2022 PHUSE US Connect

Paper

Metadata Driven Case Report Tabulation for Regulatory Submissions

Rajashekar R Thota, Associate Principal Scientist, Merck & Co., Inc., Kenilworth, NJ, USA
Rama Kudaravalli, Associate Principal Scientist, Merck & Co., Inc., Kenilworth, NJ, USA
Jacques Lanoue, Associate Director, Merck & Co., Inc., Kenilworth, NJ, USA

ABSTRACT

The complexity of clinical trials and the evolution of data standards has led pharma companies to implement metadata repositories (MDR), improving efficiency and quality in data collection, tabulation, analysis, and submission deliverables. Sponsors are investing significant effort in new technologies and tools to automate business processes end-to-end and to improve the quality of submission deliverables.

This paper will discuss the implementation of an MDR at Merck. The tool helps to automate the generation of annotated Case Report Forms and SDTM datasets. Metadata can be stored and re-used within the tool to improve the efficiency and quality of submission packages. The tool supports the integration of different SDTM data models to facilitate metadata management.

The repository tool with its artificial intelligence provides the ability to quickly set up study configurations using industry, company, or therapeutic area standards. It also provides a powerful, graphically driven capability to build mappings and transformations that allow data to be represented in standards, support data review, analysis and reporting, and submission processes. User access and provisioning of the studies provide authentication and authorization. Study data can be uploaded to the tool to customize the study programs without additional effort. Company folder systems can be integrated, and existing company macros can be uploaded. Our experience in implementing this tool and the results achieved are presented.

INTRODUCTION

A metadata-driven approach and the seamless integration of processes and people are key to efficiently creating Case Report Tabulation (CRT) for regulatory submissions that are compliant and consistent with CDISC SDTM standards. The implementation of a metadata repository (MDR) is the first step toward a metadata-driven approach; We used a Commercial Off The Shelf (COTS) MDR product and implemented tools and processes to achieve automation of the SDTM CRT deliverable generation for regulatory submissions.

METADATA CENTRIC APPROACH

At Merck, we have chosen a metadata centric approach to automate the SDTM CRT generation, with the end goal of creating SDTM CRT deliverables for each study that are compliant and consistent with SDTM, while re-using the CRT generation in the MDR for all the studies improving quality and efficiency and resulting in cost optimization.

STUDY DATA STANDARDS SYSTEM (SDSS)

The process began with the creation of metadata in the metadata repository for each CRF form.
The metadata is a set of data that describes and gives information about the study CRF form data. Metadata for all the CRF forms that collected study data is used by SDSS to generate study SDTM datasets and aCRF mappings. The global library of metadata is the master source of metadata to be used for the study setup for all studies.

