**Paper CD05**

# Using SAS® Clinical Data Integration to Implement and Manage CDISC Standards

Julie Maddox, SAS Institute Inc., Cary, NC, USA
Michael Kilhullen, SAS Institute Inc., Cary, NC, USA

## ABSTRACT
The SAS® Metadata Server is a core component of all SAS®9 solutions. It delivers the power to integrate, share, centrally manage, and leverage metadata across entire organizations. Through these capabilities, standard data models such as the CDISC Study Data Tabulation Model (SDTM) can be deployed and leveraged by all users in your organization without the need for developing additional metadata libraries or programs. In this paper, we examine the value that the SAS® Open Metadata Architecture can bring to your organization, how the SDTM data model is implemented in the metadata server, and how the metadata can be leveraged by SAS products and solutions such as SAS® Data Integration Studio.

## INTRODUCTION
SAS® Clinical Data Integration 2.1 is a new product offering from SAS that focuses on pharmaceutical industry needs for transforming, managing, and verifying the creation of industry mandated data standards such as Clinical Data Interchange Standards Consortium (CDISC). The product relies upon SAS® Data Integration Server to provide centralized metadata management using the SAS Metadata Server and the tools to visually transform data. SAS Clinical Data Integration enhances usability by adding new metadata types, plug-ins, and wizards that assist with clinically oriented tasks such as importing data standards, creating studies and submissions, and adding specialized transformations for transforming clinical data to a standard data model. It also leverages the SAS® Clinical Standards Toolkit to provide validation and conformance checking.

## METADATA REPRESENTATION OF DATA STANDARDS
SAS Clinical Data Integration implements a common metadata model for representing data standards that are based on CDISC models. The goal is to use one model to represent all standards, and allow the flexibility to turn certain attributes and properties on or off depending upon whether a specific data standard needs to collect it. For example, unlike the Study Data Tabulation Model (SDTM), the Analysis Data Model (ADaM) does not classify domains as interventions, events, or findings and does not define how custom domains can be created. Therefore, these properties and features are turned off. The wizards and property pages available in SAS Clinical Data Integration can interpret these dynamics and adjust displays accordingly. By building upon a common metadata model, pharmaceutical companies who do not standardize on CDISC (or use variations of CDISC at some point in their drug development process) can use SAS Clinical Data Integration with their own custom data standards. The only requirement is that it conforms to the SAS common data standards metadata model.



Figure 1- Data Standard Component Properties Showing Installed Domains

### IMPORTING CDISC STANDARDS

Final versions of standards that are provided by the CDISC organization are imported into SAS Clinical Data Integration through the SAS Clinical Standards Toolkit (Figure 1). The SAS Clinical Standards Toolkit is a SAS Macro approach to supporting clinical data standards in Base SAS. It supports defining data standard domains, conversion of domains between CDISC models, and validation and conformance checks. It provides periodic updates when new standards and new versions of standards are released. Once the updates are applied to the toolkit, they are automatically detected by SAS Clinical Data Integration during the import and validation processes. SAS Clinical Data Integration provides a Data Standards Metadata Import Wizard to help data standard administrators select and load metadata (Figure 2).
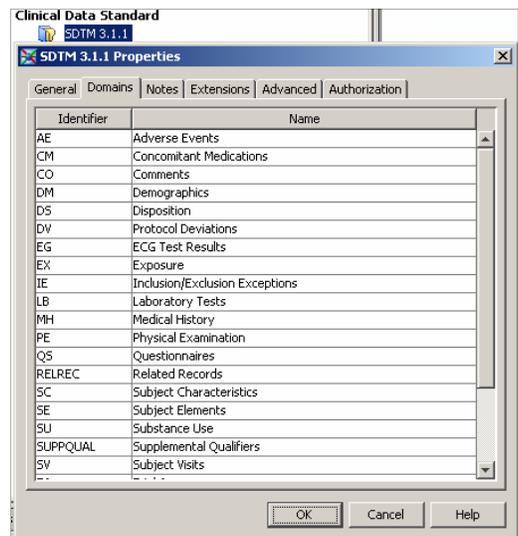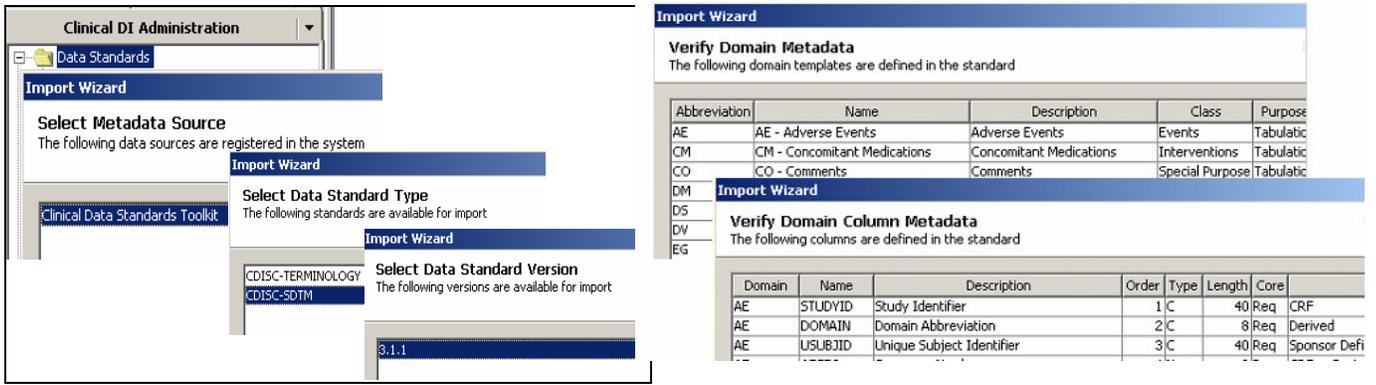
**Figure 2 - Metadata Import Sources**

