



FROM SOURCE DATA TO SDTM: ENABLING A SEMI-AUTOMATED CONVERSION

Presented by Pieter Blomme
Clinical Data Programmer
SGS Life Sciences - Clinical Research

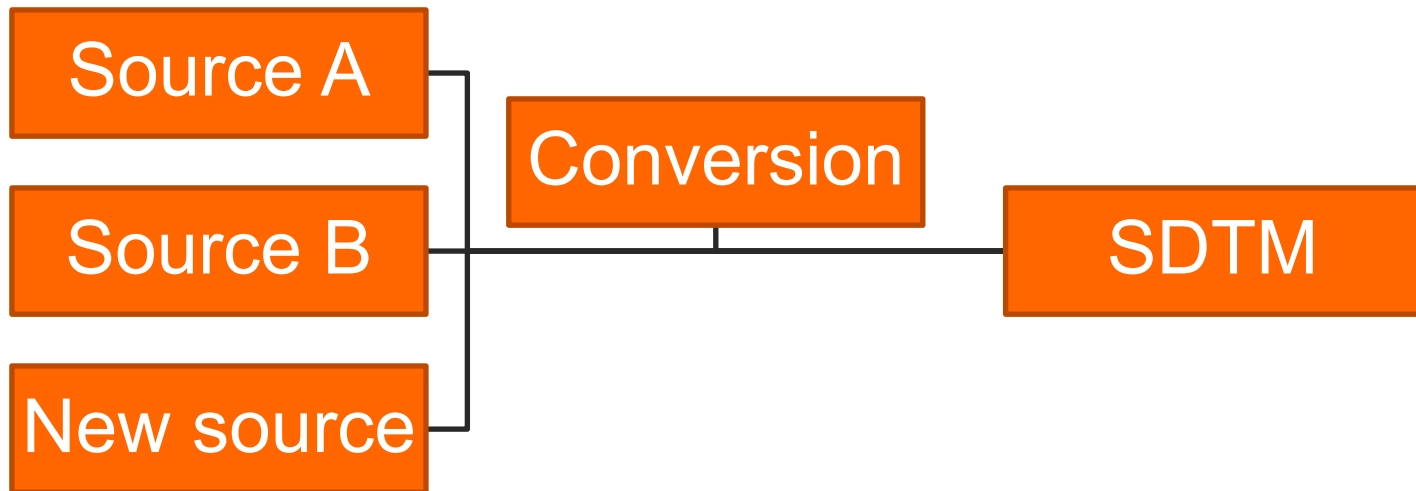
PhUSE US Connect 2019

WHEN YOU NEED TO BE SURE



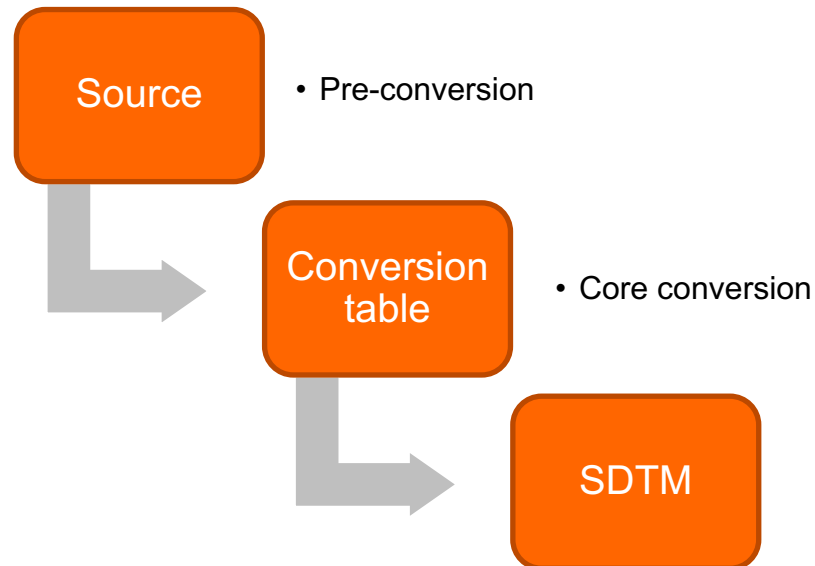
- **Introduction**
- Pre-conversion
 - Loading and formatting
 - General data pre-processing
 - Source-specific pre-processing
 - Mapping table
- SDTM conversion
- Downstream processes
- Conclusion

