



New Features in Define.xml 2.0 and Steps for SDTM Define File Generation

By: Yuan Wang
FMD K&L

Agenda

1

Background Introduction

2

Comparison of Define.xml V2.0 with V1.0

3

Steps for Creating Define.xml 2.0

4

Define.xml Validation

Background Introduction

- ❖ **1997** - The FDA **established the regulatory basis** for wholly electronic submission of data in 1997 with the publication of regulations on the use of electronic records in place of paper records (21 CFR Part 11).
- ❖ **1999** - The FDA standardized the submission of clinical and non-clinical data using the **SAS Version 5 XPORT Transport** Format and the submission of metadata using **Portable Document Format (PDF)**, respectively.
- ❖ **2005** - The **Study Data Specifications published** by the FDA included the recommendation that data definitions (metadata) be provided as a **Define-XML file**.

Background Introduction

- ❖ **2011** - The CDER Common Data Standards Issues Document stated that “**A properly functioning define.xml file is an important part of the submission of standardized electronic datasets and should not be considered optional.**”
- ❖ **2013** – **Define-XML v2.0** was **published** in March
- ❖ **2015** – **Define-XML v2.0 accepted** by FDA as the standards for study data submissions

Background Introduction

❖ What is define.xml

- It is the metadata file sent with every study in each submission, which tells the FDA what datasets, variables, controlled terms, and other specified metadata were used.
- Extension of the CDISC Operational Data Model (ODM), an XML specification to facilitate the archival and interchange of the metadata and data for clinical research
- CRT-DDS or commonly known as “define.xml”, is not just for SDTM, SEND or ADaM but can even be used for non submission, proprietary data sets.

Background Introduction

❖ For FDA

“A properly functioning define.xml file is an important part of the submission if the electronic study datasets. ... An insufficiently documented define file is a common deficiency that the reviewers have noted.”

- FDA Study Data Technical Conformance Guide

Background Introduction

❖ For FDA

- **Documents traceability** from the dataset variables back to the annotated CRFs
- Allows reviewers to **quickly familiarize** themselves with the submission data
- **Avoids unnecessary queries**
- Often **printed out** for reviewers and used as a desktop reference

Background Introduction

- ❖ **For CROs and Sponsors**
 - **A great way to communicate expected deliverables**
 - **What tests should be collected at each visit**
 - **What are accepted values**
 - **Automatic study specific validation with Pinnacle**
 - **Study specific code lists**
 - **Study specific value level metadata**

Background Introduction

❖ Define.xml Structure

- **Table of Contents:** Dataset, Description, Structure, Purpose, Keys, Location
- **Data Definition Tables:** Provides the variable level attributes, definitions, and usage information for each variable contained within each domain. (Variable, Label, Type, CT or format, Origin, Role, Comment)
- **Controlled Terminology (Code List):** Many variables have a discrete list of valid values or controlled terms associated with them. Controlled terminologies may be developed by sponsors or by an external third party such as MedDRA.

Background Introduction

❖ Define.xml Structure

- **Value Level Metadata (Value List):** Case Report Tabulation datasets require additional metadata support in order to be useful for review and analysis
- **Computational Algorithms :** Computation Method element is to be used to provide details about computational algorithms used to derive or impute variable values
- **Supporting Documents:** (aCRF, Supplemental Data Definitions, ..)

Compare Define.xml V1.0 with V2.0

❖ **Key changes include:**

- **Support for CDISC Controlled Terminology**
- **Flexible definition of the Value Level Metadata**
- **Enhances documentation of data origin or source**
- **Improved support for ADaM metadata**
- **Improved handling of comments**

Compare Define.xml V1.0 with V2.0

❖ Dataset Definitions – v1.0

Datasets for Study 1234					
Dataset	Description	Structure	Purpose	Keys	Location
DM	Demographics	Special Purpose - One record per event per subject	Tabulation	STUDYID, USUBJID	crt/datasets/1234/dm.xpt
TE	Trial Elements	Trial Design - One Record Per Element	Tabulation	STUDYID, ELEMENT	crt/datasets/1234/te.xpt
TA	Trial Arms	Trial Design - One Record per Element for each Arm	Tabulation	STUDYID, ARM	crt/datasets/1234/ta.xpt
TV	Trial Visits	Trial Design - One Record per Visit per Arm	Tabulation	STUDYID, VISIT	crt/datasets/1234/tv.xpt
SE	Subject Elements	Study Design - One Record Per Subject Element	Tabulation	STUDYID, ELEMENT	crt/datasets/1234/se.xpt
SV	Subject Visits	Study Design - One Record Per Subject Visit	Tabulation	STUDYID, VISIT	crt/datasets/1234/sv.xpt
EX	Exposure	Interventions - One record per constant dosing interval per subject	Tabulation	USUBJID, EXTRT, EXSEQ	crt/datasets/1234/ex.xpt
CM	Concomitant Medications	Interventions - One record per event per subject	Tabulation	USUBJID, CMTRT, CMSEQ	crt/datasets/1234/cm.xpt
SU	Substance Use	Interventions - One record per substance use type per subject	Tabulation	USUBJID, SUTRT, SUSEQ	crt/datasets/1234/su.xpt
AE	Adverse Events	Events - One record per event per subject	Tabulation	USUBJID, AETERM, AESEQ	crt/datasets/1234/ae.xpt
DS	Disposition	Events - One record per disposition status or protocol milestone per subject	Tabulation	USUBJID, DSTERM, DSSEQ	crt/datasets/1234/ds.xpt
MH	Medical History	Events - One record per event per subject	Tabulation	USUBJID, MHTERM, MHSEQ	crt/datasets/1234/mh.xpt
EG	ECG Test Results	Findings - One record per event per subject	Tabulation	USUBJID, EGTESTCD, EGSEQ	crt/datasets/1234/eg.xpt
IE	Inclusion/Exclusion Exceptions	Findings - One record per event per subject	Tabulation	USUBJID, IETEST, IESEQ	crt/datasets/1234/ie.xpt
LB	Laboratory Tests	Findings - One record per lab test per subject	Tabulation	USUBJID, LBTESTCD, VISITNUM, TPTNUM, LBSEQ	crt/datasets/1234/lb.xpt
PE	Physical Examination	Findings - One record per event per subject	Tabulation	USUBJID, PETEST, PESEQ	crt/datasets/1234/pe.xpt
SC	Subject Characteristics	Findings - One record per subject characteristic	Tabulation	USUBJID, SCTESTCD	crt/datasets/1234/sc.xpt
VS	Vital Signs	Findings - One record per subject per vital sign	Tabulation	USUBJID, VSTESTCD	crt/datasets/1234/vs.xpt
CO	Comments	Special Purpose - One record per comment per subject	Tabulation	STUDYID, COSEQ	crt/datasets/1234/co.xpt

