



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

Yan Lei, Yongjiang Xu, Michelle Pupek

SI04

02/24/2019

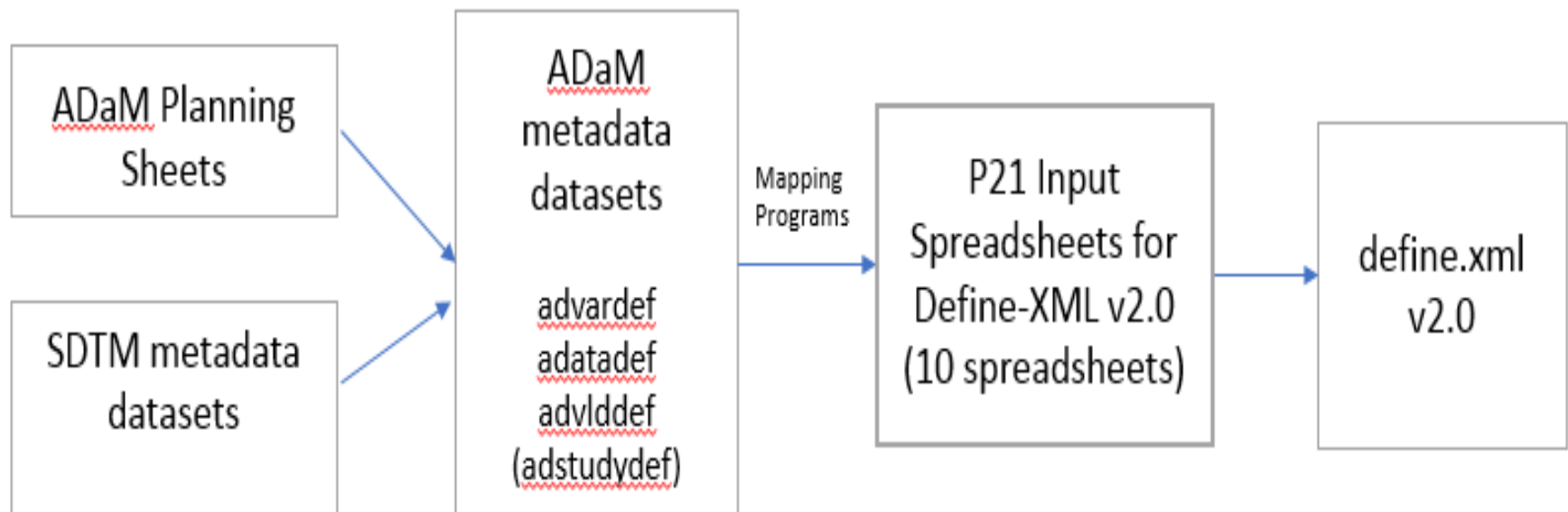


ABSTRACT

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.

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, manually populate 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.
------	----------	--	---

STEP 1. CREATE ADaM METADATA DATASETS

Sample output of ADaM dataset planning sheet:

Predecessor

DATASET	VARNAME	VARLABEL	DATATYPE	PARAMID	LNTH	DECDIG	NUMFMT	CODELST	ORIGCOM	DERVCOPY
Dataset Name	Variable Name	Variable Label.....	Variable Type	Parameter Id	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

STEP 1. CREATE ADaM METADATA DATASETS

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

ADDATADEF	ADaM datasets definition
ADVARDEF	Analysis variables definition
ADCD	Codelist definition

3. 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 created 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 macros
Macro	
%esdef2studydoc	create Study, Documents datasets
% <u>esckspchars</u>	check/translate the special characters
%esdef2dsn	create Datasets dataset
%esdef2vars	create Variables dataset
%esdef2vlllvl	create <u>ValueLevel</u> , <u>WhereClauses</u> datasets
%esdef2mthdcm	create Methods, Comments
%esdef2cd	create <u>Codelists</u> Dictionaries
%esdef2sas2xlsx	output 10 datasets to 10 tabs excel spreadsheet

STEP 2. CREATE DEFINE-XML V2.0 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

```
%*****;  
%* 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);|
```

