



PhUSE US Connect 2019

Paper SI04

Creation of ADaM Define.xml v2.0 Using SAS Program and Pinnacle 21

Yan Lei, Johnson & Johnson, Spring House, PA, USA
Yongjiang Xu, Johnson & Johnson, Spring House, PA, USA
Michelle Pupek Johnson & Johnson, Spring House, PA, USA

ABSTRACT

FDA has announced that Define-XML v2.0 is required for data submission for studies with first patient in after March 15, 2018. However, it is highly possible that multiple studies are compiled in one submission. Some of them may have been done in early years with previous Define-XML versions, while others may have implemented Define-XML v2.0. To keep the Define-XML versions consistent in a single submission, there is an urgent need for generating Define-XML v2.0 for all studies. Considering significant amount of time and effort that needs to be invested to manually update contents and structure of metadata, it is more efficient and cost effective to convert them automatically.

This paper presents an approach to create ADaM define.xml v2.0 using SAS program that read in the existing SAS metadata datasets and output a metadata Excel file in compliance with Pinnacle 21 (P21) requirement. The metadata excel file is imported to P21 then P21 exports a define.xml v2.0 for electronic submission.

INTRODUCTION

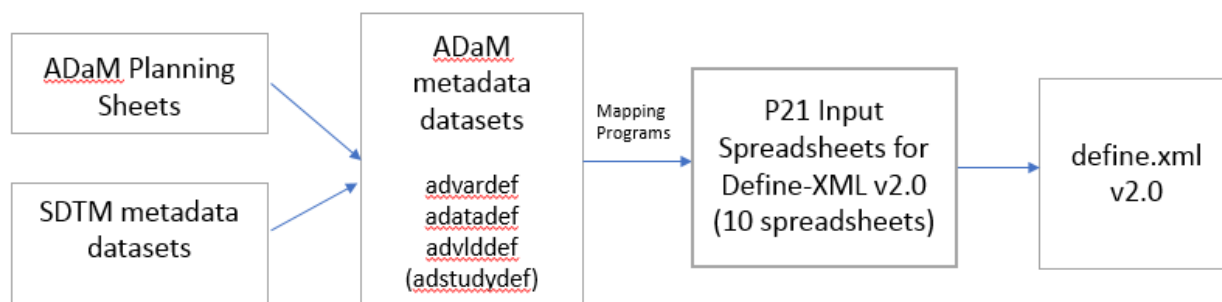
Define.xml is an essential part of a submission package. It increases the level of automation and improves the efficiency of the Regulatory Review process, making it highly desirable with every NDA submission.

Previously known as OpenCDISC, Pinnacle 21 are widely used in the pharmaceutical industry to validate clinical trial data and generate the define.xml. It features three components: Validator, Define.xml Generator and Converter. The Validator could be used to validate ADaM, SDTM, SEND datasets and validate define.xml as per XML schema developed by CDISC Define-XML Team. The Define.xml Generator can create a P21 format spec in MS Excel spreadsheet based on the validation data. Then P21 uses the spreadsheet to generate define.xml.

However, many studies have completed with metadata Specs in a different format. Furthermore, many phase II/III trials include a lot of value level metadata in ADaM data. It will require significant amount of time and effort to manually update the metadata specs to the P21 format Spec, especially to ensure the ID links between different sheets are linked correctly, for example, ValueLevel and WhereClauses.

The following steps briefly describe a process that use SAS programs/macros to read in existing SAS metadata datasets, programmatically assign the link IDs and create a P21 format Excel spreadsheet containing 10 tabs that can be uploaded to Pinnacle 21 to generate define.xml v2.

PROCESS FLOW



STEP 1. CREATE ADaM METADATA DATASETS

1. Complete ADaM datasets planning metadata sheets. If metadata had been generated following Define-XML v1.0 requirement, manually populate the define.xml v2 relevant fields that cannot be filled out programmatically, such as Variables.Origin (Predecessor vs Assigned vs Derived), Methods.Type (Computation vs Imputation).

