PhUSE US Connect 2018

Paper DS11

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

Hong Qi, Majdoub Haloui, Larry Wu, Gregory T Golm
Merck & Co., Inc., Upper Gwynedd, PA, USA

ABSTRACT

In the current FDA regulated-industry, Define-XML is an essential component of a CDISC data submission analysis package. Analysis Results Metadata (ARM) v-1.0 added to the Define-XML v2.0 supports the review of key efficacy and safety analysis results and outlines their relation to the submitted clinical trial data in CDISC ADaM format. The components of the ARM are normally gathered from SAS® programs and reports, and documented as the submission package is being prepared, which could impact the efficiency and quality. In this paper, the authors propose an effective approach to a proactive planning for information collection, storage, and display using the Pinnacle 21 Enterprise Define tool. This approach not only helps the team to reduce the cycle time for the submission package preparation but also delivers a high quality define document for the Analysis package.

INTRODUCTION

Compliance with the CDISC data standards is critical for a successful new drug submission; however, it will require additional effort at each step of the process from the ADaM datasets and report development to the submission package preparation. Understanding the regulatory requirements will help streamline the process, and a good planning of submission package in the early stage of ADaM datasets and reports design will help maintain a high quality of the submission documents.

One of the challenges we encountered when working on the submission document was gathering the details from reports and programs in order to complete the ARM of Define-XML at the last minute. It is inefficient for every project team to go through common programs, or protocol-specific programs to identify the information needed for the ARM in different submissions. It is error prone for non-developers of the programs to identify the key programming statements from SAS code at the end of the process, especially from complex programs.

The focus of this paper is to describe the ARM requirements in the updated CDISC ADaM Analysis Data Model Version 2.1 document, and propose an effective approach to proactively plan for information collection, storage, and display using the Pinnacle 21 Enterprise Define tool.

ANALYSIS RESULTS METADATA VERSION 1.0

WHAT IS ARM?

ARM stands for Analysis Results Metadata. It is a specification that describes the Analysis Data Model (ADaM) ARM extension to the Define-XML 2.0 model for the purpose of submission to regulatory agencies such as the United States Food and Drug Administration (FDA), Pharmaceuticals and Medical Devices Agency (PMDA) as well as for the exchange of analysis datasets and key results between other parties. This Analysis Results Metadata extension is based on the metadata model as described in the CDISC ADaM Analysis Data Model Version 2.1 document.

It provides the metadata needed for traceability from a result in a statistical display to the supporting data in the analysis datasets. Version 1.0 of the ARM specification was released in January 2015 and is available at https://www.cdisc.org/standards/foundational/adam. It supports the submission of ARM as defined in ADaM v2.1.

WHY ARM?

ARM provides a fundamental principle of ADaM, i.e., traceability, which enables the regulatory reviewers to understand the data flow from collection to SDTM and from ADaM to Analysis Results.

ARM describes key components of significant analysis results, providing both text and links from a result to metadata and other sources describing the analysis. It has the potential to save a significant amount of time for regulatory reviewers and thus speed up the drug approval.

ARM is an optional ADaM metadata component according to the ADaM version 2.1 document. Currently, PMDA requires ARM in the submission package, while FDA is interested in receiving it. As a best practice, the ARM should be provided to assist reviewers by identifying the critical analyses, providing links between results, documentation, and datasets, and documenting the analyses performed. However, ARM is not necessarily required or even advisable

for *every analysis* in a clinical study report (CSR) or submission. The sponsor determines which analyses should be displayed in ARM.

ARM REQUIREMENT

ARM describes the major attributes of a specified analysis result found in a CSR or submission. The metadata fields to be used to describe an analysis result, along with the metadata example for dose response analysis in Figure 1 and 2, are listed in Table 1.0.

