed custom tasks, it will be instructive to find out what goes on behind the scenes and how the three players (custom task, Enterprise Guide and SAS System) interact with each other. A simple diagram displayed in Figure 10 depicts important steps in the process of generating output by Enterprise Guide tasks and defines the elements you need to work with as a custom task developer.

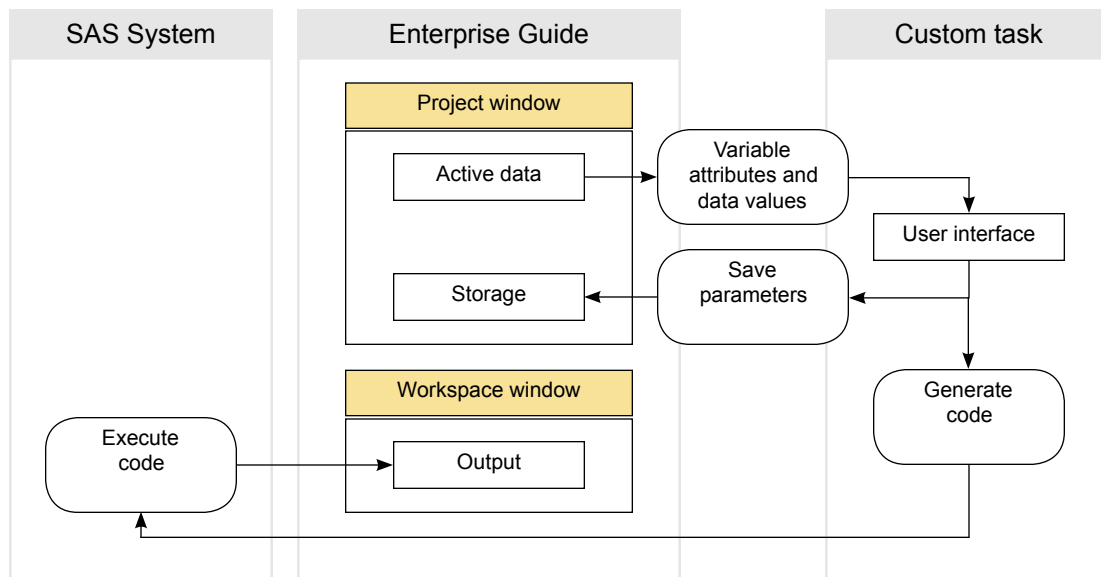


Figure 10. Behind the scenes look at custom tasks

As shown in Figure 10, one needs to go through four steps to generate output in the Enterprise Guide environment:

- **Active data set.** One data set in the Project window is designated as the active data and every time you launch a built-in or custom data task, the task creates a link that enables it to extract information (variable attributes and even data values) from the active data set.
- **User interface.** The task's user interface is designed to help you quickly select options or enter data that will later be used to generate SAS code. When you run the task, it uses the collected information to generate SAS code and passes it to Enterprise Guide. In addition, the task saves its parameters within the active project to save you time when you need to go back and make changes.
- **Code execution.** Enterprise Guide submits the code it has received from the task to the SAS System running on

your computer or a remote SAS server. The SAS System executes the code.

- Output. Once the code built by the task has been executed, Enterprise Guide takes over to bring the output or any other results back. The output created by the SAS System is displayed in the Workspace window.

If you are interested in building your own custom tasks, you need to concentrate on the first two steps of the outlined process, i.e., understand how to retrieve data set attributes and data values, build a user interface, save the parameters specified by the user and generate SAS code.

CUSTOM TASK DEVELOPMENT

Custom Enterprise Guide tasks are created using the Microsoft Component Object Model technology (Enterprise Guide 2.0) or Microsoft .NET technology (Enterprise Guide 3.0) and generally require advanced programming skills. They are typically written in Microsoft Visual Basic or C#. If you would like to take advantage of new features available in Enterprise Guide 3.0, you need to develop your tasks in Microsoft Visual Studio .NET using Visual Basic .NET or C# .NET. For example, the custom task described in this paper was written in Visual Basic .NET.

SAS Institute has provided useful information on its web site to help custom task developers get started. This information is briefly summarized below.

CREATING CUSTOM TASKS FOR ENTERPRISE GUIDE 3.0

The main source of information about custom tasks for Enterprise Guide 3.0 is the custom task web page at

<http://support.sas.com/documentation/onlinedoc/guide/release30/addins/index.htm>

This web page provides a high-level description of custom tasks, contains links to the Enterprise Guide 3.0 custom task documentation, Microsoft Visual Studio .NET template and provides source code for several simple tasks.

ENTERPRISE GUIDE 3.0 CUSTOM TASK DOCUMENTATION

You can download a zip-archive file of the Enterprise Guide 3.0 custom task documentation (SAS.EG.Addins.zip). The documentation comes in the form of a help file and provides a detailed (but rather technical) description of various elements custom developers need to implement to enable their tasks “talk” to Enterprise Guide. The help file also explains how to set up projects and create user interfaces in the Microsoft Visual Studio .NET environment. This information is useful if you have never worked with Microsoft Visual Studio .NET.

EXAMPLE PROJECTS

SAS Institute generously provides source code (mostly C# .NET and sometimes Visual Basic .NET) for 8 simple custom tasks that illustrate the power and flexibility of Enterprise Guide tasks. It is important to keep in mind that the custom tasks were created using Microsoft Visual Studio .NET 2003 and need to be slightly modified if you are planning to work in Microsoft Visual Studio .NET 2002.

MICROSOFT VISUAL STUDIO .NET TEMPLATE

Another very useful tool available on the custom task web page is a custom task template for Microsoft Visual Studio .NET 2003. The template can be downloaded as a zip-archive file (SASAddInTemplates.zip) and enables you to quickly create Visual Basic .NET or C# .NET projects for developing custom tasks. With this template, you do not need to worry about setting up numerous individual elements of Microsoft Visual Studio .NET projects and can focus on building user interfaces and SAS code generation mechanisms for your custom tasks.

LEARN MORE ABOUT ENTERPRISE GUIDE

If you are interested in learning more about Enterprise Guide, please see *A Visual Guide to Enterprise Guide*. This electronic book provides a concise review of main features of Enterprise Guide and is an ideal way to get up to speed with this powerful software package. For more information about the book, please visit the Business Intelligence SAS Users Group (BISUG)'s web site:

<http://www.bisug.org/visualguide>

Sign up for the BISUG membership to receive access to information on SAS Business Intelligence products including free Enterprise Guide tutorials, papers presented at recent SAS conferences, review of Enterprise Guide training courses and custom tasks submitted by Enterprise Guide users.

CONCLUDING REMARKS

The paper introduces user-developed tools (known as *custom tasks* and *stored processes*) that can be used with Enterprise Guide (as well as Add-in for Microsoft Office) and explains how SAS users can utilize them to increase their productivity:

1. A high-level overview of Enterprise Guide custom tasks and a comparison of custom tasks to user-written SAS macros.

2. Installation of Enterprise Guide custom tasks and their structure (these steps are illustrated using a custom task built by the author).
3. Information for users who are interested in developing their own custom tasks.

The paper demonstrates that custom tasks provide a significant improvement over user-written tools available in the regular SAS environment such as user-written SAS macros and provides information for users who would like to build their own tasks to streamline routine data management, analysis and reporting jobs.

CONTACT INFORMATION

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

Alex Dmitrienko
Eli Lilly and Company
Lilly Corporate Center
Indianapolis, IN 46285
Work phone: 317-277-1979
Fax: 913-402-7239
Email: Dmitrienko_Alex@lilly.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.