

Paper SA08

Efficient Conversion of SDTM/ADaM Specifications from Excel to Define.xml v2.1 Using Pinnacle 21 Community

Hiromi Okada, Japan Tobacco Inc., Tokyo, Japan
Seiko Yamazaki, Certara, Tokyo, Japan

ABSTRACT

Define.xml is a critical component of electronic data submissions, providing detailed metadata descriptions for clinical trial datasets, particularly those conforming to the SDTM and ADaM standards.

While many professionals develop their dataset specifications in Excel during the early stages of implementation, generating the Define.xml file can be challenging due to its technical complexity and strict formatting requirements. This is especially true for users with limited experience in XML structure and regulatory standards.

To streamline this process, we utilized Pinnacle 21 Community®, a widely adopted tool in clinical data standardization, to convert Excel-based specifications into a compliant Define.xml v2.1 file.

We also explored the use of Pinnacle 21 Enterprise®, the commercial version of the tool.

This presentation outlines the step-by-step procedure used for the conversion and shares practical tips and best practices to assist others in efficiently producing high-quality Define.xml files for regulatory submission.

INTRODUCTION

When submitting electronic data to regulatory agencies, Define.xml is a mandatory document as metadata for required datasets such as SDTM and ADaM. It is the responsibility of the data submitter to create this document appropriately. In this paper, we considered the following scenarios:

- In a trial involving a product acquired from another company, we had already obtained the SDTM and ADaM datasets and analysis reports, but the Define.xml had not been created. Therefore, it became necessary to additionally create the Define.xml for submission purposes.
- In a past trial, it was initially assumed that the study would not be included in the submission package. As a result, while the SDTM and ADaM datasets and analysis reports were prepared, the sudden decision to include them in the package necessitated the additional creation of the Define.xml.

There are several ways to create define.xml for free, but we decided to use Pinnacle 21 Community as a useful no-code tool for this paper. Our approach involves registering XPT files in Pinnacle 21 Community to generate an initial Define.xml Excel spec, which is then corrected and populated with necessary information. This updated Excel spec is subsequently used to generate the Define.xml file, followed by a validation process to ensure it meets regulatory requirements. The Define.xml file based on the Define.xml Excel spec generated by Pinnacle 21 Community after reading XPT files is incomplete because for example, some fields that should be filled may be left blank, or the necessary content may not be fully covered. Therefore, to create a submission-ready Define.xml, it is necessary to manually and appropriately update the Define.xml Excel spec. This paper highlights the importance of understanding the roles of each sheet and column in the Define.xml Excel spec, as certain columns significantly impact validation outcomes.

METHOD

The xpt files were registered with Pinnacle 21 Community, and a Define.xml Excel spec was exported. Necessary information was added to the exported Excel spec. The updated xlsx file was then re-registered with Pinnacle 21 Community to create a Define.xml file. The step-by-step process for generating Define.xml Excel spec and define.xml is illustrated in [Dutta, 2018]. Validation of the created Define.xml was performed using Pinnacle 21 Community.

RESULT

By updating the Define.xml Excel spec exported from Pinnacle 21 Community, we were able to successfully create a Define.xml file. Validation of the created Define.xml had no issues related to severity of rejection in the PMDA's validation rules. Although some errors remained, they could be explained in the Reviewer's Guide, and we determined that the minimum requirements necessary for submission were met. Although the creation of Define.xml involves numerous necessary elements and a substantial amount of information, it was crucial to understand the role of each sheet and column in the Define.xml Excel spec to accurately input the required information.

ROLES OF EACH SHEET AND COLUMN

The following provides a column-by-column summary of the content that is pre-generated in the Define.xml Excel spec upon registering the xpt file, as well as the content that needs to be reviewed and updated.

DEFINE SHEET

Column label	Row label	Notes (Input guideline)
Attribute	StudyName	Provide a simplified expression of the study name, or the protocol number if available
	StudyDescription	Provide study description
	ProtocolName	Provide a simplified expression of the protocol name, or the protocol number if available
	StandardName	Provide standard name For example: SDTMIG, ADaMIG
	StandardVersion	Provide standard version
	Language	en

Displayed in Define.xml

The following outlines the information displayed in the Define.xml when it is generated using the Define.xml Excel spec with Pinnacle 21 Community. The red text in the figures below represents the corresponding *sheet name.column label* in Define.xml.Excel.spec. However, for the define sheet, it is represented as *sheet name.rowlabel*.