TABLE 1 ANALYSIS RESULTS 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
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 (e.g., "Pre-specified in Protocol," "Pre-specified in SAP," "Data Driven," "Requested by Regulatory Agency") and the purpose of the analysis within the body of evidence (e.g., "Primary Efficacy," "Key Secondary Efficacy," "Safety").	Protocol and/or SAP Primary efficacy analysis as pre-specified in protocol
DATASET	The name of the dataset used to generate the analysis result. In most cases, this is a single dataset. However, if multiple datasets are used, they are all listed here.	Define.xml Dataset section ADQSADAS
SELECTION CRITERIA	Specific and sufficient selection criteria for analysis subset and / or numerator – a complete list of the variables and their values used to identify the records selected for the analysis.	Protocol and/or SAP ITTFL='Y' and AVISIT='Week 24' and PARAMCD='ACTOT11'
DOCUMENTATION	Textual description of the analysis performed.	Protocol and/or SAP SAP Section 10.1.1. Linear model analysis of dose response for the ADAS-Cog(11) total score change from baseline at Week 24 - missing values imputed using LOCF, Efficacy population. Used PROC GLM in SAS to produce p-value (from Type III SS for treatment dose); Independent terms in model are TRTDOSE (randomized dose: 0 for placebo; 54 for low dose; 81 for high dose) SITEGR1 (site group, as a class variable) and BASE (baseline ADAS-Cog score).

Metadata Field	Definition	Source / Metadata Example

PROGRAMMING STATEMENTS	The software programming code used to perform the specific analysis. This includes, for example, the model statement (using the specific variable names) and all technical specifications needed for reproducing the analysis (e.g., covariance structure).	Study SAS ® programs
		[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;

Example of a rendition of ARM metadata using a stylesheet (The resource that transforms the XML into navigable HTML) is provided in Figure 1 and Figure 2.

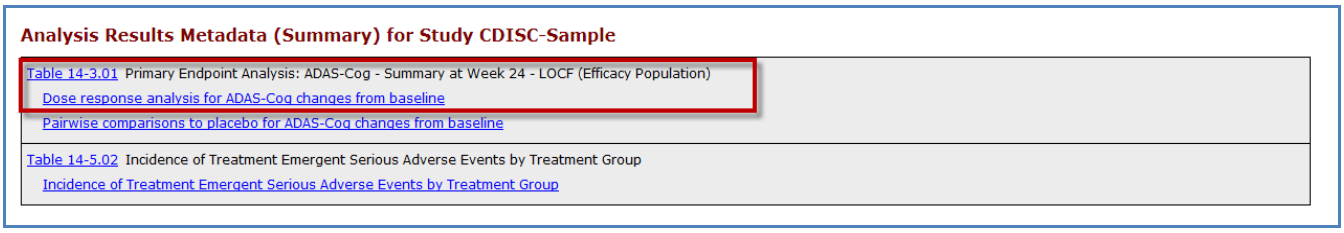


Figure 1 Analysis Results Metadata (Summary)

