

## Leveraging metadata when mapping to CDISC Standards with SAS® machine learning in a Results-as-a-Service Model (RAAS)

Preetesh V Parikh, SAS Institute Inc., Cary, North Carolina, USA

Sandeep Juneja, SAS Institute Inc., Cary, North Carolina, USA

Ben Bocchicchio, SAS Institute Inc., Cary, North Carolina, USA

### ABSTRACT

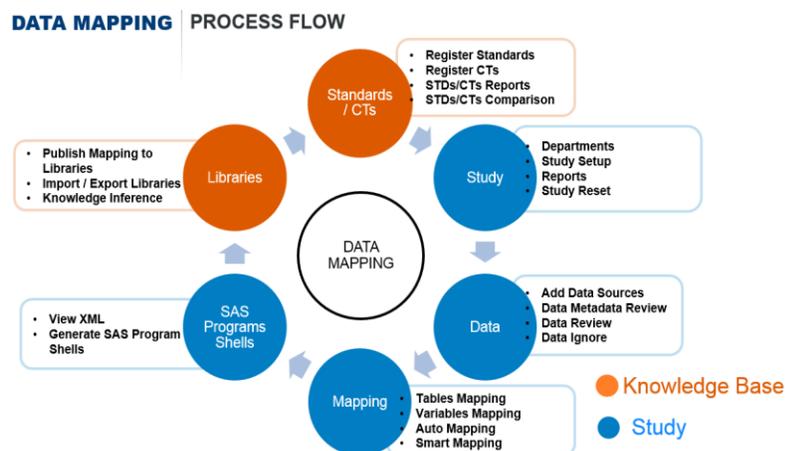
Mapping raw Electronic Data Capture (EDC) data to a Study Data Tabulation Model (SDTM) or Analysis Data Model (ADaM) standard is a very time consuming process. As the evolution of standards continues to develop, the ability to collect and reuse metadata on how raw variables get mapped into standardized variables can help reduce the time to standardize subsequent studies. The reduction in time and effort comes from the ability to reuse the mapping by analyzing the raw and standardized metadata.

Using the power of Machine Learning algorithms can further help derive and predict possible mappings. Instead of a customer having to install software locally and making the necessary configurations for it to run, SAS can offer this as a hosted solution. This means that customer can upload data and standards, run the process, and results of the process will be SAS code containing the mapping information that can be utilized in the customer environment.

### INTRODUCTION

Mapping raw data to standards is one of the most challenging process in the healthcare industry. Reusing or reapplying the information collected during mapping processes from previously mapped studies and building upon that knowledge inference is the most important part of the mapping process. Most companies struggle with building knowledge inferences and reapplying them through the data process efficiently. In addition, sometimes, there are multiple tools/programs required to go through full-cycle of data mapping. Therefore, it is difficult for the standard user to know all the different versions and tools, and use them correctly throughout the data mapping process.

The SAS Data Mapping Tool is a web-based tool that provides a user-friendly interface for everything from mapping raw data to generating SDTM standards (including domain templates). Simple User Interface (UI) and click-away concept design provides access to all the required information on a single screen. Auto-mapping and smart-mapping features in the SAS Data Mapping Tool, which are based on knowledge inference derived from machine learning algorithms, reduce time and effort for the user. This leads to improvements in quality, efficiency and consistency.



Display 1. Data Mapping Process Flow

## SAS DATA MAPPING TOOL'S SIX MAJOR COMPONENTS:

### 1. STANDARDS/CONTROLLED TERMINOLOGIES

To begin mapping, users need to register Standards/Controlled Terminologies (CTs) for use in mapping the study. Users can import or add Standard/Controlled Terminology to the tool and register it. This allows the user to build reports and provide comparisons among Standards/CTs. As more studies go through the mapping process, the tool allows users to see associated studies with the same Standards/CTs.

