

Paper PO18

Benefits of Using Bridging Metadata

Doug Janssen, Eli Lilly and Company, Indianapolis, IN

ABSTRACT

As working on analyzing clinic trial data for pharmaceutical companies, we first collect subjects' information through either electronic forms or paper forms. However, both are raw data and cannot be used for analysis purposes. Therefore, we need to build up analysis data based on the information we collected. While creating data we need to define the formats for each variable such as variable name, label, type, and length. You can input those formats in the program as part of data step. You also can input formats in Excel spreadsheet as a dataset. Whereas the later is easy for user to input, check, and modify if necessary. This paper provides an idea about how to generate analysis data through Bridging Metadata, which connects Source data (such as data location and variable name) and Target data (such as data location, variable name, label, variable type, variable length, and formats). Both Source data and Target data are created by Excel and using Bridging Metadata is not only able to reduce the cycle time of developing customized programs, but also able to increase the re-usability by enabling automation.

INTRODUCTION

Countless hours are required to reorganize, reformat, append, and merge data from one or more data sets to create a new data set. A process that uses metadata to create these new data sets has been developed that simplifies this process, reduces validation time and for future projects enables this process to be automated.

I will begin by discussing the different metadata data sets that are needed to create Target data sets from Source data sets and then present an overview of the Bridging Program which creates the Target data set.

Target Data Set Information

The first metadata data set contains information about the target data set. The variables in this metadata data set are:

Variables

- Target Name of the target data set.
- TargetOrder The order that the data set will be created.
- TargetMacro Macro to call once this program has finished moving the source data variables to the target data variables. This macro would contain any code required to do any derivations.

Example

Target	Target Order	Target Macro
SUBJINFO	1	
DISPOSIT	2	Disposit

Source / Target Information

Next a metadata data set is needed that contains information about how the source data set should be merged or appended to create the target data set. The variables in this metadata data set are:

Variables

- Source Name of the source data set.
- Target Name of the target data set.

- **Target Order** Order in which the source data sets are processed. In the example below the source data set SUM1001 is processed prior to the source data set AE1001.
- **Target Merge By** This describes how the source data sets are merged together. If this is blank that means that either the source data set is the first data set to be processed or that the source data set should be appended to the previous source data set.

Example

Source	Target	Target Order	Target Merge By
SUM1001	DISPOSIT	1	
AE1001	DISPOSIT	2	Invid Subjid Aeid

Source / Target Variable Information

This metadata data set contains information about how the target data set is populated using the source data set variables. The variables in this metadata data set are:

Variables

- **Source** Name of the source data set.
- **Target** Name of the target data set.
- **Source Variable** Name of the source variable.
- **Target Variable** Name of the target variable
- **Before Macro** Macro that the bridging program should call prior to moving the source variable to the target variable. In the example below the macro `_SASDATE` is called to convert the source variables to date formats.
- **After Macro** Macro that the bridging program should call after moving the source variable to the target variable.

Example

Source	Target	Source Variable	Target Variable	Before Macro	After Macro
AE1001	DISPOSIT	Aeid	Aeid		
AE1001	DISPOSIT	HLGTERM	HLGTERM		
AE1001	DISPOSIT	HLTERM	HLTERM		
AE1001	DISPOSIT	Invid	Invid		
AE1001	DISPOSIT	LLTCD	LLTCD		
AE1001	DISPOSIT	LLTERM	LLTERM		
AE1001	DISPOSIT	PTERM	PTERM		
AE1001	DISPOSIT	SOCTERM	SOCTERM		
AE1001	DISPOSIT	Subjid	Subjid		
AE1001	DISPOSIT	Visid	Visid		
SUM1001	DISPOSIT	AEID	AEID		
SUM1001	DISPOSIT	RSDCFIN	DS		
SUM1001	DISPOSIT	DISCDT	DSDT	<code>_SASDATE</code>	
SUM1001	DISPOSIT	DISCDT	DSDTC	<code>_SASDATE</code>	
SUM1001	DISPOSIT	DSSTAT	DSSTAT		
SUM1001	DISPOSIT	'Final'	DSTP		
SUM1001	DISPOSIT	DTHDT	DTHDT	<code>_SASDATE</code>	
SUM1001	DISPOSIT	DTHDTC	DTHDTC	<code>_SASDATE</code>	
SUM1001	DISPOSIT	FACILITY	FACILITY		
SUM1001	DISPOSIT	INVID	INVID		
SUM1001	DISPOSIT	RESPROJ	RESPROJ		
SUM1001	DISPOSIT	SDYID	SDYID		
SUM1001	DISPOSIT	SUBJID	SUBJID		