Compare Define.xml V1.0 with V2.0

❖ Dataset Definitions – v2.0

Tabulation Datasets for Study CDISC01 (SDTM-IG 3.1.2)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
TA	Trial Arms	TRIAL DESIGN	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD	ta.xpt	
TE	Trial Elements	TRIAL DESIGN	One record per planned Element	Tabulation	STUDYID, ETCD	te.xpt	
TI	Trial Inclusion/Exclusion Criteria	TRIAL DESIGN	One record per I/E criterion	Tabulation	STUDYID, IETESTCD	ti.xpt	
TS	Trial Summary	TRIAL DESIGN	One record per trial summary parameter value	Tabulation	STUDYID, TSPARMCD, TSSEQ	ts.xpt	
TV	Trial Visits	TRIAL DESIGN	One record per planned Visit per Arm	Tabulation	STUDYID, VISITNUM, ARMCD	tv.xpt	
DM	Demographics	SPECIAL PURPOSE	One record per subject	Tabulation	STUDYID, USUBJID	dm.xpt	See Reviewer's Guide, Section 2.1 Demographics Reviewers Guide
SE	Subject Elements	SPECIAL PURPOSE	One record per actual Element per subject	Tabulation	STUDYID, USUBJID, SESTDTC, SEENDTC, TAETORD, ETCD	se.xpt	
SV	Subject Visits	SPECIAL PURPOSE	One record per actual visit per subject	Tabulation	STUDYID, USUBJID, SVSTDTC, VISITNUM	sv.xpt	

Compare Define.xml V1.0 with V2.0

❖ Dataset Variable Definitions – v1.0

Demographics Dataset (DM)						
Variable	Label	Type	Controlled Terms or Format	Origin	Role	Comment
STUDYID	STUDY IDENTIFIER	text		CRF Page	Identifier	Demographics CRF Page 4
DOMAIN	DOMAIN ABBREVIATION	text		CRF Page	Identifier	DOMAIN ABBREVIATION
USUBJID	UNIQUE SUBJECT IDENTIFIER	text		CRF Page	Identifier	Demographics CRF Page 4
SUBJID	SUBJECT IDENTIFIER	text		CRF Page	Topic	Demographics CRF Page 4
RFSTDTC	SUBJECT REFERENCE START DATE/TIME	date		CRF Page	Timing	SUBJECT REFERENCE END DATE/TIME
RFENDTC	SUBJECT REFERENCE END DATE/TIME	date		CRF Page	Timing	SUBJECT REFERENCE START DATE/TIME
SITEID	STUDY SITE IDENTIFIER	text		CRF Page	Record Qualifier	Demographics CRF Page 4
INVID	INVESTIGATOR IDENTIFIER	text		CRF Page	Record Qualifier	Demographics CRF Page 4
BRTHDTC	DATE/TIME OF BIRTH	date		CRF Page	Result Qualifier	Demographics CRF Page 4
AGE	AGE IN AGEU AT REFERENCE DATE/TIME	integer		DERIVED	Result Qualifier	AGE IN AGEU AT REFERENCE DATE/TIME
AGEU	AGE UNITS	text		DERIVED	Variable Qualifier	AGE UNITS
SEX	SEX	text		CRF Page	Result Qualifier	Demographics CRF Page 4
RACE	RACE	text		CRF Page	Result Qualifier	Demographics CRF Page 4
ARMCD	ARM CODE	text		DERIVED	Result Qualifier	PLANNED ARM CODE
ARM	DESCRIPTION OF ARM	text		DERIVED	Synonym Qualifier	DESCRIPTION OF PLANNED ARM
COUNTRY	COUNTRY	text		CRF Page	Result Qualifier	COUNTRY
DMDTC	DATE/TIME OF COLLECTION	date		CRF Page	Timing	Demographics CRF Page 4
DMDY	STUDY DAY OF COLLECTION	integer		DERIVED	Timing	STUDY DAY OF COLLECTION

Compare Define.xml V1.0 with V2.0

❖ Dataset Variable Definitions – v2.0, SDTM

Demographics (DM) [Location: [dm.xpt](#)]

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
STUDYID	Study Identifier	1	text	7		Protocol	
DOMAIN	Domain Abbreviation		text	2	["DM" = "Demographics"] < Domain Abbreviation (DM) >	Assigned	
USUBJID	Unique Subject Identifier	2	text	14		Derived	Concatenation of STUDYID and SUBJID
SUBJID	Subject Identifier for the Study		text	6		CRF Page 3	
RFSTDTC	Subject Reference Start Date/Time		date		ISO8601	Derived	RFSTDTC = first date/time of study drug, for safety subject. Null for screen failures.
RFENDTC	Subject Reference End Date/Time		date		ISO8601	Derived	RFENDTC = termination date, for safety subjects. Null for screen failures.
SITEID	Study Site Identifier		text	3		CRF Page 3	
BIRTHDTC	Date/Time of Birth		date		ISO8601	CRF Page 6	
AGE	Age		integer	2		Derived	Age at Screening Date (Screening Date - Birth date). For the complete algorithm see the referenced external document. Complex Algorithms (complexalgorithms.pdf)
AGEU	Age Units		text	5		Assigned	Defaulted to YEARS
SEX	Sex		text	1	["F" = "Female", "M" = "Male", "U" = "Unknown"] < Sex >	CRF Page 6	
RACE	Race		text	40	Race	CRF Page 6	