Define-XML v1.0	Define-XML v2.0	Note
Copied from Source	Predecessor	
Derived	Derived	
Derived	Assigned	Assigned variable a value
Derived	Predecessor	Variable is just a copy from adxx.xxx

Type	Required	Allowable Values: Computation Imputation	The Method type: A Computation uses an algorithm to derive a value. An Imputation is the process of replacing missing data with substitute values.
------	----------	--	---

Sample output of ADaM dataset planning sheet:

DATASET	VARNAME	VARLABEL	DATATYPE	PARAMID	LGTH	DECDIG	NUMFMT	CODELST	ORIGCOM	DERVCOPY
Dataset Name	Variable Name	Variable Label.....	Variable Type	Parameter Ids	Length	Dec Digits	Display Format	Codelist/ Controlled Terms	Source/Derivation.....	Derived or Copied from Source Variable
ADTTE	STUDYID	Study Identifier	text						DM.STUDYID	Copied from Source
ADTTE	USUBJID	Unique Subject Identifier	text						DM.USUBJID	Copied from Source
ADTTE	PARAM	Parameter	text		200			PARAMTTE	Assigned based on codelist PARAMTTE	Assigned
ADTTE	PARAMCD	Parameter Code	text		8			PARAMCDTTE	Assigned based on codelist PARAMCDTTE corresponding to PARAM	Derived
ADTTE	AVAL	Analysis Value	integer		8				Derived as ADT - STARTDT + 1.	Derived
ADTTE	STARTDT	Time to Event Origin Date for Subject	integer	DURRESP	8		DATE9.		Imputation: Set to the earliest date that a subject was PR, VGPR, CR, or sCR by the IMWG response criteria	Derived
ADTTE	STARTDT	Time to Event Origin Date for Subject	integer	PFS OS	8		DATE9.		Set to the earliest date of the time of onset of the event by the IMWG response criteria	Derived
ADTTE	STARTDT	Time to Event Origin Date for Subject	integer	INVPFS	8		DATE9.		Set to the earliest date of the time of onset of the event by Investigator's assessment	Derived
ADTTE	CNSR	Censor	integer	DURRESP	8			CNSR	Set to 0 if subject has progressive disease per the IMWG criteria or died due to progressive disease. Otherwise	Derived



- Run SAS program to create ADaM metadata datasets based on planning sheets:

ADDATADEF	ADaM datasets definition
ADVARDEF	Analysis variables definition
ADCD	Codelist definition

- Address metadata issues if any. Ensure the special characters are removed or translated to ASCII text codes.

STEP 2. CREATE DEFINE-XML V2.0 EXCEL SPREADSHEET

Once ADaM metadata datasets are ready and free with special characters, they then are used as input datasets to create the P21 format Spec: a MS excel spreadsheet with 10 tabs.

Study	Datasets	Variables	Methods	ValueLevel	WhereClauses	Codelists	Dictionaries	Comments	Documents
-------	-----------------	-----------	---------	------------	--------------	-----------	--------------	----------	-----------

The program and macros were developed for creating the spreadsheet:

Program	
esdef2mapping	set macro parameters and call <u>marcos</u>
Macro	
%esdef2studydoc	create Study, Documents datasets
% <u>esckspchars</u>	check/translate the special characters
%esdef2dsn	create Datasets dataset
%esdef2vars	create Variables dataset
%esdef2vldlvl	create <u>ValueLevel, WhereClauses</u> datasets
%esdef2mthdcm	create Methods, Comments
%esdef2cd	create <u>Codelists</u> Dictionaries
%esdef2sas2xlsx	output 10 datasets to 10 tabs excel spreadsheet

