



Generating Define.xml v2.0 and Analysis Result Metadata
using ADaM Specifications, ADaM Datasets and TFL
Annotation

- Introduction
- Define.xml
 - Content Overview
 - Pre-requirements
 - Process
- ARM.xml
 - Content Overview
 - Pre-requirements
 - Process
- Validation
- General Information

- Case Report Tabulation Data Definition Specification format in XML format or Define.xml
- Required document when doing CDISC based electronic submission
- Provides information to help understand the collected data (SDTM) and data used for analysis (ADaM)

- Analysis Results Metadata provides information about the purpose of analyses
- Analysis Results Metadata provides traceability from analysis results to the data in the analysis datasets
- Data flow in form of analysis datasets and variables used, and analysis methods

Define.xml

- Define.xml provides metadata for
 - Datasets:
 - Name, Description, Class, Structure, Purpose, Keys, Location
Documentation
 - Variables:
 - Name, Label, Type, Length/Display Format, Controlled Terms
Source/Derivation/Comment
 - Parameter Value Level Metadata
 - Controlled Terminology
 - Analysis Derivations
 - Comments
 - External Documentation (ADRG)

- Define Stylesheet (define2-0-0.xls) for define.xml v2.0
OR
define.css, define1-0-0.xls and define-extension.xsd for
define.xml v1.0
- ADaM Datasets
 - Trimmed ADaM Datasets
 - SAS Transport Files using Trimmed ADaM Datasets
- ADaM ADS Specifications in required format

- Documentation
 - ADRG
 - SAS Codes of ADaM Datasets
 - Other documentation that is being referred in ADS Specifications

- ADaM Datasets
 - Trimming of ADaM Datasets
 - Character variable length adjusted as per data values
 - Use ads_trimming.sas to trim ADaM Datasets
 - Trimmed ADS
 - CSV file with names of ADS > 5 GB
 - Datasets > 5 GB will be divided into two or more smaller datasets with size ≤ 5 GB.
 - Generate SAS Transport Files using PROC COPY
- ADaM ADS Specifications
 - Formatting of ADaM ADS Specifications

Process – ADaM ADS Specifications

- Study Information:

Parameter	Description
StudyName	111-P-222
StudyDescription	Define for Study XYZ
ProtocolName	Define for Study XYZ
StandardName	ADaM-IG
StandardVersion	1.0
Language	en

Process – ADaM ADS Specifications

- Analysis Metadata:

Dataset	Description	Structure	Purpose	Class	KeyVariables
ADSL	Subject Level Analysis	One record per screened subject	Analysis	SUBJECT LEVEL ANALYSIS DATASET	USUBJID
ADVS	Vital Signs	One record per subject per parameter per visit	Analysis	BASIC DATA STRUCTURE	USUBJID,PARAMCD,AVISITN

Repeating	Documentation	Document
No	Refer to Analysis data reviewer's guide for further information	LF.ADRG; LF.ADSL.PGM;
Yes	SAP, Protocol	LF.ADVS.PGM;

Process – ADaM ADS Specifications

- Variable Metadata:

Dataset	Variable	Label	Type	Length	Sigdigit	Key	Format	CodeList
ADSL	STUDYID	Study Identifier	text	15				
ADSL	SITEID	Study Site Identifier	text	25				
ADSL	INVID	Investigator Identifier	text	25				
ADSL	INVNAM	Investigator Name	text	200				
ADSL	USUBJID	Unique Subject Identifier	text	25		1		
ADSL	SUBJID	Subject Identifier for the Study	text	25				
ADSL	RANDNO	Randomization Number	integer	8				
ADSL	BRTHDTC	Date/Time of Birth	datetime	19				
ADSL	AGE	Age	integer	8				
ADSL	AGEU	Age Units	text	6				AGEU

Process – ADaM ADS Specifications

- Variable Metadata:

Origin	Source	ValueList	Mandatory	FormalExpression	Document
Predecessor	DM.STUDYID		Yes		
Predecessor	DM.SITEID		Yes		
Predecessor	DM.INVID		Yes		
Predecessor	DM.INVNAM		No		
Predecessor	DM.USUBJID		Yes		
Predecessor	DM.SUBJID		No		
Derived	DS.DSREFID where DS.DSSCAT="RANDOMIZATION"		No		
Predecessor	DM.BRTHDTC		No		
Predecessor	DM.AGE		Yes		
Predecessor	DM.AGEU		Yes		

