

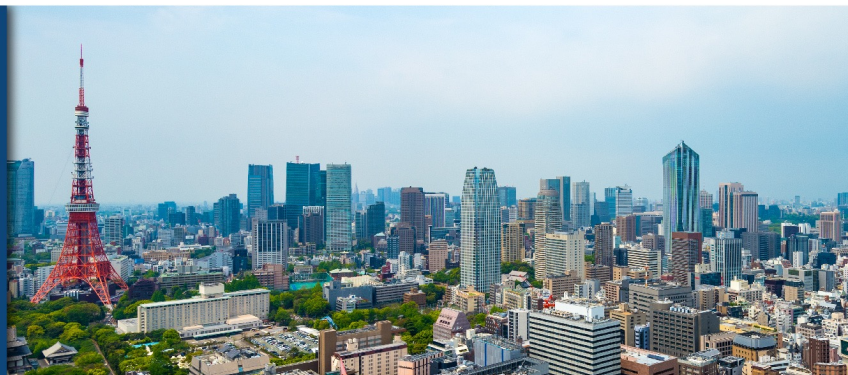


2023

JAPAN

INTERCHANGE

TOKYO | 10-11 JULY

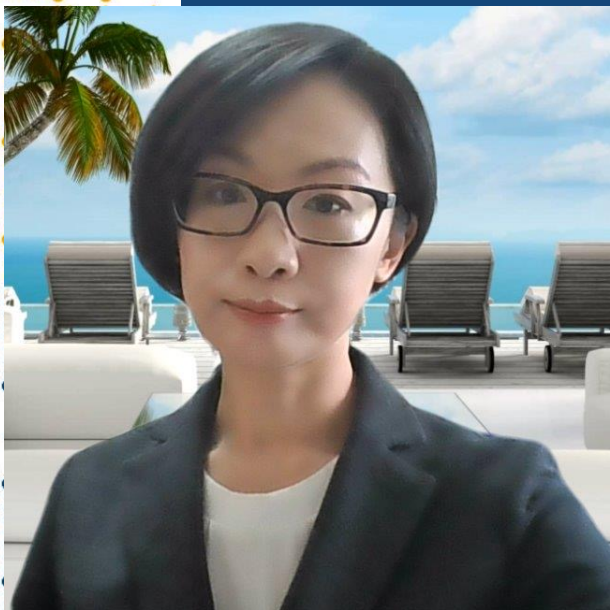


## Automating SDTM and ADaM Creation in Clinical Trials

Presented by Rie ICHIHASHI, Biostatistics and Programming, Sanofi

# Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*
- *{Please disclose any financial relationship or conflict of interest relevant to this presentation here OR}*
- *The author(s) have no real or apparent conflicts of interest to report.*



# Meet the Speaker

Rie Ichihashi

**Title:** Statistical Programmer

**Organization:** Sanofi K.K.

Rie Ichihashi is a senior statistical programmer in Biostatistics and Programming Department at Sanofi. Rie has MS in biology, has 16 years of industry experience in statistical programming and data standards. Over the years participated in setting up Sanofi global SDTM/ADaM standards. More recently, her work has been focused on end-to-end standardization and automation of e-data submission.

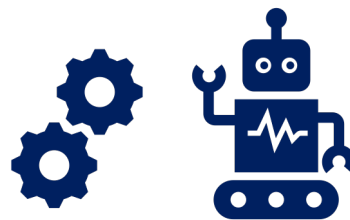
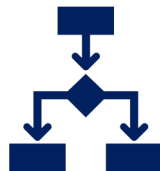


# Agenda

1. Introduction 疑似コード
2. ADaM and the concept of the **pseudo-code**
3. Nine Standard Derivation Types
4. Automation of programs
5. Benefits and challenges of **pseudo-code**

