

Paper 167-2012

## Using Custom Data Standards in SAS® Clinical Data Integration

Michael Kilhullen, SAS Institute

### ABSTRACT

SAS Clinical Data Integration is a product offering from SAS® that enables you to collect and centrally manage metadata about how clinical data is transformed to published industry standards. However, many companies already have internal standards that enable greater business process efficiencies, or use standards that are required by an external source. This paper discusses how a custom standard can be added to SAS Clinical Data Integration and used in metadata management and data mapping features to transform data to the custom standard.

### INTRODUCTION

SAS Clinical Data Integration currently supports tabular data standards such as the CDISC Study Data Tabulation Model (SDTM). While designed with many of the SDTM business rules in mind, SAS Clinical Data Integration can be configured to leverage your own custom standards, especially when they are derived from SDTM. You can determine what properties you want to manage, add your own template definitions, and define column groups that can be used to easily create custom domains. Data standards are imported into SAS Clinical Data Integration by using metadata provided by the SAS Clinical Standards Toolkit. Once imported, all the data standards management tools are available to help you maintain and improve its use. This paper presents several options for defining custom data standards and considerations for adding or modifying data standard metadata in the SAS Clinical Standards Toolkit in preparation for import into SAS Clinical Data Integration.

### FUNDAMENTAL CONCEPTS FOR DATA STANDARDS

In order to implement your custom data standards in SAS Clinical Data Integration, you must first understand a few core concepts about how standards are represented in metadata and used by the tools provided in SAS Clinical Data Integration.

#### DATA STANDARD

A data standard describes common metadata about domains that users should create when standardizing studies and submissions. It promotes consistency across your organization by centrally managing predefined domain templates, default metadata attributes, standard values for table and column properties, and a data model for creating custom domains. Data standards are managed from the **Clinical Administration** tab. If you do not see a **Clinical Administration** tab when you open SAS Data Integration Studio, then contact your system administrator and ask to be added to the CDI Administrators Group. Data Standards managed by SAS Clinical Data Integration are listed in the Data Standards folder (Figure 1).

#### DOMAIN TEMPLATE

A domain template is a predefined definition of the domain that you want your users to create. It contains column definitions and default metadata content that is copied by your users as a starting point for standardizing data.

#### CUSTOM DOMAIN

A custom domain is a domain that a user adds that is not represented as a domain template. It is created according to the business rules established by the data standard. SDTM, for example, defines the way that columns can be combined and classified as interventions, events, or findings. Like SDTM, SAS Clinical Data Integration facilitates an aggregation approach to creating a custom domain. The groups of columns that the user can pick from are represented in metadata as a column group.

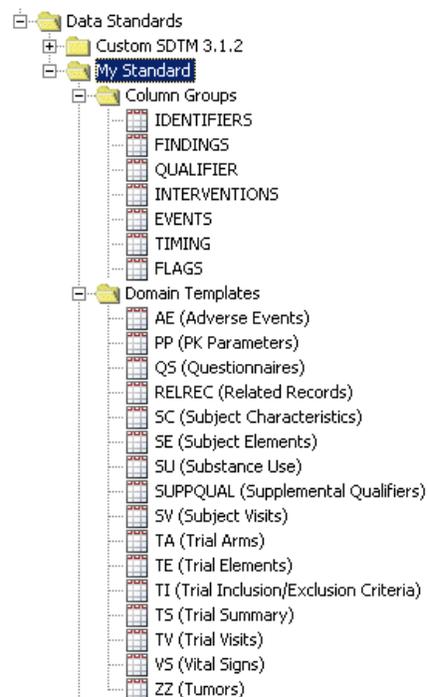


Figure 1. Contents of a Data Standard from the Clinical Administration Tab

Using Custom Data Standards in SAS® Clinical Data Integration, continued

## Column Groups

Column groups are a subset of columns that can contribute to a complete domain. Examples of column groups in SDTM are identifiers and timing. A custom domain can contain any number of columns from either column group.

## Conditional Column Groups

Conditional column groups are a set of two or more column groups where only one of the column groups can contribute to a complete domain. The best example of conditional column groups in SDTM is the classification of domains as interventions, events, and findings. When you create a new domain, it can contain columns from only one of these classifications. That is, if your domain is classified as Findings, it cannot contain columns from Interventions or Events.

## DATA STANDARD PROPERTIES

A data standard is represented by a standard set of properties (Figure 2). You can define how these properties appear in dialog boxes and set default values and constraints for each one. You can also manage the default properties that are assigned to domains and domain columns when new domains are created.

