



Single Day Event

Anil Babu Adepu,
ICON

Automate SDTM Specification Creation



Disclaimer

The opinions expressed in this presentation and on the following slides are solely those of the presenter and not necessarily those of ICON. ICON does not guarantee the accuracy or reliability of the information provided herein



Introduction

- SDTM programming specification writing involves referring database annotated CRF and SDTM annotated CRF to provide the raw dataset and variable names for programmers and providing CRF page numbers for reviewers.
- When the input is available in a machine-readable format, key components of SDTM specification writing can be automated
- In this presentation, we discuss an approach we developed to leverage the machine-readable SDTM annotations from our internal processes to automate SDTM specification writing

SDTM Specification

- Input documents for SDTM specification
 - Blank CRF with database annotations
 - aCRF – with SDTM annotations
 - Standard SDTM specification template
- Primary consumers of SDTM specification
 - SDTM programmers
 - Reviewers (Define.xml)
- What does a programmer look for in SDTM specification
 - [Raw datasets required for a domain](#)
 - [Input variables for a specific SDTM variable](#)
 - Mapping and transformation algorithms
- How does a reviewer use define.xml
 - Reviews comments/derivation logics if any
 - [Trace back to a specific question on a CRF page contributing to an SDTM variable](#)

Programmers

- Input raw datasets for a SDTM domain
- Input variables for each SDTM variable
- Transformation algorithms

VARIABLE	ADDCOM VS
VSTESTCD	When VS.HEIGHT_VSORRES then VSTESTCD='HEIGHT', When VS.WEIGHT_VSORRES then VSTESTCD='WEIGHT', When VS.BMI_VSORRES then VSTESTCD='BMI', When VS.TEMP_VSORRES then VSTESTCD='TEMP', When VS.SYSBP_VSORRES then VSTESTCD='SYSBP', When VS.DIABP_VSORRES then VSTESTCD='DIABP', When VS.PULSE_VSORRES then VSTESTCD='PULSE', When VS.RESP_VSORRES then VSTESTCD='RESP', When VS.O2SAT_VSORRES then VSTESTCD='OXYSTAT'
VSORRES	Populate using VS.HEIGHT_VSORRES, VS.WEIGHT_VSORRES, VS.BMI_VSORRES, VS.TEMP_VSORRES, VS.SYSBP_VSORRES, VS.DIABP_VSORRES, VS.PULSE_VSORRES, VS.RESP_VSORRES, VS.O2SAT_VSORRES
VSORRESU	Populate using VS.HEIGHT_VSORRES, VS.WEIGHT_VSORRES, VS.BMI_VSORRESU, VS.TEMP_VSORRES, VS.SYSBP_VSORRESU, VS.DIABP_VSORRESU, VS.PULSE_VSORRESU, VS.RESP_VSORRESU, VS.O2SAT_VSORRESU
VSDTC VSTPT	Populate using VS.VSDAT, VS.VSTIM. Use collected date and/or time variables to populate date in ISO 8601 format.

Reviewers

- Derivation/Comment associated with a variable
- CRF Page numbers associated with the specific SDTM variable

Vital Signs (VS) [Location: [vs.xpt](#)]

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment 1
STUDYID	Study Identifier	1	text	20		Protocol	Populated with [***STUDYID**].
DOMAIN	Domain Abbreviation		text	2	SDTM Domain Abbreviation	Assigned	Populated with VS.
USUBJID	Unique Subject Identifier	2	text	29		Derived	Concatenate study identifier (STUDYID) and subject identifier (SUBJID), separated by a hyphen.
VSSEQ	Sequence Number		integer	8		Derived	Sequential number identifying records within each USUBJID and generated using the key sequence of the domain.
VSGRPID	Group ID		text	8		Assigned	
VSSPID	Sponsor-Defined Identifier		text	8		Assigned	
VSTESTCD	Vital Signs Test Short Name	3	text	8	Vital Signs Test Code	CRF Page 1	

Specification

- Information for reviewers
 - page numbers
- Information for programmers
 - Raw datasets
 - Input variables