Compare Define.xml V1.0 with V2.0

❖ Dataset Variable Definitions – v2.0, ADaM

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	["Placebo", "Xanomeline Low Dose", "Xanomeline High Dose"] <ARM>	Predecessor: DM.ARM
TRT01P	Planned Treatment for Period 01	text	20	["Placebo", "Xanomeline Low Dose", "Xanomeline High Dose"] <ARM>	Predecessor: DM.ARM
TRT01PN	Planned Treatment for Period 01 (N)	integer	8	["0" = "Placebo", "54" = "Xanomeline Low Dose", "81" = "Xanomeline High Dose"] <ARMN>	Assigned: Numeric code for TRT01P which corresponds to the randomized dose

Compare Define.xml V1.0 with V2.0

❖ Value Level Metadata Definition – v1.0

Value Level Metadata							
Source Variable	Value	Label	Type	Controlled Terms or Format	Origin	Role	Comment
SCTESTCD	ALLERGY	Allergy Status	integer	YESNOUNK	CRF Page		Subject Characteristics CRF Page 4
SCTESTCD	EDLEVLN	EDUCATIONAL LEVEL-DVN	float		CRF Page		Subject Characteristics CRF Page 4
SCTESTCD	EXCLSN	EXERCISE CLASSIFICATION-DVN	float		CRF Page		Subject Characteristics CRF Page 4
SCTESTCD	RACEOTH	Text for other race	text		CRF Page		Subject Characteristics CRF Page 4
SUTRT	ALCBEERA	Beer consumed-oz/week	float		CRF Page		Substance Use CRF Page 4
SUTRT	ALCHIST2	Does the patient drink	integer	YESNO	CRF Page		Substance Use CRF Page 4
SUTRT	ALCOTHA	Spirits/liquor consumed-oz/week	float		CRF Page		Substance Use CRF Page 4
SUTRT	ALCWINEA	Wine consumption-oz/week	float		CRF Page		Substance Use CRF Page 4
SUTRT	SMKCLASS	Smoking classification	integer	SMKCLAS	CRF Page		Substance Use CRF Page 4
VSTESTCD	FRAME	Frame	float	FRAME	CRF Page		Vital Signs CRF Page 4
VSTESTCD	HTRAW	Height raw	text		CRF Page		Vital Signs CRF Page 4
VSTESTCD	WTRAW	Weight raw	text		CRF Page		Vital Signs CRF Page 4
VSTESTCD	MEANBP	Mean Blood Pressure	float		VSSTRESN		See Computational Method: COMPMETHOD.MEANBP

Compare Define.xml V1.0 with V2.0

❖ Value Level Metadata Definition – v2.0

Value Level Metadata - EG [EGORRES]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/Comment
EGORRES	EGTESTCD EQ INTP (Interpretation)	text	8	["ABNORMAL", "NORMAL"] < Interpretation: Original Results >	CRF Page 12	
EGORRES	EGTESTCD EQ PRMEAN (Summary (Mean) PR Duration)	integer	3		CRF Page 12	
EGORRES	EGTESTCD EQ QRSDUR (Summary (Mean) QRS Duration)	integer	3		CRF Page 12	
EGORRES	EGTESTCD EQ QTMEAN (Summary (Mean) QT Duration)	integer	3		CRF Page 12	
EGORRES	EGTESTCD EQ VRMEAN (Summary (Mean) Ventricular Rate)	integer	2		CRF Page 12	

Value Level Metadata - EG [EGSTRESC]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/Comment
EGSTRESC	EGTESTCD EQ QTCB (QTcB - Bazett's Correction Formula)	float	5.1		Derived	QTcB = QT interval / square root of (60 / heart rate). For the complete algorithm see the referenced external document. Complex Algorithms (complexalgorithms.pdf)
EGSTRESC	EGTESTCD EQ QTcF (QTcF - Fridericia's Correction Formula)	float	5.1		Derived	QTcF = QT interval / cubic root of (60 / heart rate). For the complete algorithm see the referenced external document. Complex Algorithms (complexalgorithms.pdf)

Compare Define.xml V1.0 with V2.0

❖ Controlled Terminology Definitions – v1.0

Controlled Terminology (Code Lists) Section	
Code Value	Code Text
FRAME, Reference Name (FRAME)	
S	Small
M	Medium
L	Large
XL	Extra large
SMKCLAS, Reference Name (SMKCLAS)	
1	NEVER SMOKED
2	SMOKER
3	EX SMOKER
YESNO, Reference Name (YESNO)	
Y	YES
N	NO
U	NOT APPLICABLE

Compare Define.xml V1.0 with V2.0

❖ Controlled Terminology Definitions – v2.0

Severity/Intensity Scale for Adverse Events [CL.AESEV, C66769]

Permitted Value (Code)	Display Value (Decode)
MILD [C41338]	Grade 1
MODERATE [C41339]	Grade 2
SEVERE [C41340]	Grade 3

Description of Planned Arm [CL.ARM]

Permitted Value (Code)
Miracle Drug 10 mg
Miracle Drug 20 mg
Placebo
Screen Failure

Planned Arm Code [CL.ARMCD]

Permitted Value (Code)	Display Value (Decode)
WONDER10	Miracle Drug 10 mg
WONDER20	Miracle Drug 20 mg
PLACEBO	Placebo
SCRNFAIL	Screen Failure

Compare Define.xml V1.0 with V2.0

❖ Computational Method and Comment Definitions – v1.0

Computational Algorithms Section	
Reference Name	Computation Method
COMPMETHOD.MEANBP	Sum all mean blood pressure readings for visits 0, 1, and 2 and divide by the total number of readings