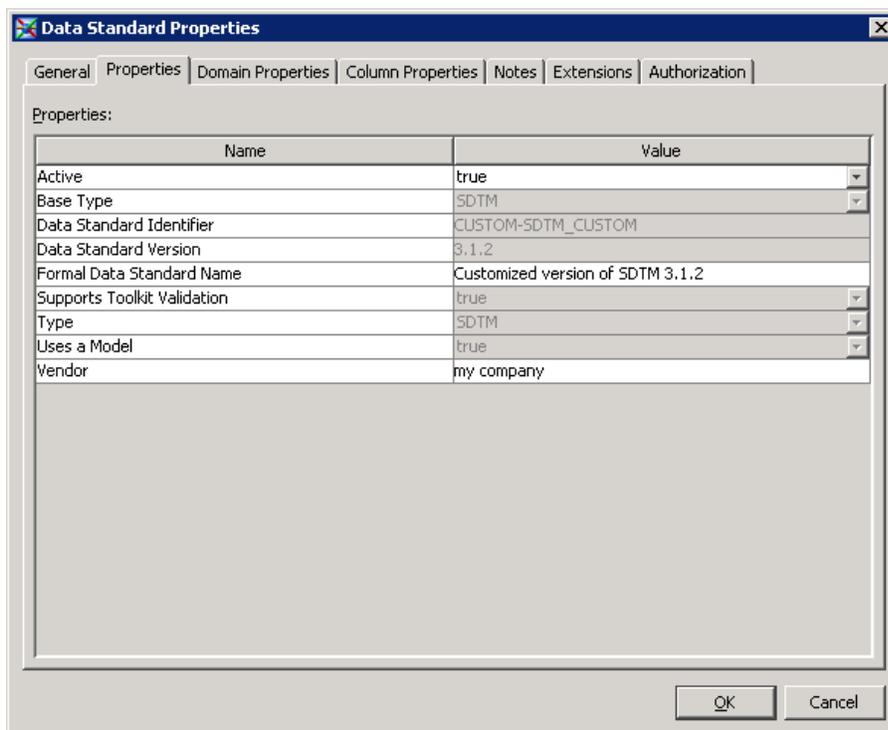


Figure 2. Data Standard Properties Managed by SAS Clinical Data Integration

## Default Domain and Domain Column Properties

SAS Clinical Data Integration also allows you to define the default properties collected for domains and domain columns (Figure 3 and Figure 4). You can choose how the properties are displayed in property dialog boxes and define the type of information collected (Boolean values, strings, integers, lookup lists, and so forth). Depending on the type of data collected, you can also establish constraints and default list contents.