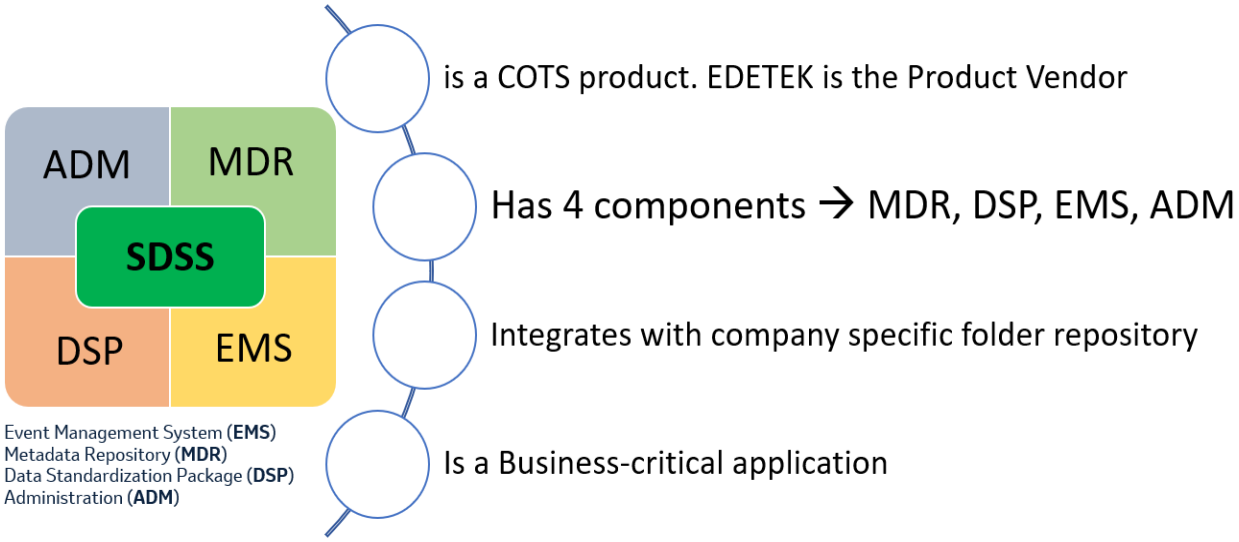

Figure 1: The components of the SDSS

Four SDSS components will be used in the automation of SDTM CRT deliverable generation.
The components of the SDSS include Event Management System (EMS), Metadata Repository (MDR), Data Standardization Package (DSP), and Administration (ADM).

SDSS COMPONENTS AND USAGE

EMS records all business and scientific events across processes associated with systems and user activities within the platform. EMS is used by all the SDTM programmers supporting studies to deliver study SDTM deliverables.

MDR is a repository that provides the ability to quickly set up study configurations using Industry, Company, or Therapeutic Area standards. It also provides a powerful, graphically driven capability to build mappings and transformations that allow data to be represented in standards supporting Data Review, Analysis, and Submission processes. MDR is used by SDTM programmers who are programming metadata for the CRF forms that will be used by all the studies that use the CRF Form.

DSP creates and maintains the code for SDTM and aCRF generation. It generates SDTM SAS program code from source datasets, eCRFs, SDTM transformation standards, Controlled Terminologies, and Value Level Metadata. It prepares and generates annotated CRFs and transforms data to SDTM and related submission documents. It is powered by AI/Machine Learning to perform CRF auto-SDTM Mapping, source to SDTM transformation, and SDTM data QC. DSP is used by all the SDTM programmers supporting studies to deliver study SDTM deliverables.

ADM component helps in maintaining the system. ADM is used by system admin users.

