PharmaSUG 2014 - PO14

Route to SDTM Implementation in In-Vitro Diagnostic Industry: Simple or Twisted

Carey Smoak, Roche Molecular Systems, Inc., Pleasanton, CA
Sofia Shamas, MaxisIT Inc., NJ
Chowdagam Chaitanya, MaxisIT Inc., NJ
Don Lim, Bluewave Consulting Inc., CA
Girish Rajeev, InVentiv Health Clinical, FL

ABSTRACT

SDTM implementation for In-Vitro Diagnostic (IVD) data found a guiding light when CDISC came up with seven new domains specific to the medical device industry. Although these domains are not tailor-designed for diagnostic data, they serve as a good starting point for us to define SDTM+, which is an extension of the SDTM standards to accommodate the large variety of IVD data. The similarity across studies following SDTM domain structures presents opportunities for code standardization and reusability. What this paper hopes to achieve is to take the reader on our journey and explain briefly the company-specific standards created, the new domains and how our data was adjusted across SDTM+ domains. The main focus will be on mapping of IVD data with company specific use-case examples.

INTRODUCTION

More and more Pharmaceutical/Device companies use CDISC SDTM as part of FDA submission. In December of 2012, seven new SDTM domains were published for use in medical device submissions (Smoak et al 2012). Yet, there hasn't been any SDTM model developed for data from clinical trials within the IVD industry. Lack of standards is accompanied with another challenge that data collection does not always follow CDASH specific standards. A use-case for IVD standards has been described elsewhere (Smoak et al 2014). This paper expands on this use-case and shows the detailed mapping of IVD data to SDTM. These standards are important because the FDA is moving towards requiring CDISC standards such as SDTM and ADaM for regulatory submissions (Smoak et al 2013).

A team of SAS® programmers from Roche Molecular Systems (RMS) worked on this project. The data presented in this paper is not the actual data from an IVD clinical trial, rather a representation of the type of data collected in an IVD study. The modified structure of SDTM+ (Figure 1.) will include the following domains –

Figure 1: Domains of SDTM+

Existing (Modified/as is)		New
Pharma	**Device**	
AE (Adverse Events)	**DI (Device Identifiers)**	**RN (Run Related)**
CM (Concomitant Medications)	**DU (Device-In-Use)**	**CH (Channel Related)**
DM (Demographics)	**DE (Device Events)**	**SM (Sample Related)**
DS (Disposition)		**SA (Sample Properties)**
DV (Protocol Deviation)		**SP (Specimen Properties)**
IE (Inclusion/ Exclusion)		
MS (Microbiological Susceptibility)		

Following SDTM+ structure achieves dual purpose of standardization of clinical data across different clinical studies as well as achieving repeatability of programming tasks across different studies.

Our journey will start with an explanation of the IVD Data that we had to work with, what follows then is the process which defined the roadmap to SDTM and then an explanation of the implementation of the process in the use-case examples.

IVD DATA EXPLAINED

Key Terms Explained:

Run – A sequential or distinct identifier, in our case is called as a run number. A run is an iteration of controls and samples that need to be tested as part of the clinical trial.

Sample – Samples are specimens (e.g., blood) collected from subjects which need to be tested to support the claims of a study.