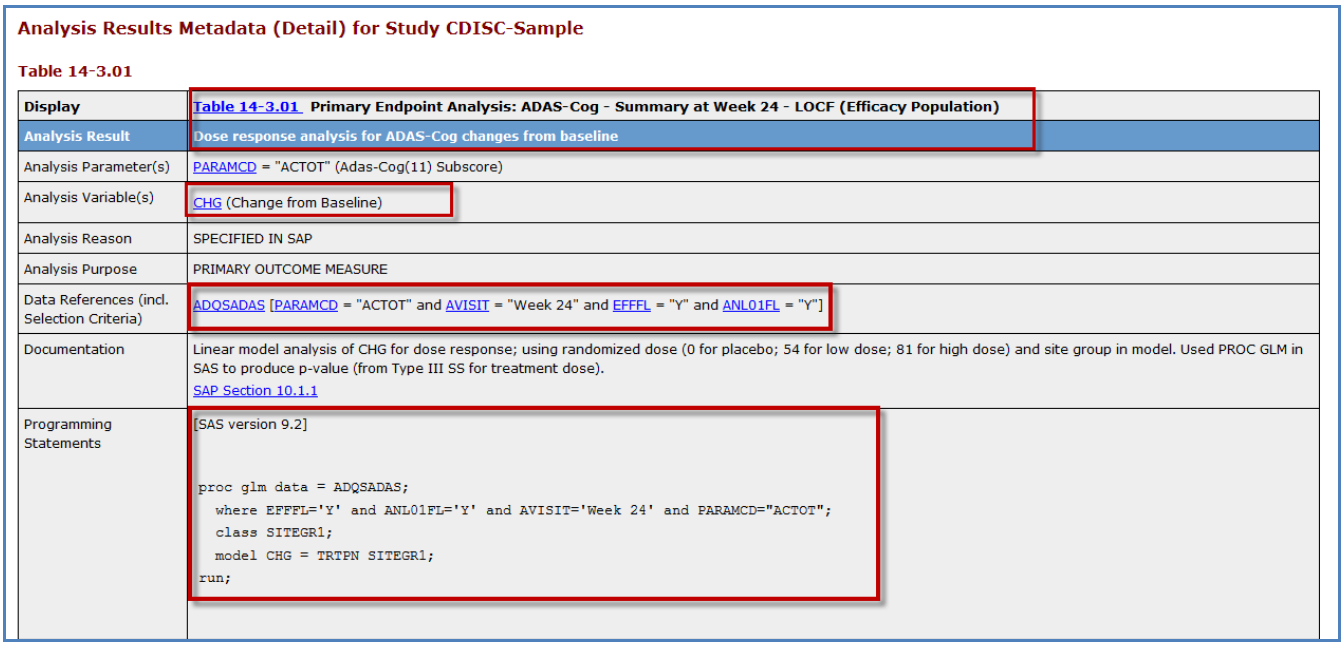
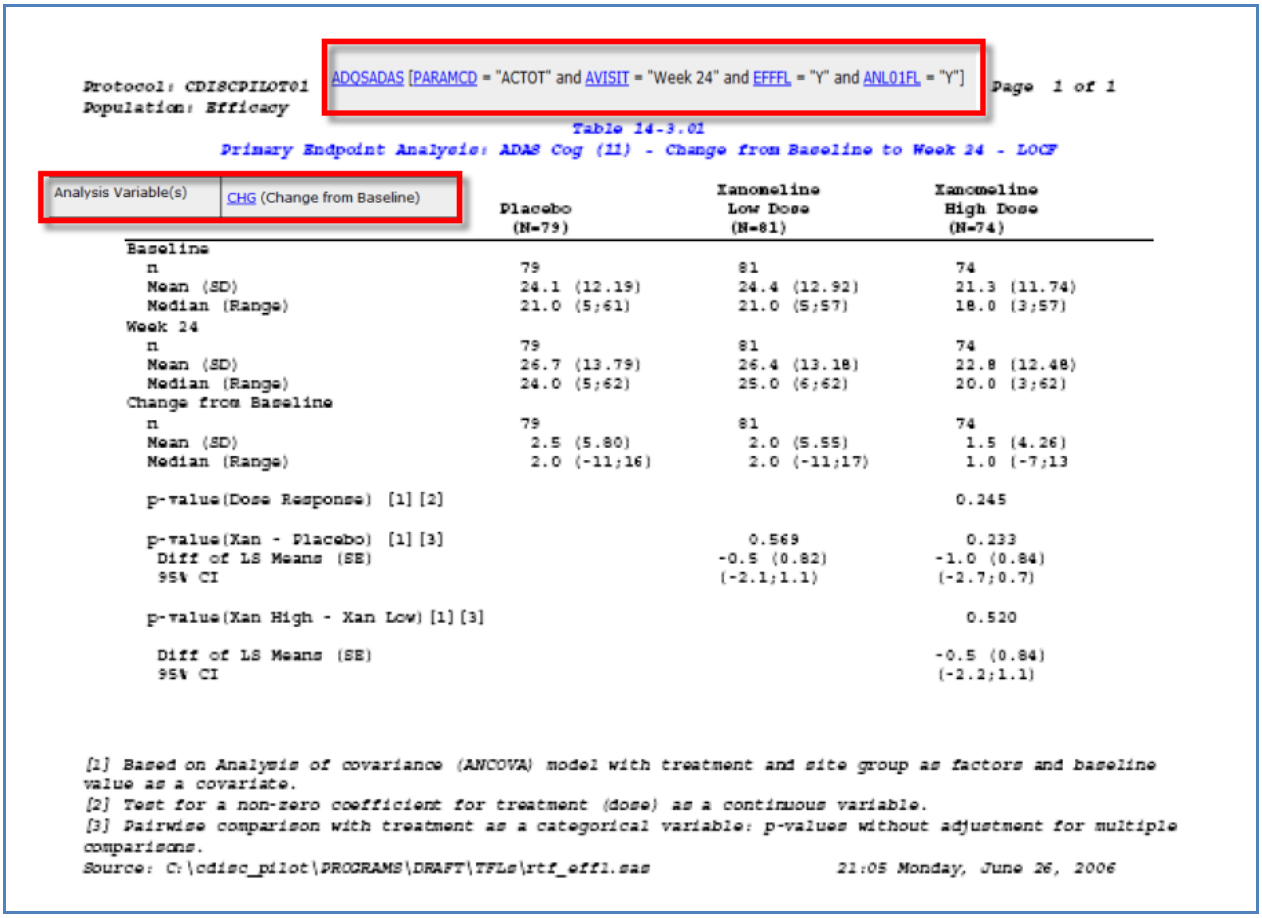