Date/Time of Define-XML document generation: 2025-07-23T17:54:27+09:00

Define-XML version: 2.1.0

Define-XML Context: Submission

Stylesheet version: 2019-02-11

Study Name

QBC4-1

Define.StudyName

Study Description

Phase 3 Clinical Study of JTE-052 Lotion - Bioequivalence Study of JTE-052 Lotion and JTE-052 Ointment in Patients with Atopic Dermatitis -

Define.StudyDescription

Protocol Name

QBC4-1

Define.ProtocolName

Metadata Name

Study QBC4-1 Data Definitions

Standards for Study QBC4-1

Standard	Type	Status	Documentation
CDISC/NCI ADaM 2016-12-16	CT	FINAL	
ADaMIG 1.3	IG	FINAL	
CDISC/NCI SDTM 2016-12-16	CT	FINAL	

DATASETS SHEET

Column label	Pinnacle 21 Community	Notes	For SDTM	For ADaM
Dataset	Output based on the xpt file			
Label	Output based on the xpt file			
Class	Populated based on the xpt file	If necessary, update the class information		
SubClass			The subclass is not defined for SDTM at this moment. Refer to CDISC documents for the updates.	The subclass is defined in Define CT, so it is necessary to define the subclass when creating the corresponding dataset. Refer to: https://evs.nci.nih.gov/ftp1/CDISC/Define-XML/Define-XML%20Terminol

Column label	Pinnacle 21 Community	Notes	For SDTM	For ADaM
				ogy.html#CL.C165635.BDSSC
Structure	Populated based on the version of Standard set in the Configuration	If necessary, update the structure information		
Key Variables	Populated based on the version of Standard set in the configuration	If necessary, update the Key variables information		
Standard	The IG version configured in Configuration is displayed in the Standard column	Validation is performed using the CDISC standard version set in the Standard column. If there are domains referencing a different IG version than the one in Configuration, they need to be corrected.		
Has No Data	When setting the XPT as the source data for the define.xml generator to create the define, the define.xml is created based on the existing XPT	If the final dataset is empty, it needs to be set to "Yes". In such cases, make sure to provide a reason in the comment column explaining why the data was not collected. Refer to: Define-XML version 2.1.pdf, 5.3.11 ItemGroupDef Element, def:HasNoData		
Repeating	Output based on the xpt file	Ensure compliance with the following defined rules and make appropriate corrections: - Set to "Yes" if the dataset may contain more than 1 record per subject or pool - set to "No" if the dataset may contain only 1 record per subject or pool - Business Rule: When IsReferenceData="Yes", set Repeating="No" Refer to: Define-XML version 2.1.pdf, 5.3.11 ItemGroupDef Element, Repeating		

Column label	Pinnacle 21 Community	Notes	For SDTM	For ADaM
Reference Data	Output based on the xpt file	Ensure compliance with the following defined rules and make appropriate corrections: Indicates whether the dataset contains reference data (not subject data) Refer to: Define-XML version 2.1.pdf, 5.3.11 ItemGroupDef Element, IsReferenceData		
Comment		Provide Comment ID when needed In cases where the "Has no data" column is marked as "Yes," there is a business rule that requires a comment explaining the reason for the absence of data The reason why the data was not collected should be noted in the Comment column Refer to: Define-XML version 2.1.pdf, 5.3.11 ItemGroupDef Element, def:HasNoData		
Developer Notes		The content entered in this column does not affect the define.xml, so it can be used for information sharing among stakeholders within the spec file for any notes		

Displayed in Define.xml SDTM

Datasets							
Dataset	Description	Class	Structure	Purpose	Keys	Documentation	Location
IA [SDTMIG 3.3]	Trial Arms Datasets.Label	TRIAL DESIGN	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD		ta.xpt
IE [SDTMIG 3.3]	Trial Elements	TRIAL DESIGN	One record per planned Element Datasets.Structure	Tabulation	STUDYID, ETCO Datasets.Key Variables		te.xpt
II [SDTMIG 3.3]	Trial Inclusion/Exclusion Criteria	TRIAL DESIGN	One record per I/E criterion	Tabulation	STUDYID, IETESTCD		ti.xpt
IS [SDTMIG 3.3]	Trial Summary	TRIAL DESIGN	One record per trial summary parameter value	Tabulation	STUDYID, TSPARMCD, TSSEQ		ts.xpt
IV [SDTMIG 3.3]	Trial Visits	TRIAL DESIGN	One record per planned Visit per Arm	Tabulation	STUDYID, VISITNUM, ARMCD		tv.xpt
CO [SDTMIG 3.3]	Comments Datasets.Dataset	SPECIAL PURPOSE	One record per comment per subject	Tabulation	STUDYID, RDOMAIN, USUBJID, COSEQ		co.xpt
DM [SDTMIG 3.3]	Demographics Datasets.Standard	SPECIAL PURPOSE	One record per subject	Tabulation	STUDYID, USUBJID		dm.xpt
SE [SDTMIG 3.3]	Subject Elements	SPECIAL PURPOSE	One record per actual Element per subject	Tabulation	STUDYID, USUBJID, ETCO, SESTDTC		se.xpt
SV [SDTMIG 3.3]	Subject Visits	SPECIAL PURPOSE	One record per subject per actual visit	Tabulation	STUDYID, USUBJID, VISITNUM		sv.xpt
CM [SDTMIG 3.3]	Concomitant Medications	INTERVENTIONS	One record per recorded intervention occurrence or constant-dosing interval per subject	Tabulation	STUDYID, USUBJID, CMTRT, CMSTDTC		cm.xpt
EC [SDTMIG 3.3]	Exposure as Collected	INTERVENTIONS	One record per protocol-specified study treatment, collected-dosing interval, per subject, per mood	Tabulation	STUDYID, USUBJID, ECTRT, ECSTDTC		ec.xpt
EX [SDTMIG 3.3]	Exposure	INTERVENTIONS	One record per protocol-specified study treatment, constant-dosing interval, per subject	Tabulation	STUDYID, USUBJID, EXTRT, EXSTDTC		ex.xpt
PR [SDTMIG 3.3]	Procedures	INTERVENTIONS	One record per recorded procedure per occurrence per subject	Tabulation	STUDYID, USUBJID, PRTRT, PRSTDTC		pr.xpt
AE [SDTMIG 3.3]	Adverse Events	EVENTS Datasets.Class	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AEDECOD, AESTDTC		ae.xpt
DS [SDTMIG 3.3]	Disposition	EVENTS	One record per disposition status or protocol milestone per subject	Tabulation	STUDYID, USUBJID, DSDECOD, DSSTDTC		ds.xpt