- Challenge
 - Different source types
 - Very different source structures
- Flexible approach needed
 - Code reusability where possible
 - Extensible for new source systems



- High-level operations for each SDTM conversion are quite similar
 - Mapping source to SDTM
 - Essential processing: visit assignment, USUBJID, ISO 8601 dates
 - Complex derivations: SE, epochs, DM domain, ...
- Opportunity for automation

- Pre-conversion
 - Hides source-specific complexity
 - Makes source available for mapping
- Conversion
 - Actual SDTM conversion
 - Universal approach, independent from source



- Introduction
- **Pre-conversion**
 - Loading and formatting
 - General data pre-processing
 - Source-specific pre-processing
 - Mapping table
- SDTM conversion
- Downstream processes
- Conclusion

- 
- Data loading
 - Formatting

- General pre-processing
- Source-specific pre-processing

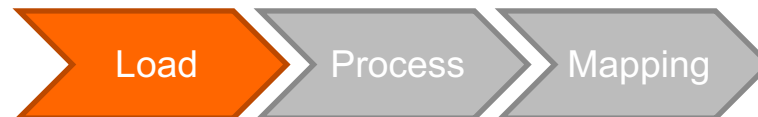
- Creating the conversion table

Types of source files

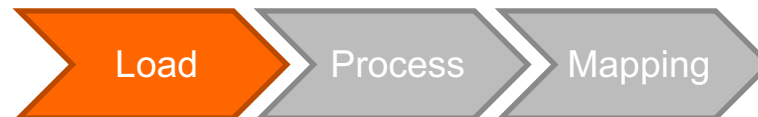
- SAS: XPT or SAS
- Text: csv, ASCII, tab delimited, ...
- Database exports

Extra complexity:

- Different SAS/Database versions
- Different character sets, eg. single byte vs Unicode



- Automatic download and load process
- Exact copy of the source in Oracle database:
 - Include all administrative fields
 - Include special characters
- Source copy is never changed directly: trace and compare
- Continually changing source system
 - Be vigilant with edge cases
 - QC of data integrity still needed
 - Living process



- Subject ID => USUBJID
- VISIT slotting
- ISO 8601



■ Problem

- Name and location different per source
- May require joining multiple tables
- Process per source is consistent

■ Solution

- Pre-conversion derives USUBJID automatically
- Hides underlying complexity from general conversion



■ VISIT slotting

- Often multiple fields
- Should be available as soon as possible
- Derived using help table to derive client-specific visit identifiers

VISITID	VISITNAME	VISIT	VISITNUM	XXTPT	XXTPTNUM	XXTPTREF
500964	SCREENING 2	SCREENING	0			
118916	SCREENING 3	SCREENING	0			
579642	SCREENING	SCREENING	0			
19333	PERIOD 1 BASELINE	PERIOD1_DAY-1	0.999			
844976	PERIOD 1 D-1	PERIOD1_DAY-1	0.999			
991851	PERIOD 1 D1 0-12H (9H)	PERIOD1_DAY1	1.001	9H	540	DOSE DAY 1
621750	PERIOD 1 D1 6H	PERIOD1_DAY1	1.001	6H	360	DOSE DAY 1
675528	PERIOD 1 D1 4H	PERIOD1_DAY1	1.001	4H	240	DOSE DAY 1
485625	PERIOD 1 D1 1H	PERIOD1_DAY1	1.001	1H	60	DOSE DAY 1
404045	PERIOD 1 D1 0-12H (11H)	PERIOD1_DAY1	1.001	11H	660	DOSE DAY 1

- Many formats => ISO 8601
- Normally consistent per source system
- Not always clear which fields are date fields
 - Metadata
 - Free text can also be date: pattern matching
 - Invalid dates: return original input (traceability)
- Pre-conversion:
 - 1 system per source
 - ISO 8601 are available from start for derivations