VARIABLE	ORIGIN	ADDCOM_VSYN	ADDCOM_VS_M
VSTESTCD	CRF Pages 17, 18, 19	When VSYN.VSPERF then VSTESTCD=VSALL'	When VS_M.WEIGHT_VSORRES then VSTESTCD=WEIGHT', When VS_M.TEMP_VSORRES then VSTESTCD=TEMP', When VS_M.SYSBP_VSORRES then VSTESTCD=SYSBP', When VS_M.DIABP_VSORRES then VSTESTCD=DIABP', When VS_M.PULSE_VSORRES then VSTESTCD=PULSE', When VS_M.RESP_VSORRES then VSTESTCD=RESP', When VS_M.O2SAT_VSORRES then VSTESTCD=OXYSTAT'
VSORRES	CRF Pages 18, 19		Populate using VS_M.WEIGHT_VSORRES, VS_M.TEMP_VSORRES, VS_M.SYSBP_VSORRES, VS_M.DIABP_VSORRES, VS_M.PULSE_VSORRES, VS_M.RESP_VSORRES, VS_M.O2SAT_VSORRES
VSORRESU	CRF Pages 18, 19		Populate using VS_M.TEMP_VSORRES, VS_M.SYSBP_VSORRESU, VS_M.DIABP_VSORRESU, VS_M.PULSE_VSORRESU, VS_M.RESP_VSORRESU, VS_M.O2SAT_VSORRESU
VSSTAT	CRF Page 17	Populate using VSYN.VSPERF	
VSDTC	CRF Pages 18, 19		Populate using VS_M.VSDAT, VS_M.VSTIM. Use collected date and/or time variables to populate date in ISO 8601 format.
VSTPT	CRF Page 19		Populate using VS_M.VSTPT



SDTM Specification Creation – Traditional Approach

- Start with standard specification template
- Refer database annotated CRFs for raw variable information
- Refer SDTM annotated CRF for SDTM variable names and page numbers
- Update the specification manually for each SDTM variable for
 - Raw dataset and variable names
 - CRF page numbers for the SDTM variable mapping

Form: Vital Signs YN?

Were vital signs assessed? Yes ☐ No ☐

Field Name	Data Type	Field Label	Units	Values
① VSPERF	\$1	Were vital signs assessed?		Y = Yes N = No

Raw input variable name

VS=Vital Signs

Form: Vital Signs YN?

Were vital signs assessed? Yes ☐ No ☐

VSSTAT=NOT DONE when VSTESTCD=VSALL when No

17 of 81

SDTM variable name and page number

SDTM Specification – Automation?

- Can something be done to reduce this manual work?
- Yes, if we have the required information available in a machine-readable format
- What do we need for SDTM Specification automation?
 - Standard SDTM specification templates
 - Machine-readable SDTM CRF annotations
 - Machine-readable raw dataset and variable names
 - Machine-readable linkage between annotation text and raw input variables
 - Machine-readable position of the annotation text
- Do we have all these machine-readable components?
 - Yes.
 - How?
 - We have an automated CRF annotation process which generates the files containing the above information



Automated(semi) CRF annotation - components

- Internal standards team creates and maintains master case report forms
- Rave EDC has a specification document called Study Design Specification (SDS)
- SDS is an excel workbook containing
 - A sheet for the list of forms and associated raw dataset names
 - A sheet for the list of fields in each form and associated variable names
 - Other information related to visit structure, controlled terminology, unit conversion factors etc
- Standards team also maintains standard annotations for master forms in excel format
- Study specific SDS file and standard annotations excel file are used as input for automation



Automated(semi) CRF annotation - process

- SAS macros are built to use SDS file and standard annotations for automation
- It's a three-step process
 1. Pre-annotate
 - A SAS macro pulls matching standard annotations for the forms and fields of a study and creates an excel file (study annotations file)
 2. Modify study level annotations in study annotations file
 3. Annotate
 - A SAS macro imports the study annotations file and creates an fdf file(Forms Data Format) containing annotations
- Associate the fdf file to blankcrf.pdf to import annotations and save as acrf.pdf

SDS from RAVE

OID	DraftFormName
SUBJ	Subject
AE ¹	Adverse Events ²
AEPC	Adverse Events [**Prostate Cancer Version**]
AEYN	Adverse Events YN?
APMHCA	Family Cancer History
APMHCAYN	Family Cancer History YN?
APMHRA	Rheumatoid Arthritis Family History
APRP	Partner Reproductive System Findings
CE	Clinical Events
CEPS	[**Clinical Event - Pre-Specified**]
CEYN	Clinical Events YN?
CM	Prior/Concomitant Medications