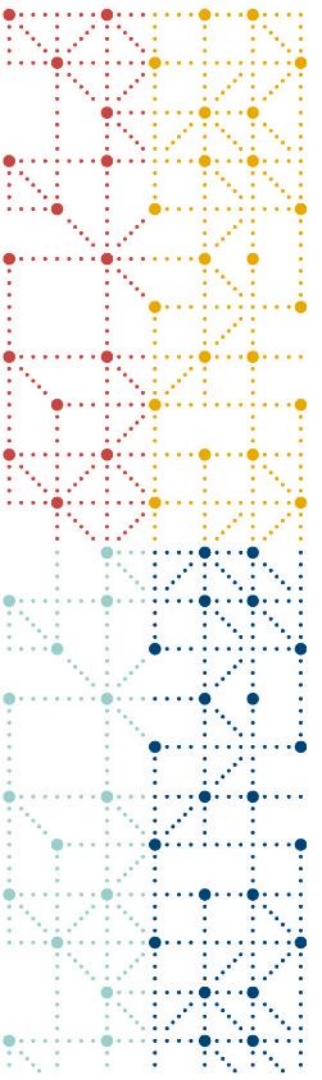
# INTRODUCTION

- Automating SDTM and ADaM in clinical trials
- Importance of the **standards and specifications**
- Concept of the **pseudo-code** and how it helps with automation  
疑似コード



Standardization

Automation



## ADaM and the concept of the pseudo-code

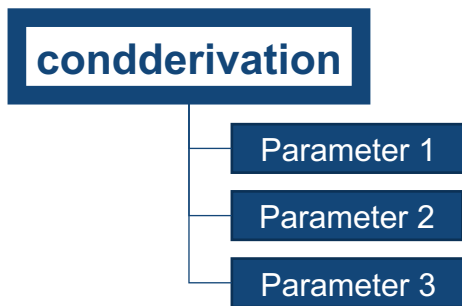
# ADAM AND THE CONCEPT OF THE PSEUDO-CODE

## Derivation specifications

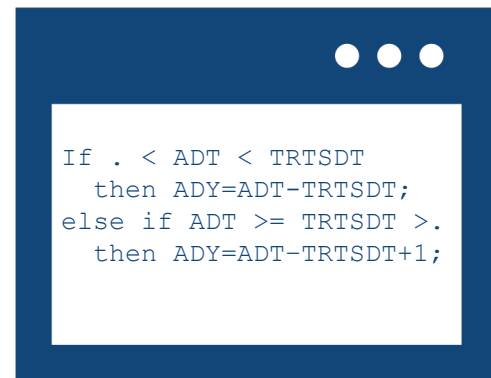


## Pseudo-code

疑似コード



## Program code



Code Generator

# ADAM AND THE CONCEPT OF THE PSEUDO-CODE

A specific way of writing the specifications to achieve automation.

For each ADaM variable in ADaM metadata, a **pseudo-code** is created with standard derivation rules.

Code Generator parses a pseudo-code to generate program code

## Variable Metadata

## 疑似コード

Dataset Name	Relative Order	Variable Name	Variable Label	Type	Controlled Terminology	Source	Derivation	Assigned	Pseudo Code	Core	Instructions for Programmers
ADSL	170	AAGEU	Analysis Age Units	text	AGEU			Assigned as "YEARS, MONTHS,	assigned	Cond	Conditionally required if AAGE is in
ADSL	180	AGEGR1	Pooled Age Group 1	text	[AGEGR1]		For AGE of [xxx] to [xxx], AGEGR1 = [xxx]		recode	Cond	Define AGE ranges in accordance with if AGE=. then AGEGR1=""
ADSL	190	AGEGR2	Pooled Age Group 2	text	AGEGRPE		For AGE of [xxx] to [xxx], AGEGR2 = [xxx]		conderivation	Cond	AGEGR2 is reserved for EudraCT : Preterm newborn - gestational age
ADSL	210	SEX	Sex	text	SEX	DM.SEX			source	Req	
ADSL	220	RACE	Race	text	RACE	DM.RACE			source	Req	If the SDTM RACE is "MULTIPLE",
ADSL	230	RACEOR	Original Race	text	RACEC	SUPPDM.RACEOR			source	Perm	
ADSL	240	RACEGRy	Pooled Race Group y	text	[RACEGRy]		For RACE of [xxx] or [xxx], RACEGRy = [xxx]		recode	Perm	Character description of a grouping used for special case when study h



