

Implementing CDISC Data Models in the SAS® Metadata Server

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

Over the past several years, SAS has been actively supporting CDISC standards in base SAS software through the CDISC procedure. PROC CDISC allows you to easily read or write XML files formatted using the Operational Data Model (ODM), or verify that data you created conforms to the Study Data Tabulation Model (SDTM). We have also been active in demonstrating how SAS products and solutions can be used to convert your data into the SDTM data model using a centralized, streamlined, metadata driven approach. Given recent enhancements to PROC CDISC, we can now demonstrate how the metadata generated during the standardization process can be used to automatically generate a Case Report Tabulation Data Definition Specification document (CRT-DDS). To understand how SAS provides this capability and the value it brings to your organization, we will present some of the key concepts behind our solution, and how the SAS 9 architecture enables a manageable corporate-wide deployment of CDISC models and metadata.

THE SAS OPEN METADATA ARCHITECTURE

Every good solution begins with a good architecture. The SAS® Intelligence Platform is the foundation for all SAS 9 technologies and solutions. It provides the infrastructure to bring data integration, business intelligence, and analytic intelligence together into a common enterprise-wide platform. At the heart of this platform is the SAS Open Metadata Architecture, a general-purpose metadata management facility that provides common metadata services to SAS and other applications. Using the metadata architecture, separate SAS applications can exchange metadata, making it easy for these applications to work together as data moves through your drug development process. The metadata architecture reduces development effort because all applications use the same, centralized metadata, which makes it the perfect solution to sharing and managing CDISC models and metadata throughout your organization. The metadata architecture consists of three components: The SAS® Metadata Server, SAS® Metadata Model, and SAS® Open Metadata Interface (Figure 1).

SAS METADATA SERVER

The SAS Metadata Server is a multi-user server that surfaces metadata from one or more repositories. The repositories can be configured with dependencies so that information from one repository can be accessed from other repositories. The focal point of this configuration is the Foundation repository which contains information about the servers, data, and relevant 3rd party software used in your organization. The foundation repository also contains a metadata authorization facility that allows you to establish user/group permissions for accessing metadata repositories and content. Finally, all metadata repositories can optionally support change management allowing a controlled, team development environment for creating and managing metadata.

SAS METADATA MODEL

The SAS Metadata Model defines the metadata types for the most commonly used objects and enables you to extend the metadata types with custom attributes and associations. This extensibility allows the SAS Metadata Model to be expanded to handle specific business/industry needs. The model metadata is represented in XML allowing easy transformation of metadata into industry standard representations.

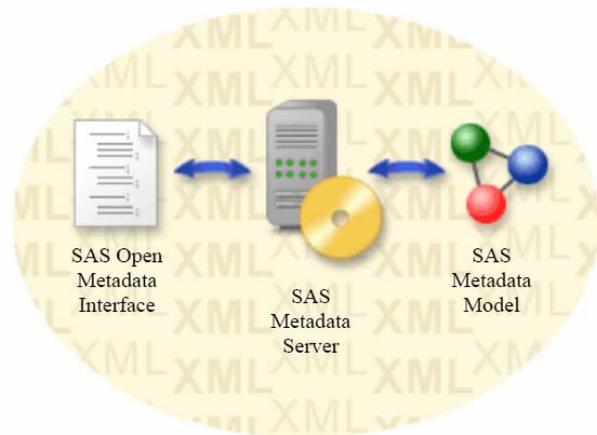


Figure 1 - The SAS Open Metadata Architecture

SAS OPEN METADATA INTERFACE

The SAS Open Metadata Interface is an application programming interface (API) that provides access to the metadata servers and content from a variety of programming environments including Java, COM/DCOM, and SAS. Using the metadata interface, you can interact with the metadata server to provide new capabilities or functionality to solve your specific business needs.

APPLYING CDISC MODELS TO THE OPEN METADATA ARCHITECTURE

The Study Data Tabulation Model (SDTM) defines the physical structure, metadata and business rules for creating standard domains of data. In addition, to effectively work with the model, we must also be able to leverage controlled terminology tables, and the base model definition that will be needed to create custom domains. In general, SDTM metadata is applied to the SAS metadata model using the PhysicalTable metadata type and several associated types such as Columns and UniqueKeys. As expected, these types collectively store metadata about a table structure. In addition, the metadata model also allows us to add custom metadata by defining extended attributes using the Extension metadata type. Extensions allow us to specifically add SDTM metadata to each table and column defined. The end result displayed in Figure 2 shows a complete domain definition stored in the metadata. Notice how the extensions are integrated into the table definition.