Process – ADaM ADS Specifications

- ValueLevel Metadata:

Dataset	Variable	Label	Type	Length	Sigdigit	Format	Origin	PARAMCD	Source	Document
ADVS	PARAM	Parameter	text	200			Derived	VSALL	PARAM= trim(left(VS.VSTEST))	
ADVS	PARAM	Parameter	text	200			Derived	BMI	PARAM= trim(left(VS.VSTEST))	
ADVS	PARAM	Parameter	text	200			Derived	DIABP	PARAM= trim(left(VS.VSTEST)) " (" trim(left(VS.VSSTRESU)) ")"	
ADVS	PARAM	Parameter	text	200			Derived	HEIGHT	PARAM= trim(left(VS.VSTEST)) " (" trim(left(VS.VSSTRESU)) ")"	
ADVS	PARAM	Parameter	text	200			Derived	PULSE	PARAM= trim(left(VS.VSTEST)) " (" trim(left(VS.VSSTRESU)) ")"	
ADVS	PARAM	Parameter	text	200			Derived	SYSBP	PARAM= trim(left(VS.VSTEST)) " (" trim(left(VS.VSSTRESU)) ")"	
ADVS	PARAM	Parameter	text	200			Derived	WEIGHT	PARAM= trim(left(VS.VSTEST)) " (" trim(left(VS.VSSTRESU)) ")"	
ADVS	PARAMN	Parameter (N)	integer	8			Derived	SYSBP	if PARAMCD="SYSBP" then PARAMN=1;	
ADVS	PARAMN	Parameter (N)	integer	8			Derived	DIABP	if PARAMCD="DIABP" then PARAMN=2;	
ADVS	PARAMN	Parameter (N)	integer	8			Derived	PULSE	if PARAMCD="PULSE" then PARAMN=3;	
ADVS	PARAMN	Parameter (N)	integer	8			Derived	HEIGHT	if PARAMCD="HEIGHT" then PARAMN=4;	
ADVS	PARAMN	Parameter (N)	integer	8			Derived	WEIGHT	if PARAMCD="WEIGHT" then PARAMN=5;	
ADVS	PARAMN	Parameter (N)	integer	8			Derived	BMI	if PARAMCD="BMI" then PARAMN=6	
ADVS	PARAMN	Parameter (N)	integer	8			Derived	VSALL	if PARAMCD="VSALL" then PARAMN=100;	

Process – ADaM ADS Specifications

- Controlled Terminology:

Codelist	Name	Datatype	Code	Decode	Rank	OrderNumber	CodeListNum	Codenum
AGEU	Age Unit	text	YEARS			1	C66781	C29848
SEX	Sex	text	M			1	C66731	C20197
SEX	Sex	text	F			2	C66731	C16576
RACE	Race	text	AMERICAN INDIAN OR ALASKA NATIVE			1	C74457	C41259
RACE	Race	text	ASIAN			2	C74457	C41260
RACE	Race	text	BLACK OR AFRICAN AMERICAN			3	C74457	C16352
RACE	Race	text	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER			4	C74457	C41219
RACE	Race	text	WHITE			5	C74457	C41261
AVISIT	AVISIT	text	Screening			1		
AVISIT	AVISIT	text	Day -2 to -1			2		
AVISIT	AVISIT	text	Day 1 to 3			3		
AVISIT	AVISIT	text	Day 8			4		
AVISITN	AVISITN	integer	1	Screening		1		
AVISITN	AVISITN	integer	2	Day -2 to -1		2		
AVISITN	AVISITN	integer	3	Day 1 to 3		3		
AVISITN	AVISITN	integer	4	Day 8		4		
PARAM_ADVS	PARAM_ADVS	text	All Vital Signs Data			7		
PARAM_ADVS	PARAM_ADVS	text	Body Mass Index			6		
PARAM_ADVS	PARAM_ADVS	text	Diastolic Blood Pressure (mmHg)			2		
PARAM_ADVS	PARAM_ADVS	text	Height (cm)			4		
PARAM_ADVS	PARAM_ADVS	text	Pulse Rate (BEATS/MIN)			3		
PARAM_ADVS	PARAM_ADVS	text	Systolic Blood Pressure (mmHg)			1		
PARAM_ADVS	PARAM_ADVS	text	Weight (kg)			5		