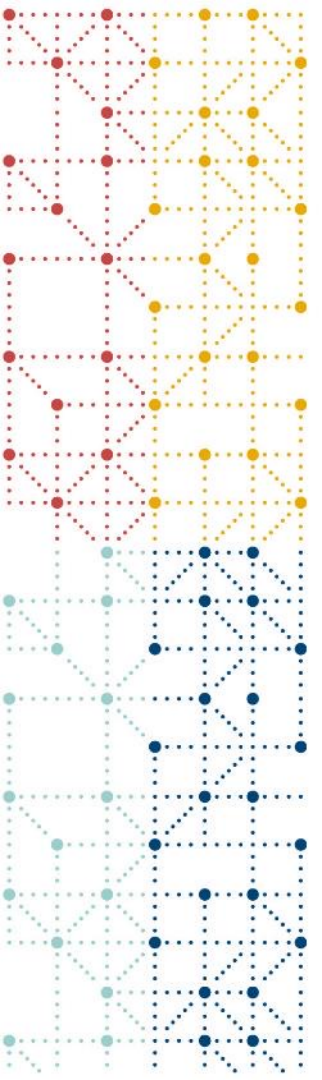
# ADAM AND THE CONCEPT OF THE PSEUDO-CODE

**Pseudo-code** is also applicable to value level derivations.

*WhereClauses, the main purpose is to provide metadata for Value Level Metadata*

疑似コード

Dataset Name	Variable Name	Where Variable	Comparator	Check Value	Value Label	Derivation	Pseudo Code
ADVS	AVAL	PARAMCD	EQ	WEIGHT	Weight (kg)	VS.VSSTRESN Where PARAMCD=VS.VSTESTCD	copy_stres
ADVS	AVAL	PARAMCD	EQ	DIABP	Diastolic Blood Pressure (mmHg)	VS.VSSTRESN Where PARAMCD=VS.VSTESTCD. An average record is created for triplicate measurements on the same date.	copy_stres, std_function(dtype_average)
ADVS	AVAL	PARAMCD	EQ	SYSBP	Systolic Blood Pressure (mmHg)	VS.VSSTRESN Where PARAMCD=VS.VSTESTCD. An average record is created for triplicate measurements on the same date.	copy_stres, std_function(dtype_average)



## Nine Standard Derivation Types

# NINE STANDARD DERIVATION TYPES

Pseudo Code	Description
source	Indicates the variable has an origin of “Predecessor” where the source variable is defined in the “Source” column. This is used mainly for SDTM predecessor variables
assigned	Assignment of a constant value for all the records in a dataset
derivation	Indicates that detailed derivation specs are available in the “Derivation” column in the “Variable Metadata” tab
recode	Indicates that detailed derivation algorithm is available in the “recode” tab
rename	Renaming a variable. This is used when a temporary variable is renamed into the ADaM dataset variable. The most common example is –DT, –DTM, and –TM variables. The ISO8601 datetime variables (e.g. EXSTDTC) in a source SDTM variable that is converted to temporary SAS date, datetime, and time variables (e.g. EXSTD, EXSTD, and EXSTT) early in the ADaM program, and renamed to suitable ADaM variable names (e.g. ADEX.ASTD, ADSL.TRTSTD, etc.)
condderivation	Indicates that detailed derivation algorithm is available in the “condderivation” tab
lookup	Simple proc sql merge of 2 datasets, using subset condition if needed
std_function	Complex derivations which involve multiple datasets and/or multiple steps of computations that cannot be classified into the derivation types described above
whereclauses	Indicates that pseudo-code for value level parameters is populated in the VLM table in the “WhereClauses” tab

# NINE STANDARD DERIVATION TYPES

- Three simple pseudo-codes “**source**”, “**derivation**”, and “**assigned**”. These pseudo-codes use the 3 corresponding variable metadata elements directly. Thus, the column which has the same name as the pseudo-code is used to pass the key element to the code generator.

