Converting CDISC Controlled Terminology to SAS Formats

Sandeep Juneja, ASG Inc., Cary, NC
Vivek Mohan, ASG Inc., Cary, NC

ABSTRACT:

CDISC Controlled Terminology (CT) is used to define and support the terminology needs of the CDISC domains. Applying CT to individual variables with different values across data sets in multiple studies can be a very cumbersome and pain-staking process. This paper describes an easy way of extracting unique values of the variable and unique values of the CT term and provides a GUI interface to carry out one-to-one mapping for CT terms and data values. This mapping allows generating customized SAS Formats which can be applied directly to raw data set variables when converting to CDISC compliant data sets. This utility not only reduces the time required to apply CT to applicable CDISC variables, but also eliminates any errors caused by manually typing in each unique value for a variable. The software/modules used in this utility are SAS, Microsoft Excel, VBA and SAS IOM.

INTRODUCTION:

CDISC CONTROLLED TERMINOLOGY (CT)

CT has been developed to standardize the unique values for certain variables in standard CDISC domains.

SAS INTEGRATED OBJECT MODULE (IOM)

SAS defines the Integrated Object Model as a set of “distributed object interfaces to SAS... IOM enables you to use industry-standard languages, programming tools, and communication protocols to develop client programs that access these services on IOM servers.”

By setting up references to the SAS IOM components in your VBA program, you can achieve most of the results that you can with SAS in batch.

THE SAS STORED PROCESS SERVER

SAS Stored Processes are slightly different than traditional SAS programs in that they must be executed on a SAS Integration Technologies server. They cannot be executed in a normal batch or interactive SAS session. The main difference between running a SAS job in a Workspace and running it as a Stored Process is that, with the former approach you can generate and submit code from the client or run code already existing on the server, while with the the Stored Process Server the code has to be accessible from the server.

Requirements:

The Mapping document is the most important document required for this utility (shown on left). Apart from the CDISC variable name and CDISC variable format, the mapping document should contain the following information at the CDISC variable level:

1. CT names must be specified, if applicable.
2. The variable name(s) and raw data set name(s) that maps to the CDISC variable must be specified in the following format <raw data set>.<variable name>(Format of Variable, if present). It can either be one-one or one-many depending upon the data sets or variables used to derive the CDISC variable.

The mapping document in conjunction with the raw data sets provides the information required to produce the CT Mapping Excel Sheet.
THE PROCESS:

1. CREATING INDIVIDUAL CT DATA SETS

Download the available CT Excel workbook from CDISC website (see link in 'More Information' section at the end of this paper) and convert each CT as a unique SAS data set. Please make sure that each SAS data set has the same name as that of the controlled terminology it represents. For e.g.: acn.sas7bdat will correspond to ACN controlled terminology.

2. GENERATING THE CT MAPPING EXCEL SPREADSHEET

Using the mapping document and the sas data sets generated for each CT, a sas program can be written to generate the excel spreadsheet that contains the following information per CDISC domain (see below)

1. Summary Information: This lists the unique CTs and list of variables to which the CT is applied.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Controlled Terminology</td>
<td>Variable List</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>ACN</td>
<td>AECN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ABEEF</td>
<td>ABEET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>BST</td>
<td>ABBST, ABสอบ, ABG2085</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>OOT</td>
<td>AECOUT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>DOC</td>
<td>AECOUT</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Variable Mapping Information: It contains the following information:
   (i) CDISC Format Name
   (ii) CDISC Variable Name
   (iii) Raw variable (Raw variable label)
   (iv) Raw Variable distinct values
   (v) Name of the CT
   (vi) Section for any user comment.

Hint: Use ExcelXP tagset along with SAS to generate the above excel workbook with one sheet per CDISC domain.

3. SAS IOM/ VBA

Establish the connection between SAS IOM & VBA using appropriate References & code.

References:

In order to use IOM from a VB project, it is necessary to add in references to the SAS link libraries. To add these, go to Project>References on the main VB menu. Scroll down the list and click to add the following references:

- SAS: Integrated Object Model (SAS System 9.1)
- SASWorkspaceManager 1.1 Type Library

This will allow VB to resolve references to the SAS objects used to connect to the server.

A brief overview of the steps used in our utility to develop the GUI interface is provided below

i). Establish Connection with SAS IOM - Below is the standard code used to establish the connection bridge between SAS & VBA.
Also, similar to the code above, snippets of code to close SAS, Submit the SAS job etc. are available on the internet.

ii). Establish the Library connection for the CT data and raw data set.

iii). Establish the connection to SAS via ADO to read in the SAS data set.

iv). Read in the unique value for CT data and for each raw data set variable and display the results in the GUI screen.

v). Select the appropriate Raw variable value and the CDISC value and implement one-one mapping.
4. GENERATING THE FORMAT.SAS PROGRAM

Once the mapping is completed, the stored SAS process can be called to create the format.sas program.

---

**CONCLUSION:**
This utility not only removes any typing errors but also drastically reduces the time required to map the variables to the proper CT values. It helps in standardizing the CDISC data sets very efficiently and quickly.
REFERENCES


ACKNOWLEDGEMENTS

We would like to thank Balaji Ayyappan and Nivethitha Manickavasagar for providing us excellent ideas and helping develop code to implement this utility.

RECOMMENDED READING

CDISC CONTROLLED TERMINOLOGY
http://www.cdisc.org/standards/terminology/index.html

SAS IOM

CONTACT INFORMATION

Your comments and questions are valued and encouraged. You can contact us at:

Sandeep Juneja  
ASG Inc.  
2000 Regency Parkway, Suite # 675  
Cary, NC 27518  
Phone: (919) 653 3982  
Email: sandeep.juneja@asg-inc.com

Vivek Mohan  
ASG Inc.  
2000 Regency Parkway, Suite # 675  
Cary, NC 27518  
Phone: (919) 653 3997  
Email: vivek.mohan@asg-inc.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