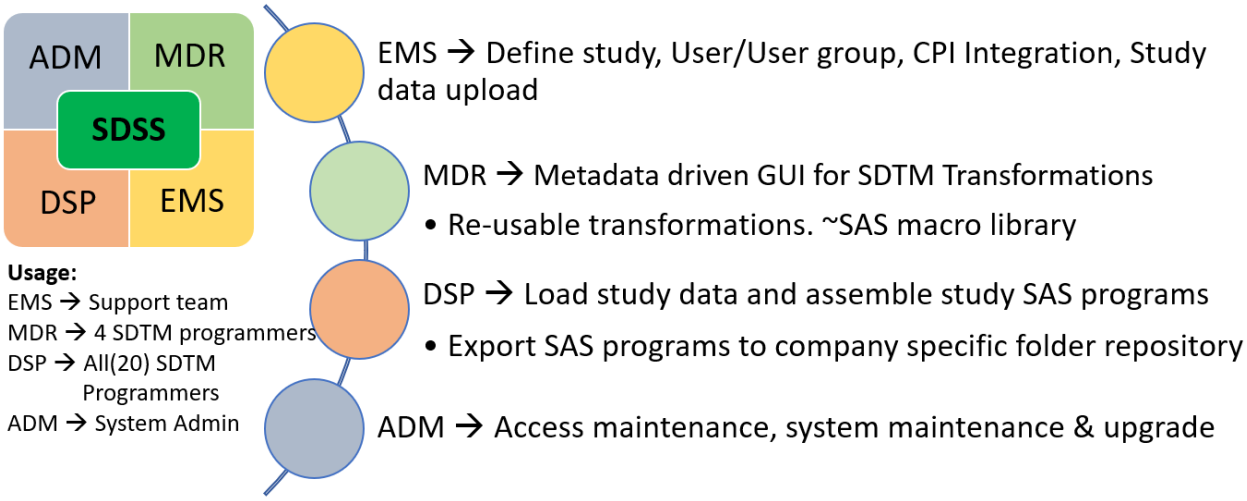

Figure 2: SDSS components and usage

SDSS FEATURES

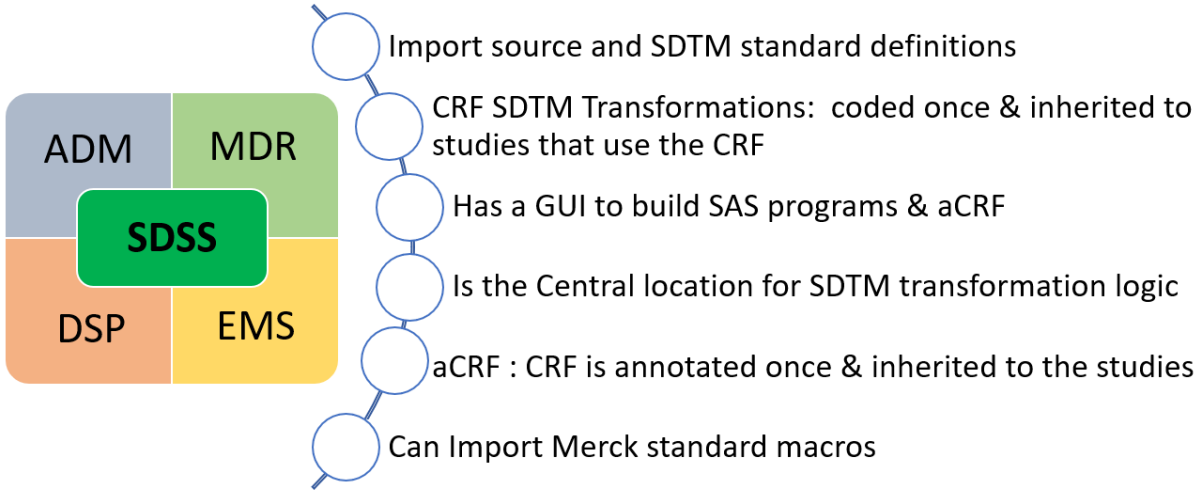

Figure 3: SDSS Features

SDSS features that allow the creation of maintaining SDTM metadata. Metadata is used to generate SDTM programs needed to generate SDTM datasets and CRF annotations. The metadata is agnostic of the study data. As part of working on the study data, study data will be loaded into SDSS and the SDTM programs that are dynamically adjusted to the study data are generated and exported to company folders. These programs are run by the existing company tools to generate study SDTM data.