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

```
%*****
%* 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 2. CREATE DEFINE-XML V2.0 EXCEL SPREADSHEET

These macros **create link IDs to link between sheets**, map the metadata to appropriate columns programmatically in the spreadsheet, avoid human errors.

C	D	E	F	J	K	L	M
Variable	Where Clause	Data Type	Length	Codelist	Origin	Pages	Method
ANRIND	ADEG.ANRIND.01.NOTIN	text	4		Assigned		
ANRIND	ADEG.ANRIND.02.HR	text	4	LNHFLG	Derived		ADEG.ANRIND.02.HR
ANRIND	ADEG.ANRIND.03.PR	text	4	LNHFLG	Derived		ADEG.ANRIND.03.PR
ANRIND	ADEG.ANRIND.04.QRS	text	4	LNHFLG	Derived		ADEG.ANRIND.04.QRS
ANRIND	ADEG.ANRIND.05.QT	text	4	LNHFLG	Derived		ADEG.ANRIND.05.QT
ANRIND	ADEG.ANRIND.06.QTCB	text	4	LNHFLG	Derived		ADEG.ANRIND.06.QTCB
ANRIND	ADEG.ANRIND.07.QTCF	text	4	LNHFLG	Derived		ADEG.ANRIND.07.QTCF
ANRIND	ADVS.ANRIND.01.NOTIN	text	6		Assigned		
ANRIND	ADVS.ANRIND.02.DIABP	text	6	LNHFLG	Derived		ADVS.ANRIND.02.DIABP
ANRIND	ADVS.ANRIND.03.SYSBP	text	6	LNHFLG	Derived		ADVS.ANRIND.03.SYSBP
ANRIND	ADVS.ANRIND.04.PULSE	text	6	LNHFLG	Derived		ADVS.ANRIND.04.PULSE
ANRIND	ADVS.ANRIND.05.TEMP	text	6	LNHFLG	Derived		ADVS.ANRIND.05.TEMP
ANRIND	ADVS.ANRIND.06.RESP	text	6	LNHFLG	Derived		ADVS.ANRIND.06.RESP

Study
Datasets
Variables
Methods
ValueLevel
WhereClauses
Codelists
...
+
◀

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

ID	Dataset	Variable	Comparator	Value
ADEG.ANRIND.01.NOTIN	ADEG	PARAMCD	NOTIN	HR, PR, QRS, QT, QTCB, QTCF
ADEG.ANRIND.02.HR	ADEG	PARAMCD	EQ	HR
ADEG.ANRIND.03.PR	ADEG	PARAMCD	EQ	PR
ADEG.ANRIND.04.QRS	ADEG	PARAMCD	EQ	QRS
ADEG.ANRIND.05.QT	ADEG	PARAMCD	EQ	QT
ADEG.ANRIND.06.QTCB	ADEG	PARAMCD	EQ	QTCB
