



Automate Analysis Results Metadata in the Define-XML v2.0

Hong Qi, Majdoub Haloui, Larry Wu, Gregory T Golm

Merck & Co., Inc., USA

Topics

- Introduction
- Analysis Results Metadata (ARM) Version 1.0
 - What is ARM?
 - Why ARM?
 - ARM Requirements
- Automate ARM Generation in Define
 - Current Process
 - Enhanced Process
- Conclusion

Introduction (1)

- Compliance with regulatory requirements
 - Successful new drug submission
 - Additional effort at each step of the process
- Understanding the regulatory requirements
 - Streamline the process
- Good planning of submission package in the early stage
 - Maintain high quality

Introduction (Cont.)

- Challenge – gathering the details for the ARM of Define-XML from reports and programs at the last minute
 - Efficiency
 - Quality
- Focus of this presentation
 - ARM requirements
 - Effective approach

What is ARM?

- A document that identifies the critical analyses
- Provides traceability from results in a statistical display to the analysis data used to create it
 - Document the analyses performed
 - Provide the regulatory reviewer with the means to trace results back to their source documents (programs, datasets, and statistical analysis plans)

What is ARM? (Cont.)

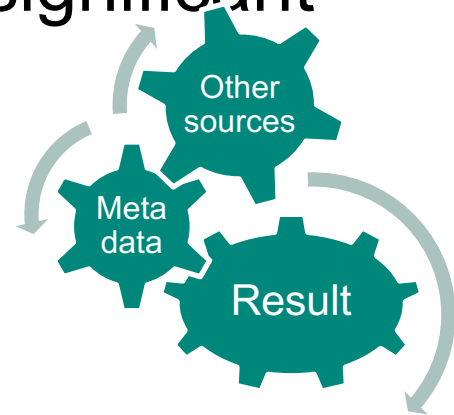
- Purpose
 - Submission to regulatory agencies
 - United States Food and Drug Administration (FDA)
 - Japan Pharmaceutical and Medical Devices Agency (PMDA)
 - Exchange of analysis datasets and key results between other parties

What is ARM? (Cont.)

- CDISC ADaM Analysis Data Model Version 2.1
- Analysis Results Metadata Specification Version 1.0 for Define-XML Version 2
 - Released in January 2015
 - <https://www.Cdisc.Org/standards/foundational/adam>

Why ARM?

- Provides a fundamental principle of ADaM – traceability
- Describes key components of significant analysis results
 - Text and links
- Save review time and expedite the drug approval
 - Identifying the critical analyses
 - Understanding and reproducing analysis results



Why ARM? (Cont.)

- A highly recommended ADaM metadata component
 - PMDA (Japan) is already asking for it ([TCG](#) section 4.1.2.1 Definition documents of datasets page 23)
 - FDA is interested
- Benefits
 - Provide ARM to assist the review
 - Critical analyses
 - Links between results, documentation, datasets
 - Documenting the analyses performed
 - Has a potential to save significant time for the regulatory reviewer and expedite the drug approval

Why ARM? (Cont.)

- Sponsor's decision on which analyses to be displayed
 - ARM focus should be on primary and key secondary analysis results

ARM Requirement

Table 1 Analysis Result Metadata fields

Metadata Field	Definition	Source / Metadata Example
DISPLAY IDENTIFIER	A unique identifier for the specific analysis display (such as a table or figure number)	Table/Figure number Table 14-3.01
DISPLAY NAME	Title of display, including additional information if needed to describe and identify the display (e.g., analysis population)	Table/Figure title Primary Endpoint Analysis: ADAS-Cog (11) - Change from Baseline to Week 24 – LOCF
RESULT IDENTIFIER	Identifies the specific analysis result within a display when there are multiple analysis results on a display.	Table/Figure Analysis of dose response
PARAM	The analysis parameter in the BDS analysis dataset that is the focus of the analysis result. Does not apply if the result is not based on a BDS analysis dataset.	Define.xml variable section ADAS-Cog (11) Total Score
PARAMCD	Corresponds to PARAM in the BDS analysis dataset. Does not apply if the result is not based on a BDS analysis dataset.	Define.xml variable section ACTOT11

ARM Requirement – Metadata Fields

Table 1 Analysis Result Metadata fields (Cont.)