Figure 2 Analysis Results Metadata (Detail)

Figure 3 shows an example of the statistical display in the CSR as the source for Figure 1 and Figure 2.



Protocol: CDISCPILOT01 ADQSADAS [PARAMCD = "ACTOT" and AVISIT = "Week 24" and EFFFL = "Y" and ANL01FL = "Y"] Page 1 of 1
Population: Efficacy

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-values without adjustment for multiple comparisons.
Source: C:\cdisc_pilot\PROGRAMS\DRAFT\TFLs\rtf_eff1.sas 21:05 Monday, June 26, 2006

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

AUTOMATE ARM GENERATION IN DEFINE

CURRENT PROCESS

In our current process, information in the ARM is identified from the reports and programs when the submission package is prepared, and ARM is manually generated as an external PDF file, analysis-results-metadata.pdf, and linked to define.xml generated by Pinnacle 21 Enterprise as shown in Figure 4. This is time consuming and error prone.

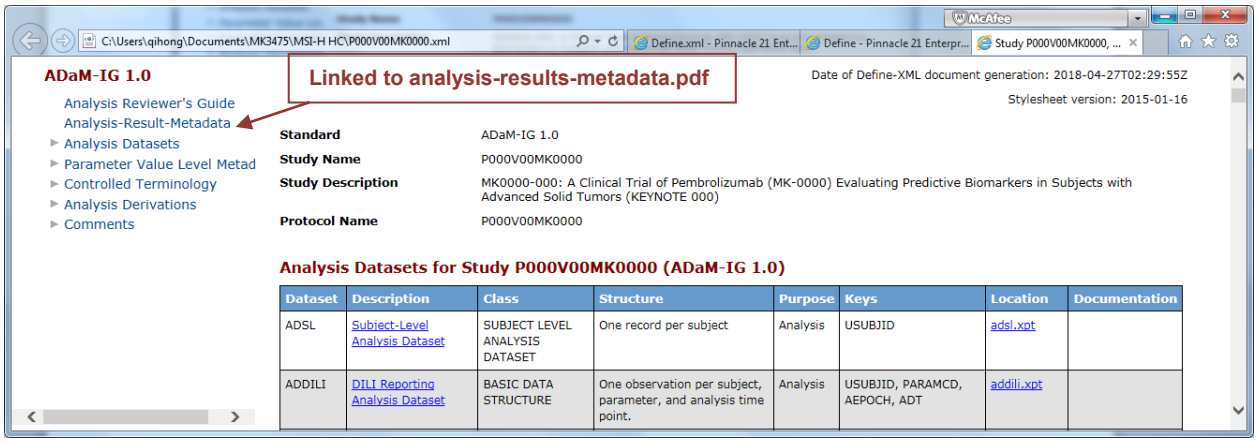


Figure 4 analysis-results-metadata.pdf linked to define.xml as an external file.