- Help tables
 - Restructuring source tables for convenience
- Problem
 - Related data may be entered on multiple forms
 - eg. Start / End of meal
 - eg. eSource questionnaires
 - Inconvenient narrow structure
- Solution
 - Mapping help table instead of source



■ Narrow questionnaire table

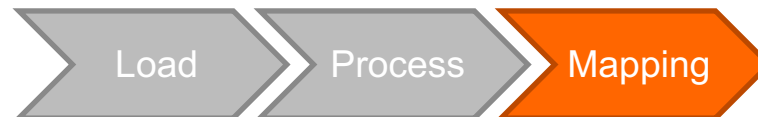
- 1 record per question
- Hard to map: requires decoding the question name

PERIOD_NAME	TRIALDAY_NAME	QUES_NM	ANS_TXT
P01	D004	GEN_SYMP + medical intervention_3	Medical intervention: NO
P01	D004	GEN_SYMP + medical intervention_3	Medical intervention: NO
P01	D004	GEN_SYMP + medical intervention_3	Medical intervention: NO
P01	D004	GEN_SYMP + medical intervention_3	Medical intervention: NO
P01	D004	GEN_SYMP + medical intervention_3	Medical intervention: NO
P01	D008	GEN_SYMP + medical intervention_3	Score : 0
P01	D008	GEN_SYMP + medical intervention_3	Score : 0
P01	D008	GEN_SYMP + medical intervention_3	Medical intervention: NO
P01	D008	GEN_SYMP + medical intervention_3	Medical intervention: NO
P01	D008	GEN_SYMP + medical intervention_3	Medical intervention: NO
P01	D008	GEN_SYMP + medical intervention_3	Medical intervention: NO

- Wide table
 - 1 record contains all questions of a single form
 - Map per variable name

	PERIOD_NAME	TRIALDAY_NAME	Details 5_7_b	Details 5_7_f
▶	P01	D001		
	P01	D002		
	P01	D003		
	P01	D004		
	P01	D008		
	P01	D015		
	P01	D085	Medical intervention: NO	Medical intervention: NO
	P01	D001		
	P01	D002		
	P01	D003		
	UNSCHEDULED	0024:00 POST		

- Structural differences between forms
- Many types of mapping
 - Multiple forms to 1 record
 - 1 form to multiple records
 - Some fields trigger new forms
- Universal solution:
 - 1 general mapping table
 - Define where each table/variable goes
 - Possibility to link to other forms if needed

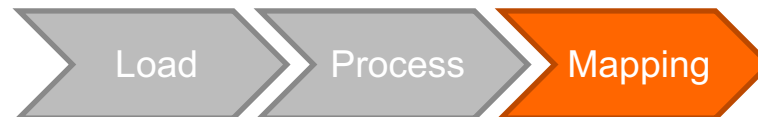




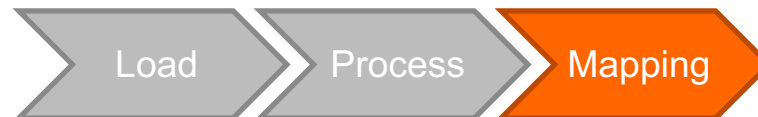
MAPPING TABLE

FORMREFNAME	CC_CONTROLREFNAME	FINDING	DOMAIN	TARGET	PARAM1	CONVERTTYPE	CODE
MD_LB_GL_904	RECORDPOSITION					SKIP	
MD_LB_GL_904	LBPERF_STD	Y	LB	LBALL		ANY	case when upper(CTS\$ITEM) in ('N','NA') then 'NOT DONE' end
MD_LB_GL_904	LBPERF_STD	N	LB	LBREASND		DECODE	DECODE(upper(CTS\$ITEM),'N','SAMPLE NOT COLLECTED','NA','NOT APPLICABLE')
MD_LB_GL_904	LBPERF_STD	Y	X1	LBALL		ANY	case when upper(CTS\$ITEM) = 'Y' then 'Y' end
MD_LB_GL_904	LBPERF_STD	N	LB	LBFAST	LBCOND_STD	PARAM1	upper(CTS\$PARAM1)
MD_LB_GL_904	LBPERF_STD	N	LB	LBCAT		CONSTANT	CHEMISTRY
MD_LB_GL_904	LBPERF_STD	SS	LB	LBCLSIG		CONSTANT	
MD_LB_GL_904	LBPERF_STD	SS	LB	CSFLAG		CONSTANT	
MD_LB_GL_904	LBPERF_STD	SS	LB	PICSFLAG		CONSTANT	
MD_LB_GL_904	LBPERF_STD	SS	LB	LBCODE		CONSTANT	
MD_LB_GL_904	LBDAT	N	LB	LBDTC	LBTIM	ANY	ISO8601_DATE upper(CTS\$PARAM1)