1. Raw dataset name
2. Form name

FormOID	FieldOID	Ordinal	PreText
VS ¹	VSDAT	1	Measurement Date
VS	VSTIM	2	Measurement Time
VS	HEIGHT_VSORRES ²	3	Height ³
VS	WEIGHT_VSORRES	4	Weight
VS	BMI_VSORRES	5	Body Mass Index
VS	BMI_VSORRESU	6	Body Mass Index Unit
VS	TEMP_VSORRES	7	Temperature
VS	VSPPOS	8	Position
VS	SYSBP_VSORRES	9	Systolic Blood Pressure
VS	SYSBP_VSORRESU	10	Systolic Blood Pressure Unit
VS	DIABP_VSORRES	11	Diastolic Blood Pressure
VS	DIABP_VSORRESU	12	Diastolic Blood Pressure Unit
VS	PULSE_VSORRES	13	Pulse
VS	PULSE_VSORRESU	14	Pulse Unit
VS	HR_VSORRES	15	Heart Rate
VS	HR_VSORRESU	16	Heart Rate Unit
VS	RESP_VSORRES	17	Respiratory Rate
VS	RESP_VSORRESU	18	Respiratory Rate Unit

1. Raw dataset name
2. Variable name
3. Question text on CRF

Study annotations excel file

1. Raw dataset name
2. Raw variable name
3. Annotation type (domain, free, variable)
4. Annotation text
5. Domain number (to identify a domain when same field is mapped to multiple domains)

Form OID	Source	Type	Annotation Text	Domain Number
VS		Domain	VS=Vital Signs	1
VS	VSDAT	Variable	VSDTC	1
VS	VSTIM	Variable	VSDTC	1
VS	HEIGHT_VSORRES	Variable	VSORRES/VSORRESU when VSTESTCD=HEIGHT	1
VS	WEIGHT_VSORRES	Variable	VSORRES/VSORRESU when VSTESTCD=WEIGHT	1
VS	BMI_VSORRES	Variable	VSORRES when VSTESTCD=BMI	1
VS	BMI_VSORRESU	Variable	VSORRESU when VSTESTCD=BMI	1
VS	TEMP_VSORRES	Variable	VSORRES/VSORRESU when VSTESTCD=TEMP	1
VS	VSPOS	Variable	VSPOS	1
VS	SYSBP_VSORRES	Variable	VSORRES when VSTESTCD=SYSBP	1
VS	SYSBP_VSORRESU	Variable	VSORRESU when VSTESTCD=SYSBP	1
VS	DIABP_VSORRES	Variable	VSORRES when VSTESTCD=DIABP	1
VS	DIABP_VSORRESU	Variable	VSORRESU when VSTESTCD=DIABP	1
VS	PULSE_VSORRES	Variable	VSORRES when VSTESTCD=PULSE	1
VS	PULSE_VSORRESU	Variable	VSORRESU when VSTESTCD=PULSE	1
VS	HR_VSORRES	Variable	VSORRES when VSTESTCD=HR	1
VS	HR_VSORRESU	Variable	VSORRESU when VSTESTCD=HR	1
VS	RESP_VSORRES	Variable	VSORRES when VSTESTCD=RESP	1
VS	RESP_VSORRESU	Variable	VSORRESU when VSTESTCD=RESP	1

Every annotation is linked to a source variable

SDTM Specification Automation - process

- Parse the annotation text into individual components
 - This enables us to link a single source variable to one or more SDTM components
- Fetching the physical page number associated with a specific annotation
 - This can be obtained from comment summary export after associating fdf file to blankcrf
- Combining the raw dataset and variable information with page numbers
- Process this data to create required columns and values for a specification using standard specification template

Parsing the annotation text

- The annotation text is parsed to extract the possible components like
 - SDTM variable name
 - SDTM variable value
 - TESTCD value etc
- Screenshot showing that the SDTM VSORRES and VSTESTCD are dependent on DIABP_VSORRES variable of raw data