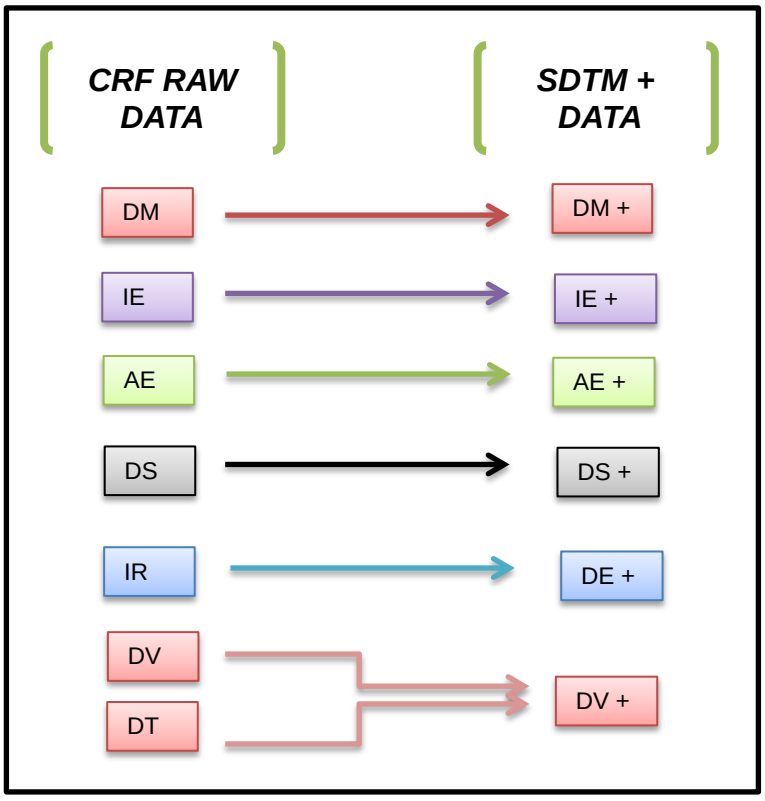
Controls – specimens with known results (e.g., positive and negative) that are processed like test samples.

Channel – Will filter the target (analyte) that will falls within the acceptable range of a particular test. For example, if a lab instrument can test for multiple targets, then it will need a channel for each target (analyte)

CRF Data:

IVD CRF data is similar to Pharma/Device data for Demographics (DM), Disposition (DS), Adverse Events (AE), Protocol Deviations (DV) and Inclusion/Exclusion Criteria (IE) for both specimen collection as well as testing of specimens.

Figure 2: CRF Data Mapped to SDTM+

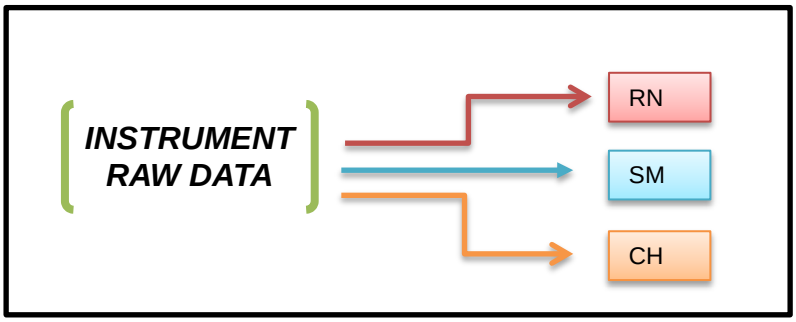


There are, however, additional domains which are unique to diagnostics such as Run Submission, which collects information related to each run. An Incident Report is another domain which is unique to diagnostics. In Incident Report collects data about incidents that occur during the conduct of clinical trial (Figure 2).

Raw Instrument Data:

IVD instrument data is the actual testing data collected in a clinical trial to support the device efficacy. This data can be classified under a Findings domain. Typically Pharma result lab data (and other data) are collected over a long period of time. However, in the IVD industry large number of results data can be collected in a very short span of time. IVD data structure is hierarchical and a run consists of a set of samples and controls which are processed together (Figure 3). An IVD lab instrument may detect more than one analyte (e.g., HIV, HBV, HCV etc.) in a run and multiple channels (one per analyte) is needed to get the test results for each analyte. Additionally. For the intent of this paper, we will look at only three of those domains.

Figure 3: Instrument Data Mapped to 3 New SDTM+ Domains



1. Mapping of RMS Incident Report CRF to the Device Event domain (DE).
2. Mapping of RMS Instrument Testing Data related to samples to Roche defined Sample Level domain (SM).
3. Mapping of RMS Instrument Testing Data at Run Level and CRF Run Submission Data to Roche defined Run Level domain (RN).