Compare Define.xml V1.0 with V2.0

❖ Computational Method and Comment Definitions – v2.0

Computational Algorithms

Method	Type	Description
Algorithm to derive AEENDY	Computation	$AEENDY = AEENDTC - RFSTDTC + 1$ if AEENDTC is on or after RFSTDTC. $AEENDTC - RFSTDTC$ if AEENDTC precedes RFSTDTC
Algorithm to derive AESTDY	Computation	$AESTDY = AESTDTC - RFSTDTC + 1$ if AESTDTC is on or after RFSTDTC. $AESTDTC - RFSTDTC$ if AESTDTC precedes RFSTDTC
Algorithm to derive the AETRTEM flag	Computation	AETRTEM = "Y" if Adverse Event was not present prior to the RFSTDTC, or it was present prior to the RFSTDTC but increased in severity during the treatment period. Null otherwise.
Algorithm to derive AGE	Computation	Age at Screening Date (Screening Date - Birth date). For the complete algorithm see the referenced external document. Complex Algorithms (complexalgorithms.pdf)
Algorithm to derive CLSIG	Computation	Only created if value qualifies as potentially clinically significant high or low based on the high and low ranges specified in the Statistical Analysis Plan.
Algorithm to derive CMENDY	Computation	$CMENDY = CMENDTC - RFSTDTC + 1$ if CMENDTC is on or after RFSTDTC. $CMENDTC - RFSTDTC$ if CMENDTC precedes RFSTDTC
Algorithm to derive CMSTDY	Computation	$CMSTDY = CMSTDTC - RFSTDTC + 1$ if CMSTDTC is on or after RFSTDTC. $CMSTDTC - RFSTDTC$ if CMSTDTC precedes RFSTDTC
Algorithm to derive DA.VISITNUM	Computation	Derived from DADTC and SVSTDTC.
Algorithm to derive DADY	Computation	$DADY = DADTC - RFSTDTC + 1$ if DADTC is on or after RFSTDTC. $DADTC - RFSTDTC$ if DADTC precedes RFSTDTC.
Algorithm to derive DASTRESC	Computation	$DASTRESC = DAORRES$

Compare Define.xml V1.0 with V2.0

❖ Computational Method and Comment Definitions – v2.0

Comments

CommentOID	Description
COM.LBREFID	Accession number
COM.RELTYPE	All values are null since this is used only when identifying a dataset-level relationship.
COM.ARMCD	Assigned based on Randomization Number. See Note 2.1 Reviewers Guide (reviewersguide.pdf)
COM.MHBODSYS	Assigned for Medical History but not Psychiatric History
COM.ARM	Assigned from TA.ARM based on ARMCD.
COM.EG.VISITNUM	Assigned from the TV domain based on the VISIT
COM.IE.VISITNUM	Assigned from the TV domain based on the VISIT
COM.PE.VISITNUM	Assigned from the TV domain based on the VISIT
COM.QS.VISITNUM	Assigned from the TV domain based on the VISIT
COM.VS.VISITNUM	Assigned from the TV domain based on the VISIT
COM.CMCLASCD	Coded to ATC level 3 Code based on CMINDC

Compare Define.xml V1.0 with V2.0

❖ Supplemental Data Definition Documents

Define-XML V2.0 enhanced links in Documentation allow sponsors to provide information in external files such as annotated CRFs, a SAS program, a data reviewer's guide (SDRG or ADRG), or SAP.

Steps for Creating Define.xml 2.0

❖ **Pre-required Files**

- **SDTM Mapping Spec**
- **SDTM Domain in XPT**
- **BlankCRF.pdf**
- **Reviewer Guide.pdf**
- **Complex Algorithm.pdf**

Steps for Creating Define.xml 2.0

❖ Pre-required Files (Cont.)

• SDTM Mapping Spec

SDTM Domain	SDTM Domain Description	SDTM Domain Structure	Program Name	General Observation Class	Source Data Used	Key Variables	Sort Order	Extract Rule
AE	Adverse Events	One record per adverse event per subject	ae.sas	Events	AE, SDTM.DM	STUDYID, USUBJID, AESTDTC, AEENDTC, AETERM, AEDECOD	STUDYID, USUBJID, AESTDTC, AEENDTC, AETERM, AEDECOD	ALL
SUPPAE	Supplemental Qualifiers for AE	One record per IDVAR, IDVARVAL, and QNAM value per subject	ae.sas	Special-Purpose Datasets	AE, SDTM.DM	STUDYID, USUBJID, IDVAR, IDVARVAL, QNAM	STUDYID, USUBJID, IDVAR, IDVARVAL, QNAM	ALL
CM	Concomitant Medications	One record per recorded medication occurrence or constant-dosing interval per subject	cm.sas	Interventions	CM, SDTM.DM	STUDYID, USUBJID, CMSTDTC, CMTRT, CMDOSTXT	STUDYID, USUBJID, CMSTDTC, CMTRT, CMDOSTXT	ALL
SUPPCM	Supplemental Qualifiers for CM	One record per IDVAR, IDVARVAL, and QNAM value per subject	cm.sas	Special-Purpose Datasets	CM, SDTM.DM	STUDYID, USUBJID, IDVAR, IDVARVAL, QNAM	STUDYID, USUBJID, IDVAR, IDVARVAL, QNAM	ALL
DM	Demographics	One record per subject	dm.sas	Special Purpose Domains	DM, DS, IC, RAND, EX, AE, RANDLISTNEW, SDTM.DM	STUDYID, USUBJID	STUDYID, USUBJID	ALL
SUPPDM	Supplemental Qualifiers for DM	One record per IDVAR, IDVARVAL, and QNAM value per subject	dm.sas	Special-Purpose Datasets	DM, DS, IC, RAND, SDTM.DM	STUDYID, USUBJID, IDVAR, IDVARVAL, QNAM	STUDYID, USUBJID, IDVAR, IDVARVAL, QNAM	ALL
DS	Disposition	One record per disposition status or protocol milestone per subject	ds.sas	Events	DS, IC, RAND, SDTM.DM	STUDYID, USUBJID, DSSTDTC, DSDECOD	STUDYID, USUBJID, DSSTDTC, DSDECOD	ALL
SUPPDS	Supplemental Qualifiers for DS	One record per IDVAR, IDVARVAL, and QNAM value per subject	ds.sas	Special-Purpose Datasets	DS, RAND, SDTM.DM	STUDYID, USUBJID, IDVAR, IDVARVAL, QNAM	STUDYID, USUBJID, IDVAR, IDVARVAL, QNAM	ALL
DV	Protocol Deviation	One record per protocol deviation per subject	dv.sas	Events	Protocol Deviation List_06June2016.xlsx	STUDYID, USUBJID, DVSTDTC, DVTERM, DVDECOD	STUDYID, USUBJID, DVSTDTC, DVTERM, DVDECOD	ALL
SUPPDV	Supplemental Qualifiers for DV	One record per IDVAR, IDVARVAL, and QNAM value per subject	dv.sas	Special-Purpose Datasets	Protocol Deviation List_06June2016.xlsx	STUDYID, USUBJID, IDVAR, IDVARVAL, QNAM	STUDYID, USUBJID, IDVAR, IDVARVAL, QNAM	ALL

Steps for Creating Define.xml 2.0

❖ Pre-required Files (Cont.)

