

Getting your skeletons out of the closet

Parag Naithani
Ferring Pharmaceuticals, Copenhagen

Purpose

CDISC has laid strict Analysis Data Model Implementation Guideline (ADaMIG), specifying ADaM standard dataset structures and variables, including naming conventions, labels, types, formats etc. A programming solution is required to effectively adhere to the standard without the risk of making mistakes and duplication of work. This paper describes a method which programmatically uses the ADaM specification to shape the structure of the final ADaM datasets.

Solution

This data standardization technique is based on two utilities.

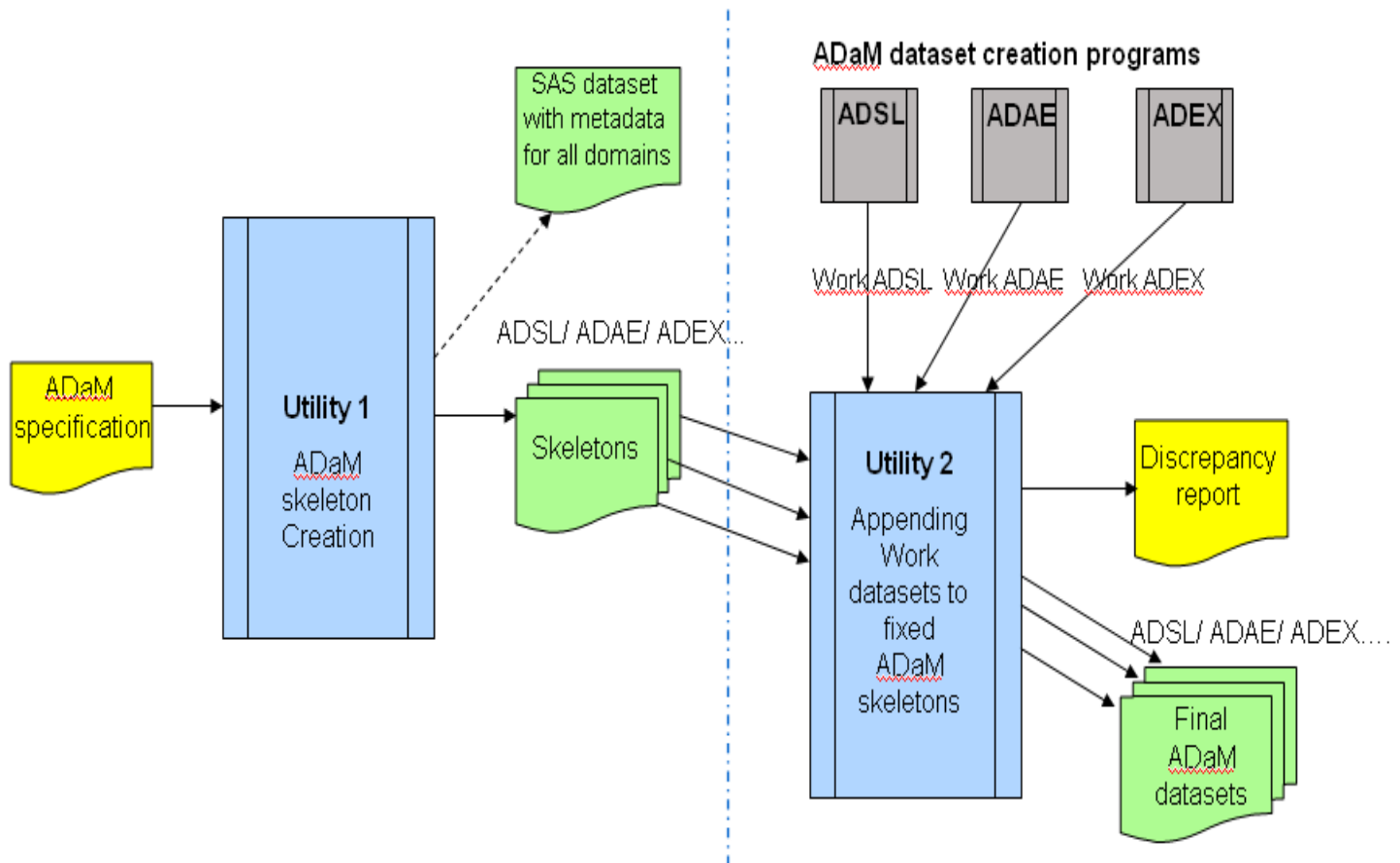
This Utility 1 produces empty SAS datasets (ADaM skeletons) in a STANDARD library from the Analysis Dataset Specification which is stored as a Microsoft Word file. This Utility also produces an excel sheet with the metadata information pertaining to the individual ADaM domain as well as the Mandatory variable section.