These macros create link IDs to link between sheets, map the metadata to appropriate columns programmatically in the spreadsheet. Name the spreadsheet as STUDYNM-def2.xlsx

esdef2mapping.sas

```
%macro createmacvar;
```

```
%global studynm studydes protnm stdnm stdver type suppdoc paramasn;
```

```
%let studynm=56021927PCR1007;
```

```
%let protnm=&studynm;
```

```
%let stdnm=ADaM-IG;
```

```
%let stdver=1.0;
```



```
%if &stdver = %then %let stdver=%str( );
%if %index(%upcase(&stdnm),ADAM)>0 %then
    %let type=ADAM;
%else
    %let type=LEGACY;
/*****
 * Set to Y if you want to set DERVCOPY="Assigned" for variables PARAM/PARAMCD/PARAMN
 * where a codelist was defined and DERVCOPY was filled as "Derived".
 * This is usually for studies using SDTM v3.1.x
 *****/
%let paramasn=N;
run;
%mend;

%createmacvar;

%*****;
%* Call macro to generate Study and Documents datasets ;
%*****;
%esdef2StudyDoc(suppdoc=adrg.pdf supplementaldefinitions.pdf,
    indirmeta=AMS_IN,
    studynm=&studynm,
    protnm=&protnm,
    stdnm=&stdnm,
    stdver=&stdver,
    type=&type);

%*****;
%* Call macro to fixed special characters in metadata input datasets ;
%*****;
%escksplchars(dsn=ams_in.advardef, var=origcom, dsnout=advardef);
%escksplchars(dsn=ams_in.addatadef, var=docnotes, dsnout=addatadef);

%*****;
%* Call macro to generate Datasets dataset ;
%*****;
%esdef2dsn(indirmeta=AMS_IN);

%*****;
%* Call macro to generate Variables dataset ;
%*****;
%esdef2vars(indirmeta=AMS_IN,
    type=&type,
    paramasn=&paramasn);

%*****;
%* Call macro to generate Value level and Where clauses datasets ;
%*****;
%esdef2v11d1v1(indirmeta=AMS_IN,
    type=&type,
    paramasn=&paramasn);

%*****;
%* Call macro to generate Methods and Comments datasets ;
%*****;
%esdef2mthdcm(indirmeta=AMS_IN);
```



```

%*****;
%* Call macro to generate Codelist and Dictionaries datasets ;
%*****;
%esdef2cd(indirmeta=AMS_IN);

%*****;
%* Call macro to output dataset to an excel spreadsheet ;
%*****;
%global amsout;

%macro setamsout;
%if (&environ eq PDEV) %then %do;
    %let AMSOUT = &rptdrv.\PDEV\&USERNAME;
%end;
%else %do;
    %let AMSOUT = &rptdrv.\&environ.\ANALYSIS\METADATA\SUBMISSION;
%end;
%mend;
%setamsout;
%esdef2sas2xlsx(datasets=Study Datasets Variables Methods ValueLevel
    WhereClauses Codelists Dictionaries Comments Documents,
    xlsxfile=%bquote (&amsout.\&studynm-def2.xlsx));

```

STEP 3. IMPORT THE SPREADSHEET TO PINNACLE 21

1. Start Pinnacle 21 and select Design Studies -> Define.xml -> New Define.xml

Project	Study	Data Package
ADaM	ADaM007	56021927PCR1007-def2-new
ADaM	ADaM007	56021927PCR1007-YL
ADaM	ADaM007	JNJ-54767414MMY3007-YL

2. Select "Import Excel Specification" to upload the spreadsheet



3. Create new data package and name the data package as same as spreadsheet - STUDYNM-def2.

4. Click next, provide the STUDYNM-def2.xlsx to be uploaded. Follow the screen instructions, upload the spreadsheet.



- It is not surprising that the first couple of uploading attempts may not be successful due to some errors (most of them are relevant to special characters or codelists). Follow P21 error messages, address errors, recreate STUDYNM-def2.xlsx and reload it.

Sample error:

Confirm

Please confirm the selections made for your define before clicking submit

Project: ADaM
Study: ADaM007
Data package: JNJ-54767414MMY3004
Standard: ADaM-IG 1.0
Define data source: Import Excel Specification
File to Import: JNJ-54767414MMY3004-def2.xlsx

Validation failed. We found **3** errors.

Codelists, 1672 The term 'Cycles 1-3' has already been defined for the code list 'TCYCGR', duplicates are not allowed

Codelists, 1673 The term 'Cycles 4-8' has already been defined for the code list 'TCYCGR', duplicates are not allowed

Codelists, 1674 The term 'Cycles >=9' has already been defined for the code list 'TCYCGR', duplicates are not allowed

- Once STUDYNM-def2.xlsx is uploaded, evaluate/address the issues as many as possible if there are issues in Issue box.

