

PHUSE US CONNECT 2025

Utility Macro for Ensuring Consistent Updates of CRF Data in SDTM

Naren Surampalli

SM09



3/18/2025

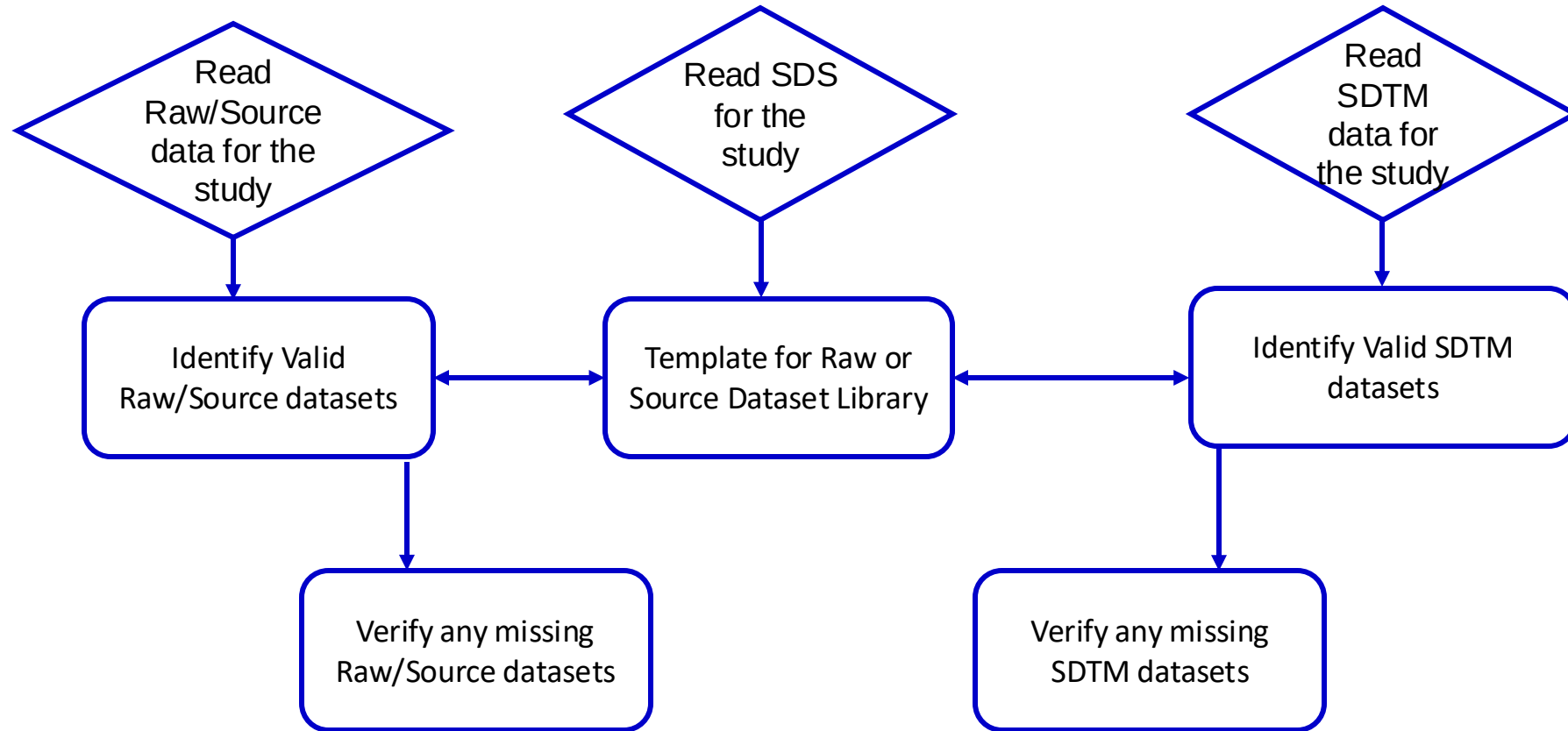
Agenda

- ❖ Introduction
- ❖ Utilizing Study Design Specifications
- ❖ Template for Source & SDTM library
- ❖ Create/Verify existence of valid Source/SDTM datasets
- ❖ Conclusion

Introduction

- The source data collected from the Case Report Form (CRF) pages is generally recorded in clinical trial databases.
- Various organizations may employ different tools to extract this source data and appropriately map it to the corresponding SDTM datasets.
- Commonly, not all CRF page data is available at the onset of a trial.
As the trial advances, additional data is gathered from growing CRF pages.
- The updated information from the new CRFs must be accurately extracted and synchronized for both the source and SDTM data.
- This presentation will outline a utility macro that incorporates multiple checks to ensure that updates to source data are reliably captured in the SDTM datasets.