The Utility 2 appends last Work.dataset created during the development of the ADaM datasets for each domain to the respective ADaM skeleton created by Utility 1 to ensure adherence to the specification document. It also creates an HTML discrepancy report with clarification on error in structure and clarification on data existing in the skeleton, but not in the Work.dataset.

Prerequisites

Analysis dataset specification should exist in a Predefined location and follow a specific ADS template requirement. A STANDARD library should be defined for storing the outputs from the Utility. User should not interact with the computer while executing the macro since it is using the user's computer clipboard. Also the Utility macro must be used on a computer having MS Word and Excel installed.

Process Overview Diagram



Subject level ADaM dataset specification

4. Analysis Dataset Specifications

4.1 ADSL - Subject Level Analysis Dataset

Analysis Dataset Metadata

Dataset Name	Dataset Description	Dataset Location	Dataset Structure	Key variables of Dataset	Class of Dataset	Documentation
ADSL	Subject demographics	..././ADSL	One record per subject randomised	USUBJID	ADSL	SAP

Analysis Variable Metadata for ADSL

Dataset Name	Variable Name	Variable Label	Variable Type	Display Format	Codelist/ Controlled Terms	Source/ Derivation
ADSL	STUDYID	Study Identifier	Char	\$20		
ADSL	SUBJID	Subject Identifier for the Study	Char	\$20		
ADSL	USUBJID	Unique Subject Identifier	Char	\$30		
ADSL	SCREENNO	Screening Number	Char	\$20		
ADSL	FASFL	Full Analysis Set Population Flag	Char	\$1	Y, N	