## *Columns for additional information for the code generator*

Dataset Name	Variable Name	Variable Label	Type	Source	Derivation	Assigned	Pseudo Code
ADSL	STUDYID	Study Identifier	text	DM.STUDYID			source
ADSL	TRTDURD	Total Treatment Duration (Days)	integer		TRTEDT - TRTSDT + 1		derivation
ADEG	AWU	Analysis Window Unit	text			Assigned as "DAYS"	assigned

# NINE STANDARD DERIVATION TYPES

- On the other hand, when the pseudo-code is “**recode**”, “**rename**”, “**condderivation**”, “**lookup**”, “**std\_function**” or “**whereclauses**”, additional information needs to be entered in another tab that has the same name as the pseudo-code.

## *Tabs for additional information for a code generator*

Dataset Name	Variable Name	Variable Label	Type	Source	Derivation	Assigned	Pseudo Code	Core	Instructions for Programmers
ADAE	AREL	Analysis Causality	text		AE.AEREL. If AEREL is missing, AREL=[xxx].		condderivation	Perm	For Pharma & Pasteur: If AEREL is derived due "Related" as per Safety Guidelines), AREL should be set to AEREL.
ADAE	ARELN	Analysis Causality (N)	integer			Numeric code for AREL	recode	Perm	The numeric code for AREL. One-to-one map to

Dataset Metadata	<b>Variable Metadata</b>	recode	rename	condderivation	lookup	std_function	WhereClause: ...	+	⋮
------------------	--------------------------	--------	--------	----------------	--------	--------------	------------------	---	---

# DETAILS ABOUT DERIVATION TYPES — “source”

- 1. Copy from a variable in the main source dataset (pseudo-code= “source”).

Variables Example	Rendered SAS® code
ADSL.STUDYID, ADSL.USUBJID, ADSL.AGE, etc.	Keep STUDYID USUBJID AGE;

## Pseudo-code “source” in the Variable Metadata tab

Dataset Name	Variable Name	Variable Label	Type	Source	Derivation	Assigned	Pseudo Code
ADSL	STUDYID	Study Identifier	text	DM.STUDYID			source
ADSL	USUBJID	Unique Subject Identifier	text	DM.USUBJID			source
ADSL	AGE	Age	integer	DM.AGE			source

[Variable Metadata](#) | [recode](#) | [rename](#) | [condderivation](#) | [lookup](#) | [std\\_function](#) ... (+)

ADSL (Subject-Level Analysis) - [ADaMIG 1.1]

Variable	Label / Description	Type	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
STUDYID	Study Identifier	text	12		Predecessor: DM.STUDYID
USUBJID	Unique Subject Identifier	text	11		Predecessor: DM.USUBJID

# DETAILS ABOUT DERIVATION TYPES — “derivation”

- 2. Simple computation which appears in define.xml as the computational algorithm is directly used as the program code (pseudo-code = “derivation”).

Variables Example	Rendered SAS <sup>®</sup> code
Total Treatment Duration (Days) (ADSL.TRTDURD).	<code>TRTDURD = TRTEDT - TRTSDT + 1;</code>

## Pseudo-code “derivation” in the Variable Metadata tab

Dataset Name	Variable Name	Variable Label	Type	Source	Derivation	Assigned	Pseudo Code
ADSL	TRTDURD	Total Treatment Duration (Days)	integer		TRTEDT - TRTSDT + 1		derivation

▶ ...
Variable Metadata
recode
rename
condderivation
lookup
std\_function
... (+)

Variable	Label / Description	Type	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
TRTDURD	Total Treatment Duration (Days)	integer	3		Derived (Source: Sponsor) TRTEDT-TRTSDT+1

# DETAILS ABOUT DERIVATION TYPES — “assigned”

- 3. Assignment of a constant value for all the records in a dataset (pseudo-code = “assigned”).

Variables Example	Rendered SAS <sup>®</sup> code
Analysis Window Unit (ADEG.AWU)	AWU = "DAYS";

## Pseudo-code “assigned” in the Variable Metadata tab

Dataset Name	Variable Name	Variable Label	Type	Source	Derivation	Assigned	Pseudo Code
ADEG	AWU	Analysis Window Unit	text			Assigned as "DAYS"	assigned

Variable Metadata | recode | rename | conderivation | lookup | std\_function

Variable	Where Condition	Label / Description	Type	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
AWU		Analysis Window Unit	text	4	Unit - AWU • "DAYS"	Assigned (Source: Sponsor) Assigned as "DAYS".