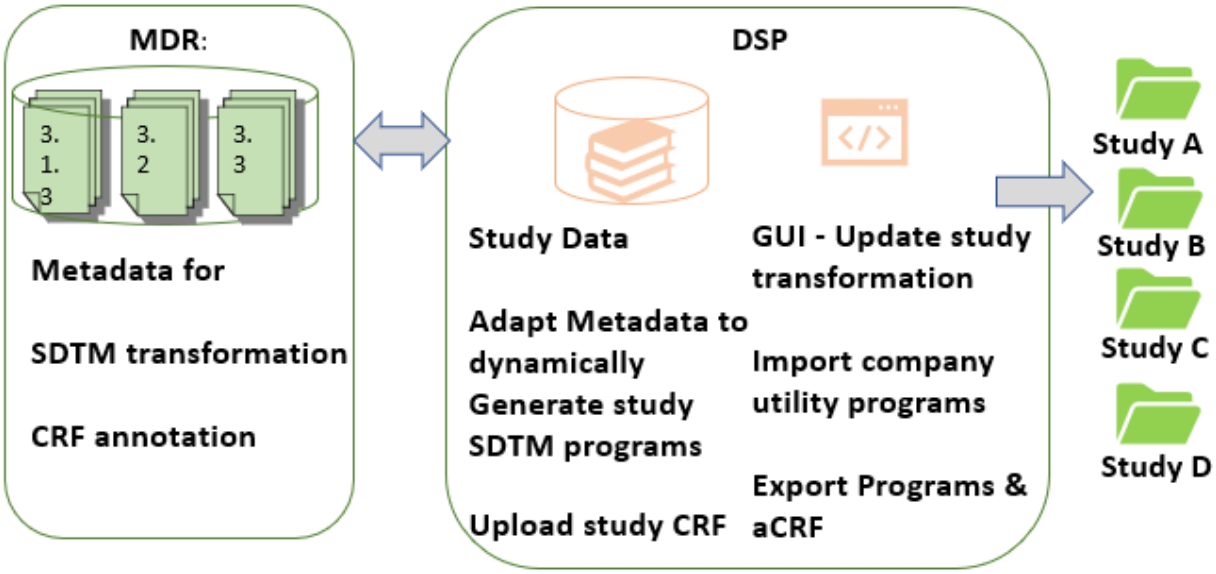

Figure 4: Metadata usage for the study

Metadata in the MDR is organized by SDTMIG versions. This allows each of the SDTMIG versions to have mappings specific to that version. This applies to both SDTM data transformation and aCRF annotations.

Study data is available in the DSP and metadata is inherited to the study. The metadata inherited from the MDR to the study can be updated in DSP to meet specific study needs. DSP allows a user to perform study-specific updates to the SDTM transformations, uploading of study CRF, generation of SDTM programs, review the SDTM data, upload existing macros specific to the study, and export SDTM programs, and aCRF. Once the programmer is satisfied with SDTM data, the SAS programs to generate SDTM can be exported and used to generate the SDTM data.