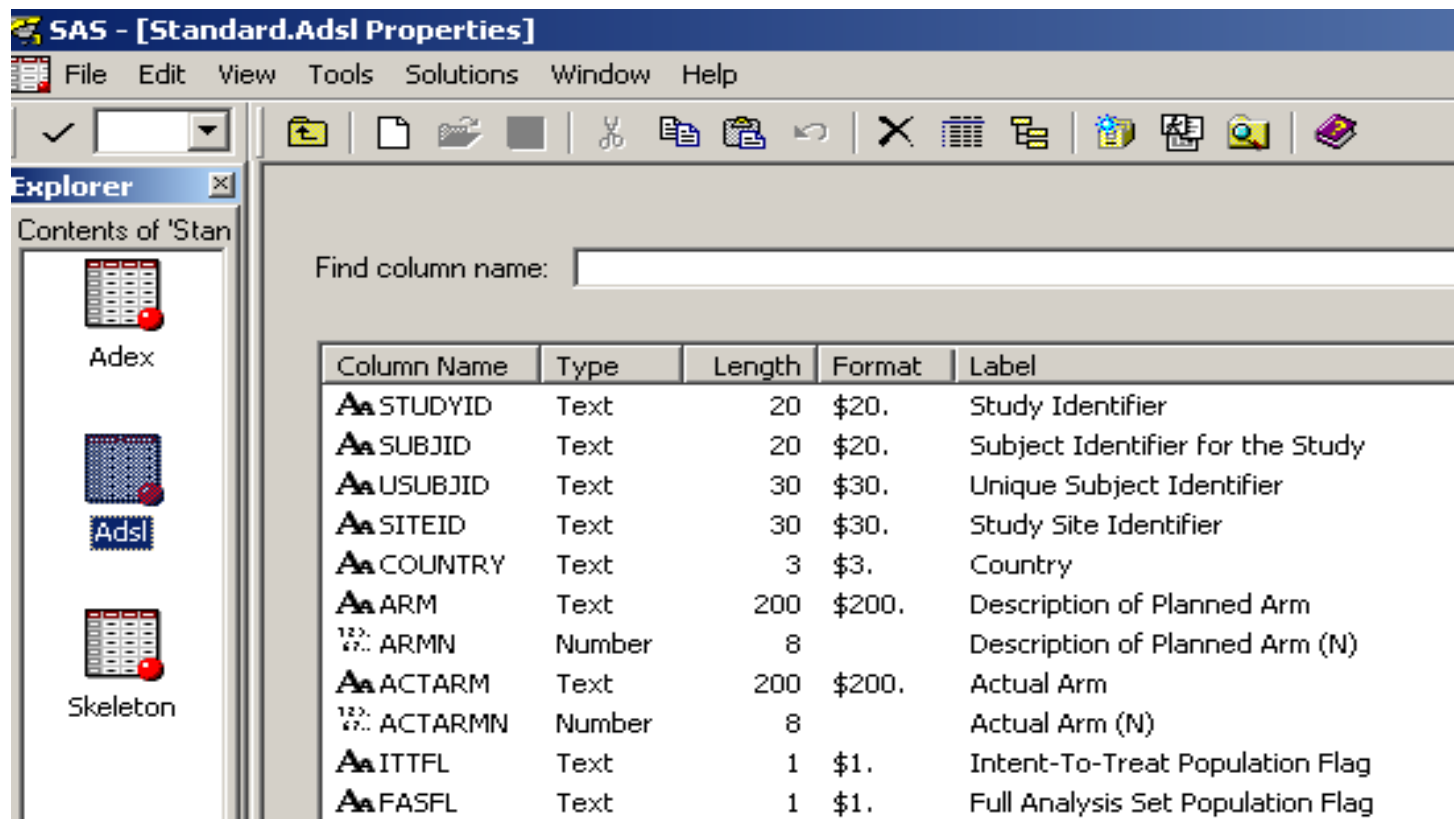
Mandatory variables section

3.5 Mandatory Variables

All datasets must include the following mandatory variables:

Variable Name	Variable Label	Variable Type	Display Format	Codelist/ Controlled Terms	Source/ Derivation
STUDYID	Study Identifier	Char	\$20		
SUBJID	Subject Identifier for the Study	Char	\$20		
USUBJID	Unique Subject Identifier	Char	\$30		
SITEID	Study Site Identifier	Char	\$30		
COUNTRY	Country	Char	\$3		
ARM	Description of Planned Arm	Char	\$200		
ARMN	Description of Planned Arm (N)	Num			
ACTARM	Actual Arm	Char	\$200		

Skeleton dataset ADSL properties



SAS - [Standard.Adsl Properties]

File Edit View Tools Solutions Window Help

Find column name:

Column Name	Type	Length	Format	Label
 STUDYID	Text	20	\$20.	Study Identifier
 SUBJID	Text	20	\$20.	Subject Identifier for the Study
 USUBJID	Text	30	\$30.	Unique Subject Identifier
 SITEID	Text	30	\$30.	Study Site Identifier
 COUNTRY	Text	3	\$3.	Country
 ARM	Text	200	\$200.	Description of Planned Arm
 ACTARM	Text	200	\$200.	Actual Arm
 ITTFL	Text	1	\$1.	Intent-To-Treat Population Flag
 FASFL	Text	1	\$1.	Full Analysis Set Population Flag

Utility 2 (Skeleton to ADaM)

Utility 2 is called at the end of each ADaM dataset creation program. This Utility appends last Work.dataset created in the program to the ADaM skeleton dataset (created by Utility 1) for the specific domain. It sorts the data according to sorting key available in the Analysis variable metadata (created by Utility 1). Also, it creates an discrepancy report

