



# Robust approach to create Define.xml v2.0

Vineet Jain



## DEFINE SYSTEM GOALS

Generic

Works across SDTM, ADaM & SEND

---

Powerful

Create define.xml, annotated CRF & define.pdf

---

Reliable

Create compliant & consistent deliverables

---

Integrable

Integrates into existing client environment

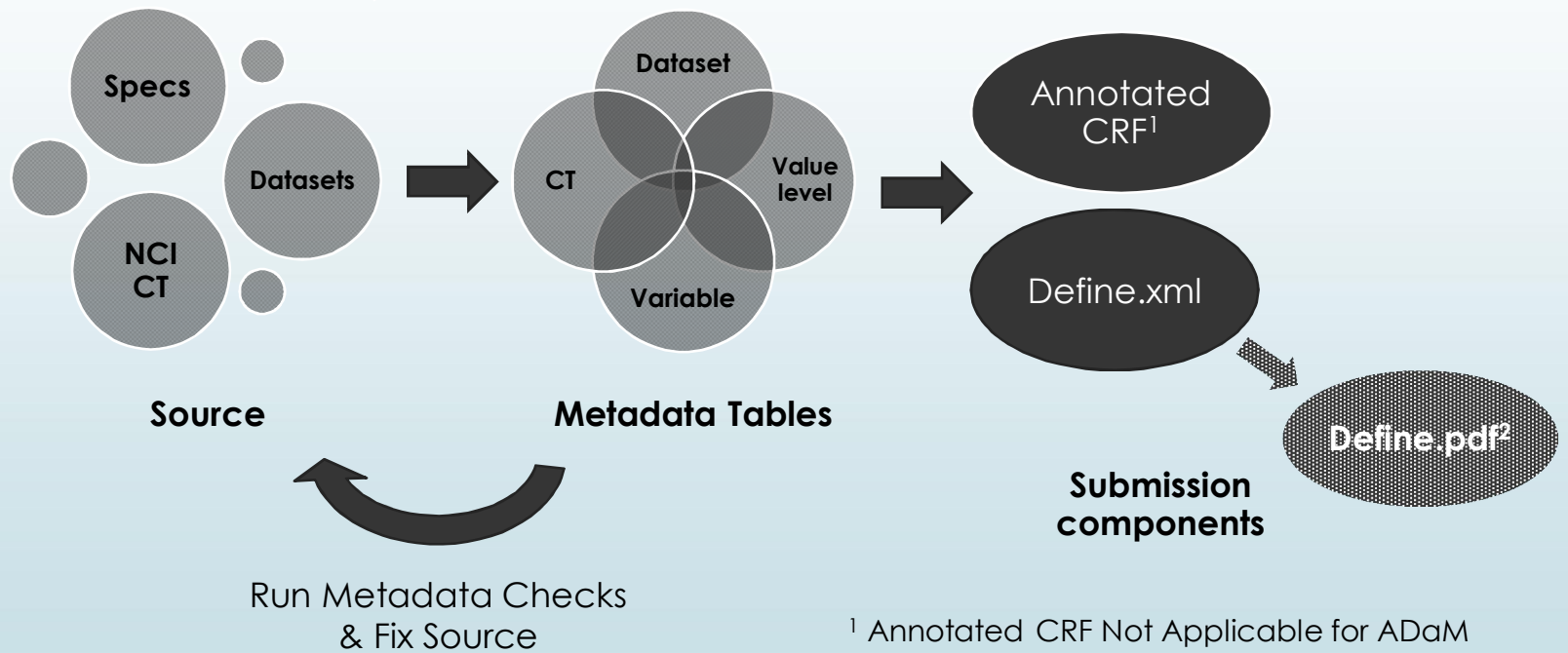
---

Efficient

Easy to use, minimal manual input &  
Automatically find issues

---

# DEFINE WORKFLOW



<sup>1</sup> Annotated CRF Not Applicable for ADaM

<sup>2</sup> Define.pdf useful but not required

# DATASET METADATA

DATASET	DOMAIN	LABEL	STRUCT	CLASS	REPEATING	ISREF	PURPOSE	COMMENT	ORDER	DOCREFI
ADSL		Subject Level Analysis Dataset	One record per subject	SUBJECT LEVEL ANALYSIS DATASET	No	No	Analysis	The source SAS data sets for ADSL are the following subject-level data SDTM datasets: DM, DS, EX, MH, SC, SV, QS, and VS. Refer Reviewer's Guide Section 1.1	1	ReviewersGuide# ND# Section1.1
ADLB		Laboratory	One record per subject, per parameter, per visit	BASIC DATA STRUCTURE	Yes	No	Analysis	Only keep randomized patients (ADSL.RANDFL = Y). Refer Reviewer's Guide Page 2	2	ReviewersGuide# PR#2
ADQS		Question	One record per subject, per parameter, per visit.	BASIC DATA STRUCTURE	Yes	No	Analysis	Only keep randomized patients (ADSL.RANDFL = Y)	3	

# VARIABLE METADATA

DATASET	VARIABLE	LABEL	LENGTH	ORDER	FMTNAME	ORIGIN	ORGDETL	MANDATORY	ROLE	DATATYPE	DISPFMT	SIGDIGIT	KEYSEQ	COMMENT	METHYP	DOCREF1
ADSL	STUDYID	Study Identifier	12	1		Predecessor	DM. STUDYID	Yes		text			1			
ADSL	USUBJID	Unique Subject Identifier	11	2		Predecessor	DM. USUBJID	Yes		text			2			
ADSL	SUBJID	Subject Identifier for the Study	4	3		Predecessor	DM. SUBJID	Yes		text						
ADSL	TRT01P	Planned Treatment for Period 01	20	18	ARM	Derived		No		text				Derived from DM.ARM	Computation	
ADSL	BMIBL	Baseline BMI	8	28		Derived		No		float		1		Derive from VS: VSSTRESN where VSTESTCD=BMI and VISITNUM=1	Computation	