This is a one-time process per model version; once they are loaded, the SAS Metadata Server manages changes and additions to the data standards. The wizard prompts you to select the standard and version from the toolkit, displays the metadata content in detail for verification, and then imports the metadata. Once imported, the standard and domain templates are surfaced through SAS Clinical Data Integration plug-ins.

**SUPPORTING CUSTOM STANDARDS**

Custom versions of standards can also be imported using the SAS Clinical Standards Toolkit. This requires that the custom standards are implemented as part of the SAS Clinical Standards Toolkit. When the metadata import wizard is run, these versions of standards are automatically included in the selections and imported. In addition, CDISC standards can be imported and then customized to meet a company standard.

## IMPLEMENTING DATA STANDARDS

Data standard domains can be implemented as part of a data standard, study, or submission. When data standard domains are implemented in a data standard, we refer to these as templates because they are used as the basis for the actual data and metadata collected by a study or submission. Studies and submissions have their own metadata representation in the SAS Metadata Server that is based on a common metadata object called a clinical component (Figure 3). In addition to having specialized metadata about the study or submission, they also catalog the metadata contents created by users (such as jobs, reports, tables, and so on), define the versions of standards that are allowed for a study or submission, and enable additional processes such as importing, exporting, and archiving.
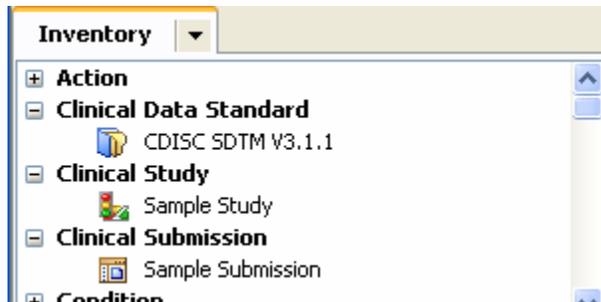


**Figure 3 - Clinical Components in the Metadata Server**

**CREATING CLINICAL COMPONENTS**

Studies and submissions are created using new object wizards. The wizard will collect basic object metadata such as name, description, responsible parties, and content location in the metadata tree. Next, metadata about the clinical component is collected. For example, a study component will collect metadata such as protocol title, indication, and phase. This is followed by defining which versions of standards can be used for the component. This information feeds other SAS Clinical Data Integration plug-ins to ensure that users are creating the proper versions of standards for a given clinical component. Finally, if any clinical component defaults were defined, the wizard will prompt you through necessary related actions.