ADEG.ANRIND.07.QTCF	ADEG	PARAMCD	EQ	QTCF
ADVS.ANRIND.01.NOTIN	ADVS	PARAMCD	NOTIN	DIABP, SYSBP, PULSE, TEMP, RESP
ADVS.ANRIND.02.DIABP	ADVS	PARAMCD	EQ	DIABP
ADVS.ANRIND.03.SYSBP	ADVS	PARAMCD	EQ	SYSBP
ADVS.ANRIND.04.PULSE	ADVS	PARAMCD	EQ	PULSE
ADVS.ANRIND.05.TEMP	ADVS	PARAMCD	EQ	TEMP
ADVS.ANRIND.06.RESP	ADVS	PARAMCD	EQ	RESP
ADVS.CRIT1.01.NOTIN	ADVS	PARAMCD	NOTIN	PULSE, SYSBP, DIABP, TEMP, RESP
ADVS.CRIT1.02.PULSE	ADVS	PARAMCD	EQ	PULSE
ADVS.CRIT1.03.SYSBP	ADVS	PARAMCD	EQ	SYSBP
ADVS.CRIT1.04.DIABP	ADVS	PARAMCD	EQ	DIABP
ADVS.CRIT1.05.TEMP	ADVS	PARAMCD	EQ	TEMP
ADVS.CRIT1.06.RESP	ADVS	PARAMCD	EQ	RESP
ADVS.CRIT2.01.NOTIN	ADVS	PARAMCD	NOTIN	PULSE, SYSBP, DIABP, TEMP, RESP



Study

Datasets

Variables

Methods

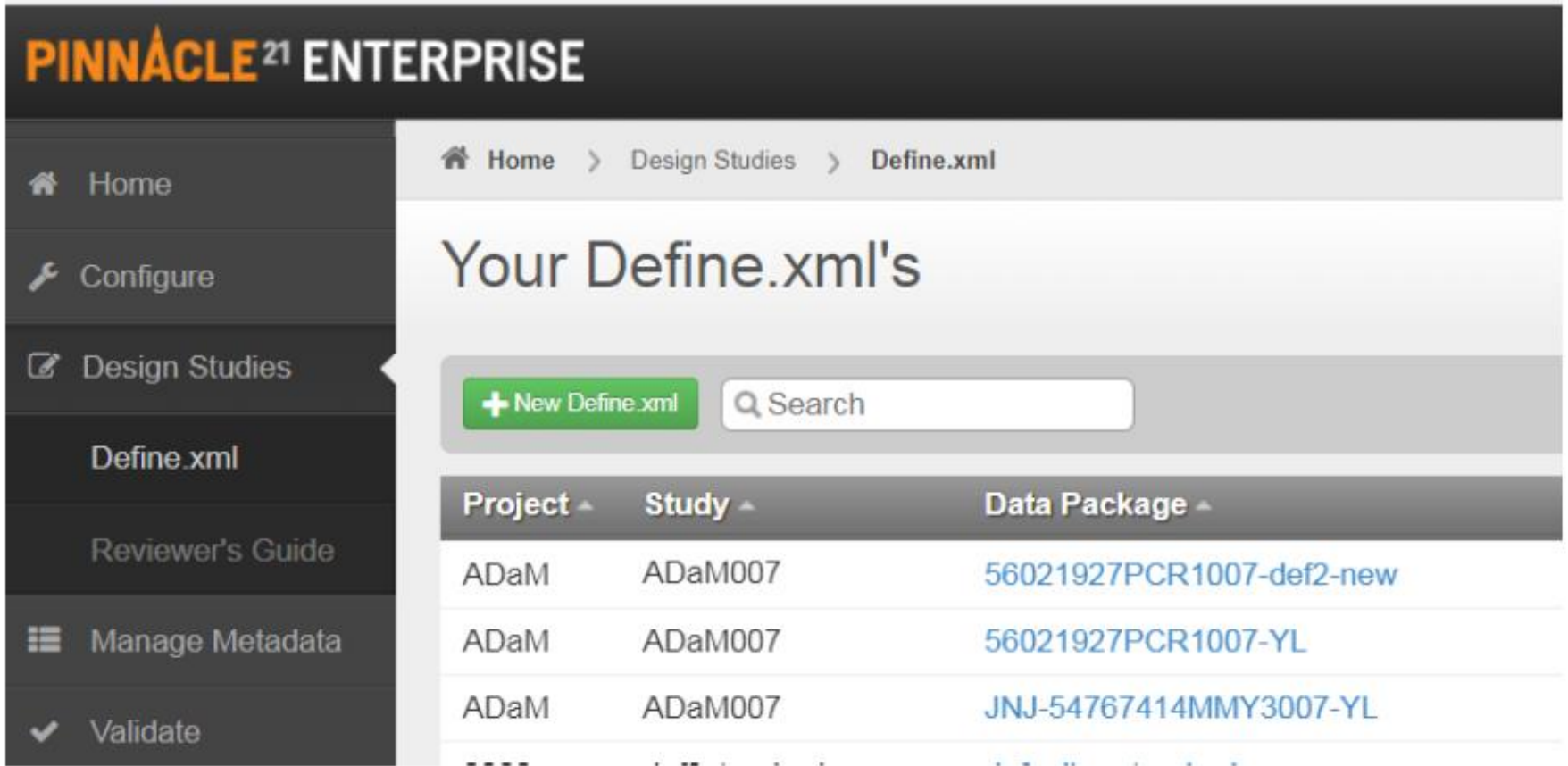
ValueLevel

WhereClauses

Codelists

STEP 3. IMPORT THE SPREADSHEET TO PINNACLE 21

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

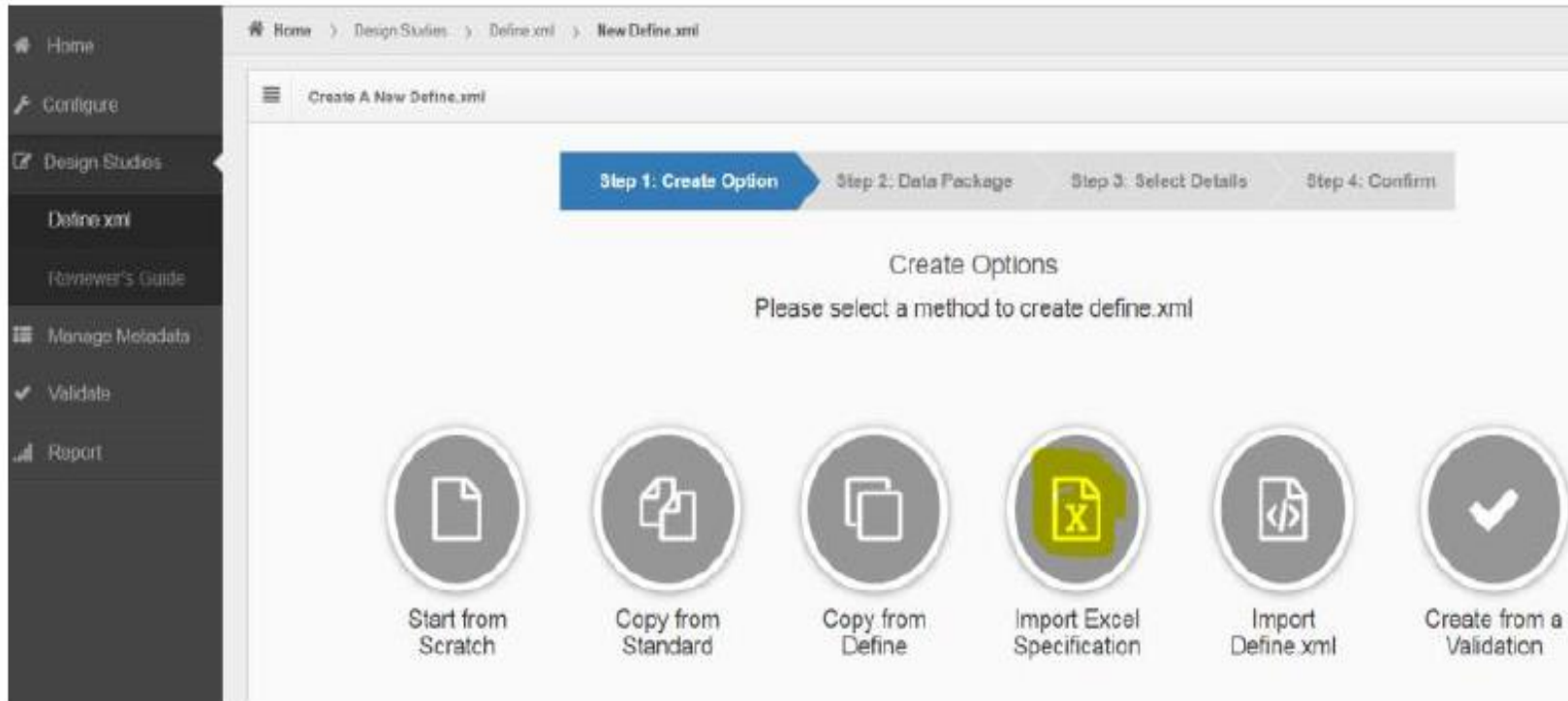


The screenshot displays the Pinnacle 21 Enterprise web application. The left sidebar contains navigation links: Home, Configure, Design Studies, Define.xml (highlighted), Reviewer's Guide, Manage Metadata, and Validate. The main content area shows the breadcrumb path: Home > Design Studies > Define.xml. Below the breadcrumb, the heading 'Your Define.xml's' is followed by a '+ New Define.xml' button and a search bar. A table lists existing Define.xml files with columns for Project, Study, and Data Package.

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

STEP 3. IMPORT THE SPREADSHEET TO PINNACLE 21

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



STEP 3. IMPORT THE SPREADSHEET TO PINNACLE 21