ENHANCED PROCESS

Aiming at an efficient submission process with high quality, our strategy is to proactively document the ARM components when the reports are being designed or generated, and to automate the ARM generation as an extension to Define.xml using Pinnacle 21 Enterprise tool when the submission package is being developed.

To gather and document the ARM components, we are building an efficient submission tool, <ARM0template.xlsm>, a macro-enabled spreadsheet with tabs containing macros that automate frequently used tasks. It contains two types of tabs, the regular Excel® tabs for documenting the study reports, and Pinnacle 21 Enterprise required tabs which are macro-enabled spreadsheet tabs that select and combine information from the study report tabs (Figure 5).

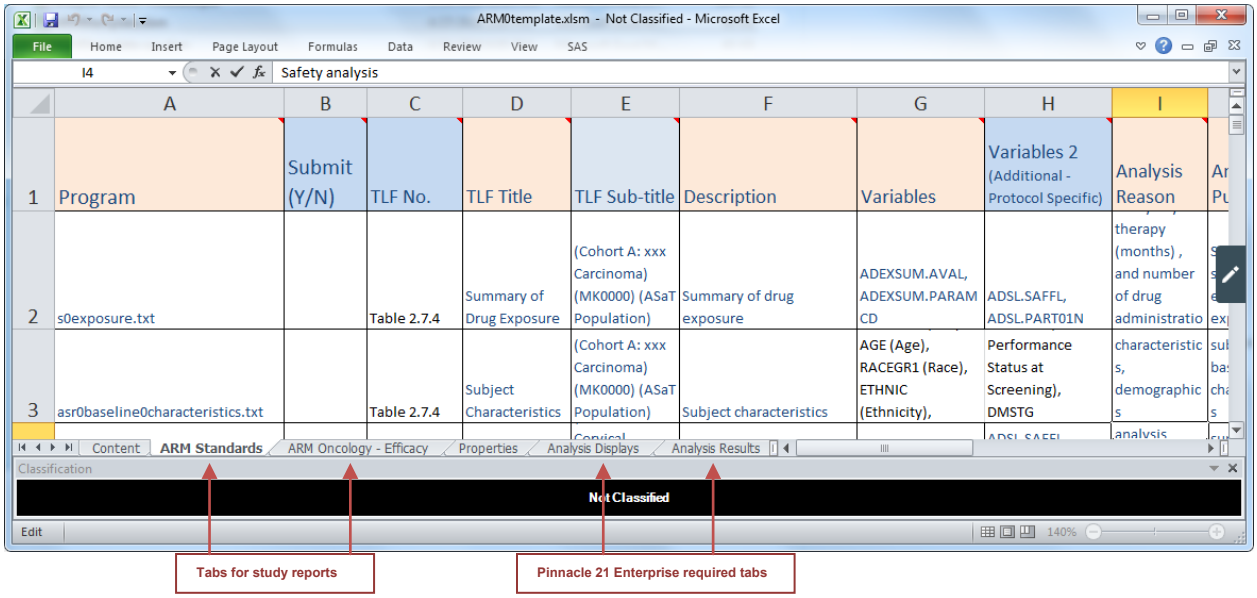


Figure 5 ARM0template.xlsm with study reports tabs and Pinnacle 21 Enterprise required tabs

Study Reports Tabs:

Tabs for study reports include those for Standards reports and for therapeutic area specific reports, i.e. tabs “ARM Standards”, and “ARM Oncology – Efficacy” in Figure 5, which contain the ARM components for the reports common to all the studies and specific to a therapeutic area, respectively. For the ease of maintenance, each study tab contains all the ARM components needed to produce the ARM as the Define-XML extension. With all the reports documented ahead of the schedule, each study team will only need to select the reports to be included in the package, and add study specific information, such as the subtitle of the reports, additional selection criteria, etc., when working on the submission, instead of spending time to read, understand, and select each report program at the last minute.

Pinnacle 21 Enterprise Required Tabs:

Pinnacle 21Enterprise required tabs, “Analysis Display” and “Analysis Criteria”, are the macro-enabled spreadsheet tabs in <ARM Template.xlsm> with the columns required by Pinnacle 21 Enterprise. Macro functions are embedded in each column to automatically link the information from the study tabs once data are entered into the study tabs. Pinnacle 21 Enterprise required tabs can be copied directly to the data specification, e.g. P000V00MK0000.xlsx, as the input file for Pinnacle 21 Enterprise (Figure 6) to generate ARM as part of the Define-XML (Figure 7).