Metadata Field	Definition	Source / Metadata Example
ANALYSIS VARIABLE	The analysis variable being analyzed	Define.xml variable section CHG
REASON	The rationale for performing this analysis. It indicates when the analysis was planned and the purpose of the analysis.	Protocol and/or SAP Primary efficacy analysis as pre-specified in protocol
DATASET	The name of the dataset(s) used to generate the analysis result.	Define.xml Dataset section ADQSADAS
SELECTION CRITERIA	Specific and sufficient selection criteria for analysis subset and / or numerator.	Protocol and/or SAP ITTFL='Y' and AVISIT='Week 24' and PARAMCD='ACTOT11'
DOCUMENTATION	Textual description of the analysis performed.	Protocol and/or SAP Linear model analysis of dose response; Refer to Figure 2 for details. SAP Section 10.1.1.
PROGRAMMING STATEMENTS	The software programming code used to perform the specific analysis including the model statement, and all technical specifications needed for reproducing the analysis	Study SAS ® programs [SAS version 9.2] Refer to Figure 2

ARM Requirement – Example of ARM

Analysis Results Metadata (Summary) for Study CDISC-Sample

[Table 14-3.01](#) Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)

[Dose response analysis for ADAS-Cog changes from baseline](#)

[Pairwise comparisons to placebo for ADAS-Cog changes from baseline](#)

Figure 1 Analysis Results Metadata (Summary)

[Table 14-5.02](#) Incidence of Treatment Emergent Serious Adverse Events by Treatment Group

[Incidence of Treatment Emergent Serious Adverse Events by Treatment Group](#)

Figure 1 Analysis Results Metadata (Summary)

ARM Requirement – Example of ARM

Analysis Results Metadata (Detail) for Study CDISC-Sample

Table 14-3.01

Display	Table 14-3.01 Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)
Analysis Result	Dose response analysis for ADAS-Cog changes from baseline
Analysis Parameter(s)	PARAMCD = "ACTOT" (Adas-Cog(11) Subscore)
Analysis Variable(s)	CHG (Change from Baseline)
Analysis Reason	SPECIFIED IN SAP
Analysis Purpose	PRIMARY OUTCOME MEASURE
Data References (incl. Selection Criteria)	ADQSADAS [PARAMCD = "ACTOT" and AVISIT = "Week 24" and EFFFL = "Y" and ANL01FL = "Y"]
Documentation	Linear model analysis of CHG for dose response; using randomized dose (0 for placebo; 54 for low dose; 81 for high dose) and site group in model. Used PROC GLM in SAS to produce p-value (from Type III SS for treatment dose). SAP Section 10.1.1
Programming Statements	[SAS version 9.2] proc glm data = ADQSADAS; where EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD="ACTOT"; class SITEGR1; model CHG = TRTPN SITEGR1; run;

Figure 2 Analysis Results Metadata (Detail)

ARM Requirement – Statistical Display in CSR

Protocol: CD18CPILOT01
Population: Efficacy

ADQSADAS (PARAMCD = "ACTOT" and AVISIT = "Week 24" and EFFFLL = "Y" and ANL01FL = "Y")

Table 14 – 3.01
Primary Endpoint Analysis: ADAS Cog (11) – Change from Baseline to Week 24 – LOCF

Analysis variable(s): **CHG** (Change from Baseline)

	Placebo (N=79)	Xanomeline Low Dose (N=81)	Xanomeline High Dose (N=74)
Baseline			
N	79	81	74
Mean (SD)	24.1 (12.19)	24.4 (12.92)	21.3 (11.74)
Median (Range)	21.0 (5;61)	21.0 (5;57)	18.0 (3;57)
Week 24			
N	79	81	74
Mean (SD)	26.7 (13.79)	26.4 (13.18)	22.8 (12.48)
Median (Range)	24.0 (5;62)	25.0 (6;62)	20.0 (3;62)
Change from Baseline			
N	79	81	74
Mean (SD)	2.5 (5.80)	2.0 (5.55)	1.5 (4.26)
Median (Range)	2.0 (-11;16)	2.0 (-11;17)	1.0 (-7;13)
p-value (Dose Response) [1] [2]			0.245
p-value (Xan - Placebo) [1] [3]		0.569	0.233
Diff of LS Means (SE)		-0.5 (0.82)	-1.0 (0.84)
95% CI		(-2.1;1.1)	(-2.7;0.7)
p-value (Xan High – Xan Low) [1] [3]			0.520
Diff of LS Means (SE)			-0.5 (0.84)
95% CI			(-2.2;1.1)

(1) Based on Analysis of covariance (ANCOVA) model with treatment and site group as factors and baseline value as a covariate.
(2) Test for a non-zero coefficient for treatment (dose) as a continuous variable
(3) Pairwise comparison with treatment as a categorical variable: p-value without adjustment for multiple comparisons.