METADATA ORGANIZATION IN MDR

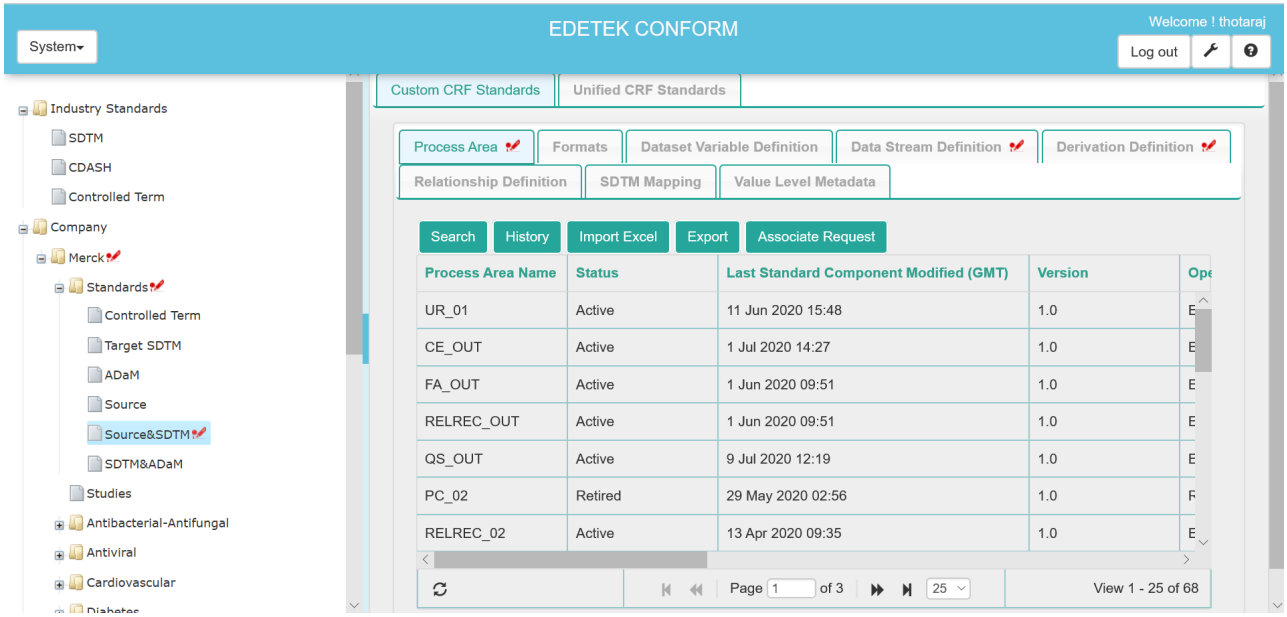

Figure 5: Metadata organization in MDR

Figure 5 illustrates the standard structure of an MDR library. Each parent level standard can be inherited and updated in the child level standard. At each level metadata for all the SDTM domains are inherited and updated as per the needs of the company. For each of the SDTMIG versions, the metadata is maintained by the SDTMIG version.

An MDR system allows the maintenance of standards at hierarchical levels within the organization. This system concurrently manages the company (organizational), therapeutic-area (TA), and study-level metadata objects. This system provides the mechanism to define the relationship among metadata objects at various levels. Every study is assigned to only one TA and the study inherits the standards from its TA standards.