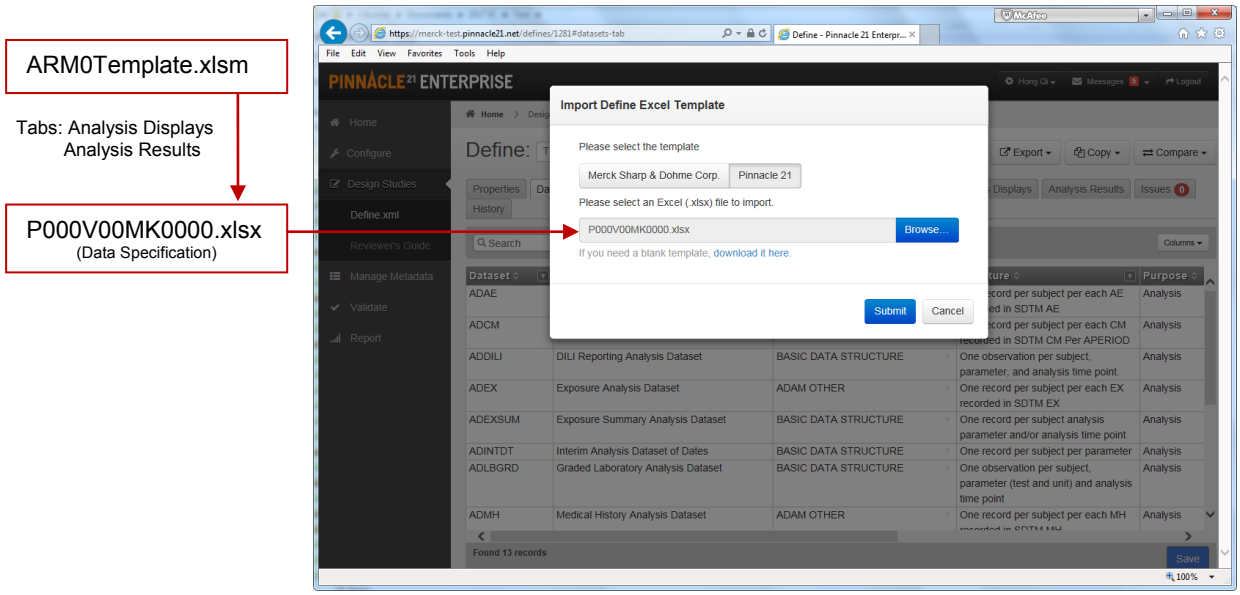


Figure 6 Pinnacle 21 Enterprise required tabs in ARM0Template.xlsm, copied to the data specification, as the input file for Pinnacle 21 Enterprise