ADaM

Datasets							
Dataset	Description	Class - SubClass	Structure	Purpose	Keys	Documentation	Location
ADSL [ADaMIG 1.3] Datasets.Dataset	Subject-Level Analysis Dataset Datasets.Label	SUBJECT LEVEL ANALYSIS DATASET Datasets.Class	One record per subject Datasets.Structure	Analysis	STUDYID, USUBJID Datasets.Key Variables		adsl.xpt
ADEX [ADaMIG 1.3] Datasets.Standard	Exposure Analysis Dataset	BASIC DATA STRUCTURE	One or more records per subject per analysis parameter per analysis timepoint	Analysis	STUDYID, USUBJID, PARAM, PARAMCD, AVAL		adex.xpt
ADLB [ADaMIG 1.3]	Laboratory Tests Analysis Dataset	BASIC DATA STRUCTURE	One or more records per subject per analysis parameter per analysis timepoint	Analysis	STUDYID, USUBJID, PARAM, PARAMCD, AVAL		adlb.xpt
ADPC [ADaMIG 1.3]	Plasma Concentration Analysis Dataset	BASIC DATA STRUCTURE	One or more records per subject per analysis parameter per analysis timepoint	Analysis	STUDYID, USUBJID, PARAM, PARAMCD, AVAL		adpc.xpt
ADSR [ADaMIG 1.3]	Skin Response Analysis Dataset	BASIC DATA STRUCTURE	One or more records per subject per analysis parameter per analysis timepoint	Analysis	STUDYID, USUBJID, PARAM, PARAMCD, AVAL		adsr.xpt
ADVS [ADaMIG 1.3]	Vital Signs Analysis Dataset	BASIC DATA STRUCTURE	One or more records per subject per analysis parameter per analysis timepoint	Analysis	STUDYID, USUBJID, PARAM, PARAMCD, AVAL	Datasets.Comment	advx.xpt
ADEXC [ADaMIG 1.3] [No Data] Datasets.Has No Data =Yes	Exclusion Data Analysis Dataset	ADAM OTHER	One record per subject per exclusion data	Analysis	STUDYID, EXCKVAL1, EXCKVAL2, EXCKVAL3, EXCKVAL4	Follow Exclusion Criteria List Exclusion Criteria List	
ADAE [ADaMIG 1.3]	Adverse Events Analysis Dataset	OCCURRENCE DATA STRUCTURE Datasets.SubClass • ADVERSE EVENT	One record per subject per adverse event	Analysis	STUDYID, USUBJID	Document ID by Comments.Document	adae.xpt

Go to the [top](#) of the Define-XML document

VARIABLES SHEET

Column label	Pinnacle 21 Community	Notes	For SDTM	For ADaM
Order	The output follows the variable order of the xpt file			
Dataset	The dataset label name from the xpt file is displayed			
Variable	The variable names from the xpt file are displayed			
Label	The variable label names from the xpt file are displayed			
Data Type	The variable attributes from the xpt file are displayed			
Length	The variable length from the xpt file is displayed	The variable length in characters Business Rules: • Length is the maximum expected variable length. It is not based on the length of the data collected • Length should be present only when DataType is "Text", "Integer", or "Float" • For ItemDef elements referenced from a ValueListDef, the Length must be less than or equal to the length of the parent variable Refer to: Define-XML version 2.1.pdf, 5.3.12 ItemDef Element, Length		
Significant Digits	The variable length from the xpt file is displayed	When the Data Type is "Float," it is necessary to specify the number of decimal places.		