**CLINICAL COMPONENT DEFAULTS**

SAS Clinical Data Integration can be configured by an administrator to create default metadata content when clinical components are created (Figure 4). This allows you to maintain consistent content between studies or submissions. For example, the administrator can define a standard metadata folder structure for studies. When the new study wizard finishes, it creates the metadata folders (Figure 5). The administrator can also define standard libraries to ensure that the correct

2

librefs are generated.  This scenario is useful when your organization implements standard reporting macros that depend on specific librefs to gain access to data and store results.
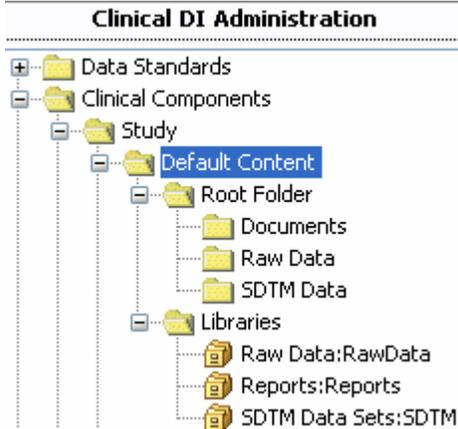


**Figure 4- Default Content Defined for Studies**
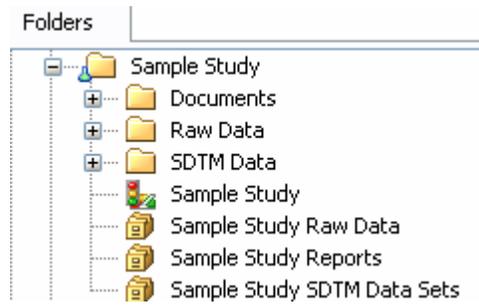


**Figure 5 - Content Created for a New Study Based on Defaults**

**CREATING STANDARD DOMAINS**

Domains are copied from the data standards using the New Standard Domain wizard.  The wizard will only display selections based on the allowed data standard versions for the clinical component that were defined during clinical component creation.  You select one or more domains from a standard and define a folder where they will be stored in the clinical component.  Next, you select the library where the physical data will be created.  The selected domains are then copied to the specified folder and assigned the appropriate library.  Once a folder contains a registered domain, only domains of the same standard version can be added to that folder.



**EDITING DOMAIN AND COLUMN PROPERTIES**

Once a domain is created, clinical metadata is accessible through standard SAS Data Integration Studio property dialogs that are enhanced by SAS Clinical Data Integration plug-ins.  Additional tabs are added where the clinical metadata is displayed (Figure 6).  If you have edit permissions, then you can change the values of the clinical metadata.  Changes made are considered study or submission specific.  That is, they

**Figure 6 - Clinical Metadata in Column Properties Page**

do not propagate back to the data standard template.  The same is true of changes made to data standard templates; they do not propagate to registered domains.  Other tools are available in SAS Clinical Data Integration to compare domains to templates and apply changes.
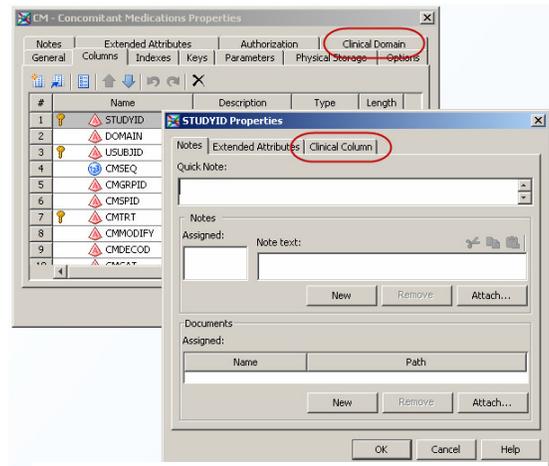
**CREATING CUSTOM DOMAINS**

Certain CDISC standards, such as SDTM, support user-defined domains (domains needed for clinical data that are not defined in the implementation guide).  In this case, the underlying data model is used to assemble the new domain.  Because the data standard installed in the metadata contains the underlying model, SAS Clinical Data Integration provides a wizard to guide you in defining the custom domain.  Data standard administrators can use the same wizard to design new domain templates in the data standard.  This helps ensure custom domains are implemented consistently across studies.  Administrators can also review and promote custom domains defined by users in studies and submissions to the data standard.