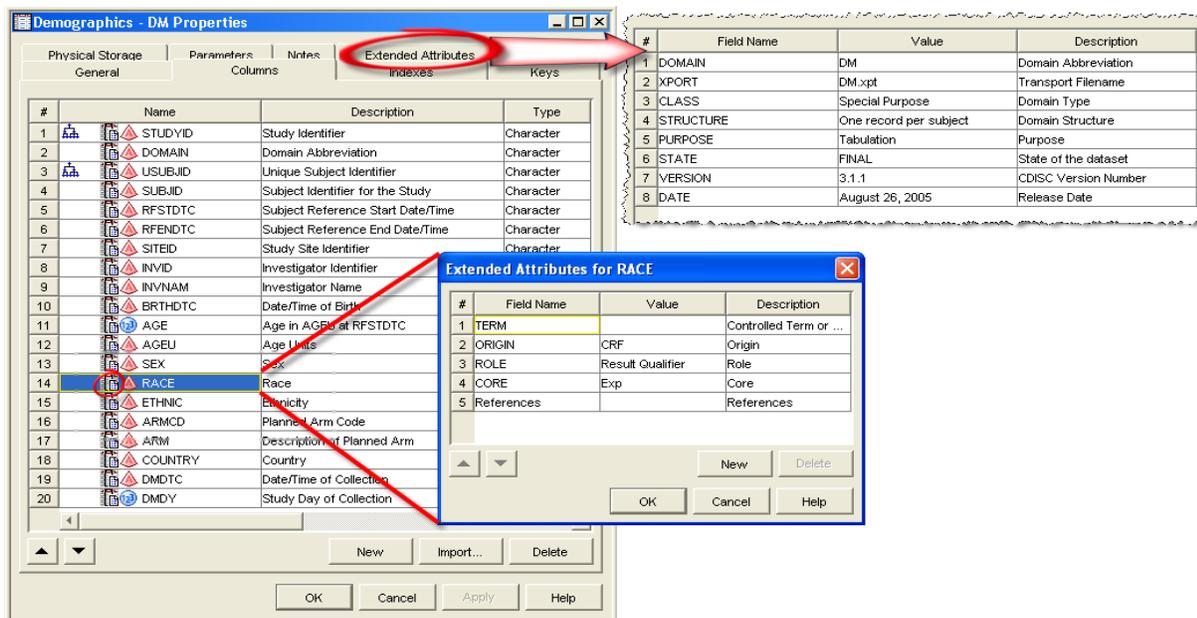


Figure 2 - Metadata Created for a SDTM Domain

ORGANIZATION AND DEPLOYMENT

In order to use the domain metadata more efficiently, we organize the models in the metadata using the Tree metadata type. Trees allow you to build a hierarchy around the metadata objects. To be consistent with the SDTM Implementation Guide, we organize domain models by classification. The complete SDTM model implementation is shown in Figure 4. The model classification variables stored in the SDTM 1.1 folder are provided to facilitate creating custom domains. The model tree is deployed to a common metadata repository. Recall that the foundation repository is the focal point of the metadata server. While this could be used, we will usually create a different repository to store the CDISC models. This allows you to keep your growing repository of business metadata separated from the core foundation information. When you are ready to start processing a study, a new repository is created for the study that depends on information from your standards repository (Figure 3). By using this configuration, business needs such as archiving and concurrent study development can be more easily managed. But most importantly, the CDISC metadata remains centralized and is common to every study defined, ensuring that all users

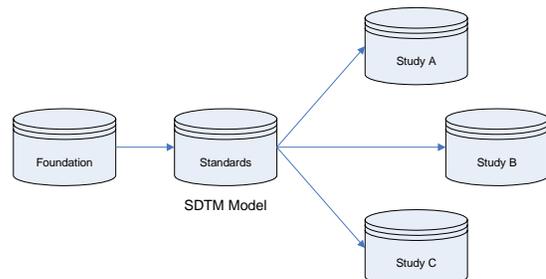


Figure 3 - Repository Relationships

in your organization are using the same model definition and enhancements. Since the CDISC metadata is subject to study specific information, you copy the model into your study repository. From this point forward, changes made in order to standardize your study data is stored as part of the study.