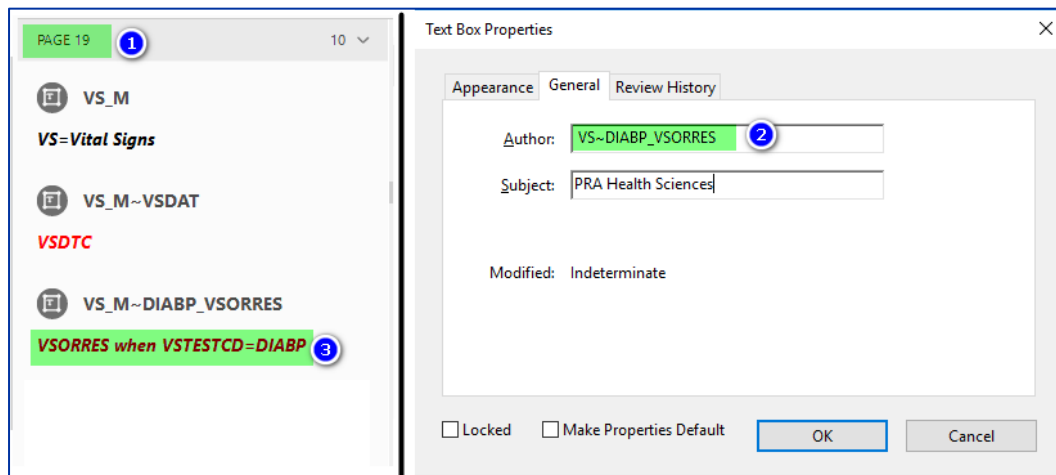
FORM_OID	SOURCE	ANNOTATION_TEXT	DOMAIN	FIELDNAM1	FIELDNAM2	FIELDNAM3
VS	DIABP_VSORRES	VSORRES when VSTESTCD=DIABP	VS	VSORRES ①	VSTESTCD	DIABP
VS	DIABP_VSORRES	VSORRES when VSTESTCD=DIABP	VS	VSTESTCD ②	VSTESTCD	DIABP

Common annotation text patterns

Annotation text patterns	Annotation text	SDTM components
Pattern 1 (Single word)	VSDTC	VS.VSDTC
Pattern 2 (Supplementary qualifier)	RACE1 in SUPPDM	SUPPDM.QNAM.RACE1
Pattern 3 (Supplementary qualifier)	LBCLSIG=N in SUPPLB	SUPPLB.QNAM.LBCLSIG
Pattern 4 (Test result)	VSORRES when VSTESTCD=WEIGHT	VS.VSORRES, VS.VSTESTCD, VLM.VS.WEIGHT
Pattern 5 (Test result)	VSSTAT=NOT DONE when VSTESTCD=WEIGHT	VS.VSSTAT, VS.VSTESTCD
Pattern 6 (Test result and unit)	VSORRES/VSORRESU when VSTESTCD=HEIGHT	VS.VSORRES, VS.VSORRESU, VS.VSTESTCD, VLM.VS.HEIGHT
Pattern 7 (Fixed text assignment conditionally)	DSTERM=COMPLETED when Yes	DS.DSTERM

Fetching the page numbers

- All the annotations we place on a pdf file are considered as 'comments'
- Each comment is associated with a physical pdf page number
- Each comment has a few properties
 - Comment (the annotation text)
 - Author (We use raw data and variable names as author)
- Comments can be exported as a summary
- The comments summary is parsed to extract
 1. Physical page number
 2. Raw dataset name and raw variable name
 3. The actual annotation text



Fetching the page numbers

- Screenshot showing the parsed comment summary containing physical page number

FORM_OID	SOURCE	ANNOTATION_TEXT	VARIABLE	PDFPAGENUM
VS	DIABP_VSORRES	VSORRES when VSTESTCD=DIABP	VSORRES	18



Combining variable and page information

FORM_OID	SOURCE	ANNOTATION_TEXT	DOMAIN	VARIABLE	PDFPAGENUM	FIELDNAM	FIELDNAM1	FIELDNAM2	FIELDNAM3
VS	DIABP_VSORRES	VSORRES when VSTESTCD=DIABP	VS	VSORRES	18	VSORRES when VSTESTCD=DIABP	VSORRES	VSTESTCD	DIABP
VS	DIABP_VSORRES	VSORRES when VSTESTCD=DIABP	VS	VSTESTCD	18	VSORRES when VSTESTCD=DIABP	VSTESTCD	VSTESTCD	DIABP

- Screenshot showing the combined raw dataset and variables information along with physical page number

Process the data to create SDTM specification

- This information is programmatically processed further in combination with standard SDTM metadata to create study specific
 - Dataset level metadata
 - Variable level metadata
 - Value level metadata