**REFRESHING DOMAINS FROM THE DATA STANDARD**

On occasion, the Data Standards Administrator may update the domain information for a standard, and notify the Study administrators to refresh the domains in their study.  The Refresh Domain plug-in compares the study domain to the domain in the standard and shows the differences (Figure 7).  The study administrator can then select the changes he wishes to incorporate into his study.
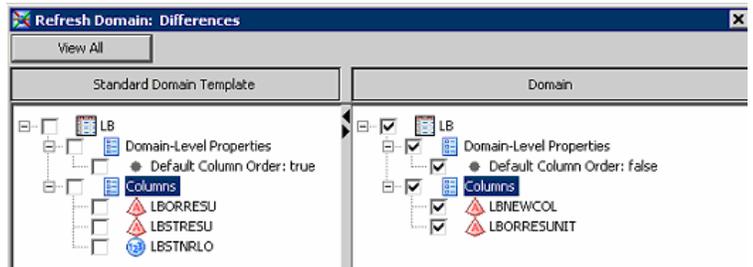


**Figure 7 – Refresh Domain Analysis**

**TRANSFORMING OPERATIONAL DATA TO A STANDARD**
The data transformation process is achieved using standard SAS Data Integration Studio functionality. To accomplish this, jobs are created where you visually design how data is transformed from one format and/or structure to another. SAS Data Integration Studio 4.2 offers greatly improved usability over previous versions including flow control, integrated error/warning reporting, job performance monitoring and statistics, and reporting. See the recommended reading section for references to other SAS Global Forum Papers related to SAS Data Integration Studio. In addition, Clinical Data Integration provides several new transformations specifically designed to validate and transform data to the CDISC standards.



**Figure 8 - Example of Using Metadata to Monitor Mapping and Validation Tasks**

Using a visual design tool to transform data is a unique experience. While this can take more time to complete compared to standard SAS programs and macros, the advantages gained far outweigh this constraint. In fact, as you accumulate metadata in the system, it can be reused to minimize the time it takes to complete your study or submission. Remember that implementing CDISC standards involves more than just creating data sets. You must also collect and manage metadata. The metadata also needs to be transferrable between the CDISC data models. When you collect and manage metadata, you can automate the conversion process. These needs are core capabilities of SAS metadata-driven products such as SAS Clinical Data Integration, SAS Data Integration Studio, SAS Management Console, and so on. Furthermore, by establishing links between steps in a job flow, data lineage that includes the transformation expressions applied to columns can be retrieved. These capabilities, in addition to the ability to analyze job performance, mean that you have better tools to develop more efficient transformation processes and automatically collect, manage and monitor the metadata required by CDISC standards (Figure 8).

**VALIDATING DOMAIN CONTENT AND STRUCTURE**
As data standard domains are implemented in studies and submissions, the structure and content can vary from the data standard. You must periodically verify that the domains maintain conformance to the data standard. To do this, the SAS Clinical Standards Toolkit is used by SAS Clinical Data Integration plug-ins to run validation checks and generate reports. SAS Clinical Standards Toolkit provides 143 unique SDTM 3.1.1 validation checks. These checks are derived from three sources: the CDISC-SDTM WebSDM™ documented checks, checks supporting loads into the Janus study data repository, and SAS-derived checks based upon SAS' data management and cleansing experiences building CDISC-SDTM domains. The validation checks are designed to enable an assessment of the consistency of data values within a specific column, between columns, across records within a specific data set, and across data sets. In addition to the SAS-supplied validation checks, the data standards administrator can create their own customized compliance checks using the Manage Compliance Checks dialog (Figure 9).



