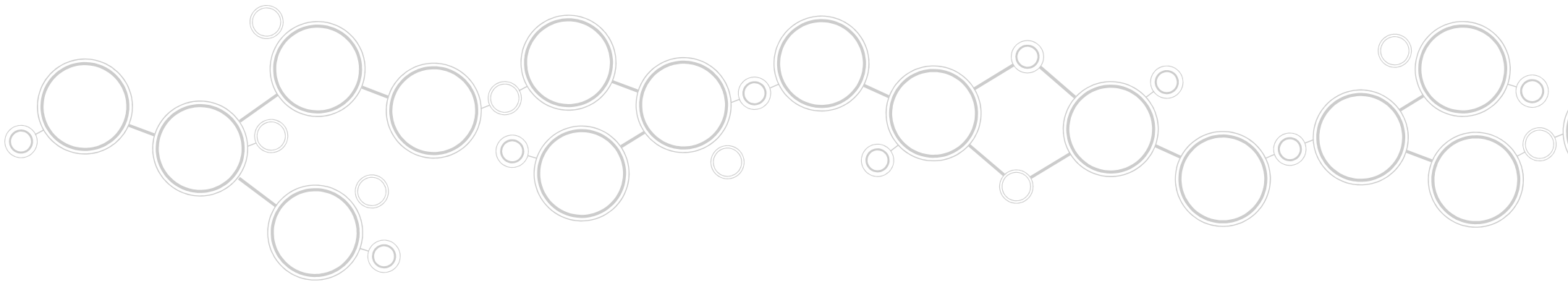




Much ADaM about Nothing – a PROC Away in a Day

Endri
Rowland Hale

PhUSE Conference
Brighton (UK), 9th - 12th October 2011



Introduction

CDISC SDTM

is the new data standard for clinical trials and submissions of clinical trial

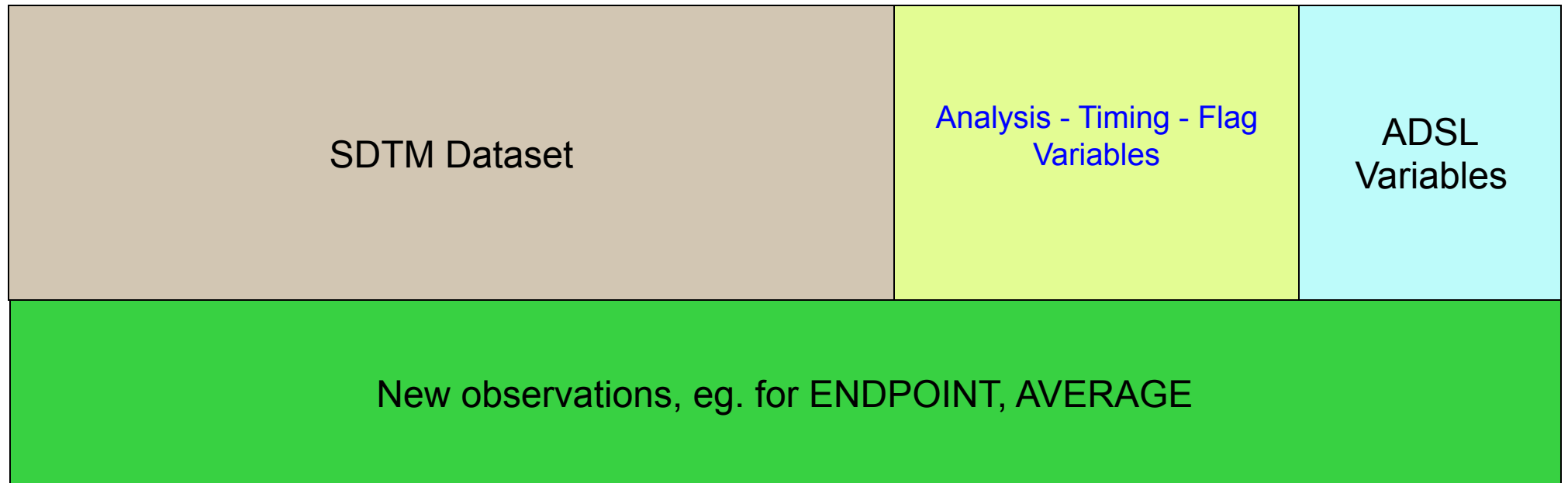
CDISC ADaM

(Analysis Data Model) datasets are an integral part of clinical study analysis

Can the creation of ADaM datasets from SDTM datasets be automated?