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

The screenshot displays the Pinnacle 21 web application interface for creating a new Define.xml file. The sidebar on the left contains the following navigation items: Home, Configure, Design Studies, Define.xml (highlighted), Reviewer's Guide, Manage Metadata, Validate, and Report. The main content area shows the 'Create A New Define.xml' wizard with a progress bar indicating four steps: Step 1: Create Option (active), Step 2: Data Package, Step 3: Select Details, and Step 4: Conf. The form is divided into three sections: Project Information, Study Information, and Data Package Information. The Project Information section contains a 'Project Name' field with the value 'ADaM'. The Study Information section contains a 'Study Name' field with the value 'ADaM007'. The Data Package Information section contains a 'Data Package Name' field with the value 'CNT01959PSO3006-def2', and four dropdown menus for 'Standard' (ADaM-IG 1.0), 'MedDRA' (20.0), 'SDTM CT' (Janssen_3.2_rel_test_22/02/2018 JANSSEN_2...), and 'ADaM CT' (2017-03-31).

Section	Field	Value
Project Information	Project Name	ADaM
	Study Name	ADaM007
Data Package Information	Data Package Name	CNT01959PSO3006-def2
	Standard	ADaM-IG 1.0
	MedDRA	20.0
	SDTM CT	Janssen_3.2_rel_test_22/02/2018 JANSSEN_2...
	ADaM CT	2017-03-31

STEP 3. IMPORT THE SPREADSHEET TO PINNACLE 21

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