Column label	Pinnacle 21 Community	Notes	For SDTM	For ADaM
Format	The display format set in the XPT dataset is outputted	It is necessary to verify that the display format is not applied to anything other than Float variables and ADaM SAS date variables Refer to: Define-XML version 2.1.pdf, 5.3.12 ItemDef Element, def:DisplayFormat		
Mandatory	For variables where retention is mandatory and null values are not allowed, the value should be set to "Yes"	Refer to: Define-XML version 2.1.pdf, 5.3.9.2 ItemRef Element, Mandatory	If a variable corresponds to Core=Required in the SDTMIG, it should be output as "Yes" This indicates that the variable is mandatory and cannot have null values according to the SDTM IG	The definition of Core=Required in the ADaMIG only mandates the inclusion of the variable and does not address the issue of null values There is no need to change it to "Yes"
Assigned Value		The content entered in this column does not affect the define.xml, so it can be used for information sharing among stakeholders within the spec file to show what value should be assigned for the variable.		
Codelist		When storing terms defined in the Codelists sheet in the variable, enter the ID defined in the Codelists sheet		
Common		keep null		
Origin		Provide the name according to Define-XML v2.1.pdf, 4.3.2.1 Origin for SDTM Datasets, 4.3.2.2 Origin for ADaM Datasets	Collected or Derived or Assigned or Protocol or Predecessor	Derived or Assigned or Predecessor
Source		Provide the name according to Define-XML v2.1.pdf, 4.3.2.1 Origin for SDTM Datasets, 4.3.2.2 Origin for ADaM Datasets	Subject or Investigator or Vendor when Origin = Collected Vendor or Sponsor when Origin = Derived Vendor or Sponsor when Origin = Assigned Sponsor when Origin = Protocol	Sponsor when Origin = Derived or Assigned

Column label	Pinnacle 21 Community	Notes	For SDTM	For ADaM
Pages			Provide acrf Page No. when Origin = Collected and Source = Investigator When entering multiple pages, insert a space between the numbers	
Method		Provide Method ID when Origin = Derived		
Predecessor			Use when a value is an exact copy of another value in an SDTM dataset Refer to: Define-XML version 2.1.pdf, 4.3.2.1 Origin for SDTM Dataset, Predecessor	Provide Dataset and Variable names when a value is an exact copy of another value in either SDTM or ADaM dataset. Refer to: Define-XML version 2.1.pdf, 4.3.2.2 Origin for ADaM Datasets
Role	Populated based on the version of Standard set in the Configuration	If necessary, update the role information		
Has No Data	Output based on the xpt file	If the final value of the variable is empty, it needs to be set to "Yes". In such cases, make sure to provide a reason in the Comment column explaining why the data was not collected. Refer to: Define-XML version 2.1.pdf, 5.3.9.2 ItemRef Element, def:HasNoData		
Comment		Provide Comment ID when needed. In cases where the "Has no data" column is marked as "Yes," there is a business rule that requires a comment explaining the reason for the absence of data. The reason why the data was not collected should be noted in the Comment column. Note that when Source is Assigned, a Comment should be present for the variable.		

Column label	Pinnacle 21 Community	Notes	For SDTM	For ADaM
		Refer to: Define-XML version 2.1.pdf, 5.3.11 ItemGroupDef Element, def:HasNoData		
Developer Notes		The content entered in this column does not affect the define.xml, so it can be used for information sharing among stakeholders within the spec file for any notes		

Displayed in Define.xml

SDTM

DM (Demographics) - [SDTMIG 3.3]							Location: dm.xpt
Related Supplemental Qualifiers Dataset: SUPPDM (Supplemental Qualifiers for DM)							
Variable	Where Condition	Label / Description	Type	Role	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
STUDYID Variables.Variable		Study Identifier Variables.Label	text	Identifier Variables.Datatype	6	Variables.Length or Variables.Format	Protocol (Source: Sponsor) Variables.Origin and Variables.Source
DOMAIN		Domain Abbreviation	text	Identifier Variables.Role	2	SDTM.Domain Abbreviation [20 Terms] Codelist.Name by Variables.Codelist	Assigned (Source: Sponsor)
USUBJID		Unique Subject Identifier	text	Identifier	13		Variables.Origin and Variables.Source Derived (Source: Sponsor) Variables.Method Concatenation of STUDYID- SUBJID.
SUBJID		Subject Identifier for the Study	text	Topic	6		Variables.Origin and Variables.Source Collected (Source: Investigator) Annotated Case Report Form [2] Variables.Pages
RFSTDTC		Subject Reference Start Date/Time	datetime	Record Qualifier		ISO 8601	Derived (Source: Sponsor) Minimum date in start date of treatment.
RFENDTC		Subject Reference End Date/Time	datetime	Record Qualifier		ISO 8601	Derived (Source: Sponsor) Maximum date in end date of treatment.