Home

Configure

Design Studies

Define.xml

Reviewer's Guide

Manage Metadata

Validate

Report

Home > Design Studies > Define.xml > 56021927PCR1007-def2-new: 56021927PCR1007

Define: ADaM > ADaM007 > 56021927...

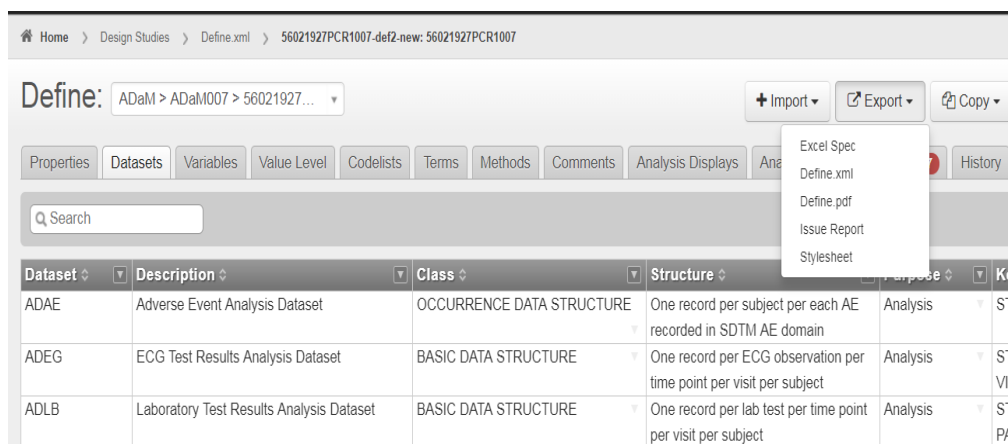
+ Import Export Copy Compare

Properties Datasets Variables Value Level Codelists Terms Methods Comments Analysis Displays Analysis Results Issues 7 History

Q Search Columns

Dataset	Description	Class	Structure	Purpose	Key Variables
ADAE	Adverse Event Analysis Dataset	OCCURRENCE DATA STRUCTURE	One record per subject per each AE recorded in SDTM AE domain	Analysis	STUDYID, USUBJID,
ADEG	ECG Test Results Analysis Dataset	BASIC DATA STRUCTURE	One record per ECG observation per time point per visit per subject	Analysis	STUDYID, USUBJID, VISITNUM, ADT, ECG
ADLB	Laboratory Test Results Analysis Dataset	BASIC DATA STRUCTURE	One record per lab test per time point per visit per subject	Analysis	STUDYID, USUBJID, PARAM, ADT, VISITN
ADPP	PK Parameter Analysis Dataset	BASIC DATA STRUCTURE	One record per PK parameter per subject	Analysis	STUDYID, USUBJID,
ADSL	Subject-level Analysis Dataset	SUBJECT LEVEL ANALYSIS DATASET	One record per subject	Analysis	STUDYID, USUBJID
ADVS	Vital Signs	BASIC DATA STRUCTURE	One record per vital sign measurement per visit per subject	Analysis	STUDYID, USUBJID, VISITNUM, ADT, VSS

- Export define.xml and stylesheet:



8. Review/check the define.xml to ensure links are linked correctly, truncations exist or not ...etc. Update planning sheets if necessary.
9. Rename the STUDYNM-def2.xml to define.xml
10. Run final P21E validation with xpt files and define.xml together, evaluate/address the issues as many as possible.
11. After update, recreate analysis metadata, xpt files and the spreadsheet, import spreadsheet using same data package and regenerate the define.xml.
12. Finalize define.xml. If define.xml is finalized in Pinnacle 21 Enterprise, the define.pdf can be output easily.