ARM.xml

- A standard template of TLF mocks annotation is needed with below tabs.
- Analysis
- WhereClause
- Documents

- Mocks
- Analysis Procedures defined by the statistician
- ADaM datasets

A	B	C	D	E	F	G	H	I	J	K	L	M
Table Number	Table Title	CSR Ref	Analysis	Metric	Reason	Purpose	Dataset	Analysis Variable	PARAMCD	Description	SAP	Code
1	Table 14.2.11.2.1	Statistical Analysis (Repeated Measures Model) of Change from Baseline in Mean Daily Average Indoxa Pain While Walking (24h recall) by Week up to Week 4 - Primary Analysis	LF Table 14.2.11.2.1/4	Difference between treatment by week	SPECIFIED IN SAP	PRIMARY OUTCOME MEASURE	ADKNEE	CHG	Y	Adjusted least squares mean	LF.SAP-SEC-10.4/2	PROC MIXED DATA=avg_pain; CLASS WEEK TRT USUBJID COUNTRY; MODEL CHG_BASE = TRT WEEK BASE_VAL TRT*WEEK COUNTRY / DDFM=KR; REPEATED WEEK/TYPE=UNSUBJECT-USUBJID; LSMEANS TRT*WEEK / PDIF CL; ESTIMATE "XXX100mg_qw_Placebo" TRT=1/0 / CL ALPHA=0.05; ESTIMATE "XXX200mg_qw_Placebo" TRT=1/0 / CL ALPHA=0.05; ESTIMATE "XXX200mg_qw_XXX100mg" TRT=0-1/1 / CL ALPHA=0.05; ODS OUTPUT LSMEANS=lm DIFFS=diff ESTIMATES=est; RUN;
2	Table 14.2.11.2.1	Statistical Analysis (Repeated Measures Model) of Change from Baseline in Mean Daily Average Indoxa Pain While Walking (24h recall) by Week up to Week 4 - Primary Analysis	LF Table 14.2.11.2.1/4	Difference between treatment by week	SPECIFIED IN SAP	PRIMARY OUTCOME MEASURE	ADKNEE	CHG	Y	Adjusted least squares mean	LF.SAP-SEC-10.4/2	PROC MIXED DATA=avg_pain; CLASS WEEK TRT USUBJID COUNTRY; MODEL CHG_BASE = TRT WEEK BASE_VAL TRT*WEEK COUNTRY / DDFM=KR; REPEATED WEEK/TYPE=UNSUBJECT-USUBJID; LSMEANS TRT*WEEK / PDIF CL; ESTIMATE "XXX100mg_qw_Placebo" TRT=1/0 / CL ALPHA=0.05; ESTIMATE "XXX200mg_qw_Placebo" TRT=1/0 / CL ALPHA=0.05; ESTIMATE "XXX200mg_qw_XXX100mg" TRT=0-1/1 / CL ALPHA=0.05; ODS OUTPUT LSMEANS=lm DIFFS=diff ESTIMATES=est; RUN;
3	Table 14-5.02	Incidence of Treatment Emergent Serious Adverse Event by Treatment Group	Incidence of Treatment Emergent Serious Adverse Event by Treatment Group	Get denominators for percentage from ADSL and counts and numerators from ADAE. Join ADAE with ADSL based on the unique subject identifier (USUBJID) keeping only records in ADAE for the numerator.	SPECIFIED IN SAP	PRIMARY OUTCOME MEASURE	ADSL			Unique count of subjects that experienced an Adverse Event by Preferred Term, System Organ Class, and Treatment Group and percentage based on the number of subjects in the safety population within each treatment group. The total number of times an event occurred was recorded by Preferred Term, System Organ Class, and Treatment Group. Fisher's exact test was used for treatment comparison of event rates.	LF Table 14-5.02.rsr	
4	Table 14-5.02	Incidence of Treatment Emergent Serious Adverse Event by Treatment Group	Incidence of Treatment Emergent Serious Adverse Event by Treatment Group	Get denominators for percentage from ADSL and counts and numerators from ADAE. Join ADAE with ADSL based on the unique subject identifier (USUBJID) keeping only records in ADAE for the numerator.	SPECIFIED IN SAP	PRIMARY OUTCOME MEASURE	ADAE	AEBODSYS AEDECOD		Unique count of subjects that experienced an Adverse Event by Preferred Term, System Organ Class, and Treatment Group and percentage based on the number of subjects in the safety population within each treatment group. The total number of times an event occurred was recorded by Preferred Term, System Organ Class, and Treatment Group. Fisher's exact test was used for treatment comparison of event rates.	LF Table 14-5.02.rsr	