1. Create an excel document which serves as a mapping document between raw and SDTM data (Figure 4).

   The Excel document is the heart of our mapping process which explains in detail how each individual variable will be mapped within a dataset. The document is created following the lines of SDTMIG.

   The first few columns in the document define the SDTM domain, including variable attributes, description, core and origin. The next set of columns will define how the CRF data values get passed on to the SDTM variables.

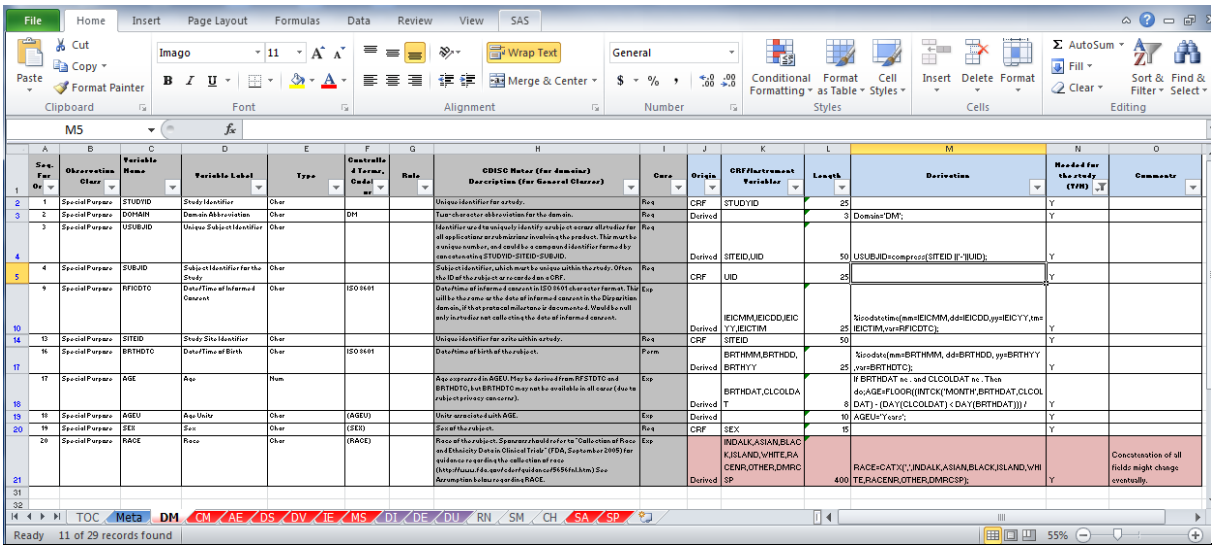
   Three kinds of derivations are followed in mapping the data:

   a. One to One mapping:
      In this scenario the CRF or the Instrument Result variable value gets carried over to the SDTM variable without modification using simple assignment statements.

   b. Derived:
      Simple Derivations or macros have been used to derive these variables.

   c. Programmatically Derived:
      This approach is followed when we have to follow complicated derivations to reach the SDTM variable or for cases where the source variable name is similar to the SDTM variable.

Figure 4: Excel Document which Serves as a Mapping Document



2. Write a SAS code which will transform raw data to SDTM using the mapping document.
   A macro was written which transformed the source to SDTM using the rules defined in the excel document. (See Appendix to see the entire code).

CASE EXAMPLES

Figure 5: Mapping of DE Domain

	IDID	IRCAT	IRINFX	IRLABACC	IRINCDD	IRINCMM	IRINCYY
1	AAA0004	Operational	Infrastructure Error	Laboratory Accident	22	Nov	2013
2	AAA0119	Instrument Error	Hardware Error		9	Dec	2013
3	AAA0274	Other			7	Jan	2014

	DOMAIN	DEIDID	DETERM	DEINFX	DELABACC	DESTDTC
1	DE	AAA0004	Operational	Infrastructure Error	Laboratory Accident	2013-11-22
2	DE	AAA0119	Instrument Error	Hardware Error		2013-12-09
3	DE	AAA0274	Other			2014-01-07

METHOD OF IMPLEMENTATION

After defining the new SDTM domains the next step was to decide which raw data gets mapped to which SDTM domain. Few of the mapped domains are illustrated below.