- **SDTM Mapping Spec**

Variable Name	Variable Label	Type	Length	Controlled Terms or Format	Origin	Core	Conversion Definition	Variable Type	Variable Order
STUDYID	Study Identifier	Char	\$20		Assigned	Req	=ABC-DEFG	SDTM	1
DOMAIN	Domain Abbreviation	Char	\$2	DOMAIN	Assigned	Req	DV	SDTM	2
USUBJID	Unique Subject Identifier	Char	\$30		Derived	Req	Concatenate the project id with subject number separated by "-".	SDTM	3
DVSEQ	Sequence Number	Num			Derived	Req	Sort by variable STUDYID, USUBJID, DVSTDTC, DVTERM, DVDECOD. Restart the numbering at each USUBJID with 1 (order sequentially).	SDTM	4
DVTERM	Protocol Deviation Term	Char	\$200		Assigned	Req	=Verbatim	SDTM	5
DVDECOD	Protocol Deviation Coded Term	Char	\$200		Assigned	Perm	=Description	SDTM	6
DVCAT	Category for Protocol Deviation	Char	\$100		Assigned	Perm	=Category	SDTM	7
EPOCH	Epoch	Char	\$20	EPOCH	Assigned	Perm	'SCREENING' if DVSTDTC is before dosing or screening failures, otherwise "TREATMENT". Missing for partial/missing dates	SDTM	8
DVSTDTC	Start Date/Time of Deviation	Char	\$20	ISO8601	Assigned	Perm	=ISO8601 (Deviation Date)	SDTM	9
DVSTDY	Study Day of Start of Deviation	Num			Derived	Perm	Calculation of DVSTDY = (Numeric version of DVSTDTC - Numeric version of DM.RFSTDTC). Note: Add 1 if calculation is greater than or equal to 0.	SDTM	10
ETCD	Element Code	Char	\$8		Derived		'SCRN' if EPOCH='SCREENING', =ARMCD if EPOCH='TREATMENT'.	SUPP	11
ELEMENT	Description of Element	Char	\$200		Derived		'Screen' if EPOCH='SCREENING', =ARM if EPOCH='TREATMENT'.	SUPP	12

Steps for Creating Define.xml 2.0

❖ **Pre-required Files (Cont.)**

- **Complex Algorithm.pdf**

- This is an external file about some additional comments descriptions to illustrate the complex algorithms for the dataset specification, such as IE full text.

Steps for Creating Define.xml 2.0

❖ Setup

- **Function**

- This is to set up the macro variables of the project drive, root directory, project and protocol, SAS options and assign libname and optional macro variables

- **User input**

- User needs to update the SDTM version if need

Steps for Creating Define.xml 2.0

❖ Step00_copy_macro_catalog

- **Function**

- This program copies SDTM define macro catalog from global library to the study area and only needs to run once

- **User Input**

- Only batch run the program, no other things need to be done

Steps for Creating Define.xml 2.0

❖ **Step01_convert_spec**

- **Functions**

- This program create spec SAS data based on the define spec in Excel

- **User Input**

- Only batch run the program, no other things need to be done

Steps for Creating Define.xml 2.0

❖ Step_02_make_toc

- **Function**

- This program created TOC excel file using spec sas data created in step 1

- **User Input**

- Only batch run the program, no other things need to be done

- **Post Editing**

- Add the **documentation** file as need

Steps for Creating Define.xml 2.0

❖ Step_02_make_toc (Cont.)