SDTM datasets → ADaM datasets

The ADaM datasets should be “Analysis-ready” or “One Statistical Procedure Away”.



General steps during the automation process

1. Import ADaM Metadata into SAS
2. Import functional core / macro libraries into SAS
3. Attach the functional core definitions to the ADaM metadata
4. The „Definition Step“
5. Macro %ADaM_Gen

Example : ADEG Metadata

Sort	Meta	Name	Label	Type	Length	Format	Informat	SourceTab	SourceVar	DerivedMeth	
100	Table	ADEG	Analysis Dataset for Electrocardiogram					SDTM.EG			
200	Var	STUDYID	Study Identifier	char	6						
300	Var	DOMAIN	Domain Abbreviation	char	2						
400	Var	USUBJID	Unique Subject Identifier	char	15						
500	Var	EGSEQ	Sequence Number	num	8						
600	Var	EGTESTCD	ECG Test or Examination Short Name	char	8						
700	Var	EGTEST	ECG Test or Examination Name	char	40						
800	Var	EGCAT	Category for ECG	char	30						
2100	Var	EGDTC	Date/Time of ECG	char	20						
2200	Var	EGENDTC	End Date/Time of ECG	char	20						
2300	Var	EGTPNUM	Planned Time Point Number	num	8						
2400	Var	EGTPT	Planned Time Point Name	char	50						
2500	Var	AVAL	Analysis Value	num	8			SDTM.EG	EGSTRESN		1st
2600	Var	CHG	Change from Baseline	num	8			DERIVED		AVAL - BASE	3rd
2700	Var	BASE	Baseline Value	num	8			SDTM.EG	EGSTRESN	Q: WHERE EGBLFL = 'Y'	2nd
2800	Var	AVAL2	Analysis Value - Converted from EGSTRES	num	8			SDTM.EG	EGSTRESC	#Char2Num	
2900	VAR	CHG1G	Change Group	char	8			DERIVED		IF (CHG/BASE) < 10 THEN chg1g = 1; ELSE chg1q=2;	

„Definition Step“:

1. AVAL , AVAL2

2. BASE

3. CHG and CHG1G

Example : ADEG Mapping Script

Macro Call:

%ADaM_Gen

(metadata = ADEG

, runall = N

, addobs = ENDPOINT # AVERAGE

);

„Definition Step“:

1st. AVAL , AVAL2

2nd. BASE

3rd. Merge BASE value

4th CHG and CHG1G

```
*****
* Program name      : ADEG_script.sas
* (This script is automatic generated by %ADaM_Gen
* Author           : Endri
* Date created      : 29.07.2011
* Study            : (Study number)
*                  : (Study title)
* Purpose          : ADEG - Analysis Dataset for Electrocardiogram
* Template         :
* Inputs           :
* Outputs          :
* Program completed : Yes/No
* Updated by       : (Name) - (Date):
*                  : (Modification and Reason)
*****/

/* Analysis Value # Analysis Value - Converted from EGSTRESC - SDTM.EG */
DATA adeg_010_eg;
  SET sdtm.eg;
  aval = egstresn;
  aval2 = %_help_c2n(var = egstresc);
RUN;

/* Baseline Value - SDTM.EG */
DATA adeg_020_base;
  SET sdtm.eg;
  base = egstresn;
  WHERE EGBLFL = 'Y';
RUN;

/* Join all */
%auto_join ( intab = adeg_010_eg
              $ adeg_020_base # base
              , outtab = adeg_30_all);

/* Change from Baseline # Change Group - DERIVED */
DATA adeg_40_chg;
  SET adeg_30_all;
  chg = AVAL - BASE;
  IF (CHG/BASE) < 10 THEN chg1g = 1;
  ELSE chg1g = 2;
RUN;

/* Adding derived observation - DERIVED*/
%calc( intab = adeg_40_chg
       , outtab = adeg_50_calc
       , calc = ENDPOINT # AVERAGE);

/* Assign label and data check */
%assign_attrib( metadata = ADEG
               , intab = adeg_50_calc);
```

Example : ADEG Mapping Script

Automatically created script :