The Incident Report (IR) CRF is mapped to the Device Events (DE) domain, which captures details about any kind of incident during the clinical trial. For example, an IR may be an event which occurred at the investigational site, such as operational, instrument, hardware or software errors.

The mapping example shown above (Figure 5) is a one-to-one mapping between the IR and DE domains with raw

values simply being carried over to the SDTM variables. All the date variables are derived in ISO8601 format from day, month and year variables (Figure 6). The code below shows the date time conversion.

Figure 6: Deriving ISO Date

```
*1. Macro For SDTM DATE ISO8601DT Format;
%macro isodate(mm= ,dd= ,yy= ,var=);
   *Find the Month in Numerics for MDY Function;
    if &mm ne ' ' then monn=month(input(cats(1,&mm,2011),date9.));
    if monn ne . and &dd ne . and &yy ne . then &var=put(mdy(monn,&dd.,&yy.),is8601da.);
%mend isodate;

*2. Macro For SDTM DATE"T"TIME ISO8601TM Format;
%macro isodatetime(mm= ,dd= ,yy= ,tm= ,var=);
   *Find the Month in Numerics for MDY Function;
    if &mm ne ' ' then monn=month(input(cats(1,&mm,2011),date9.));
    if monn ne . and &dd ne . and &yy ne . then date=put(mdy(monn,&dd.,&yy.),is8601da.);
   if &tm ne . then time=put(&tm.,E8601tm.);
   if date ne ' ' and time ne ' ' then &var=trim(date)||"T"||Trim(time);
%mend isodatetime;
```

Figure 7: Mapping of SM Domain

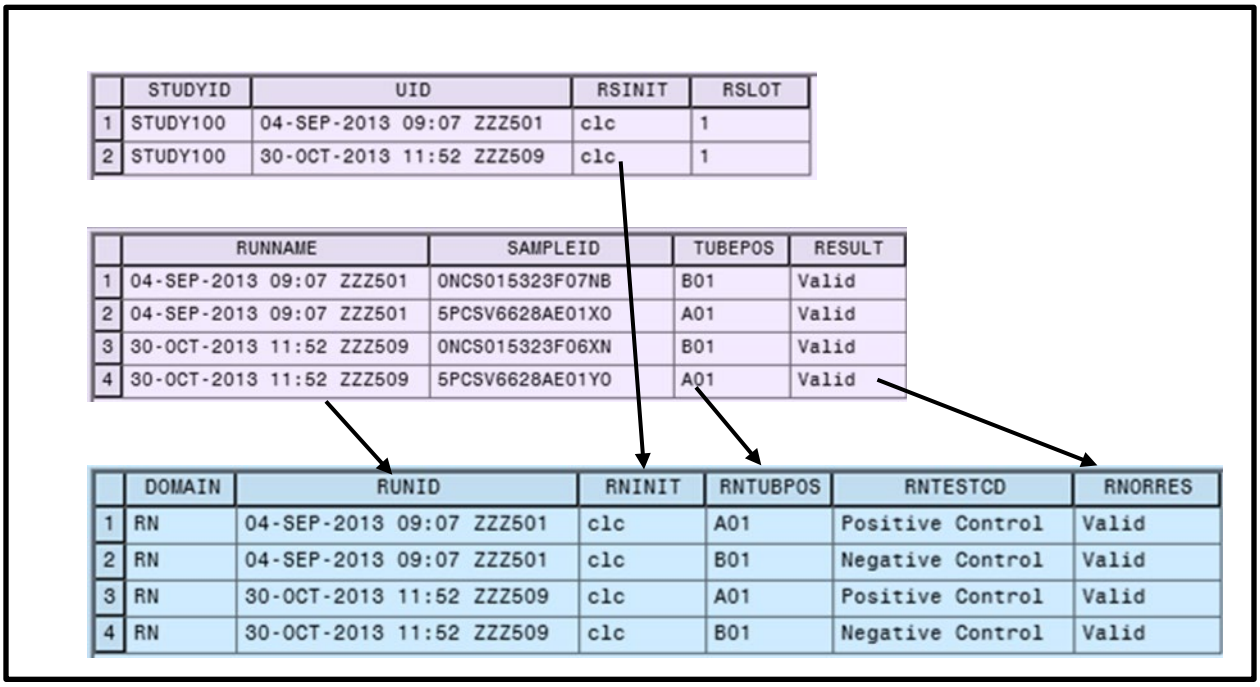
	SAMPLEID	TUBEPOS	RESULT
1	AAA0004	C01	NEG XXX-1, NEG XXX-2
2	AAA0119	E01	POS XXX-1, NEG XXX-2
3	AAA0274	F01	POS XXX-1, POS XXX-2