Using Custom Data Standards in SAS® Clinical Data Integration, continued

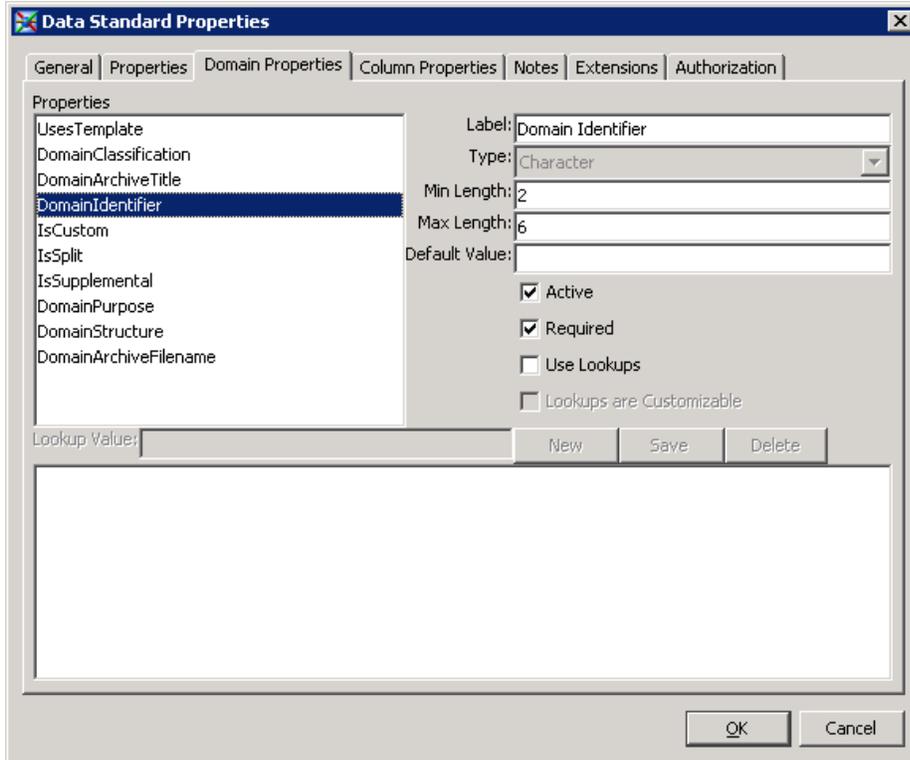


Figure 3. Default Domain Properties Managed by SAS Clinical Data Integration

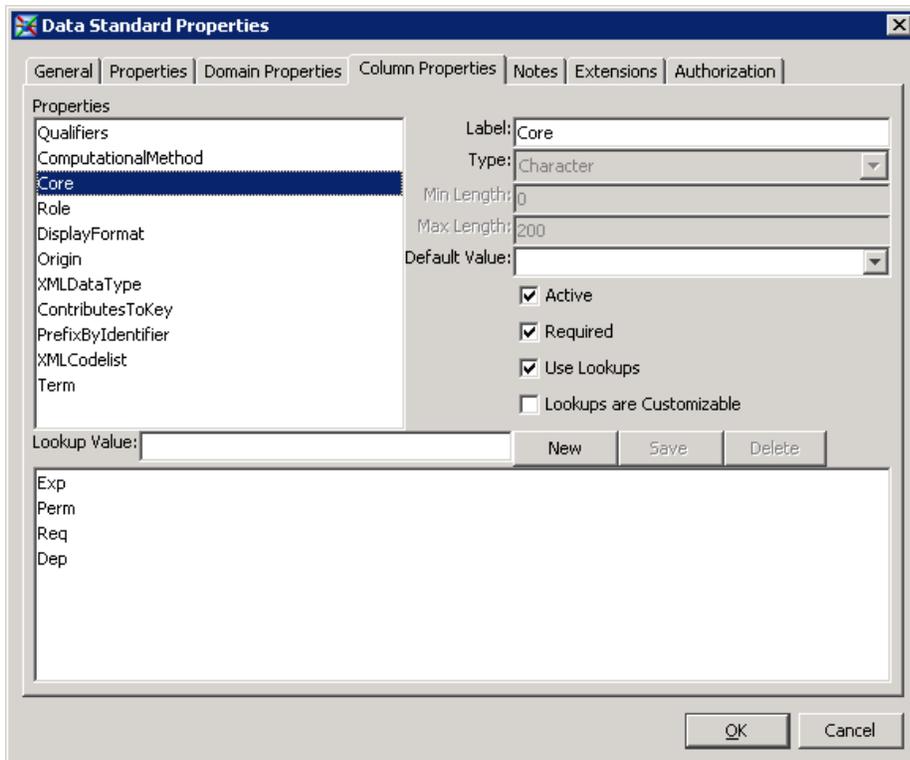


Figure 4. Default Domain Column Properties Managed by SAS Clinical Data Integration

Using Custom Data Standards in SAS® Clinical Data Integration, continued

## DEFINING CUSTOM DATA STANDARDS

SAS Clinical Data Integration allows a standard to be imported only from the SAS Clinical Standards Toolkit. Once imported, you can modify and maintain the standard using visual tools such as those described previously. SAS Clinical Standards Toolkit provides the interpretation of several data standards by SAS. When considering how to apply these standards to your own business processes, it is common to make adjustments to lengths, code lists, property descriptions, domains, domains columns, or even add new domains used by your company that are not defined as part of the CDISC SDTM Implementation Guidelines.

The most direct way of customizing a data standard is to import one that already exists in the SAS Clinical Standards Toolkit, and then use SAS Clinical Data Integration to add, edit, or delete domain and columns. Once it is imported, you can use visual tools to adjust the standards as needed. When a standard is imported, it is set to an inactive state. This allows your data standard administrator to make the necessary changes without impacting business users. Once the data standard is ready, you change the state to active. This makes it visible to users in SAS Clinical Data Integration.

If you have standards already defined in spreadsheets, SAS data sets, or external systems, you might find it easier to make bulk modifications using Base SAS® and register the standard as part of the SAS Clinical Standards Toolkit. Once registered, if it is based on SDTM, it automatically becomes visible to SAS Clinical Data Integration and can then be imported. There are two ways that you can approach this.

### Method 1: Copy and Modify SDTM in SAS Clinical Standards Toolkit and Import

This method requires you to copy an existing data standard within the SAS Clinical Standards Toolkit, modify it using Base SAS, and then import it into SAS Clinical Data Integration. The SAS Clinical Standards Toolkit User's Guide provides information for how to copy and register a data standard. Once registered, SAS Clinical Data Integration detects it and allows you to import it.

### Method 2: Build a New Data Standard in SAS Clinical Standards Toolkit and Import

If you have your data standards metadata defined in spreadsheets or SAS data sets, you can create an entirely new data standard in the SAS Clinical Standards Toolkit, and then import it into SAS Clinical Data Integration. The SAS Clinical Standards Toolkit User's Guide provides information for how to create and populate the tables required by the framework and register the data standard. Once registered, SAS Clinical Data Integration will detect it and allow you to import it. To ensure that all the tables required are created and have the proper structure, it is recommended that you follow Method 1. Then, empty the data sets and populate them with your metadata.

The details for how to copy and modify a data standard in the SAS Clinical Standards Toolkit are far beyond the scope of this paper. The remainder of this section demonstrates some of the changes that can be made with respect to the core concepts of data standards in SAS Clinical Data Integration, and how those changes are surfaced during the import of the data standard. Method 1 is used to make modifications to an existing data standard in the SAS Clinical Standards Toolkit.