- Auto header
- Auto comments
- Auto indent
- Auto join
- Standard layout

```
*****
* Program name      : ADEG_script.sas
* (This script is automatic generated by %ADaM_Gen
* Author           : Endri
* Date created      : 29.07.2011
* Study            : (Study number)
*                  : (Study title)
* Purpose          : ADEG - Analysis Dataset for Electrocardiogram
* Template         :
* Inputs           :
* Outputs          :
* Program completed : Yes/No
* Updated by       : (Name) - (Date):
*                  : (Modification and Reason)
*****/

/* Analysis Value # Analysis Value - Converted from EGSTRESC - SDTM.EG */
DATA adeg_010_eg;
  SET sdtm.eg;
  aval = egstresn;
  aval2 = %_help_c2n(var = egstresc);
RUN;

/* Baseline Value - SDTM.EG */
DATA adeg_020_base;
  SET sdtm.eg;
  base = egstresn;
  WHERE EGBLFL = 'Y';
RUN;

/* Join all */
%auto_join ( intab = adeg_010_eg
              $ adeg_020_base # base
              , outtab = adeg_30_all);

/* Change from Baseline # Change Group - DERIVED */
DATA adeg_40_chg;
  SET adeg_30_all;
  chg = AVAL - BASE;
  IF (CHG/BASE) < 10 THEN chglg = 1;
                      ELSE chglg = 2;
RUN;

/* Adding derived observation - DERIVED*/
%calc( intab = adeg_40_chg
      , outtab = adeg_50_calc
      , calc = ENDPOINT # AVERAGE);

/* Assign label and data check */
%assign_attrib( metadata = ADEG
               , intab = adeg_50_calc);
```

Functional Core – Processing

What happens when we need to :

- Sort the data first
- Carry out some prior calculation step

Simple example :

- TRT01SDT : Date of First Exposure in Period 01

Can we automate this too ?

Example : ADSL Metadata

Sort	Meta	Name	Label	Type	Length	Format	SourceTab	SourceVar	DerivedMeth
100	Table	ADSL	Subject Level Analysis Data				SDTM.DM		
200	Var	STUDYID	Study Identifier	char	6				
300	Var	USUBJID	Unique Subject Identifier	char	15				
400	Var	SITEID	Study Site Identifier	char	15				
500	Var	AGE	Age in AGEU at RFSTDTC	num	8				
600	Var	WEIGHT	Weight (kg)	num	8		SDTM.VS	VSSTRESN	Q: WHERE vstestcd = 'Weight' and visit = 'SCREENING'
700	Var	HEIGHT	Height (cm)	num	8		SDTM.VS	VSSTRESN	Q: WHERE vstestcd = 'Height' and visit = 'SCREENING'
800	Var	TRT01SDT	Date of First Exposure in Period 01	num	8	DATE9	SDTM.EX	EXSTDTC	#FirstDate

Functional Core

Example : Functional Core

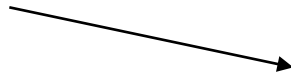
Name	Description	LibMacro	PreProcessing	Processing	PostProcessing
MinDate		%mindate(intab = ##INTAB## , outtab = ##OUTTAB## , outdate = ##NAME## , date = ##AUTO##);			
Char2Num		%_help_c2n(var = ##AUTO##)			
AutoJoin		%auto_join (intab = ##INTAB## , outtab = ##OUTTAB##);			
HelpSort				PROC SORT DATA = ##INTAB## OUT = ##OUTTAB##; BY USUBJID ##BY##;	
FirstDate				RUN;	
			#HelpSort	#MinDate	

Example : ADSL Mapping Script

Macro Call:

%ADaM_Gen
 (metadata = ADSL
 , runall = N
);

Functional Core :
#FirstDate



```
*****
* Program name       : ADSL_script.sas
* (This script is automatic generated by %ADaM_Gen
* Author            : Endri
* Date created       : 29.07.2011
* Study             : (Study number)
*                   : (Study title)
* Purpose           : ADSL - Subject Level Analysis Data
* Template          :
* Inputs            :
* Outputs           :
* Program completed : Yes/No
* Updated by        : (Name) - (Date):
*                   : (Modification and Reason)
*****/

/* SDTM.DM */
DATA adsl_010_dm;
    SET sdtm.dm (KEEP = studyid usubjid siteid age);
RUN;

/* Weight (kg) - SDTM.VS */
DATA adsl_020_weight;
    SET sdtm.vs;
    weight = vsstresn;
    WHERE vstested = 'WEIGHT' and visit = 'SCREENING';
RUN;

/* Height (cm) - SDTM.VS */
DATA adsl_030_height;
    SET sdtm.vs;
    height = vsstresn;
    WHERE vstested = 'HEIGHT' and visit = 'SCREENING';
RUN;

/* Date of First Exposure in Period 01 - SDTM.EX */
PROC SORT
    DATA = sdtm.ex
    OUT   = adsl_040_tr01sdt;
    BY    usubjid exstdtc;
RUN;

%mindate( intab    = adsl_040_tr01sdt
        , outtab   = adsl_041_mindate
        , outdate  = tr01sdt
        , date     = exstdtc);

/* Join all */
%auto_join ( intab =    adsl_010_dm
            $ adsl_020_weight # weight
            $ adsl_030_height # height
            $ adsl_041_mindate # tr01sdt
            , outtab = adsl_50_all);

/* Assign label and data check */
%assign_attrib( metadata = ADSL
              , intab    = adsl_50_all);
```