**Figure 9 - Example of Customizing Compliance Check**

The CDISC-SDTM Compliance transformation is provided for use in building validation jobs. Simply specify the data standard you wish to validate against, select the domains to be assessed and the set of compliance checks you wish to run (Figure 10).
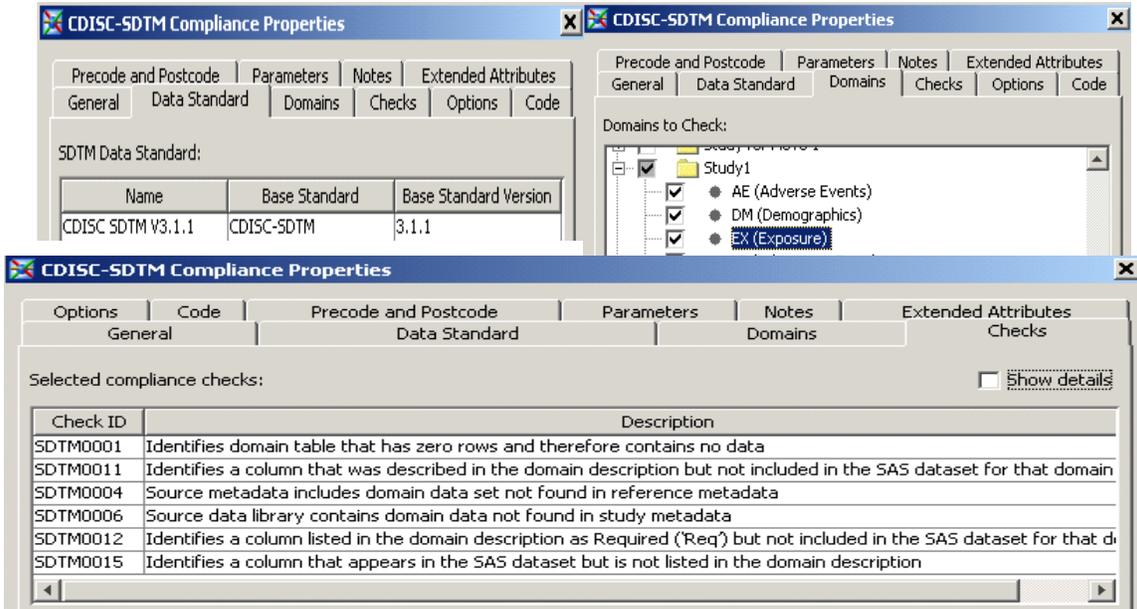
4

**Figure 10 – CDISC-SDTM Compliance Transformation Properties**

When the job executes, two datasets are produced. The Results data set contains the findings of the compliance assessment and the Metrics data set contains summary statistics on the validation process (Figure 11).
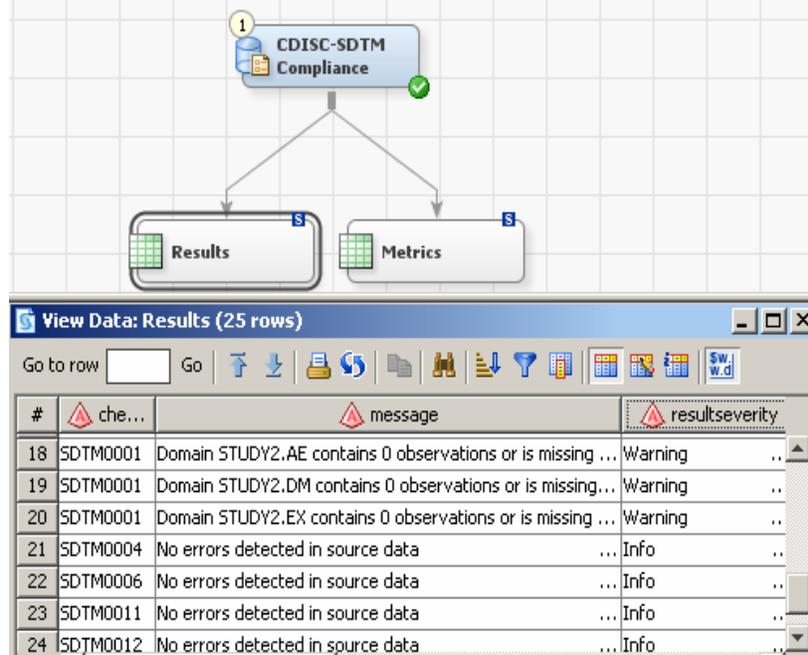