	Domain	SAMPID	POSITION	SMCAT	SMSEQ	SMTESTCD	SMORRES
1	SM	AAA0004	C01	Device-1	1	XXX-1	NEG
2	SM	AAA0004	C01	Device-1	2	XXX-2	NEG
3	SM	AAA0119	E01	Device-1	1	XXX-1	POS
4	SM	AAA0119	E01	Device-1	2	XXX-2	NEG

The SM domain is mapped from the raw instrument data and as the name suggests it captures sample level details. As shown above this domain captures details such as Sample ID, Result, Tube Position to name a few key variables
.
SAMPLEID and TUBEPOS are examples of one-to-one mapping, whereas SMTESTCD and SMORRES are programmatically derived from RESULT based on the conditions defined in the excel mapping sheet (Figure 7).

Run related information is captured both on the CRF and also in the electronic result file from the instrument. This domain was created to unify all the run related information into a single domain. This RMS domain will have two records per run number which is determined by the number of controls. It will have variables which will carry information such as run date and time, position of the controls, operator and lot related information to name a few. Variables determining the validity of the run also get saved as a part of this domain (Figure 8). Examples of one-to-one mapping include: RUNNAME: RUNID, RSINIT: RNINIT.

Figure 8: Mapping of RN Domain



	STUDYID	UID	RSINIT	RSLOT
1	STUDY100	04-SEP-2013 09:07 ZZZ501	clc	1
2	STUDY100	30-OCT-2013 11:52 ZZZ509	clc	1

	RUNNAME	SAMPLEID	TUBEPOS	RESULT
1	04-SEP-2013 09:07 ZZZ501	0NCS015323F07NB	B01	Valid
2	04-SEP-2013 09:07 ZZZ501	5PCSV6628AE01X0	A01	Valid
3	30-OCT-2013 11:52 ZZZ509	0NCS015323F06XN	B01	Valid
4	30-OCT-2013 11:52 ZZZ509	5PCSV6628AE01Y0	A01	Valid

	DOMAIN	RUNID	RNINIT	RNTUBPOS	RNTESTCD	RNORRES
1	RN	04-SEP-2013 09:07 ZZZ501	clc	A01	Positive Control	Valid
2	RN	04-SEP-2013 09:07 ZZZ501	clc	B01	Negative Control	Valid
3	RN	30-OCT-2013 11:52 ZZZ509	clc	A01	Positive Control	Valid
4	RN	30-OCT-2013 11:52 ZZZ509	clc	B01	Negative Control	Valid

PATH FORWARD

SDTM+ has been implemented in one project at RMS and continuous improvements are being done. Future steps are to define ADAM standards for RMS.