Example : Define population flag

SAP : Safety analysis set (SAF)

“A subject will be included in the SAF if he/she is randomized to a treatment group and has taken at least one unit of the study medication and has post-treatment safety data available.”

DerivedMeth:

A subject will be included in the SAF if :

- he/she is randomized to a treatment group `$DM # NOT MISSING(armcd) AND armcd NOT IN ("SCRNFAIL")$`
- `$AND$`
- has taken at least one unit of the study medication `$SDTM.EX # NOT MISSING(exstdtc)$`
- `$AND$`
- has post-treatment safety data available. `$SDTM.VS # VISIT NOT IN ("VISIT")$`

Name	Label	DerivedMeth
SAF	Safety Population Flag	A subject will be included in the SAF if : - he/she is randomized to a treatment group <code>\$DM # NOT MISSING(armcd) AND armcd NOT IN ("SCRNFAIL")\$</code> - <code>\$AND\$</code> - has taken at least one unit of the study medication <code>\$SDTM.EX # NOT MISSING(exstdtc)\$</code> - <code>\$AND\$</code> - has post-treatment safety data available. <code>\$SDTM.VS # VISIT NOT IN ("VISIT")\$</code> <code>#setflag</code>



Functional Core

Name	Description	LibMacro	PreProcessing	Processing	PostProcessing
MinDate		%mindate(intab = ##INTAB## , outtab = ##OUTTAB## , outdate = ##NAME## , date = ##AUTO##):			
Char2Num		%_help_c2n(var = ##AUTO##)			
AutoJoin		%auto_join (intab = ##INTAB## , outtab = ##OUTTAB##):			
HelpSort				PROC SORT DATA = ##INTAB## OUT = ##OUTTAB##; BY USUBJID ##BY##; RUN;	
FirstDate			#HelpSort	#MinDate	

- Overview of all macros
- Auto Text
- „Multi Processing“ to handle complex situation

Validation - Data Checks

Data checks could also be integrated in the metadata, such as ensuring values are within range or not null, or cross-checks with ADSL, etc.

Sort	Meta	Name	Label	ManualCheck
100	Table	ADEG	Analysis Dataset for Electrocardiogram	
200	Var	STUDYID	Study Identifier	
300	Var	DOMAIN	Domain Abbreviation	
400	Var	USUBJID	Unique Subject Identifier	DM.USUBJID
500	Var	EGSEQ	Sequence Number	
600	Var	EGTESTCD	ECG Test or Examination Short Name	
700	Var	EGTEST	ECG Test or Examination Name	
800	Var	EGCAT	Category for ECG	
2100	Var	EGDTC	Date/Time of ECG	
2200	Var	EGENDTC	End Date/Time of ECG	
2300	Var	EGTPTNUM	Planned Time Point Number	
2400	Var	EGTPT	Planned Time Point Name	
2500	Var	AVAL	Analysis Value	
2600	Var	CHG	Change from Baseline	
2700	Var	BASE	Baseline Value	
2800	Var	AVAL2	Analysis Value - Converted from EGSTRESC	
2900	VAR	CHG1G	Change Group	

Summary

- Can the creation of ADaM datasets be automated? Yes!
- Improve the efficiency and accuracy of ADaM dataset creation!
- Define ADaM dataset structure and variable derivation within Excel
- The Excel definitions drive the automatic generation of validation-ready SAS scripts which produce the ADaM datasets
- Incorporate data check reports to facilitate data cleaning by data management

Much ADaM about Nothing – a PROC Away in a Day

Questions ????

Authors : Endri (endri.endri@i3global.com)
Rowland Hale (rowland.hale@i3global.com)