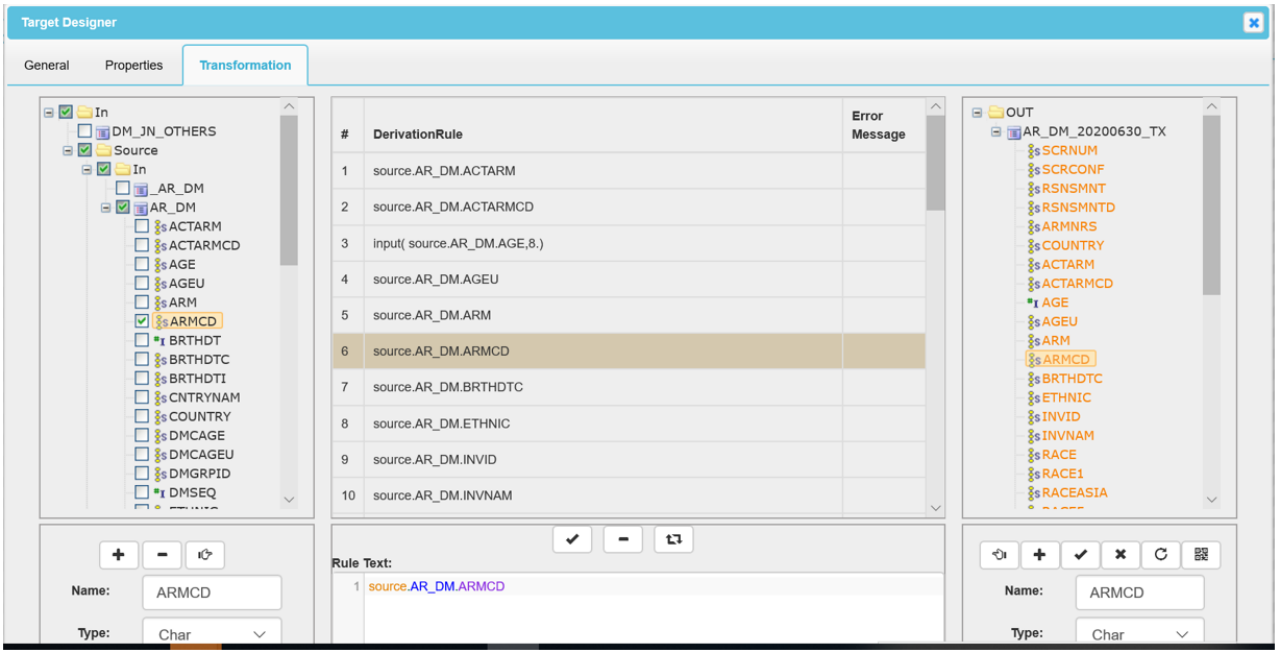

Figure 6: SDTM Transformation logic

The source data is mapped to the target for each CRF form automatically. The programmer will provide additional logic for variables, as needed.

In the MDR, SDTM mapping is provided for all the variables in each form. Once the mapping is completed, it is saved to the MDR repository and can be inherited to studies when the form is used by a study. Transformation logic for the actual variables collected on a study CRF form is inherited to the study. The rest of the variables are dropped from the study dynamically.

The programmer can update the study SDTM mapping logic as applicable. The SDTM mappings that are updated in the DSP will impact only that study.

External data such as lab data SDTM mapping is programmed similarly to CRF form mapping.

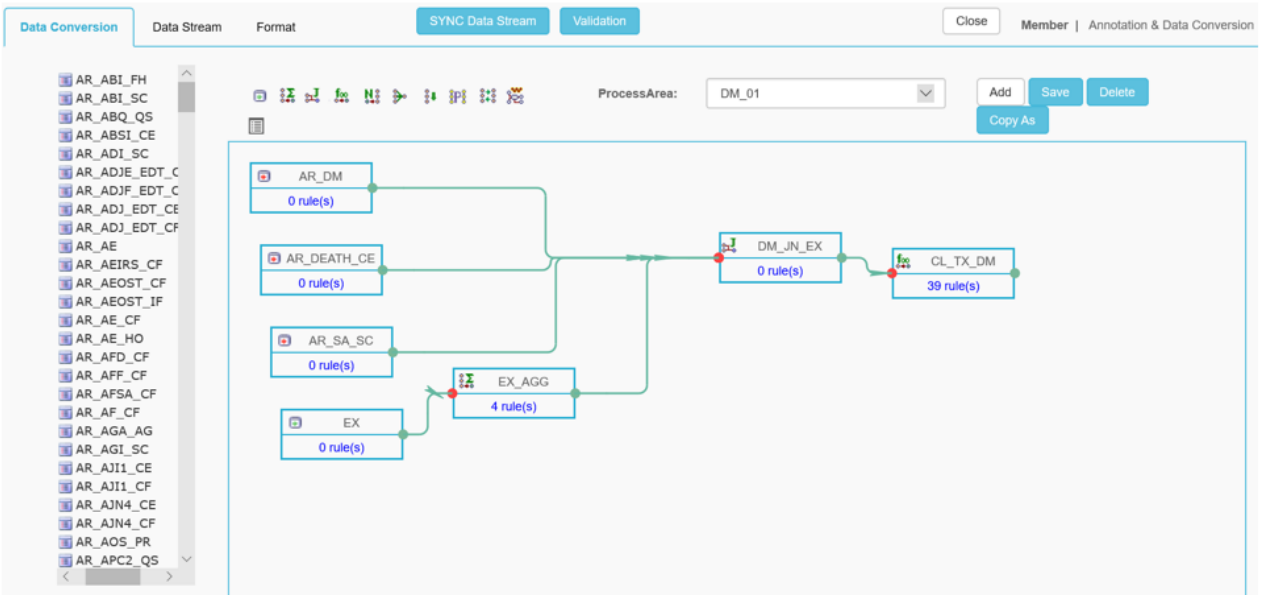

Figure 7: SDTM Transformation GUI

SDSS has a Graphical User Interface (GUI) to map the source data to the target SDTM domains. SDTM mappings for forms that collected the source data and the external data can be pulled into the SDTM domain transformation logic using this GUI.