Sample define.xml v2.0 and define.pdf

define.xml:

ADaM-IG 1.0

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

Date of document generation: 2013-03-03T14:47:00

Stylesheet version: 2013-04-24

Analysis Datasets for Study CDISC-Sample (ADaM-IG 1.0)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
ADSL	Subject-Level Analysis	SUBJECT LEVEL ANALYSIS DATASET	one record per subject	Analysis	USUBJID	adsl.xpt	Screen Failures are excluded since they are not needed for this study analysis
ADQSADAS	ADAS-Cog Analysis	BASIC DATA STRUCTURE	One record per subject per parameter per analysis visit per analysis date	Analysis	USUBJID, PARAMCD, AVISIT, ADT	adqsadas.xpt	See referenced dataset creation program and Analysis Data Reviewer's Guide, Section 2.1 adqsadas.sas Analysis Data Reviewer's Guide

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	12		Predecessor: DM.STUDYID
USUBJID	Unique Subject Identifier	text	11		Predecessor: DM.USUBJID
SUBJID	Subject Identifier for the Study	text	4		Predecessor: DM.SUBJID
SITEID	Study Site Identifier	text	3		Predecessor: DM.SITEID
SITEGR1	Pooled Site Group 1	text	3		Derived: refer to SAP, Section 7.1 - if not pooled then SITEGR1=SITEID. If pooled, SITEGR1 will be 900



define.pdf:

Subject-Level Analysis (ADSL) [Location: adsl.xml]					
Variable	Label	Type	Length / Display Format	Controlled Terms or Format	Source/Derivation/Comment
STUDYID	Study Identifier	text	12		Predecessor: DM.STUDYID
USUBJID	Unique Subject Identifier	text	11		Predecessor: DM.USUBJID
SUBJID	Subject Identifier for the Study	text	4		Predecessor: DM.SUBJID
SITEID	Study Site Identifier	text	3		Predecessor: DM.SITEID
SITEGR1	Pooled Site Group 1	text	3		Derived: refer to SAP, Section 7.1 - if not pooled then SITEGR1=SITEID. If pooled, SITEGR1 will be 900
ARM	Description of Planned Arm	text	20	ARM	Predecessor: DM.ARM
TRT01P	Planned Treatment for Period 01	text	20	ARM	Predecessor: DM.ARM
TRT01PN	Planned Treatment for Period 01 (N)	integer	8	ARMN	Assigned Numeric code for TRT01P which corresponds to the randomized dose
TRT01A	Actual Treatment for Period 01	text	20	ARM	Assigned TRT01A=TRT01P, i.e., no difference between actual and randomized treatment in this study.
TRT01AN	Actual Treatment for Period 01 (N)	integer	8	ARMN	Assigned Numeric code for TRT01A which corresponds to the randomized dose

CONCLUSION

The combination approach of SAS programs and the Pinnacle 21 eliminates the time budgeted for transforming regular specifications or specification for previous version of define.xml to generate define.xml v2 and convert define.xml v1 to define.xml v2 for complex ADaM datasets. It significantly reduces the amount of time and effort for manual updates of metadata contents and/or structure and avoids human errors.

RECOMMENDED READING

1. Study Data Technical Conformance Guide
<http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM384744.pdf>
2. CDISC Define-XML Specification Version 2.0
<https://www.cdisc.org/standards/foundational/define-xml/define-xml-v20>

ACKNOWLEDGMENTS

The authors would like to thank Dr. Baoqing Li and Kenny Yan for the great support and valuable input to this paper.

CONTACT INFORMATION

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

Name: Yan Lei
Company: Janssen Pharmaceutical Companies of Johnson & Johnson
Address: Welsh & McKean Roads
City, State ZIP: Spring House, PA 19477
E-mail: ylei5@its.jnj.com



Name: Yongjiang Xu
Company: Janssen Pharmaceutical Companies of Johnson & Johnson
Address: Welsh & McKean Roads
City, State ZIP: Spring House, PA 19477
E-mail: yxu142@its.jnj.com

Name: Michelle Pupek
Company: Janssen Pharmaceutical Companies of Johnson & Johnson
Address: Welsh & McKean Roads
City, State ZIP: Spring House, PA 19477
E-mail: mpupek@its.jnj.com

TRADEMARK INFORMATION

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.