DomainSeq	SourceDataUsed	Keys	SortOrder	Class	Purpose	Documentation	Repeating	ISREFERENCEDATA
1		STUDYID, ARMCD, TAETORD	STUDYID, ARMCD, TAETORD	TRIAL DESIGN	Tabulation		Yes	Yes
2		STUDYID, ETCD	STUDYID, ETCD	TRIAL DESIGN	Tabulation		Yes	Yes
3		STUDYID, IETESTCD	STUDYID, IETESTCD	TRIAL DESIGN	Tabulation		Yes	Yes
4		STUDYID, TSPARMCD	STUDYID, TSPARMCD	TRIAL DESIGN	Tabulation		Yes	Yes
5		STUDYID, VISITNUM	STUDYID, VISITNUM	TRIAL DESIGN	Tabulation		Yes	Yes
6	DM, DS, IC, RAND, E	STUDYID, USUBJID	STUDYID, USUBJID	SPECIAL PURPOSE	Tabulation		No	No
7	EX, DS, IC, SDTM.D	STUDYID, USUBJID, TAETORD	STUDYID, USUBJID, TAETORD, S	SPECIAL PURPOSE	Tabulation		Yes	No
8	VD, VS1, DS, IC, SD	STUDYID, USUBJID, VISITNUM	STUDYID, USUBJID, VISITNUM	SPECIAL PURPOSE	Tabulation		Yes	No
9	CM, SDTM.DM	STUDYID, USUBJID, CMSTDTC	STUDYID, USUBJID, CMSTDTC, C	INTERVENTIONS	Tabulation		Yes	No
10	EX, SDTM.DM	STUDYID, USUBJID, VISITNUM	STUDYID, USUBJID, VISITNUM, E	INTERVENTIONS	Tabulation		Yes	No
11	AE, SDTM.DM	STUDYID, USUBJID, AESTDTC	STUDYID, USUBJID, AESTDTC, A	EVENTS	Tabulation		Yes	No
12	DS, IC, RAND, SDT	STUDYID, USUBJID, DSSTDTC	STUDYID, USUBJID, DSSTDTC, D	EVENTS	Tabulation		Yes	No
13	Protocol Deviation L	STUDYID, USUBJID, DVSTDTC	STUDYID, USUBJID, DVSTDTC, D	EVENTS	Tabulation		Yes	No
14	MH, SDTM.DM	STUDYID, USUBJID, MHSTDTC	STUDYID, USUBJID, MHSTDTC, M	EVENTS	Tabulation		Yes	No
15	IE, SDTM.DM	STUDYID, USUBJID, IETESTCD	STUDYID, USUBJID, IETESTCD, I	FINDINGS	Tabulation		Yes	No
16	PG, SDTM.DM	STUDYID, USUBJID, VISITNUM	STUDYID, USUBJID, VISITNUM, L	FINDINGS	Tabulation		Yes	No
17	PE, PEYN, PEOTH	STUDYID, USUBJID, VISITNUM	STUDYID, USUBJID, VISITNUM, P	FINDINGS	Tabulation		Yes	No
18	AA, SS, CS, NA, TC	STUDYID, USUBJID, VISITNUM	STUDYID, USUBJID, VISITNUM, Q	FINDINGS	Tabulation		Yes	No
19	PG, VD, SDTM.DM	STUDYID, USUBJID, VISITNUM	STUDYID, USUBJID, VISITNUM, F	FINDINGS	Tabulation		Yes	No
20	VS, VS1, SDTM.DM	STUDYID, USUBJID, VISITNUM	STUDYID, USUBJID, VISITNUM, V	FINDINGS	Tabulation		Yes	No
21	AE, SDTM.DM	STUDYID, USUBJID, IDVAR, IDV	STUDYID, USUBJID, IDVAR, IDVA	RELATIONSHIP	Tabulation		Yes	No
22	CM, SDTM.DM	STUDYID, USUBJID, IDVAR, IDV	STUDYID, USUBJID, IDVAR, IDVA	RELATIONSHIP	Tabulation		Yes	No
23	DM, DS, IC, RAND, E	STUDYID, USUBJID, IDVAR, IDV	STUDYID, USUBJID, IDVAR, IDVA	RELATIONSHIP	Tabulation		Yes	No
24	DS, RAND, SDTM.E	STUDYID, USUBJID, IDVAR, IDV	STUDYID, USUBJID, IDVAR, IDVA	RELATIONSHIP	Tabulation		Yes	No
25	Protocol Deviation L	STUDYID, USUBJID, IDVAR, IDV	STUDYID, USUBJID, IDVAR, IDVA	RELATIONSHIP	Tabulation		Yes	No
26	EX, SDTM.DM	STUDYID, USUBJID, IDVAR, IDV	STUDYID, USUBJID, IDVAR, IDVA	RELATIONSHIP	Tabulation		Yes	No
27	IE, SDTM.DM	STUDYID, USUBJID, IDVAR, IDV	STUDYID, USUBJID, IDVAR, IDVA	RELATIONSHIP	Tabulation		Yes	No
28	PG, SDTM.DM	STUDYID, USUBJID, IDVAR, IDV	STUDYID, USUBJID, IDVAR, IDVA	RELATIONSHIP	Tabulation		Yes	No
29	MH, SDTM.DM	STUDYID, USUBJID, IDVAR, IDV	STUDYID, USUBJID, IDVAR, IDVA	RELATIONSHIP	Tabulation		Yes	No
30	PE, PEYN, PEOTH	STUDYID, USUBJID, IDVAR, IDV	STUDYID, USUBJID, IDVAR, IDVA	RELATIONSHIP	Tabulation		Yes	No
31	AA, SS, CS, NA, TC	STUDYID, USUBJID, IDVAR, IDV	STUDYID, USUBJID, IDVAR, IDVA	RELATIONSHIP	Tabulation		Yes	No
32	PG, VD, SDTM.DM	STUDYID, USUBJID, IDVAR, IDV	STUDYID, USUBJID, IDVAR, IDVA	RELATIONSHIP	Tabulation		Yes	No
33	VS, VS1, SDTM.DM	STUDYID, USUBJID, IDVAR, IDV	STUDYID, USUBJID, IDVAR, IDVA	RELATIONSHIP	Tabulation		Yes	No

Steps for Creating Define.xml 2.0

❖ Step_03_make_DDT

- **Function**
 - This program created DDT excel file using actual output dataset and specification
- **User Input**
 - Only batch run the program, no other things need to be done
- **Post Editing**
 - Update **VLM, EXTREF, EXTREFF, EXTREFF, EXTREFS** column as need.

Steps for Creating Define.xml 2.0

❖ Step_03_make_DDT (Cont.)

DomainSeq	Domain	VarSec	KeySec	DataTy	Origin	Mandat	SignificantDigits	VLM	EXTREF	EXTREFF	EXTREFS
1	TA	1	1	text	Assigned	Yes					
1	TA	2		text	Assigned	Yes					
1	TA	3	2	text	Assigned	Yes					
1	TA	4		text	Assigned	Yes					
1	TA	5	3	integer	Assigned	Yes					
1	TA	6		text	Assigned	Yes			Y	complexalgorithms.pdf	
1	TA	7		text	Assigned	No					
1	TA	8		text	Assigned	No			Y	complexalgorithms.pdf	
1	TA	9		text	Assigned	Yes					
1	TA	10		text	Assigned	Yes					
2	TE	1	1	text	Assigned	Yes					
2	TE	2		text	Assigned	Yes					
2	TE	3	2	text	Assigned	Yes					
2	TE	4		text	Assigned	Yes					
2	TE	5		text	Assigned	Yes					
2	TE	6		text	Assigned	No					
3	TI	1	1	text	Assigned	Yes					
3	TI	2		text	Assigned	Yes					
3	TI	3	2	text	Assigned	Yes			Y	complexalgorithms.pdf	
3	TI	4		text	Assigned	Yes			Y	complexalgorithms.pdf	
3	TI	5		text	Assigned	Yes			Y	complexalgorithms.pdf	

Steps for Creating Define.xml 2.0

❖ Step_03_make_DDT (Cont.)