### Data Standard Name, Version, and Folder

The first step is to copy an existing data standard in SAS Clinical Standards Toolkit. The user documentation describes the physical folders that are copied. After doing this, the data standard name and version used throughout the copied data standard needs to be changed to reflect the new data standard. For this paper, I copied the **cdisc-sdtm-3.1.2-1.4** to a new folder called **custom-sdtm-3.1.2**. The data standard name contained in these data sets is changed from CDISC-SDTM to CUSTOM-SDTM, but the version is not changed. To make these changes, I needed to edit several tables in the **custom-sdtm-3.1.2** folder as outlined in Table 1.

Folder	Data Set
custom-sdtm-3.1.2\control	standards standardsasreferences
custom-sdtm-3.1.2\metadata	class_tables class_columns reference_tables reference_columns
custom-sdtm-3.1.2\validation\control	validation_master validation_stdref

**Table 1. Toolkit Data Sets Requiring Data Standard Name Changes**

Using Custom Data Standards in SAS® Clinical Data Integration, continued

### Adding a New Domain Template

There are two tables where a domain template definition is stored in the SAS Clinical Standards Toolkit: `reference_tables` and `reference_columns`. `reference_tables` is used to define the table-level metadata while `reference_columns` is used to define the column-level metadata. To add a domain template, simply insert records as needed, ensuring that you use the new data standard name for your custom data standard. This same technique can be used to add, modify, or remove columns from existing templates. As a simple example of adding a new template, the following code adds a ZZ domain to the appropriate tables:

```
/* add custom template table */
insert into
tmp2.reference_tables(sasref,table,label,class,xmlpath,xmltitle,structure,purpose,
keys,state,date,standard,standardversion)
values("REFDATA","ZZ","Tumors","Findings","../transport/zz.xpt","Tumor SAS Transport
File","One record per tumor per subject","Tabulation","STUDYID USUBJID
ZZSEQ","Final","2012-02-01","CUSTOM-SDTM","3.1.2");

/*add custom template columns */
insert into tmp2.reference_columns(sasref,table,column,label,order,type,length,
xmldatatype,core,role,term,qualifiers,standard,standardversion,standardref,comment)
values("REFDATA","ZZ","STUDYID","Study
Identifier",1,"C",40,"text","Req","Identifier","","UPPERCASE","CUSTOM-
SDTM","3.1.2","SDTM2.2.4","Unique identifier for a study.")
values("REFDATA","ZZ","DOMAIN","Domain
Abbreviation",2,"C",8,"text","Req","Identifier","ZZ","UPPERCASE","CUSTOM-
SDTM","3.1.2","SDTM2.2.4,SDTMIG4.1.2.2,AppendixC2","Two-character abbreviation for
the domain.")
values("REFDATA","ZZ","USUBJID","Unique Subject
Identifier",3,"C",40,"text","Req","Identifier"," ","UPPERCASE","CUSTOM-
SDTM","3.1.2","SDTM2.2.4,SDTMIG4.1.2.3","Identifier used to uniquely identify a
subject across all studies for all applications or submissions involving the
product.")
values("REFDATA","ZZ","ZZSEQ","Sequence
Number",4,"N",8,"integer","Req","Identifier"," "," ","CUSTOM-
SDTM","3.1.2","SDTM2.2.4","Sequence Number given to ensure uniqueness of subject
records within a domain. May be any valid number.")
values("REFDATA","ZZ","ZZTERM","Reported Term for
Tumor",5,"C",200,"text","Req","Topic"," ","UPPERCASE","CUSTOM-SDTM","3.1.2","","Term
describing the tumor.")
values("REFDATA","ZZ","ZZORRES","Assessment Result in Original
Units",6,"C",200,"text","Exp","ResultQualifier"," ","UPPERCASE","CUSTOM-
SDTM","3.1.2","","Result of tumor assessment as originally received or collected.");
```

When a custom template is added, it appears on the appropriate wizard pages during the import. These pages depicted in Figure 5 and Figure 6 are read-only and are intended to provide a verification of metadata prior to the actual import. Once imported, the domain is selectable in SAS Clinical Data Integration for use in standardizing data.