Calling Utility 2

```
* work.ADSL ;
proc sql;
  create table ADSL as
  select a.*, case
    when b.COMPLFL = '' then 'N'
    else b.COMPLFL
  end as COMPLFL, b.DISCDT, b.REAS, b.REASSPE, b.REASCAT,
    RFXSTDT          as TRTSDT format = date9. , RFXENDT) as TRTEDT format = date9.
  from _ADSL as a left join DS_OUT as b
  on a.USUBJID = b.USUBJID
  order by USUBJID
  ;
quit;

*UTILITY 2 ;
%Utility2(inds = WORK.ADSL,
          adam_outds = ADAM.ADSL);

proc datasets nolist;
  delete _: __: ds_out vs/ memtype = DATA;
run;
quit;
```

Last Work.dataset

SAS - [Work.Ads1]													
Tools Data Solutions Window Help													
	TMP	_TMP1_	I	WEIGHTBL	ITTFL	SITEID	AGEU	ASEX	ARACE	AETHNIC	STUDYID	Subject Identifier for the Study	USUBJID
1	12NOV2012	28JAN2013	1	63	Y	8131	YEARS	Female	Asian	Not Hispanic or Latino	FEXXXX 1234	8131F181	FEXXXX1234_8131F181
2	12OCT2012	06JAN2013	2	49.5	Y	8102	YEARS	Female	Asian	Not Hispanic or Latino	FEXXXX 1234	8102F007	FEXXXX1234_8102F007
3	09NOV2012	07FEB2013	3	50.1	Y	8103	YEARS	Female	Asian	Not Hispanic or Latino	FEXXXX 1234	8103F014	FEXXXX1234_8103F014
4	26OCT2012	15JAN2013	4	49.2	Y	8105	YEARS	Female	Asian	Not Hispanic or Latino	FEXXXX 1234	8105F025	FEXXXX1234_8105F025
5	28NOV2012	19FEB2013	5	47.3	Y	8124	YEARS	Female	Asian	Not Hispanic or Latino	FEXXXX 1234	8124F141	FEXXXX1234_8124F141

Work.dataset properties

SAS - [Work.Ads1 Properties]													
File View Tools Solutions Window Help													
General Details Columns Indexes Integrity Passwords													
Find column name:													
Column Name	Type	Length	Format	Label									
123: _TMP_	Number	8	DATE9.										
123: _TMP1_	Number	8	DATE9.										
123: I	Number	8											
123: WEIGHTBL	Number	8											
123: ITTFL	Text	1											
123: SITEID	Text	30											
123: AGEU	Text	8											
123: ASEX	Text	6											
123: ARACE	Text	50											
123: AETHNIC	Text	22											
123: STUDYID	Text	20											
123: SUBJID	Text	20											
123: USUBJID	Text	30											

Subject Identifier for the Study

Final ADSL dataset

LE: Adam.Adsl]

Tools Data Solutions Window Help



	Study Identifier	Subject Identifier for the Study	Unique Subject Identifier	Study Site Identifi	Country	Description of Planned Arm	Description of Planned Arm (N)	Actual Arm	Actual Arm (N)	Intent-To-Treat Population Flag
1	FEXXXX 1234	8101F001	FEXXXX1234_8101F001	8101	JPN	Trt A	3	Trt A	3	Y
2	FEXXXX 1234	8101F002	FEXXXX1234_8101F002	8101	JPN	Trt B	1	Trt B	1	Y
3	FEXXXX 1234	8101F003	FEXXXX1234_8101F003	8101	JPN	Trt A	3	Trt A	3	Y
4	FEXXXX 1234	8101F004	FEXXXX1234_8101F004	8101	JPN	Trt C	2	Trt C	2	Y
5	FEXXXX 1234	8101F005	FEXXXX1234_8101F005	8101	JPN	Trt B	1	Trt B	1	Y
6	FEXXXX 1234	8101F487	FEXXXX1234_8101F487	8101	JPN	Trt C	2	Trt C	2	Y
7	FEXXXX 1234	8102F007	FEXXXX1234_8102F007	8102	JPN	Trt C	2	Trt C	2	Y
8	FEXXXX 1234	8102F008	FEXXXX1234_8102F008	8102	JPN	Trt A	3	Trt A	3	Y
9	FEXXXX 1234	8102F009	FEXXXX1234_8102F009	8102	JPN	Trt B	1	Trt B	1	Y
10	FEXXXX 1234	8102F010	FEXXXX1234_8102F010	8102	JPN	Trt B	1	Trt B	1	Y
11	FEXXXX 1234	8102F011	FEXXXX1234_8102F011	8102	JPN	Trt A	3	Trt A	3	Y