Variable	Label	Type	Length	DisplayFormat	Codelist	Core	Comments
STUDYID	Study Identifier	Char	14				ABC-DEFG
DOMAIN	Domain Abbreviat	Char	2		TA.DOMAIN		TA
ARMCD	Planned Arm Code	Char	5				If ARM='Pseudoephedrine' then ARM
ARM	Description of Plar	Char	21				Output one record for ARM='Pseudo
TAETORD	Planned Order of E	Num	8				If ETCD='SCRN', then TAETORD=1,
ETCD	Element Code	Char	5		ETCD		Conversion Definition refer to Compl
ELEMENT	Description of Eler	Char	21		ELEMENT		If ETCD='SCRN', then ELEMENT='S
TABRANCH	Branch	Char	35				Conversion Definition refer to Compl
TATRANS	Transition Rule	Char	1				Blank
EPOCH	Epoch	Char	9		EPOCH		If ETCD='PSE30' or ETCD='PBO' the
STUDYID	Study Identifier	Char	14				PRGO-PSE-09001
DOMAIN	Domain Abbreviat	Char	2		TE.DOMAIN		TE
ETCD	Element Code	Char	5		ETCD		One record with ETCD='SCRN'; one
ELEMENT	Description of Eler	Char	21		ELEMENT		If ETCD='SCRN', then ELEMENT='S
TESTRL	Rule for Start of Ele	Char	68				If ETCD='SCRN', then TESTRL='Info
TEDUR	Planned Duration c	Char	3		ISO8601		If ETCD='SCRN', then TEDUR='P1D

Steps for Creating Define.xml 2.0

❖ Step04_make_CT

- **Function**
 - This program creates CT.xls based on actual output dataset only output the codelist which have been specified in the DDT in previous step.
- **User Input**
 - Only batch run the program, no other things need to be done
- **Post Editing**
 - Adjust the **DECODEVAL** column, add CT as required

Steps for Creating Define.xml 2.0

❖ Step04_make_CT (Cont.)

DomainSeq	Domain	Codelist	CTOrder	CODEVAL	DECODEVAL	DATATYPE	PAIREDVAR	CODE
0		ISO3166	1	ISO3166		text		
0		ISO8601	1	ISO8601		text		
0		MedDRA	1	MedDRA	18.1	text		
0		WHODrug	1	WHODrug	June, 2015	text		
0		ND	1	NOT DONE		text		
0		NY	1	N	No	text		
0		NY	2	Y	Yes	text		
0		EPOCH	1	SCREENING		text		
0		EPOCH	2	TREATMENT		text		
0		ARM	1	Placebo	PBO	text		
0		ARM	2	Pseudoephedrine 30 mg	PSE30	text		
0		ARM	3	Screen Failure	SCRNFAIL	text		
0		ARMCD	1	PBO	Placebo	text		
0		ARMCD	2	PSE30	Pseudoephedrine 30 mg	text		
0		ARMCD	3	SCRNFAIL	Screen Failure	text		
4	TS	DOMAIN	1	TS	Trial Summary	text		
4	TS	TSPARM	1	Active Ingredient in Test Product	INGRDNT	text		
4	TS	TSPARM	2	Actual Number of Subjects	ACTSUB	text		
4	TS	TSPARM	3	Adaptive Design	ADAPT	text		
4	TS	TSPARM	4	Added on to Existing Treatments	ADDON	text		
4	TS	TSPARM	5	Clinical Study Sponsor	SPONSOR	text		
4	TS	TSPARM	6	Control Type	TCNTRL	text		
4	TS	TSPARM	7	Data Cutoff Date	DCUTDTC	text		
4	TS	TSPARM	8	Data Cutoff Description	DCUTDESC	text		
4	TS	TSPARM	9	Diagnosis Group	TDIGRP	text		
4	TS	TSPARM	10	Dose Units	DOSU	text		

Steps for Creating Define.xml 2.0

❖ Step05_make_VLM

- **Function**

- This program creates VLM.xls using DDT.xls created in step 3

- **User Input**

- Only batch run the program, no other things need to be done

Steps for Creating Define.xml 2.0

❖ Step05_make_VLM (Cont.)

DomainSeq	Domain	VarSeq	WHEREVARIABLE	WHERECOMPARATOR	WHERECHECKVALUE	Data Type	Origin	Mandatory	Significant Digits	EXTREF	EXTREFF	EXTREFS
15	IE	1	IETESTCD	EQ	EXCL03	text	CRF Page 4	Yes				
15	IE	2	IETESTCD	EQ	EXCL04	text	CRF Page 4	Yes				
15	IE	3	IETESTCD	EQ	EXCL07	text	CRF Page 4	Yes				
15	IE	4	IETESTCD	EQ	EXCL08	text	CRF Page 4	Yes				
15	IE	5	IETESTCD	EQ	EXCL13	text	CRF Page 4	Yes				
15	IE	6	IETESTCD	EQ	INCL01	text	CRF Page 4	Yes				
15	IE	7	IETESTCD	EQ	INCL02	text	CRF Page 4	Yes				
15	IE	8	IETESTCD	EQ	INCL03	text	CRF Page 4	Yes				
15	IE	9	IETESTCD	EQ	INCL04	text	CRF Page 4	Yes				
15	IE	10	IETESTCD	EQ	INCL05	text	CRF Page 4	Yes				
15	IE	11	IETESTCD	EQ	INCL06	text	CRF Page 4	Yes				
15	IE	12	IETESTCD	EQ	INCL07	text	CRF Page 4	Yes				
16	LB	1	LBTESTCD	EQ	HCG	text	CRF Page 14	Yes				
17	PE	1	PETESTCD	EQ	EARS	text	CRF Page 10	Yes				
17	PE	2	PETESTCD	EQ	EYES	text	CRF Page 10	Yes				
17	PE	3	PETESTCD	EQ	HEART	text	CRF Page 10	Yes				
17	PE	4	PETESTCD	EQ	LUNGS	text	CRF Page 10	Yes				
17	PE	5	PETESTCD	EQ	NECK	text	CRF Page 10	Yes				
17	PE	6	PETESTCD	EQ	NOSE	text	CRF Page 10	Yes				
17	PE	7	PETESTCD	EQ	OTHER	text	CRF Page 10	Yes				
17	PE	8	PETESTCD	EQ	THROAT	text	CRF Page 10	Yes				
18	QS	1	QSTESTCD	EQ	BODACHE	text	CRF Page 24	Yes				
18	QS	2	QSTESTCD	EQ	CHCONG	text	CRF Page 24	Yes				
18	QS	3	QSTESTCD	EQ	COUGH	text	CRF Page 24	Yes				
18	QS	4	QSTESTCD	EQ	CSACT	text	CRF Page 37	Yes				
18	QS	5	QSTESTCD	EQ	DIZZY	text	CRF Page 37	Yes				
18	QS	6	QSTESTCD	EQ	DROWSLEP	text	CRF Page 37	Yes				
18	QS	7	QSTESTCD	EQ	FATIG	text	CRF Page 24	Yes				
18	QS	8	QSTESTCD	EQ	FEVER	text	CRF Page 24	Yes				
18	QS	9	QSTESTCD	EQ	HARDBREA	text	CRF Page 29	Yes				
18	QS	10	QSTESTCD	EQ	HEADACHE	text	CRF Page 24	Yes				
18	QS	11	QSTESTCD	EQ	HOWCLEAR	text	CRF Page 29	Yes				
18	QS	12	QSTESTCD	EQ	NACONG	text	CRF Page 24	Yes				

Steps for Creating Define.xml 2.0

❖ Step05_make_VLM (Cont.)