**Figure 11– CDISC–SDTM Compliance Transformation and Results data set**

## GENERATING CRT-DDS

The metadata managed by SAS Clinical Data Integration can be published to CRT-DDS using the CRT-DDS Transform (Figure 12). This transform extracts metadata on the domains and study, and passes it on to the SAS Clinical Standards Toolkit for define.xml creation. The transform allows you to specify properties to control encoding, and to specify customized style sheets (Figure 13).
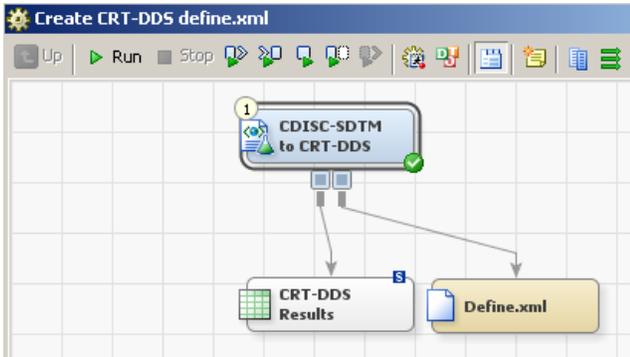
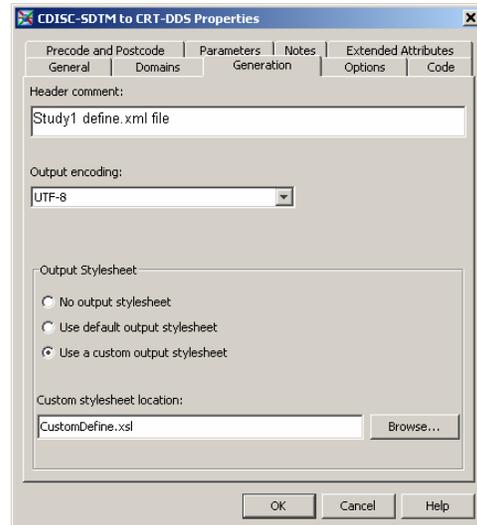**Figure 12 – Create CRT-DDS Define.xml Process Flow**



**Figure 13 – CDISC-SDTM to CRT-DDS
Transform Properties**

When the job executes, a define.xml files and a results data set
are produced.  The Results data set documents the results of the generation of the CRT-DDS file.  The define.xml file
contains summary information on each of the domain data sets (Figure 14), detailed information on each column in each
domain (Figure 15), and code list information (Figure 16).

**Datasets for Study Study1**

| Dataset | Description | Structure | Purpose | Keys |
|---------|-------------|-----------|---------|------|
| DM | Demographics | Special Purpose - One record per subject | Tabulation | STUDYID USUBJID |
| AE | Adverse Events | Events - One record per event per subject | Tabulation | STUDYID USUBJID AETERM AESTDTC |
| EX | Exposure | Interventions - One record per constant dosing interval per subject | Tabulation | STUDYID USUBJID EXTRT EXSTDTC |
| LB | Laboratory Tests | Findings - One record per lab test per time point per visit per subject | Tabulation | STUDYID USUBJID LBTESTCD VISITNUM LBTPTNUM |

**Figure 14 - Datasets Section of CRT-DDS**

**Demographics Dataset (DM)**

| Variable | Label | Type | Controlled Terms or Format | Origin | Role |
|----------|-------|------|----------------------------|--------|------|
| STUDYID | Study Identifier | text | | CRF | Identifier |
| DOMAIN | Domain Abbreviation | text | | Derived | Identifier |
| USUBJID | Unique Subject Identifier | text | | Sponsor Defined | Identifier |
| SUBJID | Subject Identifier for the Study | text | | CRF | Topic |
| RFSTDTC | Subject Reference Start Date/Time | text | | Sponsor Defined | Timing |
| RFENDTC | Subject Reference End Date/Time | text | | Sponsor Defined | Timing |
| SITEID | Study Site Identifier | text | | CRF or Derived | RecordQualifier |
| INVID | Investigator Identifier | text | | CRF or Derived | RecordQualifier |
| INVNAM | Investigator Name | text | | CRF or Derived | SynonymQualifier |
| BRTHDTC | Date/Time of Birth | text | | CRF or Derived | ResultQualifier |
| AGE | Age in AGEU at RFSTDTC | float | | CRF or Derived | ResultQualifier |
| AGEU | Age Units | text | 1 | CRF or Derived | VariableQualifier |
| SEX | Sex | text | 19 | CRF | ResultQualifier |