Final ADSL properties

SAS - [Adam.Adsl Properties]

File View Tools Solutions Window Help

✓ []

Explorer

Contents of 'Adam.Adsl'

Adsl

Find column name: []

Column Name	Type	Length	Format	Label
STUDYID	Text	20	\$20.	Study Identifier
SUBJID	Text	20	\$20.	Subject Identifier for the Study
USUBJID	Text	30	\$30.	Unique Subject Identifier
SITEID	Text	30	\$30.	Study Site Identifier
COUNTRY	Text	3	\$3.	Country
ARM	Text	200	\$200.	Description of Planned Arm
ARMN	Number	8		Description of Planned Arm (N)
ACTARM	Text	200	\$200.	Actual Arm
ACTARMN	Number	8		Actual Arm (N)
ITTFL	Text	1	\$1.	Intent-To-Treat Population Flag
FASFL	Text	1	\$1.	Full Analysis Set Population Flag
SAFFL	Text	1	\$1.	Safety Population Flag

Discrepancy report

Utility 2 creates a HTML report if there are any structural mismatches in the work dataset and the corresponding Skeleton dataset. The following errors are handled:

- Miss-match in variable type
- Length of the variable in the work dataset exceeds that of the corresponding variable in the skeleton

The Final ADaM dataset will NOT be created as these errors are critical in nature.

- Variables existing in the Skeleton dataset but not in the work dataset.

This can occur during the development phase, but not for the final data. In this case, the ADaM dataset will be created.

Discrepancy report

Output data adam.adae not created due to the following critical errors in data structure

	Skeleton		Input data		
	standard.adae		work.adae_m		
Variable	Type	Length	Type	Length	Structure Error
AESEV	C	8	C	10	Length of variable in input data exceed that of the skeleton. Variable content might be truncated. Please confirm and the specification

***Variable(s) listed below was not found in the input data work.adae_m.
Do add it/them to your input data, or remove it/them from the specification.***

Skeleton		
standard.adae		
Variable	Type	Length
STUDYID	C	20
USUBJID	C	30

Utilities run sequence

Utility 1 is run only when there is a change in the ADaM specification document. The change could if a new variable is added to any domain or a complete new domain is added to the specification. Utility 2 is run every time the ADaM dataset creation program is run.

Conclusion

Programmatically using the ADaM specification document can improve the programming accuracy and efficiency by reducing the programming time, maintenance time and validation time. There is no need for the programmer to remember which variables to keep or drop. The sequence of the variables in the Final ADaM dataset could be easily governed as the Final ADaM dataset will show the variables in the same sequence as specified in the ADaM specification. Using this technique it will be easier to maintain standards within and across different trials, thus making the data submission ready.

Since all the variables and their attributes are listed in a SAS dataset containing the Analysis variable metadata it is easy to check for inconsistencies and make necessary corrections. Also, this utility could be used by the sponsor for checking the structural inconsistencies in the ADaM datasets delivered by the CRO.

Acknowledgement

I would like to thank all my colleagues at Ferring Group who shared their experiences, patiently answered all my queries, read through the drafts and gave me constructive feedbacks.

Contact Information

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

Parag Naithani parag.naithani@ferring.com

Ferring International PharmaScience Center (IPC)
Kay Fiskers Plads 11,
DK-2300 Copenhagen