Figure 3 Example of an analysis table in CSR with some of the metadata annotations

Automate ARM Generation in Define

- Current Process
 - Identify the information needed for ARM when preparing the submission package
 - External pdf file: analysis-results-metadata.pdf
 - Linked to the define.xml

ADaM-IG 1.0

Analysis Reviewer's Guide
Analysis-Result-Metadata
▶ Analysis Datasets
▶ Parameter Value Level Metadata
▶ Controlled Terminology
▶ Analysis Derivations
▶ Comments

Standard ADaM-IG 1.0
Study Name P000V00MK0000
Study Description MK0000-000: A Clinical Trial of Pembrolizumab (MK-0000) Evaluating Predictive Biomarkers in Subjects with Advanced Solid Tumors (KEYNOTE 000)
Protocol Name P000V00MK0000

Date of Define-XML document generation: 2018-04-27T02:29:55Z
Stylesheet version: 2015-01-16

Analysis Datasets for Study P000V00MK0000 (ADaM-IG 1.0)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
ADSL	Subject-Level Analysis Dataset	SUBJECT LEVEL ANALYSIS DATASET	One record per subject	Analysis	USUBJID	adsl.xpt	
ADDILI	DILI Reporting Analysis Dataset	BASIC DATA STRUCTURE	One observation per subject, parameter, and analysis time point.	Analysis	USUBJID, PARAMCD, AEPOCH, ADT	addili.xpt	

Automate ARM Generation in Define (Cont.)

- Enhanced Process
 - Goal: an efficient submission process with high quality
 - Strategies
 - Proactively document the ARM components when the reports are being designed or generated
 - Automate the ARM generation as an extension to Define.xml using Pinnacle 21 Enterprise tool

Automate ARM Generation in Define (Cont.)

- Enhanced Process (Cont.)
 - Tool: <ARM0template.xlsm>, a macro-enabled spreadsheet
 - Study report tabs:
 - ✓ Standards reports, therapeutic area reports
 - ✓ With all the ARM components documented ahead
 - ✓ Study specific information added for the selected reports
 - Pinnacle 21 Enterprise required tabs
 - ✓ Macro-enabled spreadsheet tabs that select and combine information from the study report tabs

Automate ARM Generation in Define (Cont.)

- <ARM0template.xlsm> Sample

	A	B	C	D	E	F	G	H	I
1	Program	Submit (Y/N)	TLF No.	TLF Title	TLF Sub-title	Description	Variables	Variables 2 (Additional - Protocol Specific)	Analysis Reason
2	s0exposure.txt		Table 2.7.4	Summary of Drug Exposure	(Cohort A: xxx Carcinoma) (MK0000) (ASaT Population)	Summary of drug exposure	ADEXSUM.AVAL, ADEXSUM.PARAM CD	ADSL.SAFFL, ADSL.PART01N	therapy (months), and number of drug administration
3	asr0baseline0characteristics.txt		Table 2.7.4	Subject Characteristics	(Cohort A: xxx Carcinoma) (MK0000) (ASaT Population)	Subject characteristics	AGE (Age), RACEGR1 (Race), ETHNIC (Ethnicity),	Performance Status at Screening), DMSTG	characteristics, demographic characteristics

ARM0template.xlsm - Not Classified - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View SAS

I4 X ✓ fx Safety analysis

Content ARM Standards ARM Oncology - Efficacy Properties Analysis Displays Analysis Results