Display 2. Register Standards

### 2. STUDIES

Studies are added under departments. This keeps the study information restricted to within departments and controls user access to this information. Once the standards are defined for the study, users setup a new study by filling out the required study-related information. This includes the study start date, end date, phase, location of output template programs, and study setup file, etc. Real-time reports are available at every step of the process. A study reset option is also available. This allows users to clear content, but maintain the study structure.

Display 3. Study Setup

### 3. DATA

Users define references for data sources. Once the data sources are added, the tool can provide snapshots of the metadata associated with the raw data. This allows the user to familiarize themselves with the data before mapping. By default, the tool also shows the first 100 records from selected raw datasets. Users can restrict extraneous variables from displaying, while mapping, by using the Ignore functionality of the tool, thereby allowing them to focus on only the mapping variables.

**DATA DATA SOURCES**

1

**Add Data Source**

Name:

Description:

Path:

**Add Cancel**

Define Data Sources

2

General: Modules: Mapping:

Studies > Nicah1 > Mapping Home > Raw Metadata

Show Charts: No Added: All Status: All Source: All Table / Dataset: All # 100 View Clear

Raw Metadata: 100 records of 112 total

Status	Source	Table / Dataset	Variable	Order	Label	Type	Length	Format	Last Modified
Raw	ADVERSE	SUBJID	SUBJID	1	Subject ID	num	8	BEST12	19-MAR-17 10:09:58 EDT
Raw	ADVERSE	AEDECOD	AEDECOD	2	AE Decode from Dictionary	char	21	SF21	19-MAR-17 10:09:58 EDT
Raw	ADVERSE	AEACTTRT	AEACTTRT	3	Action Taken due to Event	num	8	BEST12	19-MAR-17 10:09:58 EDT
Raw	ADVERSE	AESV	AESV	4	Severity/Intensity of Event	num	8	BEST12	19-MAR-17 10:09:58 EDT
Raw	ADVERSE	AEOU	AEOU	5	Outcome of Event	num	8	BEST12	19-MAR-17 10:09:58 EDT
Raw	ADVERSE	AECOM	AECOM	6	Comment	num	8	BEST12	19-MAR-17 10:09:58 EDT
Raw	ADVERSE	STUDVID	STUDVID	7	Study ID	char	8	SF8	19-MAR-17 10:09:58 EDT
Raw	ADVERSE	TRTGRP	TRTGRP	8	Treatment Group	char	8	SF8	19-MAR-17 10:09:58 EDT
Raw	ADVERSE	BODYSYS	BODYSYS	9	Body System	char	8	SF8	19-MAR-17 10:09:58 EDT

Data Metadata

Display 4. Data Metadata

### 4. MAPPING

Once the raw data is input into the tool, users link source datasets to destination domain mappings. It can be one-to-many or many-to-many mappings. There is also an option to ignore the raw data source completely if it is not used within the study. After identifying tables, users proceed with variable mapping. If a user is mapping the first study, then the process is manual. However, if at least one study is published, then the user can use the auto-mapping or smart-mapping tools. The auto-mapping tool allows users to use the knowledge inference from published libraries to populate tables mapping and variables mapping. If similarly sounding or spelled variables do not map through auto-mapping, users can run smart-mapping on variables to get suggestions based on variables used in other studies. Auto-mapping and smart mapping is only in effect once users have published at least one study into the library. If users pick more than one study to scan for auto-mapping or smart-mapping, the order that information is retrieved is based on ranking. Lower ranked studies get precedence over higher ranked study in terms of obtaining information through the scanning process. At minimum, users need to have the required variables with the mapping definitions. Otherwise, the tool turns those variables a RED color with the designated “Not Mapped” status.

