PharmaSUG 2019 - Paper AD-208

Automate the process for Define-XML v2.0 with Analysis Results Metadata

Lei Jing, Chao Wang, FMD K&L Inc., Fort Washington, PA, USA

ABSTRACT

Currently, Analysis Results Metadata (ARM) is a required component in PMDA data submission package. ARM assists the reviewer by providing traceability from key efficacy and safety analysis results to analysis datasets and dataset related elements, which adds significant value to a regulatory submission as well.

However, the process of developing ARM in Define-XML may be very time-consuming if it is handled manually. This paper presents an effective approach to integrate ARM automatically into existing Define-XML v2.0. A macro is designed to convert all required information from ARM metadata into valid XML syntax and then insert the XML codes into existing Define-XML. This automating process will reduce the development cycle and increase package quality.

INTRODUCTION

The Japanese Pharmaceuticals and Medical Devices Agency (PMDA) have mentioned in their Technical Conformance Guide on Electronic Study Data Submissions (2015, [9]):

> "In order for the review of clinical study data to progress smoothly, it is important that the relationship between the analysis results shown in the application documents and the analysis datasets is easily understandable. Therefore, the definition documents of the ADaM datasets should preferably include Analysis Results Metadata, which shows the relationship between the analysis results and the corresponding analysis dataset and the variables used, for the analyses performed to obtain the main results of efficacy and safety and clinical study results that provide the rationales for setting of the dosage and administration".

The PMDA mentions further in the same document:

> "For the format of the Analysis Results Metadata, the applicant should refer to the Analysis Results Metadata specification for Define-XML by CDISC to the extent possible, but if it is difficult to include it into the definition document, it is possible to submit it as a separated file in PDF format".

Since ARM is required component in PMDA data submission package. Many completed studies did not have the ARM included in the past and it adds great value if we can integrate ARM in the existing define.xml for these studies.

STARTING WITH ANALYSIS RESULTS METADATA

Analysis Results Metadata is a specification that describes the major attributes of a specified analysis result found in a clinical study report or submission. It provides traceability for a given analysis result to the specific ADaM data that were used as input to generating the analysis result; it also provides information about the analysis method used and the reason the analysis was performed. Analysis Results Metadata is an optional ADaM metadata component according to the ADaM version 2.1 document. However, best practice is that it assists the reviewer by identifying the critical analyses, providing links between results, documentation, and datasets, and documenting the analyses performed.

The purpose of the Define-XML Analysis Results Metadata extension is to support the interchange of CDISC ADaM key Analysis Results Metadata for clinical research applications in a machine-readable format. An important use case for adding Analysis Results Metadata to Define-XML 2.0 is to support the review of analysis results and their relation to submitted clinical trial data in CDISC ADaM format.

The key components in ADaM Analysis Results Metadata are:

- Analysis Display metadata definitions
  - Analysis Result metadata definitions
    - Analysis parameter(s)
    - Analysis dataset(s)
      - Analysis variable(s)
      - Selection criteria
    - Documentation
    - Programming statements

See below figure shows Define-XML Analysis Results Metadata, which from CDISC Analysis Results Metadata (ARM) v1.0 for Define-XML v2.0 package.

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 1. Define XML 2.0 Display of Analysis Results Metadata

AUTOMATE PROCESS OF ADDING ARM IN DEFINE XML

Before click the automation button, three documents need be prepared: Define.xml, ARM excel file, SAS Macro. The process of adding ARM in define XML 2.0 has three steps as below:

