Paper 168-2009

Using SAS® Clinical Data Integration Server to Implement and Manage CDISC Standards
Michael Kilhullen, SAS Institute Inc., Cary, NC

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 Server 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 Server 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 Server 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 Server 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 Server 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 Server through the SAS Clinical Standards Toolkit. 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 Server during the import and validation processes.
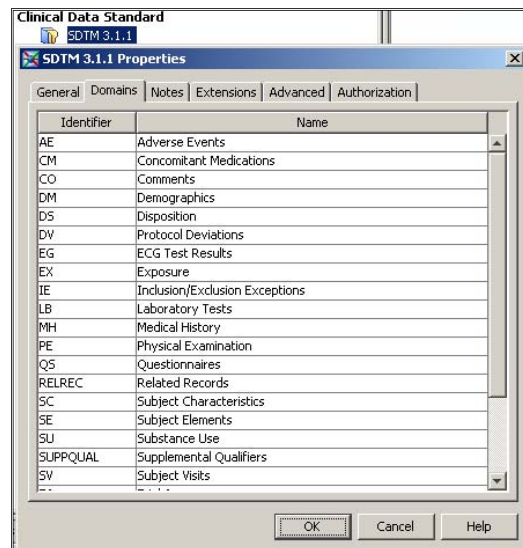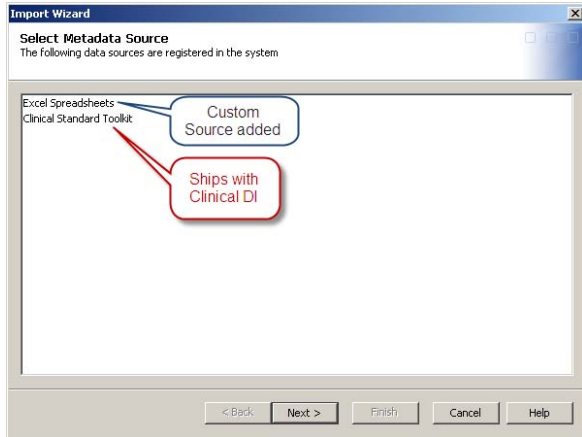
Figure 2 - Metadata Import Sources

SAS Clinical Data Integration Server provides a Data Standards Metadata Import Wizard to help data standard administrators select and load metadata (Figure 2). 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 (Figure 3). Once imported, the standard and domain templates are surfaced through SAS Clinical Data Integration Server 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.
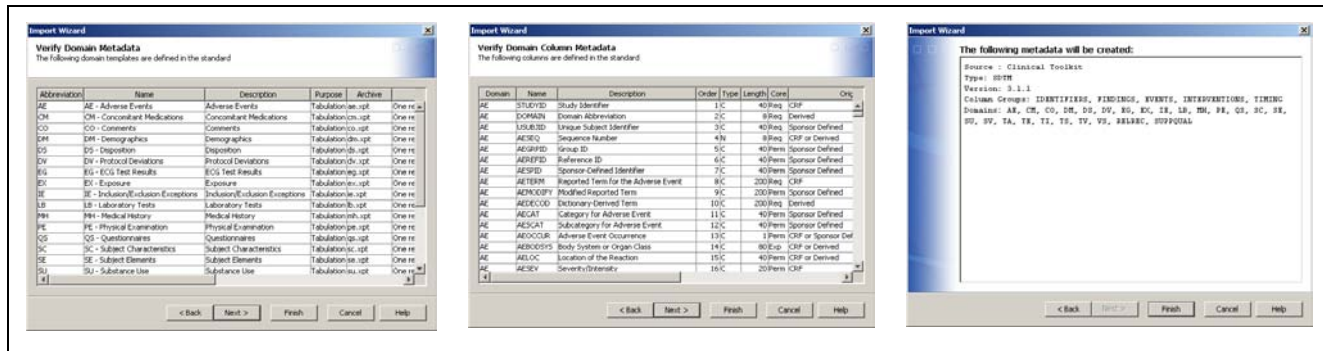