- Administrative and technical fields are filtered out
 - Labels, metadata, ...
 - Not on aCRF, so not needed in SDTM
- Mapping Tool: user interface
- Mapping table is structurally same for all trials => Allows copying



- Result: Conversion table
 - Contains all source fields
 - Associated mapping
 - VISIT, USUBJID information per record
- Core conversion starts from this table
- Source system independent



■ Example: PE - mapping

MDFCID	FORMREFNAME	CC_CONTROLREF...	FI...	DO...	TARGET	PARM1	CODE
2593.00	PHYSICAL EXA...	SKIN AND MUCOU...	Y	PE	SKINMM	SKIN/MUCOUS MEM...	case when upper(CTS\$ITEM) = 'ABNORMAL' the...
2593.01	PHYSICAL EXA...	SKIN AND MUCOU...	N	PE	PEDTC	SKIN AND MUCOUS ...	upper(CTS\$PARM1)
2593.02	PHYSICAL EXA...	SKIN AND MUCOU...	N	PE	PEREASND	SKIN AND MUCOUS ...	DECODE(upper(CTS\$ITEM), 'NOT DONE', upper(...
2593.03	PHYSICAL EXA...	SKIN AND MUCOU...	SS	PE	PECLSIG	SKIN/MUCOUS MEM...	substr(upper(CTS\$PARM1),1,1)
2593.04	PHYSICAL EXA...	SKIN AND MUCOU...	CS	PE	COVAL	SKIN AND MUCOUS ...	case when upper(CTS\$ITEM) = 'NOT DONE' the...
2593.05	PHYSICAL EXA...	SKIN AND MUCOU...	N	PE	PESTAT		CASE WHEN UPPER(CTS\$ITEM) in ('NOT DONE', '...

■ Example: PE – conversion table

SOURCEID	FORMREFN...	CC_CONTROL...	USUBJID	FINDING	DOMAIN	TARGET	CODE	VISIT	VISITNUM
637166£637166_124-...	PHYSICAL E...	SKIN AND MU...	dummy	Y	PE	SKINMM	NULL	PERIOD1_DAY-1	0.999
637166£637166_124-...	PHYSICAL E...	SKIN AND MU...	dummy	SS	PE	PECLSIG		PERIOD1_DAY-1	0.999
637166£637166_124-...	PHYSICAL E...	SKIN AND MU...	dummy	N	PE	PESTAT		PERIOD1_DAY-1	0.999
637166£637166_124-...	PHYSICAL E...	SKIN AND MU...	dummy	N	PE	PEDTC	2016-11-05T09:00	PERIOD1_DAY-1	0.999
637166£637166_124-...	PHYSICAL E...	SKIN AND MU...	dummy	N	PE	PEREASND		PERIOD1_DAY-1	0.999
637166£637166_124-...	PHYSICAL E...	SKIN AND MU...	dummy	CS	PE	COVAL	NAP - NAP	PERIOD1_DAY-1	0.999
637166£637166_218-...	PHYSICAL E...	SKIN AND MU...	dummy	Y	PE	SKINMM	ABNORMAL	UNSCHEDULED_VISIT	99
637166£637166_218-...	PHYSICAL E...	SKIN AND MU...	dummy	SS	PE	PECLSIG		UNSCHEDULED_VISIT	99
637166£637166_218-...	PHYSICAL E...	SKIN AND MU...	dummy	N	PE	PEREASND		UNSCHEDULED_VISIT	99
637166£637166_218-...	PHYSICAL E...	SKIN AND MU...	dummy	N	PE	PEDTC	2016-10-16T09:00	UNSCHEDULED_VISIT	99
637166£637166_218-...	PHYSICAL E...	SKIN AND MU...	dummy	N	PE	PESTAT		UNSCHEDULED_VISIT	99
637166£637166_218-...	PHYSICAL E...	SKIN AND MU...	dummy	CS	PE	COVAL	NAP - NAP	UNSCHEDULED_VISIT	99