Using Custom Data Standards in SAS® Clinical Data Integration, continued

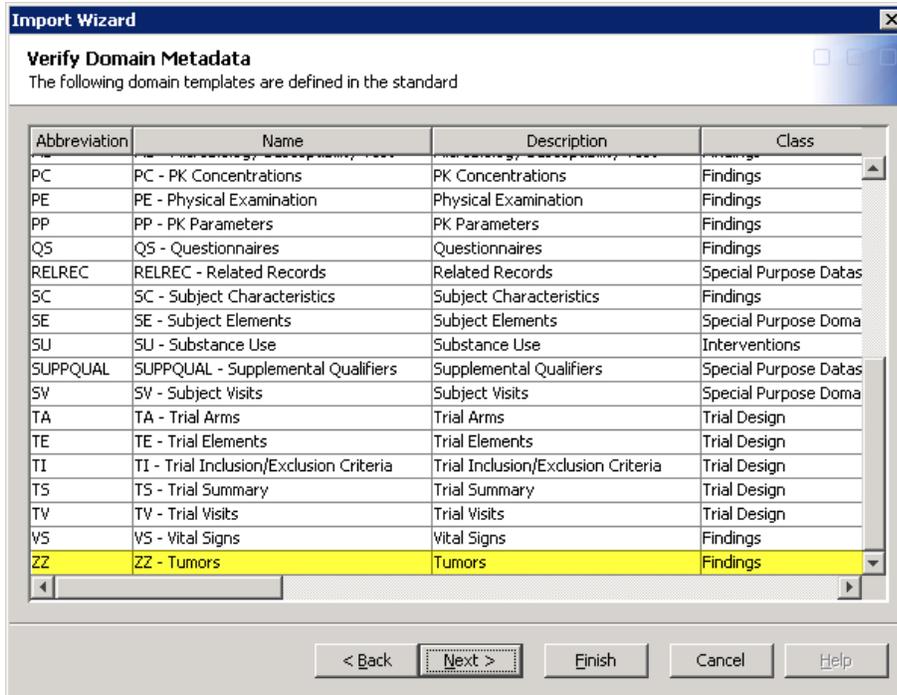


Figure 5. Custom Domain Added to reference\_tables

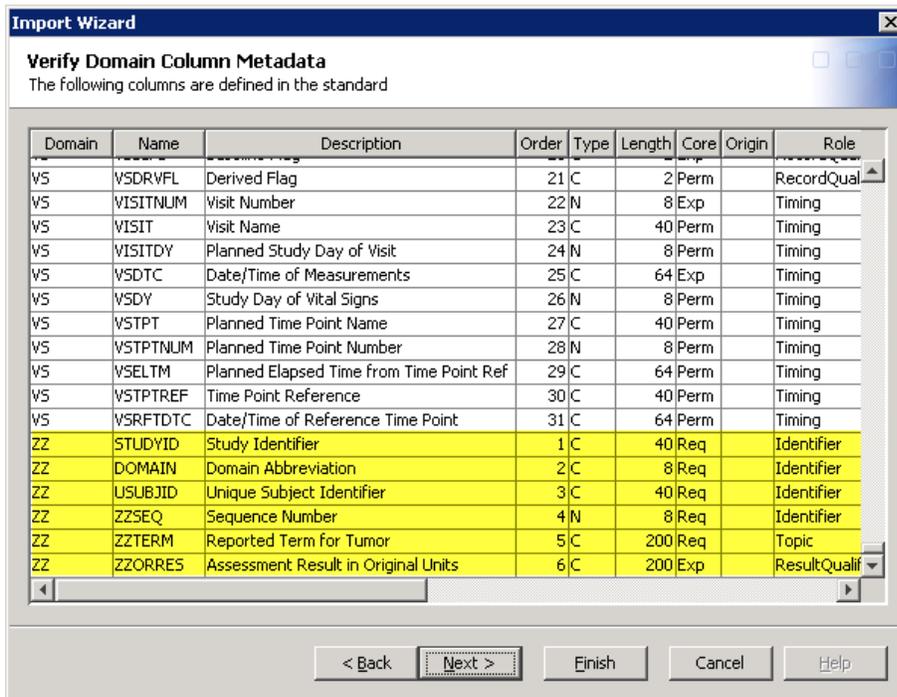


Figure 6. Custom Domain Added to reference\_columns

### Customizing the Column Groups

Column Groups are defined in the class\_tables and class\_columns data sets. Class\_tables stores metadata about the classification tables while class\_columns stores metadata about classification columns. These metadata are directly related to the information that facilitates creating new domains. If you examine the data sets for SDTM, you will recognize the classification of information by identifiers, interventions, events, findings, and timing.

Using Custom Data Standards in SAS® Clinical Data Integration, continued

For our custom data standard, we want to consider two changes to the classification:

- adding an additional group of optionally selectable columns
- adding a group representing a new classification of domains

The latter example is not practical considering the SDTM data standard model, but it serves as a good example to consider when modeling your internal data standards.