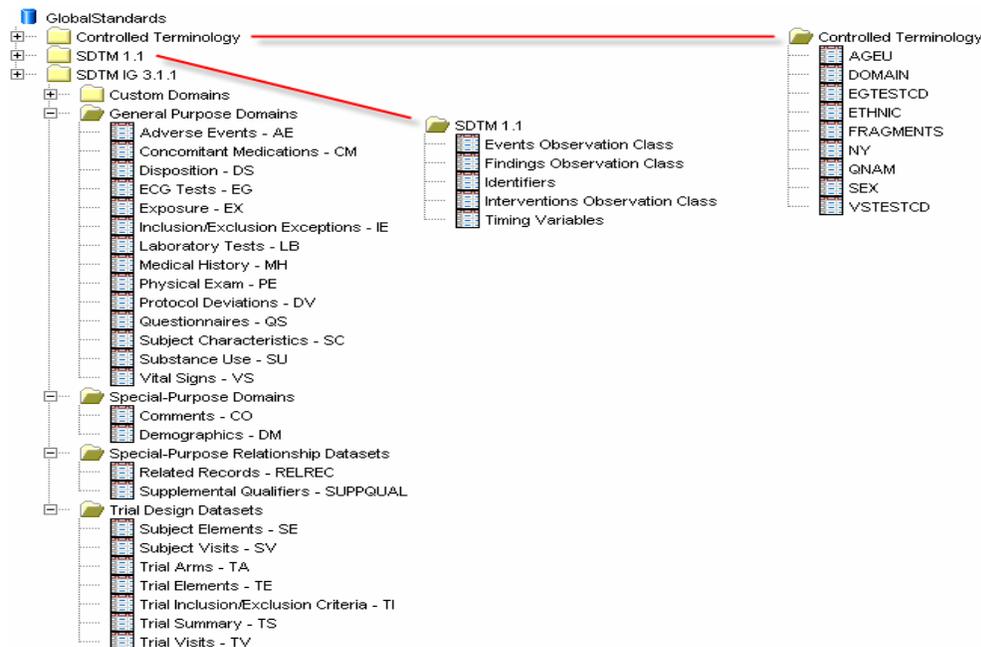


Figure 4 - SDTM Metadata Tree

PUTTING THE METADATA TO WORK: SAS DATA INTEGRATION STUDIO

SAS Data Integration Studio is a visual design tool for building, implementing and managing data integration processes regardless of data sources, applications or platforms. The drag-and-drop workflow design interface reduces programming time and training needs, and automatically captures and manages standardized metadata from any source. SAS Data Integration Studio includes many ready-to-use metadata driven transformations that eliminate the need to manage thousands of lines of custom code over the life cycle of projects. The library of transformations can also be expanded through a wizard driven custom transformation generator. Any SAS code you currently use can be easily adapted to seamlessly integrate into the metadata driven framework. Because of all this metadata driven capability, SAS Data Integration Studio brings intelligence into the standardization process by automatically identifying column inconsistencies, maintaining traceability, enforcing change management, and evaluating impact analysis throughout your data workflow.

SAS Data Integration Studio combined with the SDTM metadata becomes a streamlined, metadata driven approach for creating SDTM data. As you use the features of SAS Data Integration Studio, which is now driven by SDTM Metadata, you are proactively managing CDISC metadata while creating your final data. Figure 5 shows a typical transformation process and highlights the following benefits of working with the SDTM model in SAS Data Integration Studio:

- **Direct Access to CDISC Metadata** – as discussed previously, the CDISC metadata appears as extensions of the table object. You can directly access the metadata from the process diagram. When metadata changes are required, SAS Data Integration Studio provides the ability to edit the metadata.
- **Visual Transformation and Mapping** – each step in the process allows you to specify how the data flows through the process. When data needs to be transformed into something else, expressions can be added. Since the expressions are stored in the metadata, the finite details about how data was created can be surfaced from the metadata.
- **Leveraging Trial Design and Controlled Terminology Datasets** – because these data exist as data stores, you can access their content for use in the process. In Figure 5, controlled terms for VSTESTCD are being verified, and additional information found in the controlled terminology table is added to the data stream. Likewise, Trial Visits is being used to expand visit codes found in the source data.
- **Integrated Term Management** – as terms are being used in the data stream, issues encountered such as unknown terminology can be copied to exception tables. Because these tables are also registered in the

metadata, any users in your organization who are responsible for managing controlled terminology or trial design data can be immediately alerted to the details of the problem in order to respond more quickly to resolving discrepancies.