- Introduction
- Pre-conversion
 - Loading and formatting
 - General data pre-processing
 - Source-specific pre-processing
 - Mapping table
- **SDTM conversion**
- Downstream processes
- Conclusion

- Automated pre-conversion:
 - Download new source
 - Load
 - Apply mapping table and generate conversion table
- SDTM conversion loops through conversion table
 - Creates domain records with USUBJID, VISIT info
 - Updates associated fields, eg. –STAT, --DTC
 - Result: semi-converted SDTM domains
- Derivations: independent of source



SDTM CONVERSION

■ Example: PE – semi-converted

STUDYID	DOMAIN	USUBJID	PESEQ	PETESTCD	PETEST	PEORRES	PESTRESC	PESTAT	PEREASND	PEBLFL	VISITNUM	VISIT	EPOCH	PEDTC	PEDY	PETPT	PETPTNUM
dummy	PE	dummy		SKINMM	Skin And ...	NORMAL					0	SCREENING		2017-11-08T13:01			
dummy	PE	dummy		SKINMM	Skin And ...	NORMAL					0.999	PERIOD1_DAY-1		2017-11-21T10:16			
dummy	PE	dummy		SKINMM	Skin And ...	NORMAL					1.003	PERIOD1_DAY3		2017-11-24T09:28		48H	2880
dummy	PE	dummy		SKINMM	Skin And ...	NORMAL					1.008	PERIOD1_DAY8		2017-11-29T08:40		PRE-DOSE	-1
dummy	PE	dummy		SKINMM	Skin And ...	NORMAL					1.015	PERIOD1_DAY15		2017-12-06T08:55		PRE-DOSE	-1
dummy	PE	dummy		SKINMM	Skin And ...	NORMAL					1.022	PERIOD1_DAY22		2017-12-13T08:20		PRE-DOSE	-1
dummy	PE	dummy		SKINMM	Skin And ...	NORMAL					1.028	PERIOD1_DAY28		2017-12-19T10:23			
dummy	PE	dummy		SKINMM	Skin And ...	NORMAL					1.031	PERIOD1_DAY31		2017-12-22T09:15		48H	2880
dummy	PE	dummy		SKINMM	Skin And ...	NORMAL					1.036	PERIOD1_DAY36		2018-01-03T11:31			
dummy	PE	dummy		SKINMM	Skin And ...	NORMAL					0	SCREENING		2017-11-08T12:20			
dummy	PE	dummy		SKINMM	Skin And ...	NORMAL					0.999	PERIOD1_DAY-1		2017-11-21T09:53			

- Introduction
- Pre-conversion
 - Loading and formatting
 - General data pre-processing
 - Source-specific pre-processing
 - Mapping table
- SDTM conversion
- **Downstream processes**
- Conclusion

■ QC Tool

- Validation of conversion programming
- Runs on SDTM => 1 tool for all source systems

SGS

preprod_biomme CDP New Tab Switch app Feedback Logout

Studies Patients Steps Columns Issues Listings Codelists Releases Links

Here you get an overview of all QClistings.