ADaM

ADSL (Subject-Level Analysis Dataset) - [ADaMIG 1.3]					Location: adsl.xpt 
Variable	Label / Description	Type	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
STUDYID	Study Identifier <i>Variables.Variable Variables.Label Variables.Data Type</i>	text	6	<i>Variables.Length or Variables.Format</i>	Predecessor: DM.STUDYID <i>Variables.Origin: Variables.Predecessor</i>
USUBJID	Unique Subject Identifier	text	13		Predecessor: DM.USUBJID
SUBJID	Subject Identifier for the Study	text	6	<i>Codelists.Name by Variables.Codelist</i>	Predecessor: DM.SUBJID
SITEID	Study Site Identifier	text	2 [6 Terms]	<i>Site</i>	Predecessor: DM.SITEID
AGE	Age	integer	8		Predecessor: DM.AGE
AGEU	Age Units	text	5	<i>Age Unit</i> • "YEARS" = "Year" <i>Codelists.Term = Codelists.Decoded Value</i>	Predecessor: DM.AGEU
AGEGR1	Pooled Age Group 1	text	12	<i>Pooled Age Group 1</i> • "< 20" • "20 <= - < 30" • "30 <= - < 45" • "45 <="	Derived (Source: Sponsor) <i>Variables.Origin and Variables.Source: Variables.Method</i> if .<AGE<20 then "< 20" else if 20<=AGE<30 then "20 <= - < 30" else if 30<=AGE<45 then "30 <= - < 45" else if 45<=AGE then "45 <=" else null
AGEGR1N	Pooled Age Group 1 (N)	integer	8	<i>Pooled Age Group 1(N)</i> • 1 = "< 20" • 2 = "20 <= - < 30" • 3 = "30 <= - < 45" • 4 = "45 <="	Assigned (Source: Sponsor) Numeric code for AGEGR1 <i>Variables.Origin and Variables.Source: Variables.Comment</i>
SEX	Sex	text	1	<i>Sex</i> • "F" • "M" • "U"	Predecessor: DM.SEX
RACE	Race	text	5	<i>Race</i> • "ASIAN"	Predecessor: DM.RACE
UCONDT	Underlying Condition	text	40	<i>MedDRA</i> MedDRA 26.1	<i>Dictionaries.Name by Variables.Codelist</i> Assigned (Source: Sponsor)

VALUELEVEL SHEET

This sheet requires all columns to be filled in manually because Pinnacle 21 Community does not output values for it.

Column label	Notes	For SDTM	For ADaM																																	
Order	Provide the value level order																																			
Dataset	Provide the dataset name																																			
Variable	Provide the variable name																																			
Where Clause	Provide the conditional statement For example, as shown below: <table><tr><th>Dataset</th><th>Variable</th><th>Where Clause</th></tr><tr><td>LB</td><td>LBORRES</td><td>LBTESTCD EQ WBC</td></tr><tr><td>LB</td><td>LBORRES</td><td>LBTESTCD EQ RBC</td></tr><tr><td>LB</td><td>LBORRES</td><td>LBTESTCD EQ HGB</td></tr><tr><td>LB</td><td>LBORRES</td><td>LBTESTCD EQ HCT</td></tr><tr><td>LB</td><td>LBORRES</td><td>LBTESTCD EQ NEUTLE</td></tr><tr><td>LB</td><td>LBORRES</td><td>LBTESTCD EQ LYMLE</td></tr><tr><td>LB</td><td>LBORRES</td><td>LBTESTCD EQ MONOLE</td></tr><tr><td>LB</td><td>LBORRES</td><td>LBTESTCD EQ EOSLE</td></tr><tr><td>LB</td><td>LBORRES</td><td>LBTESTCD EQ BASOLE</td></tr><tr><td>LB</td><td>LBORRES</td><td>LBTESTCD EQ PLAT</td></tr></table>	Dataset	Variable	Where Clause	LB	LBORRES	LBTESTCD EQ WBC	LB	LBORRES	LBTESTCD EQ RBC	LB	LBORRES	LBTESTCD EQ HGB	LB	LBORRES	LBTESTCD EQ HCT	LB	LBORRES	LBTESTCD EQ NEUTLE	LB	LBORRES	LBTESTCD EQ LYMLE	LB	LBORRES	LBTESTCD EQ MONOLE	LB	LBORRES	LBTESTCD EQ EOSLE	LB	LBORRES	LBTESTCD EQ BASOLE	LB	LBORRES	LBTESTCD EQ PLAT		
Dataset	Variable	Where Clause																																		
LB	LBORRES	LBTESTCD EQ WBC																																		
LB	LBORRES	LBTESTCD EQ RBC																																		
LB	LBORRES	LBTESTCD EQ HGB																																		
LB	LBORRES	LBTESTCD EQ HCT																																		
LB	LBORRES	LBTESTCD EQ NEUTLE																																		
LB	LBORRES	LBTESTCD EQ LYMLE																																		
LB	LBORRES	LBTESTCD EQ MONOLE																																		
LB	LBORRES	LBTESTCD EQ EOSLE																																		
LB	LBORRES	LBTESTCD EQ BASOLE																																		
LB	LBORRES	LBTESTCD EQ PLAT																																		
Label		Provide the label name for SUPP domain and non-standard variables Refer to: Define-XML version 2.1.pdf, 5.3.9.1 Description Element	Provide the variable label name																																	
Data Type	Provide the value level attribute																																			
Length	As stated in Variables.Length																																			
Significant Digits	As stated in Variables.SignificantDigits																																			
Format	As stated in Variables.SignificantDigits																																			
Mandatory	As stated in Variables.Mandatory																																			