Figure 3 - Sample Displays from Metadata Importer

SUPPORTING THIRD PARTY SOURCES

If data standards are defined in another source such as Microsoft Excel, Oracle databases, comma-delimited files, and so on, the Data Standards Metadata Import Wizard can be enhanced by programmers to access the source and import the metadata. This requires writing a Java plug-in that feeds the necessary information to the Data Standards Metadata Import Wizard through a standard application programming interface. The plug-in is automatically detected and displayed as a metadata source (Figure 2). Details on how to do this are included in the SAS Clinical Data Integration Server documentation.

IMPLEMENTING DATA STANDARDS

Data standard domains can be implemented only 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 4). 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.

CREATING CLINICAL COMPONENTS

Studies and submissions are created using new object wizards. The wizard will collect standard object metadata such as name, description, responsible parties, and content location in the
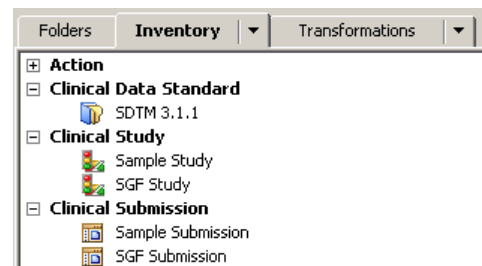

Figure 4 - Clinical Components in the Metadata Server

metadata tree. Next, metadata about the component is collected. For example, a study component will collect metadata such as protocol title, short 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 Server 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 Server can be configured by an administrator to create default metadata content when clinical components are created (Figure 5). 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 6). The administrator can also define standard libraries to ensure that the correct 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.
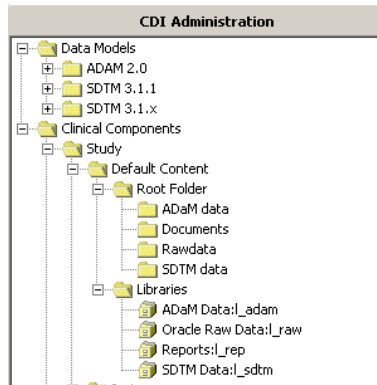

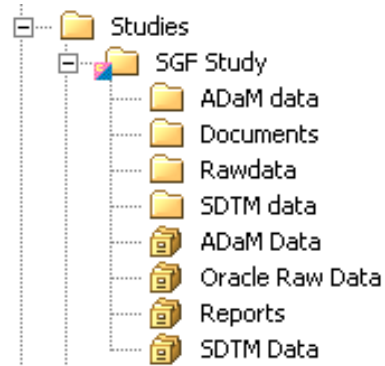**Figure 5- Default Content Defined for Studies**

Figure 6 - Content Created for a New Study Based on Defaults

REGISTERING DATA STANDARD DOMAINS

Domains are copied from the data standards using the register 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 registered, clinical metadata is accessible through standard SAS Data Integration Studio property dialogs that are enhanced by SAS Clinical Data Integration Server plug-ins. Additional tabs are added where the clinical metadata is displayed (Figure 7). 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 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 Server 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

Figure 7 - Clinical Metadata on Standard Properties Page

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 Server
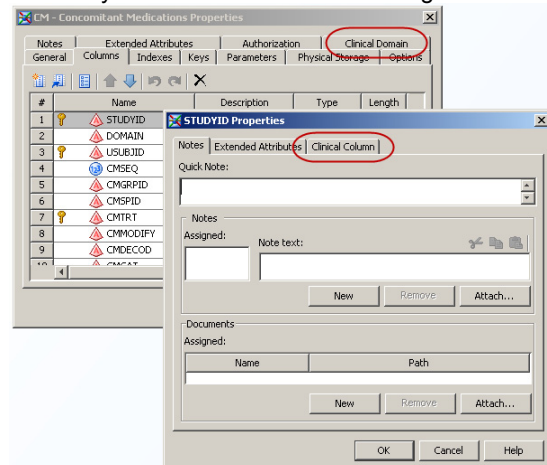
provides a wizard to guide you in defining the custom domain. Data standard administrators can use the same wizard to design new domain templates. Alternatively, they can identify and promote custom domains defined by users in studies and submissions to the data standard.