Q Go Actions

Listingtype 4 NOK Records, OK Records, CHANGED Records, NEW Records

Listingtype : DOMAIN 8

	ListingId	Ground	Stat	GoStep	Status Ok	Status NOK	Status Changed	Totalrecords	Status New
	SDTH_AE	3	OK	23	4	0	0	4	0
	SDTH_OI	3	OK	24	6	0	0	6	0
	SDTH_OA	3	EMPTY	25	0	0	0	0	0
	SDTH_OI	3	OK	26	4	0	0	4	0
	SDTH_OS	3	OK	27	21	0	0	21	0
	SDTH_OV	3	EMPTY	28	0	0	0	0	0
	SDTH_BO	3	OK	29	3	0	0	3	0
	SDTH_BO	3	OK	30	147	0	0	147	0
	SDTH_EX	3	OK	31	4	0	0	4	0
	SDTH_FA	3	OK	32	4	0	0	4	0
	SDTH_IE	3	OK	33	8	0	0	8	0
	SDTH_LB	3	OK	34	1201	0	0	1201	0
	SDTH_UH	3	OK	35	18	0	0	18	0
	SDTH_UJL	3	OK	36	28	0	0	28	0
	SDTH_FE	3	OK	37	180	0	0	180	0

■ All2One Tool

- Data cleaning
- Runs on SDTM => Rules are more versatile and reusable

The screenshot displays the All2One Tool interface. The main window shows 'Rulepackage Statistics at 5/02/2019 - 20:46' with the following data:

Category	Count
N° of Rules	259
N° of Checks	109
N° of Listings	150
N° of rules, state NEW	0
N° of rules, state INVALID	0
N° of rules, state INVALID - No Output	0
N° of rules, state INVALID - Corrupt	0
N° of rules, state VALID	0
N° of rules, state PENDING	0
N° of rules, state ACCEPTED	230
N° of rules, state REJECTED	0
N° of rules, state DATED	1
N° of rules, state DELETED	28
N° of rules without SQL	
N° of rules, code review not yet performed	
N° of non-deleted rules with SQL in conflict	

An overlay window titled 'Rule Package' is open, showing a table of rules. The table has columns: detail, history, pre-def, output, panel, rulename, and description. The rules listed are:

detail	history	pre-def	output	panel	rulename	description
				AE	AE_001_L	Check that the following items are completed for each adverse event: ([AETERM]), seriousness ([AESER]), pattern of the event ([AEPATT]), regarding investigational medication ([AEACN]), relation to investigational medication ([AESEV]) and start date ([AESTDTC]). Reported: [AESPID], [AETERM], [AESTDTC], [AESER], [AEPATT], [AEACN], [AESEV], [AESTDTC#].
				AE	AE_001_S	Check if end date of AE ([AENDTC]) < minimum performed study age [ECMOOD] = 'PERFORMED', that action taken regarding study agent. Reported: [AETERM], [AESTDTC], [AENDTC], [AEACN], [ECSTDTC#].
				AE	AE_002_S	Check if start date of AE ([AESTDTC]) < minimum performed administrative date where [ECMOOD] = 'PERFORMED' or > maximum performed administrative date where [ECMOOD] = 'PERFORMED' that medication ([AEREL]) = 'NOT RELATED'. * When [ECENDTC] is not completed, replace with current date. Reported: [AETERM], [AESTDTC], [AEREL], [ECSTDTC#MIN], [ECENDTC].
				AE	AE_003D_L	Check that the end date of AE ([AENDTC]) <= trial termination date [DSSCAT] of the record in [ZZ_DS] where [EVENT] = 'TRIAL TERMINATION'. Reported: [AETERM], [AESTDTC], [AENDTC], [DSSTDTC], [X1SPID#], [X1ORRES#ZZ].
				AE	AE_006_L	Check if adverse event outcome ([AEOUT]) = 'FATAL' that the AE is cc...

Buttons at the bottom of the 'Rule Package' window: Refresh, Close.

At the bottom right of the main window: No QC sync | No Live sync

- Pre-conversion leads to significant decrease in trial specific work
 - Quick delivery of SDTM
 - Reduced workload
 - Implement new source systems easily
- Core philosophy
 - Pre-conversion: source specific, but trial independent
 - Mapping: linked to both source and trial specifications, but largely reusable
 - SDTM conversion: source independent, SDTM and client specific derivations



THANK YOU FOR YOUR ATTENTION



Questions ?

Meet us at the SGS booth 13

CONTACT US

clinicalresearch@sgs.com

EUROPE: +32 15 27 32 45

AMERICAS: + 877 677 2667

WWW.SGS.COM/CRO



JOIN THE SCIENTIFIC COMMUNITY

CONNECT ON LINKEDIN

www.sgs.com/LinkedIn-Life