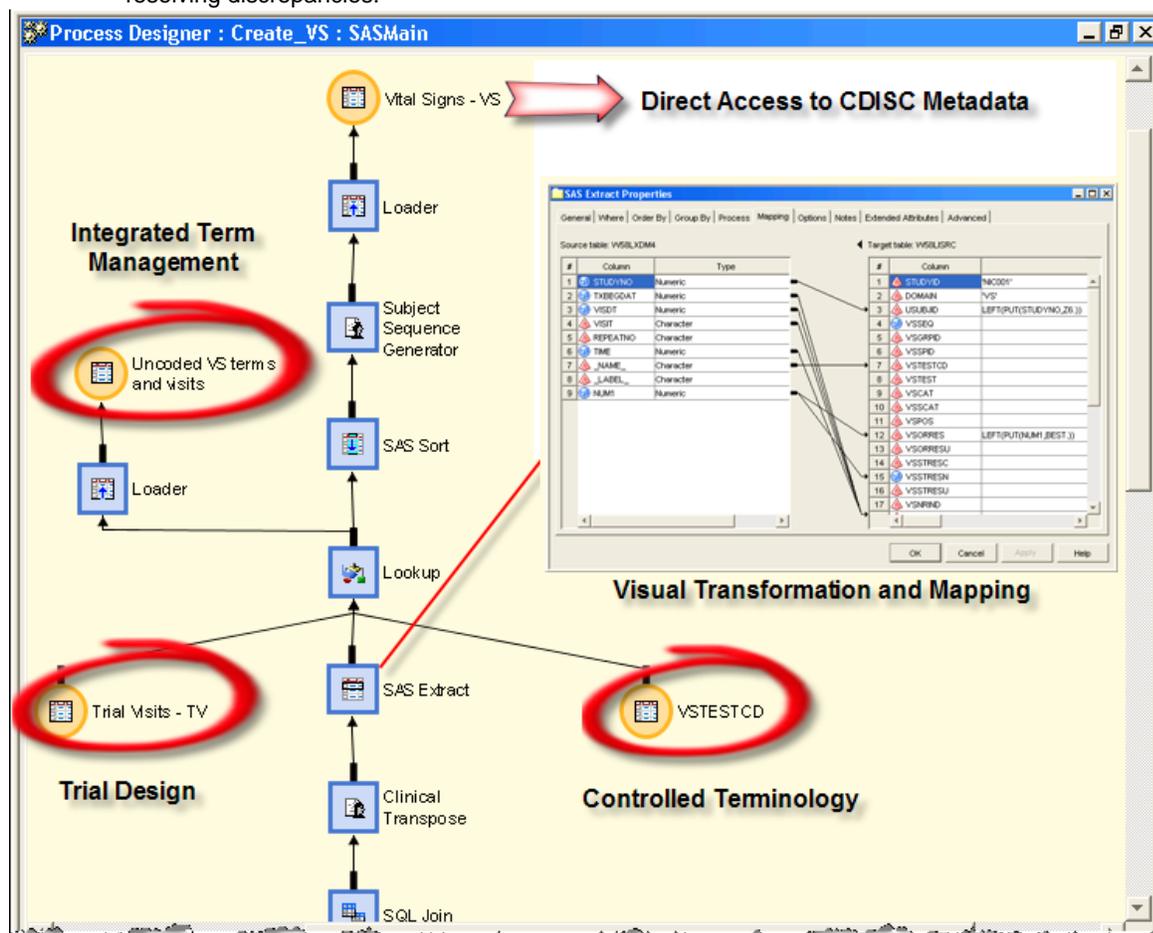


Figure 5 – Key Benefits of Working with SDTM in SAS Data Integration Studio

In addition to these specific benefits, you also inherit the following general benefits by using SAS Data Integration Studio:

- **Controlled Development** – When your metadata repository is configured to use change management, SAS Data Integration Studio requires users to check-out metadata in order to make any changes. During check-in, audit level metadata is created. Change management also reduces the likelihood of users duplicating effort.
- **Integrated Documentation** – All metadata objects allow you to create notes about the work you are doing. These notes are automatically added into the code generated by SAS Data Integration Studio producing better documented programs.
- **Extensibility** – SAS Data Integration Studio provides a plug-in architecture that allows you to add new capabilities specific to your business needs.
- **Data Flow and Relationships** – the visual interface of SAS Data Integration Studio is also metadata driven. As you build processes to standardize your data, the sequence of steps used is maintained in the metadata.