Classification

Not Classified

Edit

140%

Study report tabs

Pinnacle 21 Enterprise required tabs

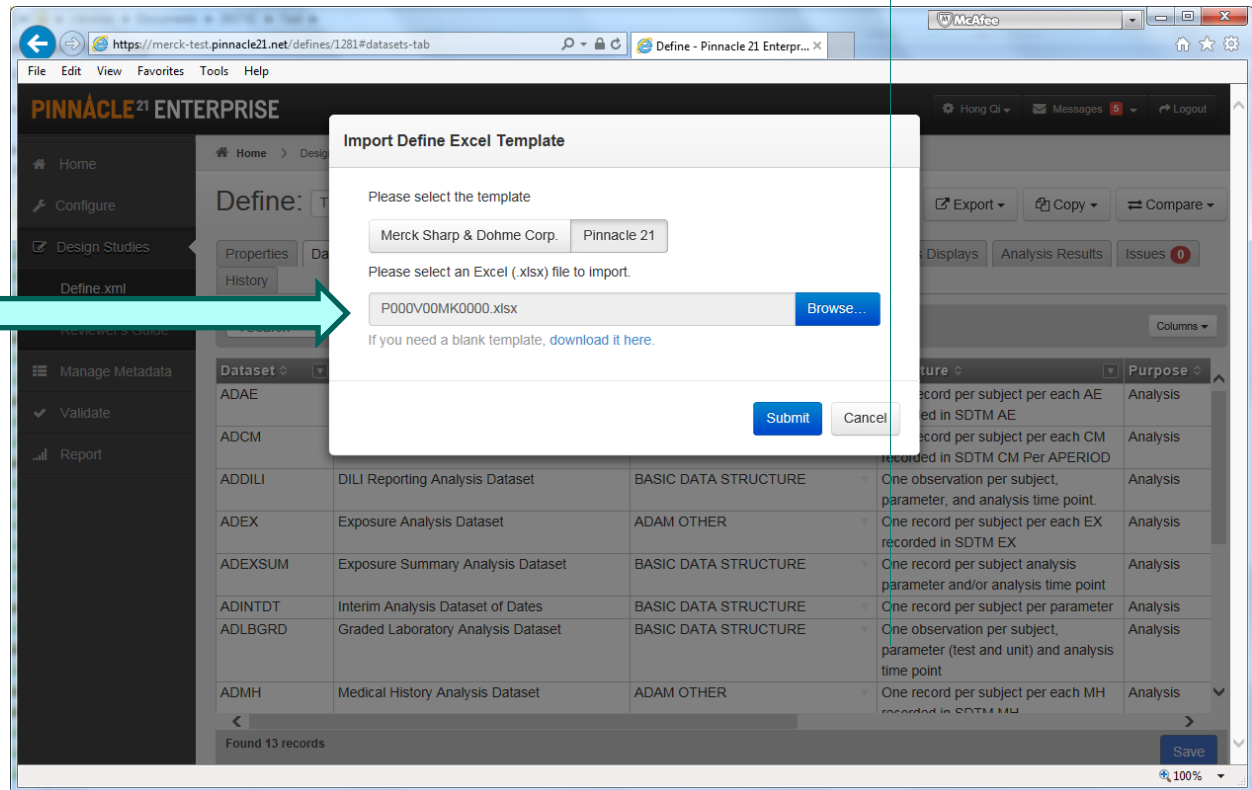
Automate ARM Generation in Define (Cont.)

- Pinnacle 21 Enterprise required tabs in ARM0Template.xlsm, copied to the ADaM data specification, as the input file for Pinnacle 21 Enterprise

ARM0Template.xlsm

Tabs: Analysis Displays
Analysis Results

P000V00MK0000.xlsx
(Data Specification)



Automate ARM Generation in Define (Cont.)

- ARM generated by Pinnacle 21 Enterprise as part of Define.xml

The screenshot shows a web browser window displaying the Define-XML document for Study P000V00MK0000. The document is titled "ADaM-IG 1.0" and includes a sidebar with navigation links. The main content area displays the "Analysis Results Metadata (Summary) for Study P000V00MK0000" and the "Analysis Results Metadata (Detail) for Study P000V00MK0000".

Analysis Results Metadata (Summary) for Study P000V00MK0000

Table	Description
Table 10-2	Subject Characteristics (ASaT Population)
Table 11-1	Summary of Best Overall Response Based on IWG per Central Review (MASaT Population)

Analysis Results Metadata (Detail) for Study P000V00MK0000

Table 10-3

Display	Table 10-3 Subject Characteristics (ASaT Population)
Analysis Result	Subject Characteristics for all subjects as treated population
Analysis Parameter(s)	
Analysis Variable(s)	RACEGR1 (Pooled Race Group 1) REGION (Region)
Analysis Reason	Baseline analysis
Analysis Purpose	Report subject baseline and disease characteristics for all treated subjects (safety population)
Data References (Incl. Selection Criteria)	ADSL [TRTFL = "Y"]
Documentation	Display baseline variables from ADSL
Programming Statements	[SAS Version 9.4], #0baseline0char0asatgen.txt For each analysis variable yyyy, proc sql; select distinct yyyy, count(distinct usubjid) as n from xx.adsl where adsl.trtfl="Y" group by yyyy; quit; Note: xx is the libname where analysis datasets are residing

Conclusion

- It is critical to fully understand the CDISC requirements and proactively plan the data collection for ARM in order to produce the ARM with high quality and cycle time reduction.
- Establishing an ARM template file to collect and store ARM components for both standards reports and therapeutic area specific reports ensures the effective implementation and governance of CDISC data standards across studies.

Acknowledgement

- Ellen Asam
- Mary Varughese
- Sapan Patel
- Amy Gillespie