The following code is an example of adding our new column groups.

```
/* add custom class table metadata */
insert into tmp2.class_tables(sasref,table,label,class,purpose,state,date,standard,
standardversion)
values("REFDATA","FLAGS","Flags","All Classes","Tabulation","Final","2012-02-
01","CUSTOM-SDTM","3.1.2")
values("REFDATA","QUALIFIER","Qualifiers","Qualifiers-General","Tabulation",
"Final","2012-02-01","CUSTOM-SDTM","3.1.2");

/* add custom class columns metadata */
insert into
tmp2.class_columns(sasref,table,column,label,order,type,length,xmldatatype,
xmlcodelist,core,role,qualifiers,standard,standardversion,comment)
values("TEMPLATE","FLAGS","SAFEFLG","Safely analysis
flag",1,"C",2,"text","NY","Perm","RecordQualifier","UPPERCASE","CUSTOM-SDTM",
"3.1.2","used to flag safety population")
values("TEMPLATE","FLAGS","TERMFLG","Early termination
flag",2,"C",2,"text","NY","Perm","RecordQualifier","UPPERCASE","CUSTOM-SDTM",
"3.1.2","used to flag early termination")
values("TEMPLATE","QUALIFIER","__AGEGRP","Age
group",1,"C",2,"text","","","Perm","RecordQualifier","UPPERCASE","CUSTOM-SDTM",
"3.1.2","used to define age grouping")
values("TEMPLATE","QUALIFIER","__NOTE","Special
notation",2,"C",2,"text","","","Perm","RecordQualifier","UPPERCASE","CUSTOM-SDTM",
"3.1.2","used to add special notation")
values("TEMPLATE","QUALIFIER","__QRY","Outstanding
Queries",3,"C",2,"text","","","Perm","RecordQualifier","UPPERCASE","CUSTOM-SDTM",
"3.1.2","outstanding queries");
```

Next, we must define how these column groups are used by the data model. To do this, we must modify information in the standardlookup data set in the custom-sdtm-3.1.2\control folder. If you examine an SDTM data standard already defined in the SAS Clinical Standards Toolkit, the pertinent records in standardlookup are shown in Table 2.

SAS libref	CST table name	CST column name	Unique CST column value	Default CST column value	Must CST column value be nonnull?	CST column value order
refmeta	class_tables	table	IDENTIFIERS	Y	Y	1
refmeta	class_tables	table	FINDINGS	N	Y	2
refmeta	class_tables	table	INTERVENTIONS	N	Y	2
refmeta	class_tables	table	EVENTS	N	Y	2
refmeta	class_tables	table	TIMING	N	Y	3

**Table 2. Column Group Definition in the SAS Clinical Standards Toolkit**

The three significant columns in this data set are the CST table name, CST column name, and CST column value order. To register our custom class tables, you must first insert records for the new column groups using the same values for CST table name and CST column name as the existing records, as shown in the following code:

```
/* add custom model tables */
insert into tmp1.standardlookup(sasref,table,column,value,default,nonnull,order)
values("refmeta","class_tables","table","FLAGS","N","Y",4)
values("refmeta","class_tables","table","QUALIFIER","N","Y",2);
```

Using Custom Data Standards in SAS® Clinical Data Integration, continued

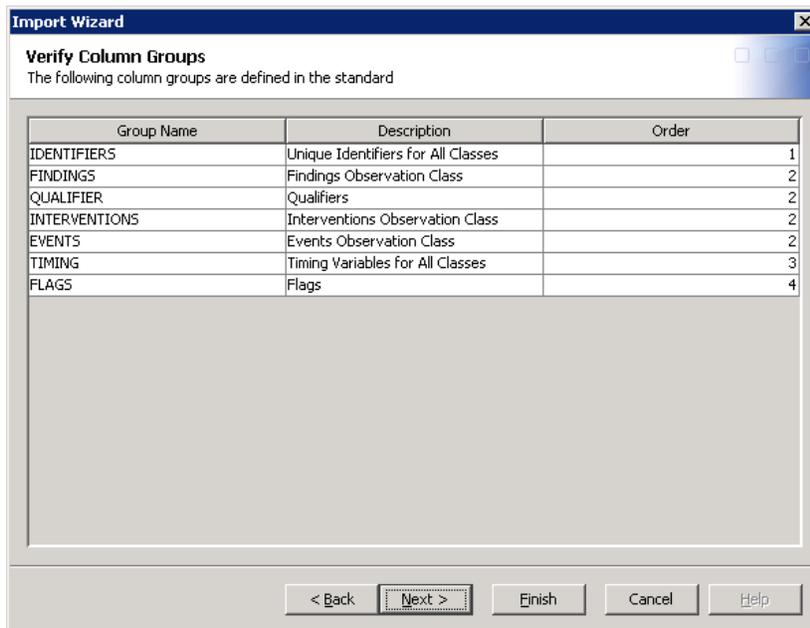
Lastly, you must adjust the value of the CST column value order. The values in this column define two things: the order that the columns will be added to a custom domain, and which tables are considered “conditional”. That is, only columns from one of the tables can contribute to the final domain. After running the code above, the data set now appears as shown in Table 3.