Step 1: Prepare ARM.xlsx and get existing Define.xml ready

Step 2: Run SAS macro to generate final new define.xml with ARM

Step 3: Check final output

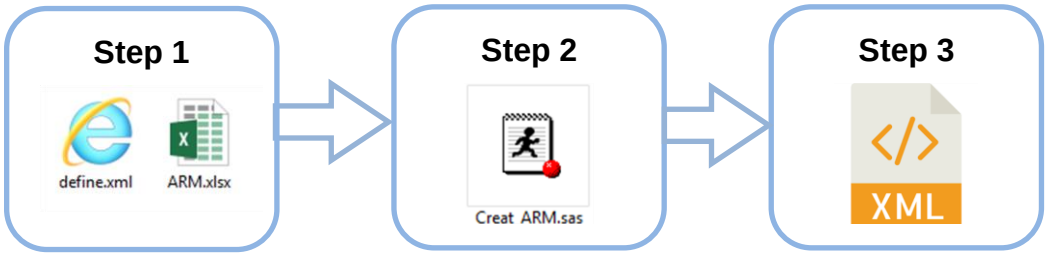

Figure 2. Process of adding ARM in existing define.xml

STEP 1: ARM EXCEL FILE

Based on the lists of required or optional elements of ARM for define.xml version 2.0, which specified in CDISC “Analysis Results Metadata v1.0 for Define-XML v2.0”. An ARM excel was designed to collect these key components, which include DisplayName, ResultName, ParameterCode, Parameter, Variable, Reason, Purpose, Data, DataCriteria, Documentation, SASVersion and Code.

For each collected item, the details description shown as following table:

ARM Attributes	Description
DisplayName	A brief description of that analysis table. Usually it is table’s title.
ResultName	Result name of efficacy analysis endpoints.
ParameterCode	Corresponds to PARAM
Parameter	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.
Variable	The variable name that contains the value for the analysis.
Reason	The reason of the analysis. e.g., “Primary Efficacy,” “Key Secondary Efficacy,” “Safety”.
Purpose	The purpose of the analysis. e.g., Efficacy outcome measure.
Data	The ADaM dataset name
DataCriteria	Specific and sufficient selection criteria for analysis subset or a complete list of the variables and their values used to identify the records selected for the analysis.
Documentation	Textual description of the analysis performed.
SASVersion	SAS version number, such as 9.3, or 9.4
Code	Key SAS code used for analysis.

Table 1. Description of ARM table’s each attribute

Following shows an example of a table’s information in ARM excel file.

	A	B	C	D
1	DisplayNo	ResultID	Attributes	Value
2	Table 14.3.1	1	DisplayName	Percent Change in mEASI Score and Results of the Mixed Effect Model at the End of Treatment (FAS Population)
3	Table 14.3.1	1	ResultName	Least squares means with 95% confidence intervals
4	Table 14.3.1	1	ParameterCode	MEASI
5	Table 14.3.1	1	Parameter	mEASI Score
6	Table 14.3.1	1	Variable	AVAL
7	Table 14.3.1	1	Reason	Primary efficacy analysis
8	Table 14.3.1	1	Purpose	Primary Efficacy Endpoints
9	Table 14.3.1	1	Data	ADEASFAS
10	Table 14.3.1	1	DataCriteria	FASFL = "Y" and FASRFL = "Y" and PARAMN = 1 and ANL01FL = "Y" and AVISITN = 9999
11	Table 14.3.1	1	Documentation	Analysis of covariance will be used, based on the linear mixed effects model, to test for any differences between the least squares mean for the 0.5% group (μ1) and the least squares mean for the placebo group (μ2) at a one-sided 2.5% significance level.
12	Table 14.3.1	1	ExternalReference	SAP.pdf
13	Table 14.3.1	1	ExternalReferenceName	SAP Section 10.1.1
14	Table 14.3.1	1	ReferenceLocation	Page 10
15	Table 14.3.1	1	SASVersion	SAS version 9.4
16	Table 14.3.1	1	Code	proc mixed data = work.efmain; class TRT01PN (ref = "1") SITEGR1N ; model PCHG = TRT01PN BASE / s cl ; random SITEGR1N ; lsmeans TRT01PN / diffs cl; ods output LSMeans = work.LSMeans Diffs = work.Diffs Tests3 = work.T3 ; run ;
17				

ARM DataCriteria

Figure 3. Screen print of Excel Spreadsheet “ARM” used for ARM information collection.