The screenshot shows the Pinnacle 21 web interface for creating a new Define.xml file. The left sidebar contains navigation links: Home, Configure, Design Studies, Define.xml (selected), Reviewer's Guide, Manage Metadata, Validate, and Report. The main content area is titled 'Create A New Define.xml' and features a progress bar with four steps: Step 1: Create Option, Step 2: Data Package, Step 3: Select Details (highlighted), and Step 4: Confirm. Below the progress bar, the text reads 'Import Define Excel Specification' and 'Please select an Excel (.xlsx) file to import.' A file input field shows the filename '56021927PCR1007-def2.xls' with a 'Browse...' button next to it. Below the input field, there is a link: 'If you need a blank template, download it here.' At the bottom of the screen, there are 'Back' and 'Next' buttons.

STEP 3. IMPORT THE SPREADSHEET TO PINNACLE 21

5. 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

STEP 3. IMPORT THE SPREADSHEET TO PINNACLE 21

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

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

Define: ADaM > ADaM007 > 56021927...

+ Import Export Copy Compare

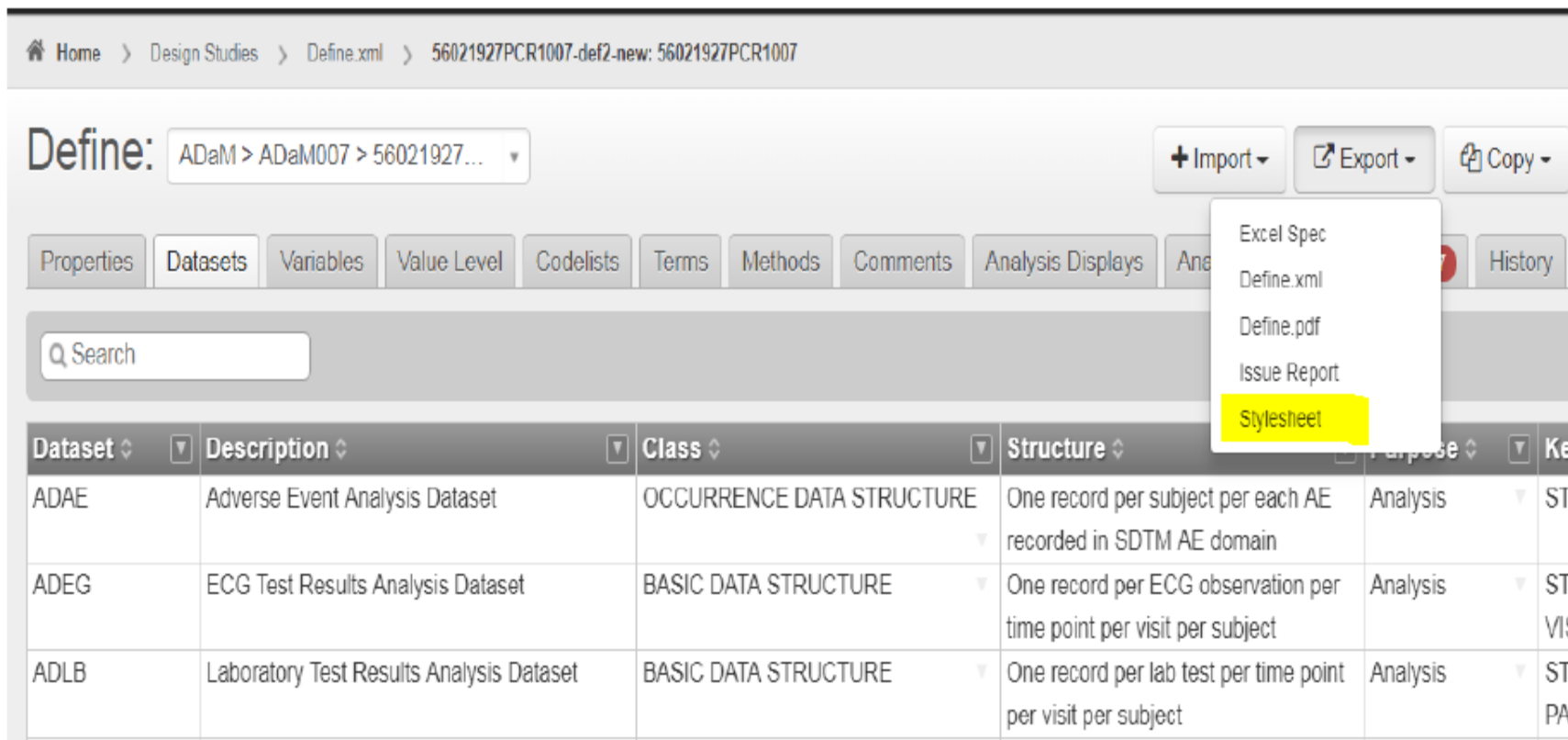
Properties Datasets Variables Value Level Codelists Terms Methods Comments Analysis Displays Analysis Results Issues 7 History

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, EGS
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

STEP 3. IMPORT THE SPREADSHEET TO PINNACLE 21

7. Export define.xml and stylesheet:



The screenshot shows the Pinnacle 21 Define.xml editor interface. The breadcrumb navigation at the top reads: Home > Design Studies > Define.xml > 56021927PCR1007-def2-new: 56021927PCR1007. The 'Define:' dropdown menu is set to 'ADaM > ADaM007 > 56021927...'