SAS libref	CST table name	CST column name	Unique CST column value	Default CST column value	Must CST column value be nonnull?	CST column value order
refmeta	class_tables	table	IDENTIFIERS	Y	Y	1
refmeta	class_tables	table	INTERVENTIONS	N	Y	2
refmeta	class_tables	table	FINDINGS	N	Y	2
refmeta	class_tables	table	EVENTS	N	Y	2
refmeta	class_tables	table	QUALIFIER	N	Y	2
refmeta	class_tables	table	TIMING	N	Y	3
refmeta	class_tables	table	FLAGS	N	Y	4

**Table 3. Customized column Group Definition in the SAS Clinical Standards Toolkit**

As you can see, FLAGS has been added to the end meaning that columns selected from it will be added to the end of the domain being created. QUALIFIER has been added to the “2” CST column order. This means that when QUALIFIER is selected, only columns from it can be added to the new domain. Columns in INTERVENTIONS, EVENTS, or FINDINGS cannot be selected.

As before, this information is presented to the user during import to verify that it has been set up correctly before performing the import. Figure 7 shows the column group and order information; Figure 8 shows the column information.



**Figure 7. Column Groups Displayed in SAS Clinical Data Integration during Import**

Using Custom Data Standards in SAS® Clinical Data Integration, continued

Domain	Name	Description	Order	Type	Length	Core	Origin
TIMING	__TPT	Planned Time Point Name	13	C	40	Perm	Timin
TIMING	__TPTNUM	Planned Time Point Number	14	N	8	Perm	Timin
TIMING	__ELTM	Planned Elapsed Time from Time Point Ref	15	C	64	Perm	Timin
TIMING	__TPTREF	Time Point Reference	16	C	40	Perm	Timin
TIMING	__RFTDTC	Date/Time of Reference Time Point	17	C	64	Perm	Timin
TIMING	__STRF	Start Relative to Reference Period	18	C	20	Perm	Timin
TIMING	__ENRF	End Relative to Reference Period	19	C	20	Perm	Timin
TIMING	__EVLINT	Evaluation Interval	20	C	64	Perm	Timin
TIMING	__STRTPT	Start Relative to Reference Time Point	21	C	40	Perm	Timin
TIMING	__STTPT	Start Reference Time Point	22	C	40	Perm	Timin
TIMING	__ENRTPT	End Relative to Reference Time Point	23	C	40	Perm	Timin
TIMING	__ENTPT	End Reference Time Point	24	C	40	Perm	Timin
FLAGS	SAFEFLG	Safely analysis flag	1	C	2	Perm	Reco
FLAGS	TERMFLG	Early termination flag	2	C	2	Perm	Reco
QUALIFIER	__AGEGRP	Age group	1	C	2	Perm	Reco
QUALIFIER	__NOTE	Special notation	2	C	2	Perm	Reco
QUALIFIER	__QRY	Outstanding Queries	3	C	2	Perm	Reco

Figure 8. Column Group Column Information Displayed in SAS Clinical Data Integration during Import

### Registering the Custom Standard

Before SAS Clinical Data Integration can import the custom standard, it needs to be registered in the SAS Clinical Data Standards Toolkit. Edit the registerstandard.sas program found in the **Programs** folder of the copied standard. Change the macro variables `__thisStandard`, `__thisStandardVersion`, and `__thisDirWithinStandards` to reflect changes made to the copied data standard. Submit the program and verify that no errors were encountered. After making this change, the custom data standard is automatically detected by SAS Clinical Data Integration as shown in Figure 9.

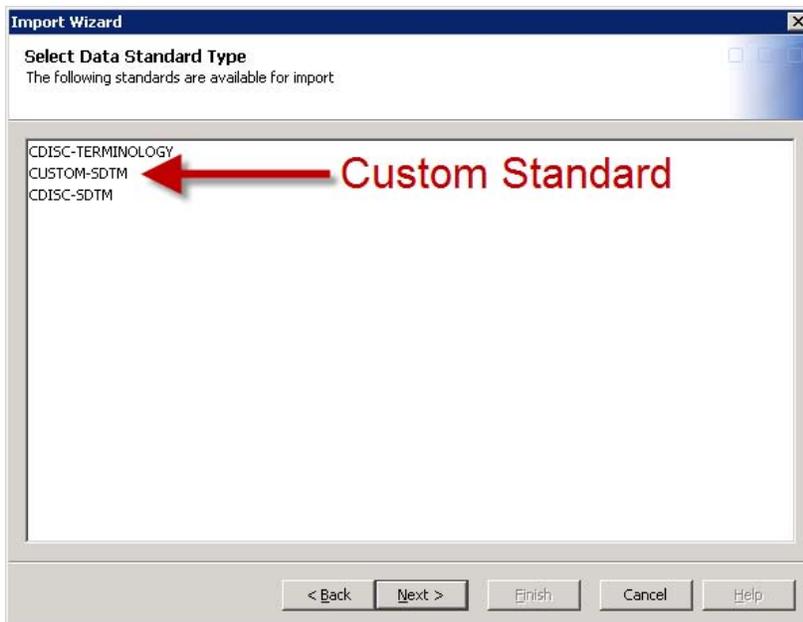


Figure 9. Custom Data Standard Appearing in SAS Clinical Data Integration

### IMPORTING DATA STANDARD INTO SAS CLINICAL DATA INTEGRATION

Once your custom data standard is added to the SAS Clinical Standards Toolkit, the Metadata Importer automatically detects it and allows it to import. Please note that in SAS Clinical Data Integration 2.3, only SDTM and Controlled

Using Custom Data Standards in SAS® Clinical Data Integration, continued

Terminology standards can be imported. The data standard contents are displayed in the **Clinical Administration** tab in Figure 1. Notice the new column groups and ZZ domain template.