Leveraging metadata when mapping to CDISC Standards with SAS® machine learning in a Results as a Service Model (RAAS), continued

**TABLES TABLES MAPPING**

Mapping Tables Mapping

1

Add Many:  
Table / Dataset(s):  
Domain:  
Ignore:

ADVERSE  
CHEM  
CONMED  
DEMO  
ECG  
FUBE  
HEMA  
TOTAL  
VITAL

AE  
CM  
DM  
DS  
EG  
EX  
HE  
LB  
MH  
PE

Identify Raw Datasets → CDISC Domain

3 Tables Auto-Mapping

2

• One - One  
• Many – One  
• Ignore

Tables to Domains Mapping: 26 record(s)

Table / Dataset	Domain	Ignore
ADVERSE	AE	X
CONMED	CM	X
DEMO	DM	X
HEMA	LB	X
CHEM	LB	X
FUBE	LB	X
TOTAL		X

Select to Scan for Auto Mapping:

ID	Name	Description	Ranking	0	1	2	3	4	5	6
1	SAS EDC									
2	Nic_Mapping_Library									
3	Type 1 Diabetes Map	Mapping of Type 1 Diabetes								
4	Prostate									
5	prostate_mapping_std	Prostate Mapping Standard								
6	Huber's Mapping									

Tables to Domain Mappings:

Rule	Table / Dataset	To Domain	Ignore
SAS EDC	DEMO		Yes
Nic_Mapping_Library	CHEM	LB	
Nic_Mapping_Library	CHEM	SUPPLB	
Nic_Mapping_Library	CHEM	SC	
Nic_Mapping_Library	CHEM	PE	

Display 5. Tables Mapping

**VARIABLES VARIABLES MAPPING**

Mapping Tables Mapping Variables Mapping

Status: All Domain: AE Core: All

Data Browse: RawADVERSE View Metadata View Data Raw Controlled Terminology SDTM CT 2015-12-18

Mappings: 41 reviewed

Clear All Mappings

Domain: AE

Pre Processing Code:

Order	Status	Variable	Format	Role	Core	Origin	Length	Description / Notes	Pseudo?	Mapping / Pseudo Code
1	Mapped	STUDYID Char Study Identifier		Identifier	Required	CRF	200			RawADVERSE STUDYID
2	Mapped	DOMAIN Char Domain Abbreviation	AE	Identifier	Required	ASSIGNED	200			Domain=AE
3	Mapped	USUBJID Char Unique Subject Identifier		Identifier	Required	DERIVED	200			RawADVERSE SUBJID
4	Not Mapped	AESEQ Num Sequence Number		Identifier	Required	ASSIGNED	8			
5	Not Mapped	AEGRPID Char Group ID		Identifier	Permissible Drop	DERIVED	200			