# VALUELEVEL METADATA

DATASET	VARIABLE	LABEL	DATATYPE	DISPfmt	SIGDIGIT	LENGTH	FMTNAME	ORIGIN	ORGDETL	COMMENT	METHYP	WHERE1	WHERE2	ORDER	DOCREF1
ADQS	AVAL		float	5.2	2	8	ACIT1	Derived		QS.QSSTRESN where QSTESTCD= PARAMCD		PARAMCD IN 'ACITM01', 'ACITM02', 'ACITM03'		1	
ADQS	AVAL		Integer	2		8		Derived		QS.QSSTRESN where QSTESTCD= PARAMCD		PARAMCD IN 'ACITM04' ... 'ACITM14'		2	
ADQS	AVAL		Integer	2		8		Derived		Sum of ADAS scores for items 1, 2, 4..14	Computation	PARAMCD EQ 'ACTOT'	ANL01FL EQ 'Y'	3	
ADQS	CHG		float	5.2	2	8		Derived		AVAL - BASE	Computation	PARAMCD EQ 'ACTOT'		1	

# CONTROLLED TERMINOLOGY METADATA

FMTNAME	FMTLAB	FMTTYPE	VALUE	DECODE	NCIFMT	NCIITEM	ORDER	DATATYPE	DICTNM	DICTVER	RANK
AGEGRP	Age Group	CT	<65				1	text			1
AGEGRP	Age Group	CT	65-80				2	text			2
AGEGRP	Age Group	CT	>80				3	text			3
YNONLY	No Yes Response	FORMAT	N	No	C66742	C49487		text			
YNONLY	No Yes Response	FORMAT	Y	Yes	C66742	C49488		text			
YONLY_N	Yes Response (N)	FORMAT	1	Yes				integer			
AEDICT	AE Dictionary	DICT						text	MEDDRA	15.0	



# METADATA QC

## Well Formed

- All Metadata variables present
- Variables have valid values
- Values are printable & parsable

## Consistent With Data/Standards

- Accurately represent the source data
- NCI CT correctly used wherever applicable

## Consistent Within

- E.g. origin/ type in Valuelevel metadata consistent with parent variable
- E.g. Reference to a CT, missing in CT metadata



# DEFINE.XML CREATION

Metadata  
Tables

Options

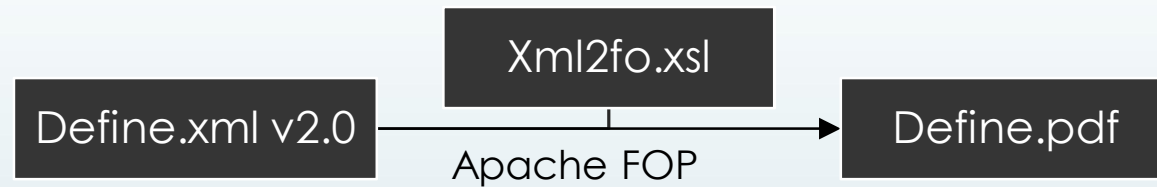
Define.xml

- Visual QC
- OpenCDISC xml Checks
- Validate against Schema

Issues

Enhance programmatic  
checks to detect issues  
upfront

# DEFINE.PDF CREATION



- ▶ Apache FOP: Free Open source software
  - ▶ Easy to use & quickly renders xml to pdf using the XSL file
  - ▶ Identifies issues in define.xml, e.g. checks all internal hyperlinking.
- ▶ XML2fo.xsl: Stylesheet file to define formatting
  - ▶ Based on CDISC's v2.0 stylesheet file
  - ▶ creates pdf almost identical to define.xml with bookmarks & links

# CRF ANNOTATION



**Sponsor:** ABC Company

**Protocol:** ABC-123 STUDYID

**Subject:** SUBJID

Auto Text

**Demographics**

DM = Demographics

SC = Subject Characteristics

Auto Color

Race	<span style="border: 1px solid black; padding: 2px;">RACE</span>	<input type="radio"/> White	<input type="radio"/> Other	<span style="border: 1px solid black; padding: 2px;">RACEOTH in SUPPDM</span>
SEX		<input type="radio"/> Male	<input type="radio"/> Female	<span style="border: 1px solid black; padding: 2px;">SEX</span>
Initials		<span style="border: 1px solid black; padding: 2px;">SCORRES when SCTESTCD=SUBJINIT</span>		
Informed Consent Date	____/____/____			<span style="border: 1px solid black; padding: 2px;">RFICDTC</span>
HEIGHT	____ cm	<span style="border: 1px solid black; padding: 2px;">VSORRES when VSTESTCD = HEIGHT</span>		
WEIGHT	____ Kg	<span style="border: 1px solid black; padding: 2px;">VSORRES when VSTESTCD = WEIGHT</span>		

Auto Size

Auto Font Size

VS = Vital Signs



# SYSTEM SUMMARY

## ► Robust

- All manual metadata entered upfront & centrally in data specs
- Programmatic checks ensure quality and consistency
- Submission deliverables end up consistent with each other

## ► Ease of use

- Common tool/process for SDTM, ADaM, SEND, define.xml/pdf, annotated CRF
- No need to dig into xml files
- Macros can be used independently or integrated with other systems
- Customizable & light-weight SAS macros (300-600 lines of code)



# THANK YOU

Download Detailed  
Paper & free Code:

- [www.linkedin.com/in/vineet7878](http://www.linkedin.com/in/vineet7878)

Contact Info:

- Email: [Vineet7878@gmail.com](mailto:Vineet7878@gmail.com)
- 📞: 908-654-3761