## INTEGRATION OF CUSTOM STANDARD IN SAS CLINICAL DATA INTEGRATION

The custom data standard functions like any other data standard in SAS Clinical Data Integration. Domain templates are available to users when standardizing a study. Domain and column properties are available in property tables and can be adjusted as needed. The more interesting integration point is the column group definitions. SAS Clinical Data Integration integrates our new column groups in the Custom Domain Wizard. Since we added a conditional column group (QUALIFIER), it appears in the list of conditional groups (Figure 10). Having selected it, the column selection screen shows these columns in addition to FLAGS, which was added as a column group for our custom data standard (Figure 11).

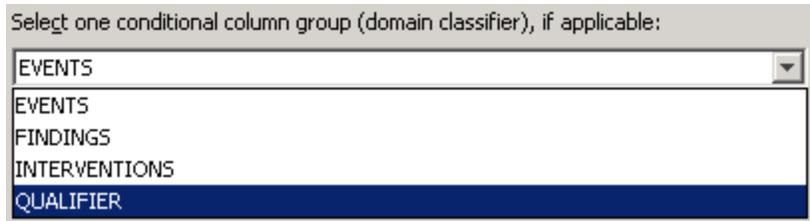


Figure 10. Conditional Column Group in Custom Domain Wizard

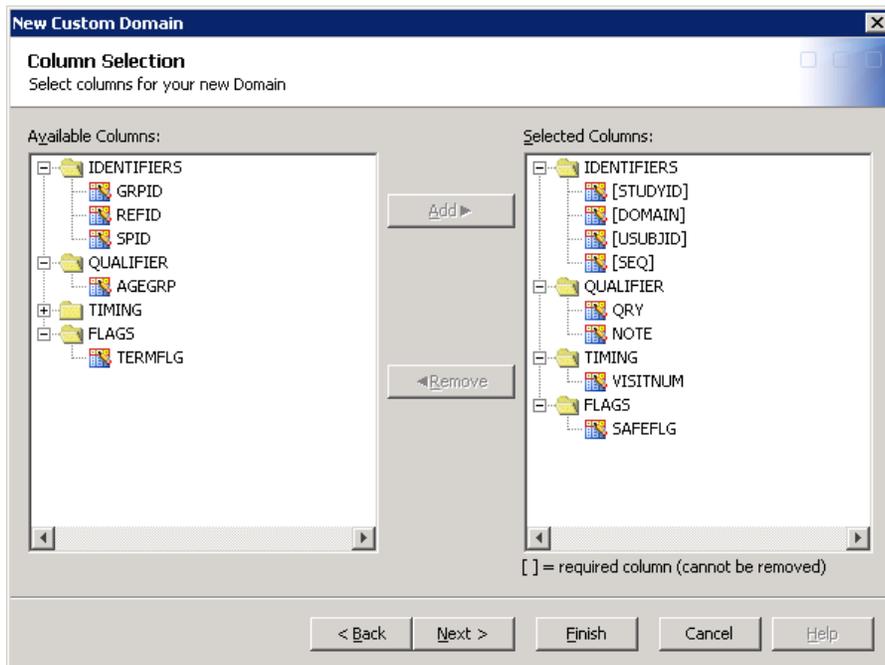


Figure 11. Custom Column Groups Displayed in the Custom Domain Wizard

## CONCLUSION

Custom data standards are supported by SAS Clinical Data Integration as long as they are based on tabulated data standards. You can define your own new data standard or modify an existing data standard. If you do not have a custom data standard defined, SAS Clinical Data Integration allows you to import existing standards from the SAS Clinical Standards Toolkit. If you have well-defined custom data standards, you can define them in the SAS Clinical Standards Toolkit and then import them into SAS Clinical Data Integration. Support for other types of standards such as Analysis Data Standards will be supported in future releases of SAS Clinical Data Integration.

Using Custom Data Standards in SAS® Clinical Data Integration, continued

## RECOMMENDED READING

- *SAS® Clinical Data Integration 2.3: User's Guide*
- *SAS® Clinical Standards Toolkit 1.4: Getting Started*
- *SAS® Clinical Standards Toolkit 1.4: User's Guide*

## CONTACT INFORMATION

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

Michael Kilhullen  
SAS Institute  
62 Edison Road  
Stewartsville, NJ 08886  
908-760-6528  
michael.kilhullen@sas.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.