CONCLUSION

With the evolving nature of the SDTM standards and new kind of data that is being analyzed in the IVD industry the route to form a standard for the industry has just started taking shape. Our first big challenge at RMS was to identify existing domains and define new ones. The second big challenge was then to figure out which variable goes into which domain. This mapping process had to be defined for various kinds of instrument data and study specific CRF data. The continued goal is to now improve the existing process and develop new tools to validate and help us with the ever changing dynamics of IVD data.

ACKNOWLEDGEMENTS

We would like to gratefully acknowledge the hard work of the SAS Programming Team for Roche Molecular Systems in Pleasanton, California. The SAS Programming Team also included: Mansi Singh, Smitha Krishnamurthy and Swarna Umesh. This project would not have been possible without their dedication and countless hours of work. The team who worked on these standards was initially led by Mario Widel who now works for Eli Lily. We are grateful for Mario's leadership in getting this project initiated.

REFERENCES

Smoak C, Singh M, Krishnamurthy S, Truong S. 2014. Forging New SDTM Standards for In-Vitro Diagnostic (IVD) Devices: A Use-Case. Proceedings of the Annual Conference of the Pharmaceutical Industry SAS User Group. San Diego, California, June 1-4, 2014.

Smoak C, Howard K, Wood F, Facile R. 2013. Data Standards Will Be Required: Challenges for Medical Device Submissions. Proceedings of the Annual Conference of the Pharmaceutical Industry SAS Users Group. Chicago, Illinois, May 12-15, 2013. http://www.lexjansen.com/pharmasug/2013/DS/PharmaSUG-2013-DS08.pdf

Smoak C, Wood F, Facile R, Howard K. 2012. Seven New SDTM Domains for Medical Devices. Proceedings of the Annual Conference of the Pharmaceutical Industry SAS Users Group, San Francisco, California, May 13-16, 2012. http://www.lexjansen.com/pharmasug/2012/DS/PharmaSUG-2012-DS05.pdf#!

CONTACT INFORMATION

Name:	Carey Smoak
Enterprise:	Roche Molecular Systems, Inc
Address:	4300 Hacienda Drive
City, State ZIP:	Pleasanton, CA 95134
Work Phone:	(925) 730 8033
Fax:	
E-mail:	carey.smoak@roche.com

Name:	Sofia Shamas
Enterprise:	MaxisIT Inc.
Address:	203 Main Street
City, State ZIP:	Metuchen, NJ 08840
Work Phone:	(732) 494 2005
Fax:	
E-mail:	sofia.shamas@contractors.roche.com

Name:	Chaitanya Chowdagam
Enterprise:	MaxisIT Inc.
Address:	203 Main Street
City, State ZIP:	Metuchen, NJ 08840
Work Phone:	(732) 494 2005
Fax:	
E-mail:	Chaitanya.chowdagam@contractors.roche.com

Name:	Dongkwan Lim
Enterprise:	Bluewave Consulting Inc.
Address:	1438 Menton Street
City, State ZIP:	Danville, CA 94506
Work Phone:	510-396-4089
Fax:	
E-mail:	bluewaveconsulting@gmail.com

Error! Hyperlink reference not valid.

Name:	Girish Rajeev
Enterprise:	InVentiv Health Clinical
Address:	1025 Greenwood Blvd, Suite 285
City, State ZIP:	Lake Mary, FL 32746
Work Phone:	1-407-548-0665
Fax:	1-888-513-4285
E-mail:	girish.rajeev@contractors.roche.com