	A	B	C	D	E	F
1	**Table Name**	**Order**	**Dataset**	**Variable**	**WhereClause**	**Value**
2	Table 14.3.1	1	ADEASFAS	FASFL	EQ	Y
3	Table 14.3.1	1	ADEASFAS	FASRFL	EQ	Y
4	Table 14.3.1	1	ADEASFAS	PARAMN	EQ	1
5	Table 14.3.1	1	ADEASFAS	ANL01FL	EQ	Y
6	Table 14.3.1	1	ADEASFAS	AVISITN	EQ	9999
7	Table 14.3.1	1	ADEASFAS	PARAMCD	EQ	MEASI

ARM DataCriteria

Figure 4. Screen print of Excel Spreadsheet "DataCriteria" used for ARM information collection.

STEP 2: RUN SAS MACRO

Basically the macro "*Integrate_ARM_In_Define*" has three parts. First one is to read existing define.xml and arm.xlsx into sas dataset. Second part is to write XML syntax based on arm.xlsx into a sas dataset, then combine old define sas dataset with arm sas dataset. The third part outputs final sas dataset back to XML file.

Following is the XML file source code structure. Make sure insert arm xml code within block <MetaDataVersion></MetaDataVersion > and under block <arm:AnalysisResultDisplays></arm:AnalysisResultDisplays>.

```
<?xml version="1.0" encoding="UTF-8"?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.3"
  xmlns:def="http://www.cdisc.org/ns/def/v2.0"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:arm="http://www.cdisc.org/ns/arm/v1.0"
  ODMVersion="1.3.2" FileType="Snapshot" FileOID="CDISC-Sample"
  CreationDateTime="2014-03-28T11:07:23:00"
  Originator="CDISC ADaM Metadata Team">
  <Study OID="cdisc01">
    <GlobalVariables>
      <StudyName>CDISC Sample</StudyName>
      <StudyDescription>CDISC-Sample Data Definitions</StudyDescription>
      <ProtocolName>CDISC-Sample</ProtocolName>
    </GlobalVariables>
    <MetaDataVersion OID="MDV.CDISC01.ADaMIG.1.0.ADaM.2.1"
      Name="Study CDISC-Sample, Data Definitions"
      Description="Study CDISC01, Data Definitions"
      def:DefineVersion="2.0.0"
      def:StandardName="ADaM-IG"
      def:StandardVersion="1.0">
        < Supplemental Data Definitions (def:SupplementalDoc) >
        < Value Level Metadata (def:ValueListDef) >
        < Where Clause Definitions (def:WhereClauseDef) >
        < Domain Level Metadata (ItemGroupDef) >
        < Variable Level Metadata (ItemDef) >
        < Controlled Terminology Metadata (CodeList) >
        < Computational Algorithms (MethodDef) >
        < Comments (def:CommentDef) >
        < Referenced Documents (def:leaf) >
        < Analysis Results Metadata (arm:AnalysisResultDisplays) >
    </MetaDataVersion>
```

Source: Analysis Results Metadata (ARM) v1.0 for Define-XML v2.0

Figure 5. Define-XML Document Structure.

There are 3 arguments in this macro:

```
%Integrate_ARM_In_Define
(
	Orgxml  = studyxx_define.xml,
	Armfile = arm.xlsx,
	Newxml  = ARM_define.xml
)
```