DOING MORE WITH THE METADATA

SAS Data Integration Studio clearly adds value to proactively leverage and manage the CDISC standards while standardizing your data. However, standardized data is not the only output of this process. In addition to the CDISC models, you have created an incredible amount of valuable metadata that we can leverage further.

IMPACT ANALYSIS

SAS Data Integration Studio is capable of performing impact analysis on the information you provided. Suppose you discover an issue with a RFSTDTC column in the DM domain. Rather than parsing through lines and lines of code, you can run an impact analysis on the variable to see how it was created throughout the entire standardization process. Figure 7 displays the output of the impact analysis. Notice that this analysis takes you all the way back to the source columns used in the original data sets, and summarizes how the column was processed through each step.

VALIDATION

During study data processing, the model metadata can also be used to verify that the data you created conforms to the standard model definition. Many of the business rules applied to the models can be validated by using the metadata. Figure 6 shows how the metadata can be leveraged to do this. When this process is implemented in the metadata, it can be added to any visual flow so that validation executes automatically as part of the standardization process.

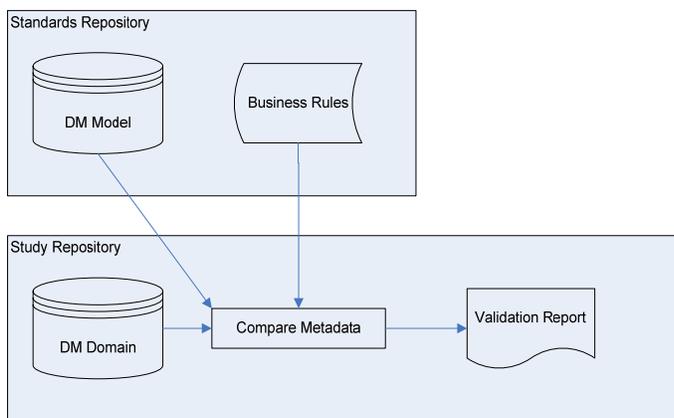


Figure 6 - Leveraging Metadata to Validate the DM Domain

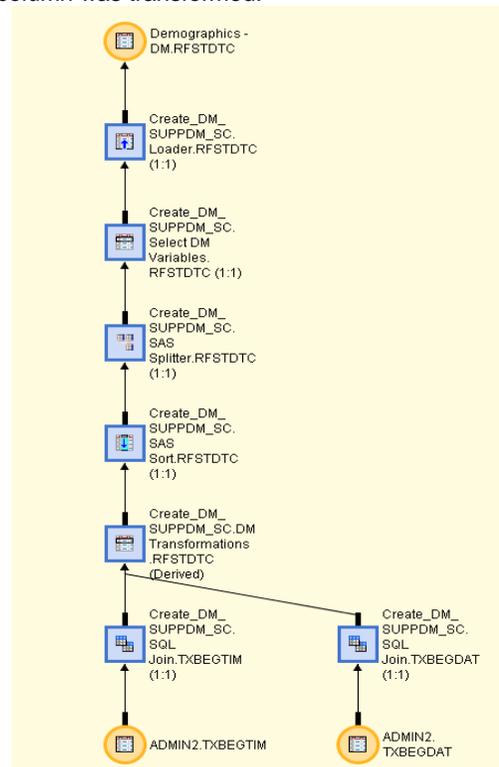


Figure 7 - Impact Analysis on RFSTDTC

METADATA REPORTING

The SAS Metadata Server supports standard XML exchange through the SAS Metadata Interface. This feature allows you to extract metadata and reformat it into meaningful reports. The metadata interface gives you the flexibility of running the report from a variety of interfaces. But once again, by taking this capability and integrating it into SAS Data Integration Studio, you can automate the extraction process, and easily add your own reports. Figure 8 shows how the generic extraction of metadata can be executed in an iterative manner. For each SDTM domain produced, the metadata is extracted then used to document the library as well as create a report detailing the change records associated to each table. A sample from the report is shown in Figure 9.