# DETAILS ABOUT DERIVATION TYPES — “recode”

- 4. Recode each value of a source variable to another value (pseudo-code = “recode”).

Variables Example	Rendered SAS® code
Pooled Age Group 1 (ADSL.AGEGR1) Pooled Age Group 1 (N) (ADSL.AGEGR1N)	<pre>if . &lt; AGE &lt; 60 then do;     AGEGR1 = "&lt; 60 years ";     AGEGR1N = 1; end; else if 60 &lt;= AGE &lt;= 70 then do;     AGEGR1 = "60-70 years";     AGEGR1N = 2; end; else if 70 &lt; AGE then do;     AGEGR1 = "over 70 years";     AGEGR1N = 3; end;</pre>

# DETAILS ABOUT DERIVATION TYPES — “recode” (cont.)

## Pseudo-code “recode” in the Variable Metadata tab

Dataset Name	Variable Name	Variable Label	Type	Source	Derivation	Assigned	Pseudo Code
ADSL	AGEGR1	Pooled Age Group 1	text		For AGE under 60,		recode
ADSL	AGEGR1N	Pooled Age Group 1 (N)	integer			Numeric code for AGEGR1	recode

... ata Variable Metadata recode rename condderivation lookup std\_function WI ...

## Corresponding “recode” tab

Dataset Name	Variable Name	Source Variable	Source Variable Value	Output Value (Char)	Output Value (Num)
ADSL	AGEGR1 AGEGR1N	AGE	(null:60)	< 60 years	1
ADSL	AGEGR1 AGEGR1N	AGE	[60:70]	60-70 years	2
ADSL	AGEGR1 AGEGR1N	AGE	(70	over 70 years	3

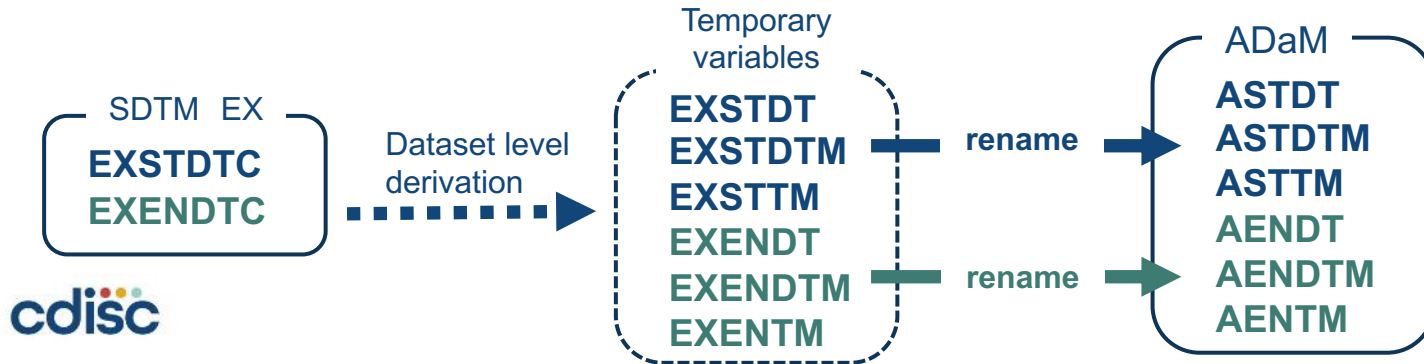
... Variable Metadata recode rename condderivation lookup std\_function WI



# DETAILS ABOUT DERIVATION TYPES — “rename”

- 5. Rename a variable (pseudo-code = “rename”).
- This is used when a temporary variable is renamed in the ADaM dataset. The most common example is --DT, --DTM, and --TM variables.