Target Variable Information

This metadata data set contains information about how the target data set is populated using the source data set variables. The variables in this metadata data set are:

Variables

- Target Name of the target data set.
- Variable Target variable name
- Vlabel Variable label
- Vtype Variable type (C = Character, N = Numeric)
- Vlen Variable length
- Vformat Format name

Example

Target	Variable	Vlabel	Vtype	Vlen	Vformat
DISPOSIT	AEID	Event Identifier	N	8	
DISPOSIT	HLGTERM	High Level Group Term	C	250	
DISPOSIT	HLTERM	High Level Term	C	250	
DISPOSIT	INVID	Investigator ID	N	8	
DISPOSIT	LLTCD	Lowest Level Term Code	N	8	
DISPOSIT	LLTERM	Lowest Level Term	C	250	
DISPOSIT	PTERM	Preferred Term	C	250	
DISPOSIT	SOCTERM	System Organ Class Term	C	250	
DISPOSIT	DS	Subject Disposition	N	8	
DISPOSIT	DSDT	Subject Disposition Date	N	8	
DISPOSIT	DSDT	Subject Disposition Date Character	C	10	
DISPOSIT	DSTP	Subject Disposition Type	C	20	
DISPOSIT	DTHDT	Subject Death Date	N	8	
DISPOSIT	SUBJID	Subject Identifier	N	8	
DISPOSIT	VISID	Visit Identifier	N	8	

Bridging Program

Once the metadata has been completed then the Bridging Program can be run to create the target data sets. The Bridging program performs the following processes in the following order:

1. Read in the Target metadata data set information to determine the order in which the target data sets should be created.
2. Create a macro variable for the metadata variable TARGET.
3. Perform a %DO loop to process all of the Target data sets.
 - Read in the Source / Target metadata information to determine the order in which the source data should be read to create this particular Target data set.
 - Create macro variables for the metadata variables SOURCE and TARGET MERGE BY.
 - Perform a %DO loop to process all of the Source inputs for this Target data set.
 1. Merge the Source / Target Variable metadata with the Target Variable metadata. Only use the observations where the variable SOURCE is equal to the current Source data set that is being processed.
 2. Create macro variables for each of the variables that are present in the merge of the Source / Target Variable metadata with the Target Variable metadata.
 3. The following code is done

```
Data _temp1;
Set &Source;
%do c = 1 %to &NumSource;
  %if &&VType&c = N %then %do;
    attrib &&TargetVar&c length=&&VLen&c...;
  %end;
  %if &&VType&c = C %then %do;
    attrib &&TargetVar&c length= $ &&VLen&c...;
```

```
%end;
```

```
&&TargetBeforeMacro&c.
```

```
&&TargetVar&c = &&SourceVar&c;
```

```
&&TargetAfterMacro&c.
```

```
%end;
```

- Append or merge data from _temp1 data step with any other Source / Target data that was created previously.

Conclusion

Using Bridging Metadata in the creation of Target data sets saves time and resources with its first use but where the use of bridging metadata really saves time and resources is with the use of bridging metadata that can be created programmatically. In a future paper I will discuss how SAS® programs can be written that would read in the PROC CONTENTS of the source data and by using bridging metadata from previous projects, the bridging metadata can be completed with little intervention.

REFERENCES

SAS® Institute (1999) Software: Reference, Version 8,
Copyright 1999 by SAS® Institute Inc., Cary, NC, USA

CONTACT INFORMATION

Your comments and questions are valuable and appreciated. Authors can be reached at

Doug Janssen
Eli Lilly and Company
Lilly Corporate Center
Indianapolis, IN 46285 U.S.A.
Email: janssen_douglas@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.