Column label	Notes	For SDTM	For ADaM
Assigned Value	As stated in Variables.Assigned Value		
Codelist	As stated in Variables.Codelist		
Origin	As stated in Variables.Origin		
Source	As stated in Variables.Source		
Pages	As stated in Variables.Pages		
Method	As stated in Variables.Method		
Predecessor	As stated in Variables.Predecessor		
Comment	As stated in Variables.Comment		
Developer Notes	As stated in Variables.DeveloperNotes		

CODELISTS SHEET

This sheet requires all columns to be filled in manually because Pinnacle 21 Community does not output values for it.

Column label	Notes
ID	Provide Codelist ID that makes the unique within the codelist
Name	Provide Codelist name
NCI Codelist Code	Provide NCI codeslist code or null when no applicable codelist
Data Type	Provide the variable attribute
Terminology	Provide a CT version when referring to NCI codes
Comment	Provide Comment ID when needed
Order	Provide the Codelist order Your codelist's terms will be displayed they are specified here
Term	Provide the Codelist term
NCI Term Code	Provide NCI term code or null when no applicable codelist
Decoded Value	Provide the Codelist Decoded Value

DICTIONARIES SHEET

This sheet requires all columns to be filled in manually because Pinnacle 21 Community does not output values for it.

Column label	Notes
ID	Provide unique ID
Name	Provide the full name of the dictionary
Data Type	Provide Data Type of the values in the dictionary
Dictionary	"CDISC Submission Value (=CodedValue) as listed in the NCI/CDISC Controlled Terminology, CodeList Dictionary Name, NCI Code C66788. This CodeList is extensible Samples: COSTART, ICD, LOINC, MedDRA, SNOMED, WHOART, WHODD Refer to: Define-XML version 2.1.pdf, 5.3.13.4 ExternalCodelist Element, Dictionary
Version	The version designator of the external codelist

METHODS SHEET

This sheet requires all columns to be filled in manually because Pinnacle 21 Community does not output values for it.

Column label	Notes
ID	Provide method ID that makes the unique within the methods sheet
Name	Provide Method name
Type	Provide either "Computation" or "Imputation" Refer to: Define-XML version 2.1.pdf, 5.3.14 MethodDef Element, Type
Description	Provide description that explains the algorithm used to derive the variable in human-readable text
Expression Context	The environment/programming language and version in which the code is written. The link will lead to the relevant line in the Method section, and it will be displayed in the following format after the expression set as the description: Formal Expression [<i>Expression Context</i>] <i>Expression Code</i>
Expression Code	The executable programming code
Document	If the details mentioned in the Method are referenced in a document, provide the corresponding DocumentsID
Pages	If the details mentioned in the Method are referenced in a document, provide the corresponding page of the document

COMMENTS SHEET

This sheet requires all columns to be filled in manually because Pinnacle 21 Community does not output values for it.

Column label	Notes
ID	Provide Comment ID that makes the unique within the comments sheet
Description	Provide description that explains how values are assigned.
Document	If the details mentioned in the Comment are referenced in a document, provide the corresponding DocumentsID
Pages	If the details mentioned in the Comment are referenced in a document, provide the corresponding page of document

DOCUMENTS SHEET

This sheet requires all columns to be filled in manually because Pinnacle 21 Community does not output values for it.

Column label	Notes
ID	Provide Document ID that makes the unique within the documents sheet
Title	Provide the name of the document
Href	Provide URL that can be used to identify the location of a document

ANALYSIS DISPLAYS SHEET

This sheet requires all columns to be filled in manually because Pinnacle 21 Community does not output values for it.

Column label	Notes
ID	Provide AnalysisID that makes the unique within the analysis displays sheet and the analysis results sheet
Title	Provide the title for display in define
Document	Provide the corresponding DocumentsID for the analysis displays

Column label	Notes
Pages	Enter the pages within the Document identified in the Document column on which the Analysis Display is discussed

ANALYSIS RESULTS SHEET

This sheet requires all columns to be filled in manually because Pinnacle 21 Community does not output values for it.