**Figure 15 – Domain Section of CRT-DDS**

6

**Figure 16 – Code list Section of CRT-DDS**

## DATA STANDARDS ADMINISTRATION

SAS Clinical Data Integration provides several plug-ins for the Data Standards Administrator to manage the data standards and to monitor how the data models are being implemented in the studies and submissions.  The Analyze Model Usage tool allows a standards administrator to view information on how a model is being implemented across several studies. Information presented includes which domains have been implemented in the studies, and which columns of the data standard columns are implemented in the study domains (Figure 17).  From this analysis, the data standards administrator can determine if changes to the data standard should be promoted, or if additional compliance checks should be added to validate the standards in place.
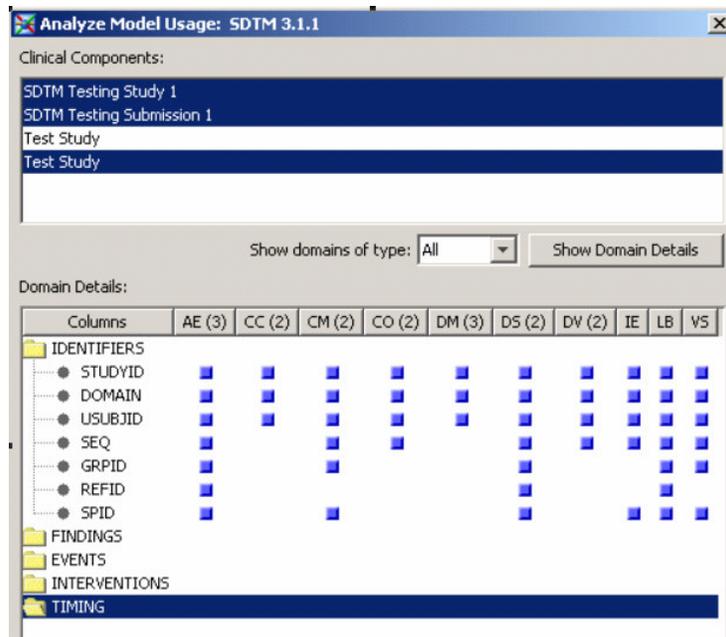


**Figure 17 – Analyze Data Model Usage**

SAS Clinical Data Integration also provides a plug-in to manage controlled terminology.  The administrator creates terminology packages containing the terminology tables (Figure 18).  When a new study or submission is created, the study administrator selects the controlled terminology package which will be used for this study.  This information is utilized by the CDISC-SDTM Compliance transformation and the SDTM to CRT-DDS transformation.
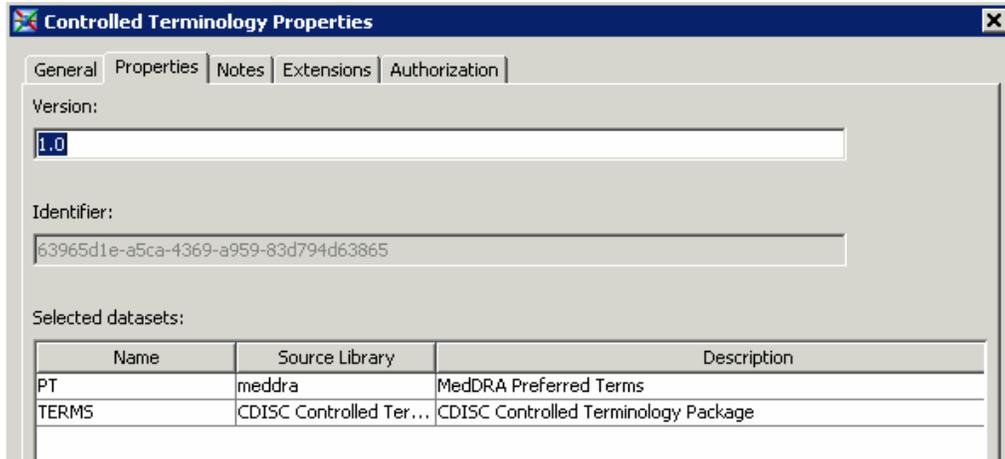
**Figure 18 – Controlled Terminology**

**CONCLUSION**

This paper has shown several key features of SAS Clinical Data Integration 2.1 related to implementing and managing CDISC standards.  By centrally collecting and managing metadata, you can reap the following benefits:

- expedite data standardization through reuse of metadata
- automate setup of studies and transformation processes
- customize validation and conformance checks
- monitor the consistency, and use of standards within an organization

Whether you are using published CDISC standards or company specific standards these tools will help enable the enforcement and assessment of compliance.

**RECOMMENDED READING**

Recommended reading lists go after your acknowledgments.

- Hunley, Eric, Gary Mehler, and Nancy Rausch. 2009. "What's New in SAS® Data Integration Studio 4.2" Proceedings of the SAS Global Forum 2009 Conference. Cary, NC: SAS Institute Inc.

**CONTACT INFORMATION**

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

Julie Maddox
SAS Institute Inc.
SAS Campus Drive
Cary, NC, 27519, USA

(919) 531-9083

julie.maddox@sas.com
http://www.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.

Brand and product names are trademarks of their respective companies.