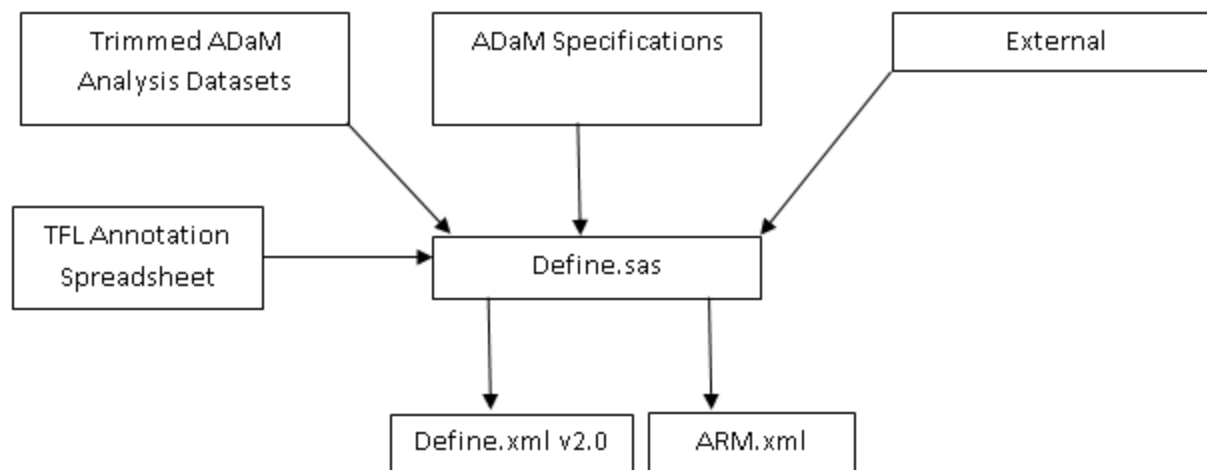
•WhereClause

Table Number	Analysis	Dataset	Variable	Value
Table 14.2.1.1.2.1	Difference between treatment by week	ADKNEE	PARAMCD	MIKPW1
Table 14.2.1.1.2.1	Difference between treatment by week	ADKNEE	FASFL	Y
Table 14.2.1.1.2.1	Difference between treatment by week	ADKNEE	AVISITN	3
Table 14.2.1.1.2.1	Difference between treatment by week	ADKNEE	AVISITN	4
Table 14.2.1.1.2.1	Difference between treatment by week	ADKNEE	AVISITN	5
Table 14.2.1.1.2.1	Difference between treatment by over 4 weeks	ADKNEE	PARAMCD	MIKPW1
Table 14.2.1.1.2.1	Difference between treatment by over 4 weeks	ADKNEE	FASFL	Y
Table 14.2.1.1.2.1	Difference between treatment by over 4 weeks	ADKNEE	AVISITN	5
Table 14.2.1.1.2.1	Difference between treatment by over 4 weeks	ADKNEE	AVISITN	6
Table 14.2.1.1.2.1	Difference between treatment by over 4 weeks	ADKNEE	AVISITN	7
Table 14-5.02	Incidence of Treatment Emergent Serious Adverse Events by Treatment Group	ADSL	SAFFL	Y
Table 14-5.02	Incidence of Treatment Emergent Serious Adverse Events by Treatment Group	ADAE	TEAE	Y
Table 14-5.02	Incidence of Treatment Emergent Serious Adverse Events by Treatment Group	ADAE	AESER	Y

•Documents

ID	Title	href
LF.SAP-SEC-10.4	SAP Section 10.4	sap.pdf
LF.Table 14.2.1.1.2.1	Table 14.2.1.1.2.1	csr.pdf
LF.Table 14-5.02.sas	Table 14-5.02.sas	at14-5-02-sas.txt