Variable	Label	Type	Length	DisplayFormat	Codelist	Core	Comments
IEORRES	Are currently experiencing a	Char	1			Req	
IEORRES	Are experiencing symptoms	Char	1			Req	
IEORRES	Are exhibiting signs or symp	Char	1			Req	
IEORRES	Are experiencing a fever 103	Char	1			Req	
IEORRES	Have need of additional me	Char	1			Req	
IEORRES	Male or female subjects, ag	Char	1			Req	
IEORRES	Subjects have an onset of c	Char	1			Req	
IEORRES	Subjects have at least two o	Char	1			Req	
IEORRES	Subjects can swallow oral ta	Char	1			Req	
IEORRES	Findings from the medical hi	Char	1			Req	
IEORRES	Subject and legally authorize	Char	1			Req	
IEORRES	Subject's legally authorized	Char	1			Req	

Steps for Creating Define.xml 2.0

❖ Step05_make_VLM (Cont.)

- **Post Editing**

- **Edit Origin column for CRF page# as the VLM has broken down the source variable to value level under each category. The Origin inherited from specification may not be accurate at value level.**
- **Add WHERE conditions as need as adjust the comments.**
- **Check the code list as Y, if need**

Steps for Creating Define.xml 2.0

❖ Step06_make_defineXML

- **Function**

- This program generates define.XML using TOC,DDT, CT, and VLM which have been created previously

- **User Input**

- User needs to update macro parameters in macro as needed. Batch run the program.

- **Post Editing**

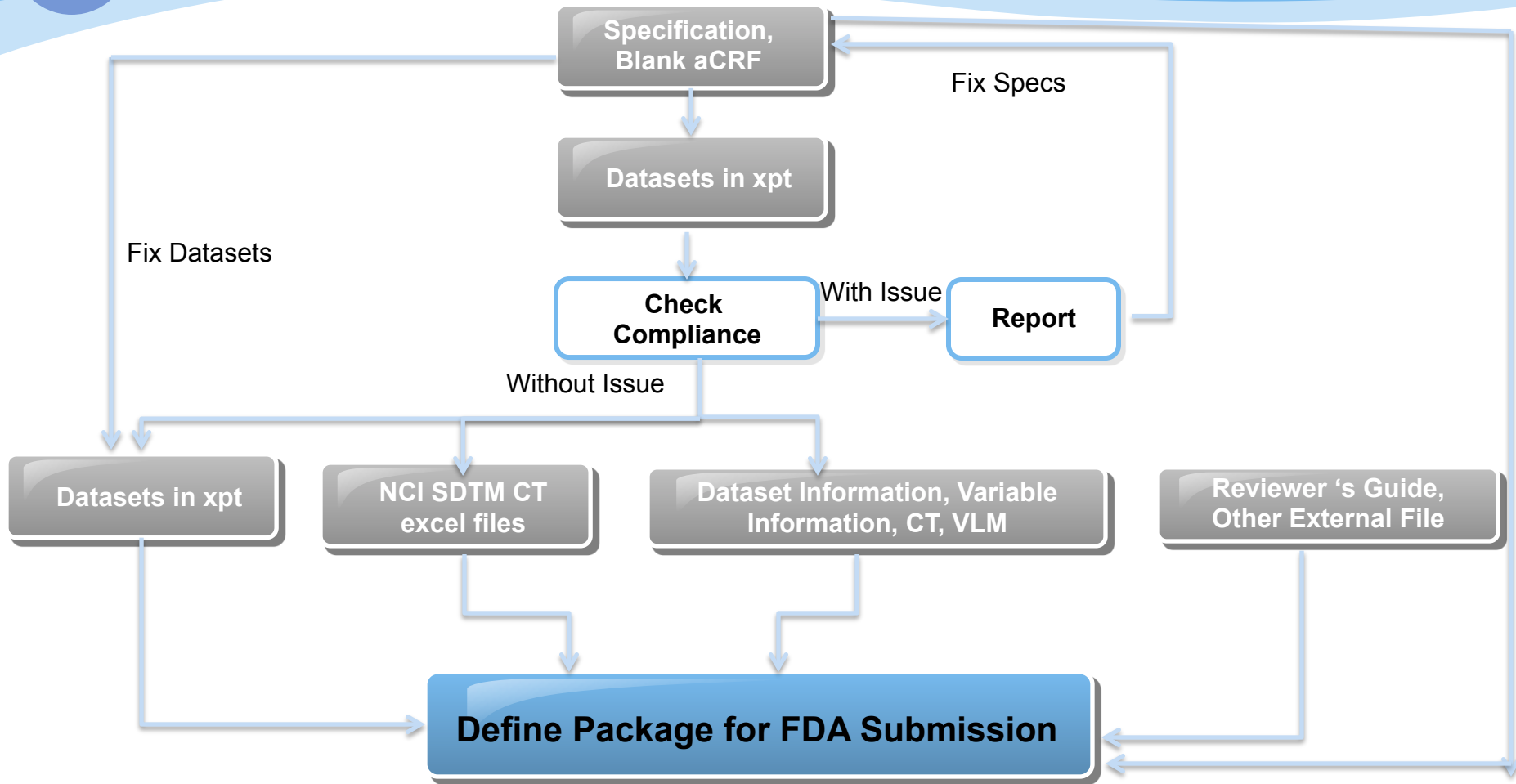
- User do not need to apply any post-editing

Define File Validation

❖ Validation

- **Make sure the ANNOTATED CASE REPORT FORM link to the first page of aCRF**
- **Check the Comment is filled for all variables with the Origin as “Derived”**
- **For xml, check ORIGIN with multiple page links are working correctly, links to open external supplemental documents are all working properly.**
- **.....**
- **Pinnacle Check, no warning, no error**

Conclusion and Questions



Conclusion and Questions





Thank You !