TRADEMARKS

Appendix – SAS Code to Import Mapping Excel and create SAS conditions to map to SDTM+ Domains

```
*-----------------------------------------------------------
List of MACROs to derive standard SDTM Variables:
-----------------------------------------------------------;
*1. Macro For SDTM DATE ISO8601DT Format;
%macro isodate(mm= ,dd= ,yy= ,var=);
*Find the Month in Numerics for MDY Function;
if &mm ne ' ' then monn=month(input(cats(1,&mm,2011),date9.));
if monn ne . and &dd ne . and &yy ne . then
&var=put(mdy(monn,&dd.,&yy.),is8601da.);
%mend isodate;

*2. Macro For SDTM DATE"T"TIME ISO8601TM Format;
%macro isodatetime(mm= ,dd= ,yy= ,tm= ,var=);
  *Find the Month in Numerics for MDY Function;
  if &mm ne ' ' then monn=month(input(cats(1,&mm,2011),date9.));
  if monn ne . and &dd ne . and &yy ne . then
date=put(mdy(monn,&dd.,&yy.),is8601da.);
if &tm ne . then time=put(&tm.,E8601tm.);
if date ne ' ' and time ne ' ' then &var=trim(date)||"T"||Trim(time);
%mend isodatetime;

*-----------------------------------------------------------
MAPPING MACRO:
-----------------------------------------------------------;
%macro MAPPING(File= ,ESHEET= ,INDATA= ,SDTMOUT= );


*-------------------------------------------------------------------------
Read in the Excel file
*-------------------------------------------------------------------------;
     filename xcel "&FILE.";
     proc import out=mapxls(Drop=H) datafile=xcel dbms=xlsx replace;
          sheet="&ESHEET.";
          getnames=no;
          mixed=yes;
     run;

     data map;
          set mapxls(where=(N='Y'));
          call symput('n', left(put(_n_,3.)));
     run;
     %put Number of Variables Required for this domain = &n;

*-----------------------------------------------------------------
Read in the Excel file and create structure of SDTM Data from Standard
*-----------------------------------------------------------------;
     proc sql noprint;
          select C, D, E, L, J, M, K into
          :var1   - :var&n,
```

```
            :label1 - :label&n,
            :type1  - :type&n,
            :len1   - :len&n,
            :orig1  - :orig&n,
            :derive1 - :derive&n,
            :crf1    - :crf&n
            from map;
      quit;

      *Create Attrib Dataset based on Attribs given in the excel;
      proc sql noprint;
      create table attrib
            (%do i=1 %to &n;
                  %if &i ne &n %then
                        &&var&i &&type&i(&&len&i) label="&&label&i", ;
                  %else &&var&i &&type&i(&&len&i) label="&&label&i"  ;
            %end;
            );
      quit;

*-----------------------------------------------------------------
Create the final dataset with the mapping rules given in the Excel
*-----------------------------------------------------------------;
      data &SDTMOUT.;
            merge attrib &INDATA.;
            %do i=1 %to &n;
            %put &i = &&orig&i;
            %if %upcase(&&orig&i)=CRF or
                  %upcase(&&orig&i)=INSTRUMENT %then %do;
                        &&var&i=&&crf&i;
                  %end;
                  %else %if %upcase(&&orig&i)=DERIVED %then %do;
                        %put &&derive&i;
                        &&derive&i;
                  %end;
            %else %if %upcase(&&orig&i)=PGDERIVED %then %do;
                        %let var2=&&var&i..1;
                        %put &var2;
                        &&var&i=&var2;
                  %end;
                  keep &&var&i;
            %end;
      run;
%mend MAPPING;


*Example Call;
%MAPPING(File   = %str(\Metadata.xlsx),
                  ESHEET = DM,
                  INDATA = SRCDATA.DM,
                SDTMOUT = SDTM.SDTM_DM);
```

```
Macro Parameters:
FILE   : Metadata Excel File Path (i.e &tpath\Metadata.xlsx)
ESHEET : Metadata Excel Sheet Name (DM,SM,CH etc.,)
INDATA : This parameter can be Raw Data (Merge of All CRFs) or Post
Raw Data files(Instrument data)Note: Please Mention the Library for
the Rawdata eg:(Rawdata.test or Work.test)
SDTMOUT : SDTM Output into \Source Data Folder.
```