Please see below for some key parts of the macro:

```
/************************************************************
Step 1: Readin original define.xml and ARM Excel file
************************************************************/
filename xmlname "&root.&orgxml.";

data xmlprg;
	length xmlcode $2000.;
	infile xmlname truncover;
	input xmlcode $2000.;
run;

…

/************************************************************
Step 2: Generate XML Syntax for ARM
************************************************************/
data part2(keep=xmlcode) ;
	length xmlcode $2000;
	set arm1 end = last;
	by displayno;
	xmlcode = '<!-- ****************************** -->'; output;
	xmlcode = '<!-- ' || strip(displayno) || ' -->'; output;
	xmlcode = '<!-- ****************************** -->'; output;
	if _N_ = 1 then do;
	xmlcode = '<arm:AnalysisResultDisplays>'; output;
	end;
	xmlcode = '<arm:ResultDisplay OID="RD.'
	                || strip(table_no2) || '" Name="'
	                || strip(table_no1) || '">'; output;
	xmlcode = '<Description>'; output;
	xmlcode = '   <TranslatedText xml:lang="en">' || strip(DisplayName)
	                || '</TranslatedText>'; output;
	xmlcode = ' </Description>'; output;
	xmlcode = ' <arm:AnalysisResult' || ' OID="AR.'
	             || strip(table_no2) || '"'; output;


	…


	xmlcode = ' </arm:AnalysisResult>'; output;
	xmlcode = ' </arm:ResultDisplay>'; output;
	if last then do;
	       xmlcode = '   </arm:AnalysisResultDisplays>'; output;
	end;
run;
/*Create Data criteria part in xml code*/
data part3;
	length xmlcode $2000;
	set whereclause end=last;
	by order;
	retain i;
```

```
        if index(upcase(Table_name), 'TABLE') then do;
                table1 = tranwrd(strip(Table_name), ' ', '_');
                table2 = tranwrd(strip(table1), '.', '_');
        end;
        if _n_ = 1 then do;
                i=1;
                xmlcode = '<def:WhereClauseDef OID="WC.' ||
                strip(table2) || '">'; output;
        end;

        …

        if last then do;
                xmlcode = '</def:WhereClauseDef>'; output;
        end;
        keep xmlcode;
run;

Data xmlarm;
        set part:;
        keep xmlcode;
run;

/*************************************************************
Step 3: Output dataset back to xml
*************************************************************/
data _null_;
        file "&root.&newxml.";
        set xmlarm;
        put xmlcode;
run;
```

Actually section <arm:AnalysisResultDisplays></arm:AnalysisResultDisplays> can be anywhere within section <MetaDataVersion></MetaDataVersion > .

FINAL RESULT

Once finish running macro, a new define.xml will be generated. XML code looks as following:

```
def:DefineVersion="2.0.0"
def:StandardName="ADaM-IG"
def:StandardVersion="1.1">
<def:SupplementalDoc>
<def:DocumentRef leafID="LF.ADRG"/>
<def:DocumentRef leafID="LF.complexalgorithms"/>
<def:DocumentRef leafID="LF.SAP"/>
</def:SupplementalDoc>
<!-- ****************************** -->
<!-- Table 14.3.1 -->
<!-- ****************************** -->
<arm:AnalysisResultDisplays>
<arm:ResultDisplay OID="RD.Table_14_3_1" Name="Table 14.3.1">
<Description>
<TranslatedText xml:lang="en">Percent Change in mEASI Score and Results of the Mixed Effect Mc
</Description>
<arm:AnalysisResult OID="AR.Table_14_3_1"
ParameterOID="IT.ADEASFAS.PARAMCD" AnalysisReason="Primary efficacy analysis" AnalysisPurpose=
<Description>
<TranslatedText xml:lang="en">Least squares means with 95% confidence intervals</TranslatedTex
</Description>
<arm:AnalysisDatasets>
<arm:AnalysisDataset ItemGroupOID="IG.ADEASFAS">
<def:WhereClauseRef WhereClauseOID="WC.Table_14_3_1"/>
<arm:AnalysisVariable ItemOID="IT.ADEASFAS.AVAL"/>
</arm:AnalysisDataset>
</arm:AnalysisDatasets>
<arm:Documentation>
<Description>
<TranslatedText xml:lang="en">Analysis of covariance will be used, based on the linear mixed e
2.5% significance level.</TranslatedText>
</Description>
</arm:Documentation>
<arm:ProgrammingCode Context="SAS version 9.4">
<arm:Code>
proc mixed data = work.efmain;
    class TRT01PN ( ref = "1" ) SITEGR1N ;
    model PCHG = TRT01PN BASE / s cl ;
    random SITEGR1N ;
    lsmeans TRT01PN / diffs cl;
    ods output
        LSMeans = work.LSMeans
        Diffs   = work.Diffs
        Tests3  = work.T3 ;
run ;
</arm:Code>
</arm:ProgrammingCode>
</arm:AnalysisResult>
</arm:ResultDisplay>
</arm:AnalysisResultDisplays>
```