COPYING DOMAINS FROM OTHER STUDIES

On occasion, you might know of a custom domain created for another study that has not been promoted to a data standard. Rather than recreate it, SAS Clinical Data Integration Server can copy it as long as it is based on the same data standard version used by your study. In doing this, metadata settings are preserved and the necessary metadata relationships to the new study are automatically adjusted.

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.

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. If you can 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 Server, 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 and manage the metadata required by CDISC standards (Figure 8).

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

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,
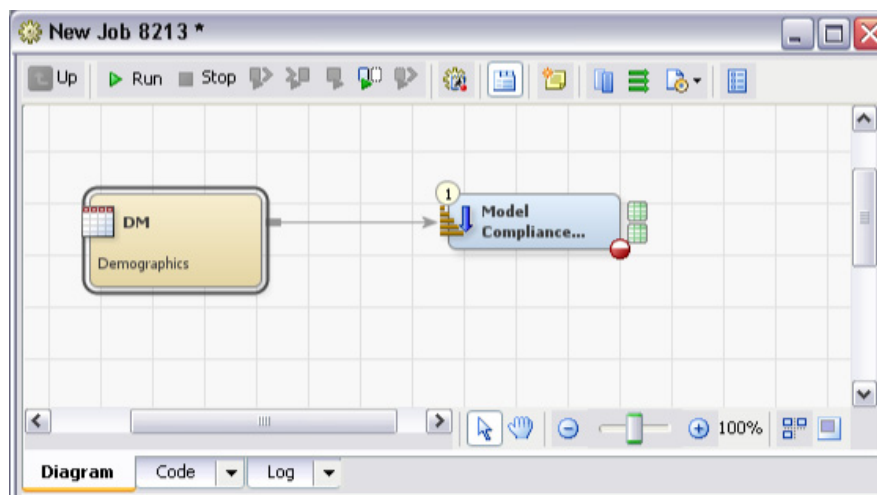

Figure 9 - Model Compliance Transformation in Job

the SAS Clinical Standards Toolkit is used by SAS Clinical Data Integration Server plug-ins to run validation checks and generate reports. Transformations are provided that you can add to a job. Simply add the domains to be validated and make selections from the model compliance properties. The model compliance transformation allows you to add as many checks as desired. It also offers filtering capabilities to help find the necessary checks.
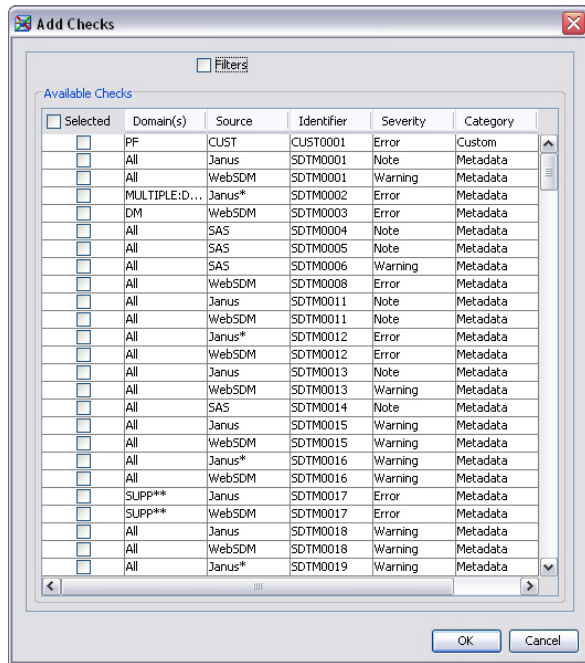