Functions such as filter and join can be performed on the source data in the process to transform all the source data to the SDTM domain. Most SAS data manipulation functions can be performed using SDSS. When study SDTM SAS programs are generated and exported to the folder, they retain all the transformations performed in the GUI. Figure 7 shows the high-level transformation flow for an SDTM domain.

SDSS METADATA USAGE DURING LIFE OF A STUDY

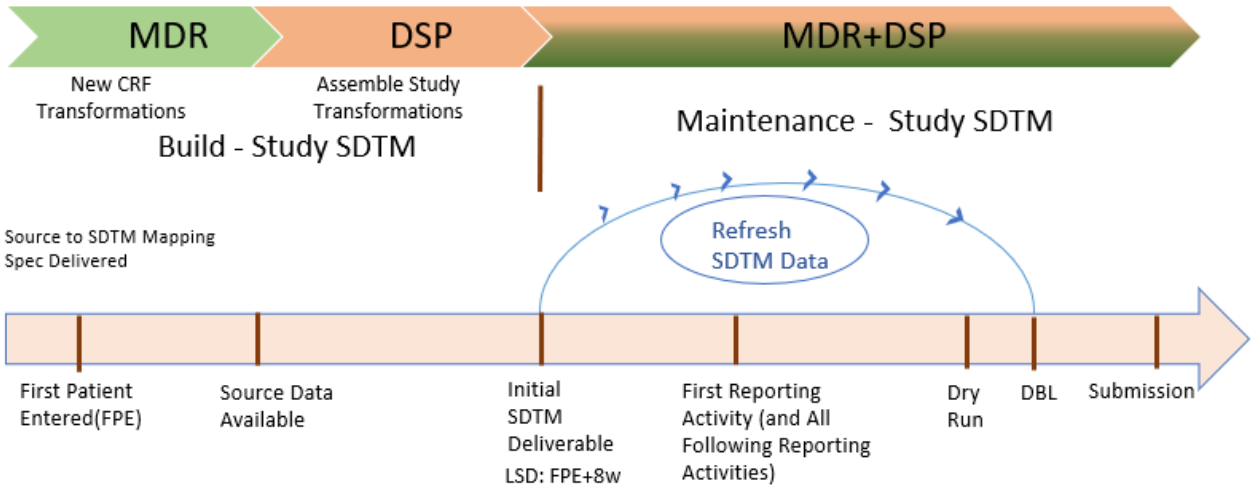

Figure 8: SDSS metadata usage during life of a study

Figure 8 illustrates the creation and usage of metadata during the life of a study. We create metadata in the MDR when a form is used for the first time. Once the form metadata is available in SDSS-MDR, we assemble the study in DSP by inheriting the study forms when study data is received. The SDTM programs are exported to the folder for running and generating SDTM data.

GENERATION OF SDTM DATASETS

1. Source data definition and the target SDTM data definition files are loaded into MDR. For a CRF form, source to target mappings are programmed for all variables and saved in MDR. Using the GUI, all the CRF forms that contribute source data to an SDTM domain are pulled in. The joining and merging logic are implemented and saved to the MDR.
2. In the DSP, a study solution is created. Metadata from the MDR is inherited to the study. Study CRF forms determine which domains are inherited to the study. If needed, the study solution logic such as mapping, merge, and filter can be updated but usually not required.
3. Study source data is loaded to EMS. This data is available to DSP. DSP will compare the study collected data and all variables in the metadata. The transformation logic for the study variables is kept and the rest of the variables are dropped from the study solution.
4. DSP will generate the SDTM SAS programs and transforms the source data to SDTM data. The study programmer reviews the SDTM data and makes any changes if needed. Usually, no changes would be needed.
5. Study programmers can export the programs to company folders. The programs are validated according to internal SOPs.
6. Programmers can use the SAS programs to generate the SDTM data.
7. SDTM data is validated with internal QC programs and the COTS validation tool.
8. If the SDTM mapping logic is changed CRF form, it is changed and saved to the MDR. The updated CRF form mappings are inherited to the studies that use the CRF form.
9. If the source data for the study changed, steps 3, 4, 5,6 are performed in that order.