With the “define2-0-0.xsl”, the final define.xml looks like:

ADaM-IG 1.1

Reviewers Guide
Complex Algorithms
Statistical Analysis Plan
Analysis Results Metadata
Analysis Datasets
Parameter Value Level Metadata
Controlled Terminology
Analysis Derivations
Comments

Date of Define-XML document generation: 2019-06-05T20:45:07

Stylesheet version: 2015-01-16

Standard	ADaM-IG 1.1
Study Name	CDISC-Sample
Study Description	CDISC-Sample Data Definition
Protocol Name	CDISC-Sample
Metadata Name	Study CDISC-Sample Data Definitions
Metadata Description	CDISC-Sample Data Definition

Analysis Results Metadata (Summary) for Study CDISC-Sample

Table 14.3.1 Percent Change in mEASI Score and Results of the Mixed Effect Model at the End of Treatment (FAS Population)
Least squares means with 95% confidence intervals

Analysis Results Metadata (Detail) for Study CDISC-Sample

Table 14.3.1

Display	Percent Change in mEASI Score and Results of the Mixed Effect Model at the End of Treatment (FAS Population)
Analysis Result	**Least squares means with 95% confidence intervals**
Analysis Parameter(s)	PARAMCD = "MEASI" (mEASI Score)
Analysis Variable(s)	AVAL (Analysis Value)
Analysis Reason	Primary efficacy analysis
Analysis Purpose	Primary Efficacy Endpoints
Data References (incl. Selection Criteria)	ADEASFAS [FASFL = "Y" and FASRFL = "Y" and PARAMN = 1 and ANL01FL = "Y" and AVISITN = 9999 and PARAMCD = "MEASI"]
Documentation	Analysis of covariance will be used, based on the linear mixed effects model, to test for any differences between the least squares mean for t for the placebo group (μ2) at a one-sided 2.5% significance level.
Programming Statements	[SAS version 9.4] proc mixed data = work.efmain; class TRT01PN (ref = "1") SITEGR1N ; model PCHG = TRT01PN BASE / s cl ; random SITEGR1N ; lsmeans TRT01PN / diffs cl; ods output LSMeans = work.LSMeans Diffs = work.Diffs Tests3 = work.T3 ; run ;

Figure 6. Screen print of ARM (Details) add in existing define.xml version 2.0

CONCLUSION

Nowadays, the methods to add ARM features in existing Define-XML are still limited and need handle manually in the end. This paper presents an automated method using one macro to convert the collected information in Excel spreadsheet to valid xml syntax, and embeds the ARM into existing define.xml. By clicking a simple button to generate ARM define.xml 2.0 could be very beneficial for CRO/pharma if they already have studies with existing define.xml but are not able to include ARM by just using SAS program.

REFERENCES

1. Jeff Xia, Shunbing Zhao, Anjana Grandhi 2018. “A Simple Method for Integrating Analysis Result Metadata into Existing Define.xml 2.0.” PharmaSUG 2018 Proceedings.
2. Lex Jansen, 2017. Creating Define-XML version 2 including Analysis Results Metadata with the SAS® Clinical Standards Toolkit. PharmaSUG 2017 Proceedings.
3. Analysis Results Metadata Specification (Version 1.0) for Define-XML (Version 2)
4. Pharmaceuticals and Medical Devices Agency, Technical Conformance Guide on Electronic Study Data Submissions, April 27, 2015 (http://www.pmda.go.jp/files/000215100.pdf)

CONTACT INFORMATION

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

Lei Jing
FMD K&L Inc.
1300 Virginia Dr, Fort Washington, PA 19034
Lei.jing@klserv.com

Chao Wang
FMD K&L Inc.
1300 Virginia Dr, Fort Washington, PA 19034
Chao.wang@klserv.com