Generated Variable Level Metadata

VARIABLE	ORIGIN	ADDCOM_VSYN	ADDCOM_VS_M
VSTESTCD	CRF Pages 17, 18, 19	When VSYN.VSPERF then VSTESTCD='VSALL'	When VS_M.WEIGHT_VSORRES then VSTESTCD='WEIGHT', When VS_M.TEMP_VSORRES then VSTESTCD='TEMP', When VS_M.SYSBP_VSORRES then VSTESTCD='SYSBP', When VS_M.DIABP_VSORRES then VSTESTCD='DIABP', When VS_M.PULSE_VSORRES then VSTESTCD='PULSE', When VS_M.RESP_VSORRES then VSTESTCD='RESP', When VS_M.O2SAT_VSORRES then VSTESTCD='OXYSTAT'
VSORRES	CRF Pages 18, 19		Populate using VS_M.WEIGHT_VSORRES, VS_M.TEMP_VSORRES, VS_M.SYSBP_VSORRES, VS_M.DIABP_VSORRES, VS_M.PULSE_VSORRES, VS_M.RESP_VSORRES, VS_M.O2SAT_VSORRES
VSORRESU	CRF Pages 18, 19		Populate using VS_M.TEMP_VSORRES, VS_M.SYSBP_VSORRESU, VS_M.DIABP_VSORRESU, VS_M.PULSE_VSORRESU, VS_M.RESP_VSORRESU, VS_M.O2SAT_VSORRESU
VSSTAT	CRF Page 17	Populate using VSYN.VSPERF	
VSDTC	CRF Pages 18, 19		Populate using VS_M.VSDAT, VS_M.VSTIM. Use collected date and/or time variables to populate date in ISO 8601 format.
VSTPT	CRF Page 19		Populate using VS_M.VSTPT

- Variable Level Metadata

- Origin column
- One column for each input raw dataset with input variables

Generated Value Level Metadata

DOMAIN	VARIABLE	PARAMCD	ORIGIN	ADDCOM
VS	VSORRES	BMI	CRF Page 18	Populate using VS.BMI_VSORRES
VS	VSORRES	DIABP	CRF Pages 18, 19	Populate using VS.DIABP_VSORRES, VS_M.DIABP_VSORRES
VS	VSORRES	RESP	CRF Pages 18, 19	Populate using VS.RESP_VSORRES, VS_M.RESP_VSORRES
VS	VSORRES	SYSBP	CRF Pages 18, 19	Populate using VS.SYSBP_VSORRES, VS_M.SYSBP_VSORRES
VS	VSORRES	WEIGHT	CRF Pages 18, 19	Populate using VS.WEIGHT_VSORRES, VS_M.WEIGHT_VSORRES
SUPPDM	QVAL	PRECONYN	CRF Page 81	Populate using PRECONYN.PRECONYN
SUPPDM	QVAL	RACEOTH	CRF Page 6	Populate using DMPD.RACEOTH

- Value level metadata is generated for
 - Findings domains
 - Supplemental domains

Generated template code

- This approach can further be enhanced to create code template for SDTM programming
 - A section for each raw dataset used
 - Populate SDTM variable name on left hand side
 - Populate the raw variable names on right hand side

```
data CM01;
  set CM CMAS CMSC CMSCYN indsname=_x_;
  _src=scan(_x_,2,'.');

  if _src='CM' then do;
    cmcat=cmcat;
    cmdose=cmdstxt;
    cmdosfrq=cmdosfrq,frqoth;
    cmdostxt=cmdstxt;
    cmdosu=cmdosu,doseuoth;
    cmendtc=cmendat;
    cmenrtpt=cmongo;
    cmentpt=cmongo;
    cmindc=cmindc;
    cmroute=cmroute,routeoth;
    cmspid=cmspid;
    cmstdtc=cmstdat;
    cmtrt=cmtrt;

    *QNAMS mapping;

    cmreas=cmreas,cmreassp;
  end;

  if _src='CMAS' then do;
    cmcat=cmcat;
    cmdose=cmdstxt;
    cmdosfrq=cmdosfrq,frqoth;
```



Value generated

- All CRF datapoints get into SDTM specification
- CRF page numbers will be correct and specific even at value level metadata
- Consistency in variable naming conventions and specification language

Thank You



**The Global Healthcare
Data Science Community**

Contact Channels

📞 UK +44 1843 609600
✉ office@phuse.global
🌐 phuse.global

Social Media

🐦 [@phusetwitta](https://twitter.com/phusetwitta)
📘 [/phusebook](https://facebook.com/phusebook)
▶ [/phusetube](https://youtube.com/phusetube)
🌐 [/company/phuse](https://linkedin.com/company/phuse)