- Run define.sas tool to generate the documents



- Validate define.xml
 - Run Pinnacle 21
 - All the issues/errors/warnings identified should be resolved prior to final package or a valid explanation should be provided explaining why these errors can't be resolved.
 - Check define.xml Source/Derivation/Comment section
 - Check ALL the external documentation hyper-links
 - Perform all the checks given in ADaM Electronic Submission Material Checklist

U:\phuse_boston\define.xml Study 111-P-222, Data Defi... x

File Edit View Favorites Tools Help

Convert Select

ADaM-IG 1.0

Date of document generation: 2015-08-20T09:22:12

Stylesheet version: 2013-04-24

- Analysis Data Reviewers Guide
- Analysis Datasets
- Parameter Value Level Metada
- Controlled Terminology
- Analysis Derivations
- Comments

Analysis Datasets for Study 111-P-222 (ADaM-IG 1.0)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
ADSL	Subject Level Analysis	SUBJECT LEVEL ANALYSIS DATASET	One record per screened subject	Analysis	USUBJID	adsl.xpt	SAP, Protocol d0_adsl.sas
ADVS	Vital Signs	BASIC DATA STRUCTURE	One record per subject per parameter per visit	Analysis	USUBJID, PARAMCD, AVISITN	adv.sas	SAP, Protocol d1_adv.sas

Go to the [top](#) of the define.xml

Subject Level Analysis (ADSL) [Location: [adsl.xpt](#)]

Variable	Label	Type	Length / Display Format	Controlled Terms or Format	Source/Derivation/Comment
STUDYID	Study Identifier	text	8		Predecessor: DM.STUDYID
SITEID	Study Site Identifier	text	2		Predecessor: DM.SITEID
INVID	Investigator Identifier	text	2		Predecessor: DM.INVID
INVNAM	Investigator Name	text	23		Predecessor: DM.INVNAM
USUBJID	Unique Subject Identifier	text	15		Predecessor: DM.USUBJID
SUBJID	Subject Identifier for the Study	text	6		Predecessor: DM.SUBJID

Table 14.2.1.1.2.1

Display	Table 14.2.1.1.2.1 Statistical Analysis (Repeated Measures Model) of Change from Baseline in Mean Daily Average Index Knee Pain While Walking (24h recall) by Week up to Week 4 - Primary Analysis
Analysis Result	Change from Baseline in Mean Daily Average Index Knee Pain While Walking
Analysis Parameter(s)	PARAMCD = "MIKPW1"
Analysis Variable(s)	CHG (Change from Baseline)
Analysis Reason	SPECIFIED IN SAP
Analysis Purpose	PRIMARY OUTCOME MEASURE
Data References (incl. Selection Criteria)	ADKNEE [PARAMCD = "MIKPW1" and AVISITN IN (4, 5, 6, 7) and FASFL = "Y"]
Documentation	Adjusted least squares mean SAP Section 10
Programming Statements	[SAS version 9.4] <pre> PROC MIXED DATA=adknee; CLASS avisitn trt01pn usubjid country ; MODEL chg = trt01pn avisitn base trt01pn*avisitn country / DDFM=KR; REPEATED avisitn/TYPE=UN SUBJECT= USUBJID; LSMEANS trt01pn*avisitn / PDIFF CL ; LSMEANS trt01pn / PDIFF CL; ESTIMATE "100 mg vs. Placebo" trt01pn -1 1 0 / CL ALPHA=0.05; ESTIMATE "200 mg vs. Placebo" trt01pn -1 0 1 / CL ALPHA=0.05; ESTIMATE "200 mg vs. 100 mg" trt01pn 0 -1 1 / CL ALPHA=0.05; ODS OUTPUT LSMEANS=lsm DIFFS=diff ESTIMATES=esti ; RUN; </pre>

- This data driven approach using ADaM Specifications, ADaM Datasets and TFL Annotation is a time-saving
- Straightforward and easy way of generating Define.xml v2.0 and ARM.xml
- Programmers don't need to have XML knowledge for generating Define.xml or ARM.xml via this method.

- Location of Reference Material:
 - <http://cdisc.org/define-xml>
 - <https://www.pinnacle21.com/>

Thank You!

Authors:

Mark Crangle

ICON Clinical Research

Mark.Crangle@iconplc.com

Hima Bhatia

ICON Clinical Research

Hima.Bhatia@iconplc.com