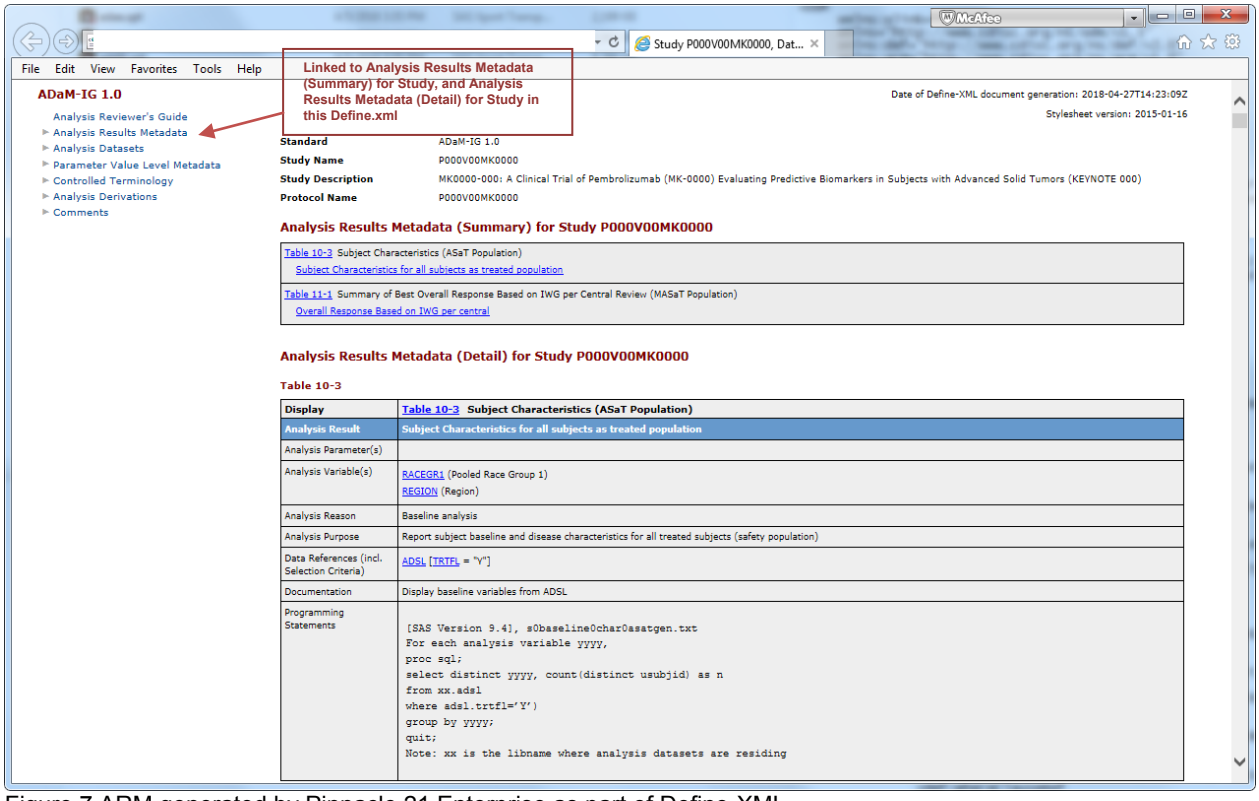


Figure 7 ARM generated by Pinnacle 21 Enterprise as part of Define-XML

CONCLUSION

Many approaches have been explored to generate ARM, such as using a word template file, SAS macro, or Pinnacle 21 Enterprise tool, etc., across the pharmaceutical industry. No matter which approach is used, 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.

REFERENCES

CDISC Analysis Data Model Version 2.1, analysis_data_model_v2.1.pdf

CDISC Analysis Results Metadata Specification Version 1.0 for Define-XML Version 2, Analysis Results Metadata v1.0 for Define-XML v2.pdf

ACKNOWLEDGMENTS

The authors thank the Merck Sharp & Dohme BARDS Scientific Programming management for the opportunity allowing us to explore the idea of automating ARM in Define-XML v2.0. Thanks Ellen Asam for her guidance on the theme of this paper; thanks Mary Varughese, and Sapan Patel for reviewing the paper and providing valuable comments.

CONTACT INFORMATION

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

Hong Qi
Merck & Co. Inc.
351 North Sumneytown Pike, UG1CD-44
PO Box 1000
North Wales, PA 19454-2505

Work Phone: 267-305-7589
Fax: 267-305-6474
Email: hong_qi@merck.com

Majdoub Haloui
Merck & Co. Inc.
351 North Sumneytown Pike, UG1CD-44
PO Box 1000
North Wales, PA 19454-2505

Work Phone: 267-305-2629
Fax: 267-305-6474
Email: majdoub.haloui@merck.com

Larry Wu
Merck & Co. Inc.
126 E. Lincoln Avenue
PO Box 2000
Rahway, NJ 07065-0900

Work Phone: 732-594-4329
Fax: 732-594-6075
Email: xi.wu1@merck.com

Gregory T. Golm
Merck & Co. Inc.
351 North Sumneytown Pike, UG1CD-44
PO Box 1000
North Wales, PA 19454-2505

Work Phone: 267-305-7652
Fax: 267-305-6474
Email: gregory_golm@merck.com