Figure 10 - Add Checks Interface on Compliance Transformation

GENERATING CRT-DDS

The metadata managed by SAS Clinical Data Integration Server can be published to CRT-DDS using the CRT-DDS Transform (Figure 11). This transform extracts metadata and passes it on to the SAS Clinical Data Standards Toolkit for define.xml creation. The transform allows you to specify properties to control encoding, log level processing, and stylesheets (Figure 12). When executed, the transform produces a define.xml document (Figure 13 and Figure 14).
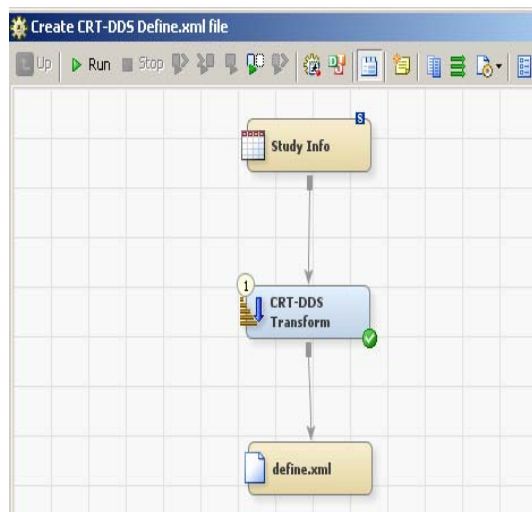
Figure 11 - Create CRT-DDS Define.xml File Process Flow
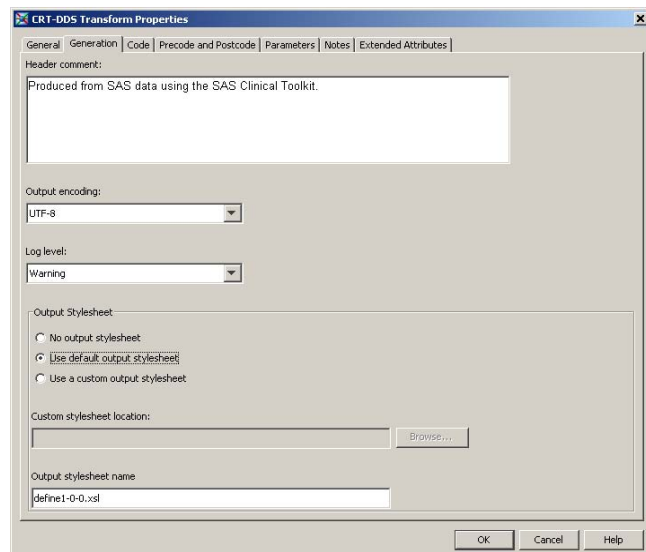
Figure 12 - CRT-DDS Transform Properties

Figure 13 - Datasets Section of CRT-DDS

Figure 14 – Domain Section of CRT-DDS

CONCLUSION

This paper has shown several key features of SAS Clinical Data Integration Server 2.1 related to implementing and managing CDISC standards. By centrally collecting and managing metadata, it can be used to automate setup and transformation processes, reuse metadata objects to expedite data standardization, feed validation and conformance checking, and improve the administration, consistency, and use of standards within an organization.

ACKNOWLEDGMENTS

Julie Maddox, SAS Software Developer, for reviewing this paper and providing content related to CRT-DDS generation

Geoff Mann, SAS Software Developer, for providing content related to validating domain content and structure

RECOMMENDED READING

- 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.

- Villiers, Peter. 2009. "Supporting CDISC Standards in Base SAS Using the SAS Clinical Standards Toolkit." *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:

Michael Kilhullen
62 Edison Road
Stewartsville, NJ 08886
908-760-6528
michael.kilhullen@sas.com
http://www.sas.com