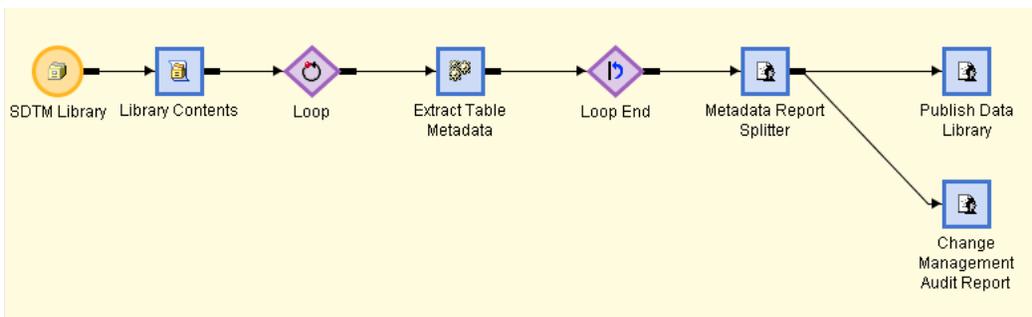


Figure 8 – Using SAS Data Integration Studio to Perform Metadata Extraction and Reporting

Metadata Report: List of SDTM Domains					
Version: 3.1.1 Publish Date: August 26, 2005					1
18:37 Tuesday, March 14, 2006					
CLASS	Dataset	Description	Location	Structure	Purpose
Events	AE	Adverse Events - AE	AE.xpt	One record per event per subject	Tabulation
	DS	Disposition - DS	DS.xpt	One record per disposition status or protocol milestone per subject	Tabulation
Findings	SC	Subject Characteristics - SC	SC.xpt	One record per subject characteristic	Tabulation
	VS	Vital Signs - VS	VS.xpt	One record per vital sign measurement per subject	Tabulation
Interventions	CM	Concomitant Medications - CM	CM.xpt	One record per medication intervention episode per subject	Tabulation
Special Purpose	DM	Demographics - DM	DM.xpt	One record per subject	Tabulation
Special Purpose Relationship	SUPPDM	Supplemental Qualifiers - SUPPDM	SUPPDM.xpt	One record per qualifier value	Tabulation

Figure 9 - Sample Metadata Report

PROC CDISC INTEGRATION: STREAMLINING CRT-DDS

Based on the content and figures presented in this paper, it may already be apparent to you that in using SAS Data Integration Studio with the CDISC data model to standardize your data, you have already collected much of the information that is needed for the Case Report Tabulation Data Definition Specification document (CRT-DDS). To easily generate this document, the next release of the CDISC procedure will support creating the CRT-DDS document. By using a process similar to Figure 8, the metadata needed for the CRT-DDS is automatically extracted and passed to PROC CDISC. A page from the document is displayed in Figure 10.

Dataset	Description	Structure	Purpose	Keys	Location
Demographics	Demographics	Special Purpose - One record per subject	Tabulation	STUDYID, USUBJID	DM
Concomitant Medications	Concomitant Medications	Interventions - One record per medication intervention episode per subject	Tabulation	STUDYID, USUBJID, CMTRT, CMSTDTC	CM
Adverse Events	Adverse Events	Events - One record per event per subject	Tabulation	STUDYID, USUBJID, AETERM, AESTDTC	AE
Disposition	Disposition	Events - One record per disposition status or protocol milestone per subject	Tabulation	STUDYID, USUBJID, DSSTDTC	DS
Subject Characteristics	Subject Characteristics	Findings - One record per subject characteristic	Tabulation	STUDYID, USUBJID, SCTESTCD	SC

Blank Case Report Form ([blankcrf.pdf](#))

Annotated Case Report Form ([ACROAnnotatedCRFs.pdf](#))

Supplemental Data Definitions Document ([supplementaldatadefinitions.pdf](#))

Go to the top of the [define.xml](#)

Date of document generation (2006-04-10T15:52:31)

Figure 10 - CRT-DDS Generated by SAS Data Integration Studio using PROC CDISC

CONCLUSION

This paper discusses the business value and process improvements gained by using the SAS Metadata Architecture to centrally deploy and manage the CDISC data models in your organization. SAS Data Integration Studio provides a meta-data driven approach that integrates the metadata management demands of CDISC into the data standardization process, resulting in a more efficient, streamlined, reusable process. Since the metadata collected reflects the needs of other CDISC models, the metadata can be further leveraged to provide better validation, automation, and generation of key submission deliverables such as CRT-DDS.

ACKNOWLEDGMENTS

Julie Maddox, SAS Pharmaceutical Software Development

RECOMMENDED READING

The Value of Integrated Metadata: SAS® Open Metadata Architecture
A SAS Whitepaper available at <http://www.sas.com/whitepapers/>

CONTACT INFORMATION

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

Michael Kilhullen
SAS Institute Inc.
62 Edison Road
Stewartsville, NJ 08886
Work Phone: (908) 760-6528
E-mail: michael.kilhullen@sas.com
Web: 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.
Other brand and product names are trademarks of their respective companies.