MIGRATION TO NEW VERSION

Migration to a new version of the tool needs to be planned with an objective of minimum impact on existing studies on the system. If the changes in the new version are significant compared to the previous version, the business verification can be significant work. For the minimum business disruption for the business activities, the following mitigation activities are performed

1. Business users' time is essential for migration. Obtain the time before the migration starts.
2. Review the new version capabilities & identify any gaps in supporting the business requirements
3. Mitigate any gaps by either with tools or updating business process as applicable
4. Establish a migration to strategy by identifying what needs to be migrated and what is not needed.
5. Establish the business verification process and tools needed.
6. Obtain necessary approvals if the study data is needed for testing migration tool efficacy. As these approvals will go through a lot of checks and documentation, they can be time-consuming.
7. Establish the user training needs and plan to create appropriate training material. Often this process can take time, agreeing on the plan is important to train users in time so there is minimum disruption to the business activities.
8. Plan an early adopter study on the new version before migrating all the studies to the new version
9. To minimize the impact on ongoing study deliverables
    a. Plan for SDSS downtime to be minimum and acceptable to the business
    b. Communicate the downtime to all the business users
    c. Plan business deliverables around the migration times
    d. Plan a dry run for studies on the new version before the actual deliverables

GENERATION OF SDTM aCRF

1. CRF is loaded to the MDR.
2. The file with annotations for all the CRF questions is loaded to the MDR. This file will have all the annotations and offsets on where the annotation is expected to appear on the CRF.
3. The study CRF with all the unique CRF forms is loaded to the study solution in the DSP.
4. DSP identifies each of the questions in the study CRF and places the annotation on the page.
5. DSP ignores the questions that are not present in the study CRF.
6. The aCRF is exported from the DSP.

CHALLENGES

The biggest challenge was setting up the process and procedures for the programmers to use. System efficiency improvements were implemented. With system upgrades, the SDTM programming team needs to perform SDSS user acceptance testing (UAT). SDTM programmers are using this approach for most of our studies excluding the legacy studies.

OPPORTUNITIES

As the tool continues to be developed, we expect a robust integrated system that efficiently produces high-quality deliverables and we are investigating the opportunity to bring define.xml, cSDRG, and perhaps ADaM deliverables to be performed with this tool.

CONCLUSION

The use of the SDSS tool to automate SDTM CRT deliverable production, using a metadata-driven approach, has paid off big time, with significantly increased efficiency and quality. We achieved a level of 90% metadata reusability across trials, with a 33% reduction in programming resources. While the road to the metadata-centric approach hasn't been an easy one, we are reaping benefits in more ways than we had foreseen. At the same time, this is a work in progress with constant improvements being made considering changing needs of internal studies, data from external studies, and requirements from regulatory agencies.

CONTACT INFORMATION

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

Rajashekar Thota rajashekar.thota@merck.com
Rama Kudaravalli rama.kudaravalli@merck.com
Jacques Lanoue jacques.lanoue@merck.com

TERMINOLOGY

SDSS	Study data standards system
MDR	Metadata repository
DSP	Data standardization package
EMS	Event management system
ADM	Administration
CRT	Case report tabulation
COTS	Commercial off the shelf
UAT	User acceptance testing
SDTM	Study data tabulation model
SDTMIG	Study data tabulation model implementation guide
cSDRG	Clinical study data reviewer's guide
ADaM	Analysis dataset model
CRF	Case report form
aCRF	Annotated case report form
QC	Quality control
SAS	Statistical analysis system programming language
GUI	Graphical user interface
TA	Therapeutic area
AI	Artificial intelligence