Variables Hide Used Variables | SUPP

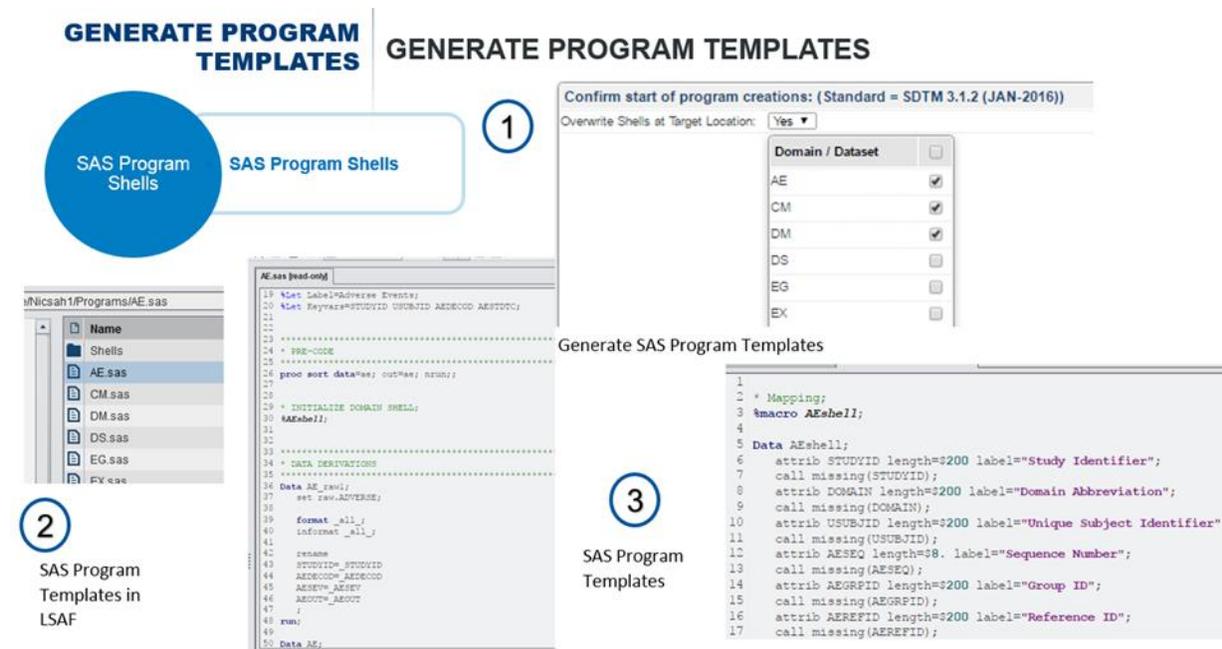
Dataset	Variable Name
RawADVERSE	SUBJID (Add to SUPP)
RawADVERSE	AEDECOD (Add to SUPP)
RawADVERSE	AEACTRT (Add to SUPP)
RawADVERSE	AESEV (Add to SUPP)
RawADVERSE	ABOUT (Add to SUPP)
RawADVERSE	AECOM (Add to SUPP)
RawADVERSE	STUDYID (Add to SUPP)
RawADVERSE	TRTRGP (Add to SUPP)
RawADVERSE	BODYSYS (Add to SUPP)

Variables Mapping

Display 6. Variables Mapping

5. SAS PROGRAM SHELLS

The SAS Data Mapping Tool allows users to generate SAS Program Shells. These shells contain the automated SAS code, based on the mapping information. It is a non-validated code, which may need further modification depending on the users' needs.



## Display 7. Program Templates

## 6. PUBLISH LIBRARIES

Once a user is through the mapping process, he or she can publish mapping information to a library. As users go through more and more studies, these libraries grow and the user can leverage this information in making mapping decisions for future studies.

Throughout these mapping components, real-time reports are available from the variable-level to the standards level. With the help of these reports, users can stay well-informed about the study progress, allowing for improved resource allocation. The tool is not limited to only SDTM. It can support ADaM or any other industry or custom standards as well.

### Results-As-A-Service (RaaS) Model:

The SAS Data Mapping Tool can be used by customers as a RAAS model in which the customer provides the data and specifications to SAS and SAS delivers the final mapping solutions using this tool.

## CONCLUSION

The SAS Data Mapping Tool is an easy-to-use tool, which provides functionalities required to map raw data to defined standards all in one place. The ability to leverage metadata and reuse knowledge inference makes it a very powerful tool.

## CONTACT INFORMATION

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

<p>Preetesh V. Parikh SAS Institute Inc. 100 SAS Campus Drive Cary, North Carolina 27513 (919) 531-3737 <a href="mailto:Preetesh.parikh@sas.com">Preetesh.parikh@sas.com</a></p>	<p>Sandeep Juneja SAS Institute Inc. 100 SAS Campus Drive Cary, North Carolina 27513 (919) 531-0541 <a href="mailto:Sandeep.juneja@sas.com">Sandeep.juneja@sas.com</a></p>	<p>Ben Bocchicchio SAS Institute Inc. 100 SAS Campus Drive Cary, North Carolina 27513 (919) 531-3704 <a href="mailto:Ben.Bocchicchio@sas.com">Ben.Bocchicchio@sas.com</a></p>
--	--	---

**Leveraging metadata when mapping to CDISC Standards with SAS® machine learning in a Results as a Service Model (RAAS),** *continued*

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.