. On the right, there are buttons for '+ Import', 'Export', and 'Copy'. Below these is a tabbed interface with 'Properties', 'Datasets', 'Variables', 'Value Level', 'Codelists', 'Terms', 'Methods', 'Comments', 'Analysis Displays', and 'Analysis'. The 'Datasets' tab is active. A search bar is located below the tabs. A dropdown menu is open from the 'Export' button, showing the following options: 'Excel Spec', 'Define.xml', 'Define.pdf', 'Issue Report', and 'Stylesheet'. The 'Stylesheet' option is highlighted in yellow. Below the menu is a table with the following data:

Dataset	Description	Class	Structure	Purpose	Key
ADAE	Adverse Event Analysis Dataset	OCCURRENCE DATA STRUCTURE	One record per subject per each AE recorded in SDTM AE domain	Analysis	STI
ADEG	ECG Test Results Analysis Dataset	BASIC DATA STRUCTURE	One record per ECG observation per time point per visit per subject	Analysis	STI VIS
ADLB	Laboratory Test Results Analysis Dataset	BASIC DATA STRUCTURE	One record per lab test per time point per visit per subject	Analysis	STI PAI

STEP 3. IMPORT THE SPREADSHEET TO PINNACLE 21

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

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)

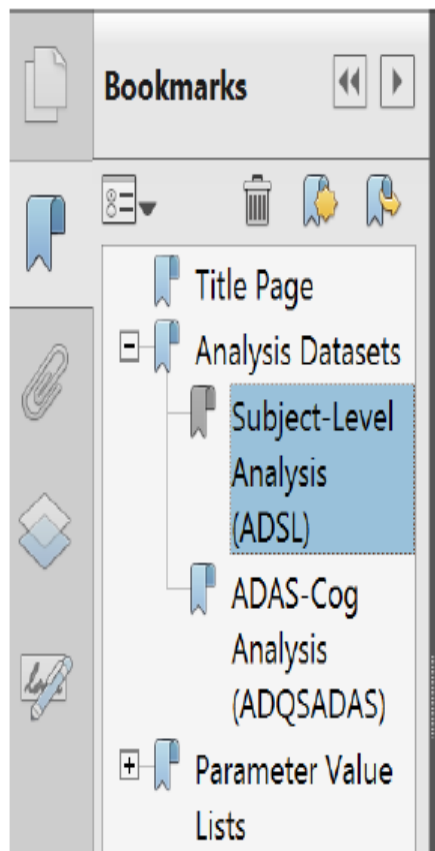
Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
ADSL	Subject-Level Analysis	SUBJECT LEVEL ANALYSIS DATASET	one record per subject	Analysis	USUBJID	adsl.xpt	Screen Failures are excluded since they are not needed for this study analysis
ADQSADAS	ADAS-Cog Analysis	BASIC DATA STRUCTURE	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

Sample define.pdf



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
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.

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