Column label	Notes
Display	Provide the ID set in the Analysis Displays sheet. This value will be displayed as the table number in the ARM report
ID	A unique ID for the report selected in Display. When creating ARM for multiple sections in a single report, set an ID that makes each row unique"
Description	Provide description of Analysis Result for display in define
Variables	Specify variables used for the analysis at the two levels: dataset name and variable name When listing multiple variables, separate them with commas
Reason	Extensible ADaM Controlled Terminology (Name: Analysis Reason, CDISC Submission Value: ANLREAS, Code: C117744) "SPECIFIED IN PROTOCOL" "SPECIFIED IN SAP" "DATA DRIVEN" "REQUESTED BY REGULATORY AGENCY" The rationale for performing this analysis. It indicates when the analysis was planned. Business rule: If a specific analysis was prespecified in both protocol and SAP, specify "SPECIFIED IN SAP" as the analysis reason. Refer to: Analysis Results Metadata v1.0 for Define-XML v2.pdf, 5.3.5 arm:AnalysisResult Element, AnalysisReason
Purpose	"Extensible ADaM Controlled Terminology (Name: Analysis Purpose, CDISC Submission Value: ANLPURP, Code: C117745): PRIMARY OUTCOME MEASURE", SECONDARY OUTCOME MEASURE", EXPLORATORY OUTCOME MEASURE" The purpose of the analysis within the body of evidence (e.g., section in the clinical study report). Refer to: Analysis Results Metadata v1.0 for Define-XML v2.pdf, 5.3.5 arm:AnalysisResult Element, AnalysisPurpose
Selection Criteria	Specify the conditions used for analysis ex. ADQSADAS[PARAMCD IN ("ACITM01 ACITM02 ACITM03") and AVISIT EQ "Week 24" and EFFFL EQ Y and ANL01FL EQ Y]
Join Comment	Explain how you created the report using two or more datasets. (For example, when creating an adverse events report using ADAE and ADSL, ADSL's SAFFL=Y was used as the denominator and ADAE's TRTEMFL=Y was used as the numerator, etc.)
Documentation	Describe the analysis performed in written form, rather than using programming language
Documentation Refs	Specify the document and page where additional context or detailed explanations regarding the analysis can be found. Enter the ID set in the Documents sheet and put the page number in parentheses ex. SAP-SEC-10.1.1 (4)

Column label	Notes
Programming Context	Provide the name and version of the programming language in which the analysis is described in the Programming Code column
Programming Code	Provide executable programming code using the correct language and version specified in the Programming Context column
Programming Document	When providing a document relating to programming for the display, provide the corresponding DocumentsID
Pages	Provide the corresponding page of document specified in Programming Document column

Displayed in Define.xml

Analysis Results Metadata - Summary	
Analysis Displays.ID Table14.2.1.1 Percent Change in mEASI Score and Results of the ANCOVA at the End of Treatment	Analysis Displays.Title Percent Change in mEASI Score and Results of the ANCOVA at the End of Treatment
Analysis Displays.ID Table14.2.2.1 Descriptive Statistics for mEASI Score	Analysis Displays.Title Descriptive Statistics for mEASI Score

Analysis Results Metadata - Detail	
Table14.2.1.1	
Display	Percent Change in mEASI Score and Result of the ANCOVA at the End of Treatment Analysis Displays.Title
Analysis Result	Percent Change in mEASI Score and Results of the ANCOVA at the End of Treatment Analysis Results.Description
Analysis Parameter(s)	PARAMCD = "MEASI" (mEASI Score) Analysis Results.Selection Criteria
Analysis Variable(s)	ADSR_PCHG (Percent Change from Baseline) Analysis Results.Variables
Analysis Reason	SPECIFIED IN SAP Analysis Results.Reason
Analysis Purpose	PRIMARY OUTCOME MEASURE Analysis Results.Purpose
Data References (incl. Selection Criteria)	ADSR [PARAMCD = "MEASI" and PROTRFL = "Y" and PROTRFL = "Y" and AVISIT = "EOT" and DTYPE = "PPS"] Analysis Results.Selection Criteria
Documentation	The PPS will be used. An analysis of covariance based on the model presented below will be performed on the percent change in mEASI score at the EOT to estimate the mean percent change from baseline in mEASI score in each treatment group and its 95% CI, and the intergroup difference (JTE-052 lotion 0.5% - JTE-052 ointment 0.5%) and its 95% CI Statistical Analysis Plan [22] Analysis Results.Documentation Refs Analysis Results.Documentation
Programming Statements	[SAS version 9.4] Analysis Results.Programming Context PROC GLM DATA = D_ADSR ; CLASS TRTPN ; model PCHG = BASE TRTPN / SS3 ; lsmeans TRTPN / CL pdiff=CONTROL('2') ALPHA = 0.05 ; ODS OUTPUT LSMeans=Rs1t_LSMeans LSMeanCL=Rs1t_LSMeanCL LSMeanDiffCL = Rs1t_LSMeanDiffCL ; quit; Analysis Results.Programming Code

Go to the [top](#) of the Define-XML document

DISCUSSION

Although define.xml can be easily created using the P21 Community Define.xml Generator and by addressing any issues identified during the validation, this does not signify the completion of the define.xml creation process. There are several points to pay attention to before finalizing define.xml.