Flow of the Macro



SDS = Study data specifications

Utilizing Study Design Specifications

Study design specifications (SDS) includes the list of CRF forms, and their visits required for a study, which eventually help in building a database by central designer. The first place to get list of all CRF forms and their form names is SDS. Usually, SDS is in an excel format, which can be easily read into SAS.

Form Title /Assessment	Form Short Title / CRF	Form RefName
INFORM SCREENING	SCREEN	D_INFSCR001
INFORM ENROLLMENT	ENROLL	D_INFENR001
INFORMED CONSENT	IC	DS001
DEMOGRAPHY	DEMOG	DM001
TRIGGER FORM	TRIG	R_TRIG001
TRIGGER FORM - COHORT SELECTION	COHORT SEL	R_TRIG001_1
INCLUSION/EXCLUSION CRITERIA	INEX	IE002
DATE OF VISIT	DOV	SV001
DISPOSITION - SCREENING	DISP SCR	DS001_4
MEDICAL HISTORY	MEDHX	MH001
CONCOMITANT PRIOR MEDICATIONS - VACCINATIONS	VAX HX	CM001_3
RANDOMIZATION	RAND	DS001_3
VITAL SIGNS	VITALS	VS001
TREATMENT - VACCINATION - INVESTIGATIONAL PRODUCT	VACIN TRT	EC001_2
SUMMARY OF REACTOGENICITY	SUMREACT	CESOR019

Template for Source & SDTM library

- After reading the form names from SDS into SAS dataset, next create a template for raw or source dataset library.
- This can be achieved by using the SDS created dataset as input and mapping all the form names to the relevant transformed raw/source dataset names.
- The newly created dataset will include the form names captured in a variable, expected mapped source dataset names and their relevant expected mapped SDTM names.
- This shell can be independent of therapeutic area and can include all expected source data.
- This step is done to ensure CRF form names get their relevant transformed raw/source datasets and also helps us to check valid records for relevant form names.

Template for Source & SDTM library

Table 1:

Obs	form	memname	sdtmname
1	AE001	SM_AE001	AE
2	AE001_1	SM_AE001	AE
3	BEETRK001_2	SM_BEETRK001	BE/CO
4	BEETRK001	SM_BEETRK001	BE/CO
5	BEETRK001_5	SM_BEETRK001	BE/CO
6	BEETRK001_3	SM_BEETRK001	BE/CO
7	BEETRK001_4	SM_BEETRK001	BE/CO
8	CECISR004	SM_CECISR004	FACE
9	CECISR004	SM_CECISR004	CE
10	CECISR004_1	SM_CECISR004	FACE
11	CECISR004_1	SM_CECISR004	CE
12	CECISR004_2	SM_CECISR004	FACE
13	CECISR004_2	SM_CECISR004	CE
14	CEIDT008	SM_CEIDT008	CE
15	CESD0006	SM_CESD0006	FACE

Identify Valid Records In Source Dataset

- Read the source/raw datasets in scope for the study.
- Align the source/raw data with SDS and source data template. Keep only the datasets in scope for the study.
- Identify valid records in source data using the field called notdonestate.
- Valid record (notdonestate=0) or invalid record (notdonestate=1).
- The goal is checking only those source datasets with valid records and ensure the information is captured in their relevant SDTM datasets

Identify Valid Records In Source Dataset

Table 2:

Raw datasets with valid records - CRF

Obs	memname
1	SM_AE001
2	SM_BEETRK001
3	SM_CECISR004
4	SM_CEIDT008
5	SM_CESOD006
6	SM_CM001
7	SM_CVECH001
8	SM_DD002
9	SM_DM001
10	SM_DS001

Table 3:

Raw datasets with no valid records - CRF

Obs	memname
1	SM_PRTFD002

Identify Valid Records In SDTM Dataset

- Identify valid raw datasets and compare against the expected SDTM datasets based on the template library.
- Check if all expected datasets are extracted.
- Flag datasets that are either not extracted or extracted but does not have a valid source.
- Please note that there could be valid SDTM datasets extracted based on non-CRF data specifications like immunogenicity and reactogenicity

Identify Valid Records In SDTM Dataset

Table 4:

Raw datasets with relevant SDTM datasets with valid records - CRF

Obs	memname	sdtmname
1	SM_AE001	AE
2	SM_BEETRK001	BE/CO
3	SM_CECISR004	CE
4	SM_CECISR004	FACE
5	SM_CEIDT008	CE
6	SM_CESOD006	CE
7	SM_CESOD006	FACE
8	SM_CM001	CM
9	SM_DM001	DM
10	SM_DS001	DS
11	SM_EC001_2	EC
12	SM_EC001_2	EX
13	SM_FAFUC009	SV
14	SM_FAVSD011	CE
15	SM_FAVSD011	FACE

Table 5:

Raw datasets present and its relevant SDTM datasets not present

Obs	memname	sdtmname
1	SM_CVECH001	CV
2	SM_DD002	DD
3	SM_EG001	EG
4	SM_CESOD006	FAAE
5	SM_MH001	MH
6	SM_PR001	PR

Identify the SDTM Datasets with missing CRF data

- In clinical trial study designs, we typically see multiple CRF pages collected to capture the same information but for different time periods.
- For example, disposition form is captured in different CRF pages for screening, randomization, treatment period and follow-up. All these CRF pages gets mapped to one SDTM called DS domain.
- For an ongoing study, not all CRF pages gets updated for the participant. Only during final CSR, we typically see the complete data.
- The macro checks if any new CRF page is captured in source but not captured in SDTM datasets as well.

Identify the SDTM Datasets with missing CRF data

Table 6 lists the different CRF pages for the same unique form. Table 7 lists the unique SDTM and their number of observations.

Table 6

Multiple CRF forms for the same domain						
Obs	FORMREFNAME	NOTDONESTATE	COUNT	memname	sdtmname	sum
1	AE001	0	41	SM_AE001	AE	41
5	CM001	0	1	SM_CM001	CM	1
6	CM001_2	0	648	SM_CM001	CM	649
7	CM001_3	0	685	SM_CM001	CM	1334
8	CM001_4	0	41	SM_CM001	CM	1375
9	DM001	0	354	SM_DM001	DM	354
10	DS001	0	354	SM_DS001	DS	354
11	DS001_3	0	352	SM_DS001	DS	706
12	DS001_4	0	353	SM_DS001	DS	1059
13	DS001_5	0	1	SM_DS001	DS	1060
14	DS001_6	0	5	SM_DS001	DS	1065
15	DS001_7	0	5	SM_DS001	DS	1070
16	X_DS_SLOPD001	0	353	SM_SLOPD	DS	1423
17	EC001_2	0	477	SM_EC001_2	EC	477
18	EC001_2	0	477	SM_EC001_2	EX	477
23	MH001	0	269	SM_MH001	MH	269
24	SV001	0	513	SM_SV001	SV	513

Table 7

Number of observations for each SDTM domain

Obs	sdtmname	nobs
1	AE	41
3	CM	1375
4	DM	354
5	DS	1070
7	EC	477
8	EX	477
11	MH	269
12	SV	513

Conclusion

- ❖ The macro discussed in this presentation helps the study teams to ensure the valid raw/source and SDTM datasets are accurately captured for each study and ensure we report all data in the CSRs accurately.
- ❖ This macro should be executed after each extraction done by the study team to ensure all relevant source/SDTM datasets are captured, and any missing datasets should be extracted or investigated to ensure no data is missed.
- ❖ The macro doesn't check non-CRF source vs expected SDTM data but gives a summary of expected non-CRF SDTM datasets

Acknowledgements

Liping Zhang (Therapeutic Area Lead, Exec. Director, Vaccines)

Richa Kala (Group Lead, Sr Director, Vaccines)

Questions ?

Thank You

Contact Information

Naren Surampalli – naren.surampalli@pfizer.com