The characteristics of a "high-quality define.xml" can be considered as follows:

1. Accuracy: The structure of datasets, variables, code lists, and other metadata are accurately described.
Example:
 - If variable attributes in a dataset are updated, that needs to be reflected in the corresponding variable metadata in define.xml.
2. Consistency: The information within the define.xml is consistent with the actual datasets and other related documents (e.g., data specifications, analysis plans).
Example:
 - If date imputation was planned in a statistical analysis plan, and it was implemented in an ADaM dataset, corresponding date imputation flag variables must be created.
 - When aCRF page number are updated, those pages are reflected in define.xml
3. Clarity: Definitions of variables and code lists are clear, allowing users to easily understand them.
Example:
 - If two independent programmers can understand the spec clearly enough that they can produce the same results.
 - Those regulatory reviewers can understand what you planned and what actually you did.
4. Completeness: All necessary metadata is included, with no omissions.
Example:

- Place the external documents into a specific file directory structure, click on the link and the document will open.
5. Standard Compliance: It adheres to CDISC standards (e.g., SDTM, ADaM).
Example:
 - Required variables are present in datasets and you are not creating another variable even though the standard variable is already defined in its implementation guide.
 - Use a standard variable for different intentions.
 6. Readability: The XML file is properly formatted, making it readable and understandable by humans when necessary.
 7. Validation: The XML has technically correct syntax and does not produce errors when using validation tools.

When using Pinnacle 21 Community to register xpt files and to update the output Excel files, attention must be paid to points 1-5 above. However, points 6-7 are taken care of by the software, so users do not need to worry about them.

How should we address points 1-5? This is not a problem limited to the creation of define.xml. It is a common challenge shared with the issue of how to create SDTM and ADaM with high quality. In creating SDTM and ADaM, accurately describing the data structure, variables, code lists, and metadata of the source data and the data being created, in a manner consistent with related documents (such as study protocols and analysis plans), clearly contributes to the efficiency of stakeholders (specification creators, dataset creators, analysts, etc.). However, this is also the area where those involved in the work often struggle daily. In many cases, efficiency is sought by creating internal standards or using materials from past studies, but customization for each study and adaptation to CDISC standard revisions are indispensable, leading many stakeholders to engage in repeated trial and error.

By effectively utilizing Pinnacle 21 Community, not only can points 6-7 be naturally addressed, but points 4-5 can also be resolved by using it in conjunction with the validator.

Simply clearing the validation rules of the authorities does not easily achieve "high quality," but understanding the roles of each sheet and column explained here can lead to securing room for addressing points 1-2 when updating specs, as well as considering point 3.

P21 Community is free desktop software, and Pinnacle 21 Enterprise is a commercial version of the cloud tool. The following shows the major differences between them. Since Pinnacle 21 Enterprise offers greater functionality and reduces the need for manual updates, users may benefit from creating define.xml more efficiently to achieve those points described above.

	P21 Community	P21 Enterprise
Creation Method	Generate from excel spec	6 different options including generating from xpt files used for validation, importing.xlsx, importing xml, copying study, copying standard, or starting with a blank template and manually entering)
Preview	N/A	Can see the preview of define.xml by clicking "Preview" button
Validation	Need to run a validation including define.xml	Checks run automatically in define designer and the user can click on "view issues" which direct to the cells where the issues need to be resolved.
Compare	N/A	Can compare to previous versions, other standards, or other defines
Merge	N/A	Merge xpt info used for validation with the define file.
aCRF import	N/A	Page numbers are populated in Pages column in Variable and Value levels for SDTM define by importing acrf.pdf.
Versioning	N/A	Included in the system

CONCLUSION

The insights gained from this paper are particularly relevant for scenarios where Define.xml files are not readily available. However, dataset specifications can be obtained in other formats. By comprehensively understanding the roles of each sheet and column, the process of updating Define.xml Excel spec and generating a Define.xml using Pinnacle 21 Community becomes straightforward.

This research focuses on files compatible with Define.xml version 2.1, offering valuable insights that may also assist in updates from version 2.0. Through this paper, we aim to demonstrate that Pinnacle 21 Community provides a user-

friendly approach to Define.xml creation, enabling efficient and consistent outputs even for those who do not regularly engage in this process.

REFERENCES

Dutta, P. (2018). Generating Define.xml from Pinnacle 21 Community. PharmaSUG Paper AD-29 .
CDISC Define-XML Specification Version 2.1
Analysis Results Metadata Specification Version 1.0 for Define-XML Version 2

CONTACT INFORMATION

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

Author Name: Hiromi Okada

Company: Japan Tobacco Inc.

Email: hiromi.okada@jt.com

Author Name: Seiko Yamazaki

Company: Certara

Email: seiko.yamazaki@certara.com