Variables Example	Rendered SAS <sup>®</sup> code
Analysis Start Date (ADEX.ASTDT) Analysis Start Datetime (ADEX.ASTDTM) Analysis Start Time (ADEX.ASTTM)	<pre>rename  EXSTDT = ASTDT         EXSTDTM = ASTDTM         EXSTTM = ASTTM;</pre>



# DETAILS ABOUT DERIVATION TYPES — “rename” (cont.)

## Pseudo-code “rename” in the Variable Metadata tab

Dataset Name	Variable Name	Variable Label	Type	Source	Derivation	Assigned	Pseudo Code
ADEX	ASTDTM	Analysis Start Datetime	integer		or EX.EXSTDTC		dataset_derivation(DTC) rename
ADEX	ASTDT	Analysis Start Date	integer		date portion of		dataset_derivation(DTC) rename
ADEX	ASTTM	Analysis Start Time	integer		time portion of		dataset_derivation(DTC) rename

Variable Metadata | recode | rename | condderivation | lookup | std\_function | Where( ... +

## Corresponding “rename” tab

Dataset Name	Variable Name	Source Variable Name
ADEX	ASTDTM ASTDT ASTTM	EXSTDTCM EXSTDTC EXSTTM

Variable Metadata | recode | **rename** | condderivation



# DETAILS ABOUT DERIVATION TYPES — “condderivation”

- 6. Derivation of a variable based on if-then conditions of the dependent variable(s) (pseudo-code = “condderivation”).

Variables Example	Rendered SAS <sup>®</sup> code
Analysis Relative Day (ADY). Derivations are different depends whether ADT is before TRTSDT or on/after TRTSDT	<pre>if . &lt; ADT &lt; TRTSDT then ADY = ADT - TRTSDT; else if ADT &gt;= TRTSDT &gt;. then ADY = ADT - TRTSDT + 1;</pre>

# DETAILS ABOUT DERIVATION TYPES — “condderivation” (cont.)

## Pseudo-code “condderivation” in the Variable Metadata tab

Dataset Name	Variable Name	Variable Label	Type	Source	Derivation	Assigned	Pseudo Code
ADEX	ADY	Analysis Relative Day	integer		[ADT-TRTSDD+1 if ADT is on or after the treatment start date, ADT-TRTSDD if ADT precedes the treatment start date]		condderivation

... Variable Metadata recode rename condderivation lookup std\_function Where( ... + : <

## Corresponding “condderivation” tab

Dataset Name	Variable Name	Derivation Condition	Output Variable Value	Source Variable or Formula
ADEX	ADY	. < ADT < TRTSDD		ADT - TRTSDD
ADEX	ADY	ADT >= TRTSDD > .		ADT - TRTSDD + 1

Dataset Metadata Variable Metadata recode rename condderivation lookup std\_function W



# SDTM AND PSEUDO-CODE

- Pseudo-code

SDTM specific pseudo-code for mapping of collected data



Pseudo-code common with ADaM for derived variables

- Source to Target

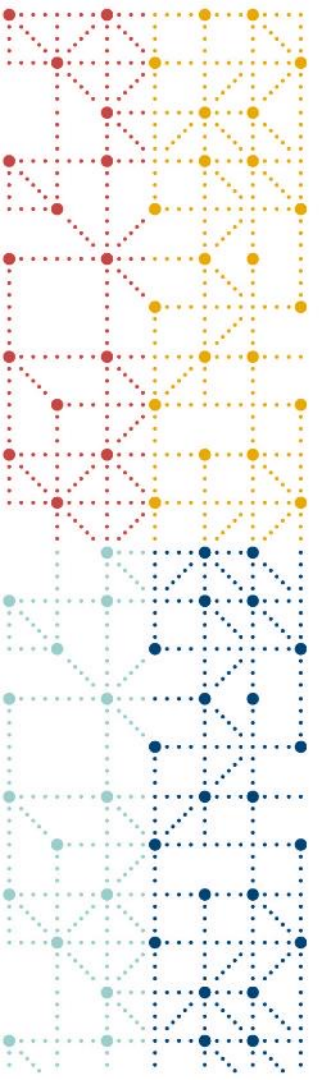
- ADaM: Pseudo-code is defined for each **target** ADaM variable
- SDTM: Pseudo-code is defined for each **source** CRF field

- Structure change

WEIGHT_VSORRES	HEIGHT_VSORRES	TEMP_VSORRES
62.0	160	36



VSTESTCD	VSORRES
WEIGHT	62.0
HEIGHT	160
TEMP	36

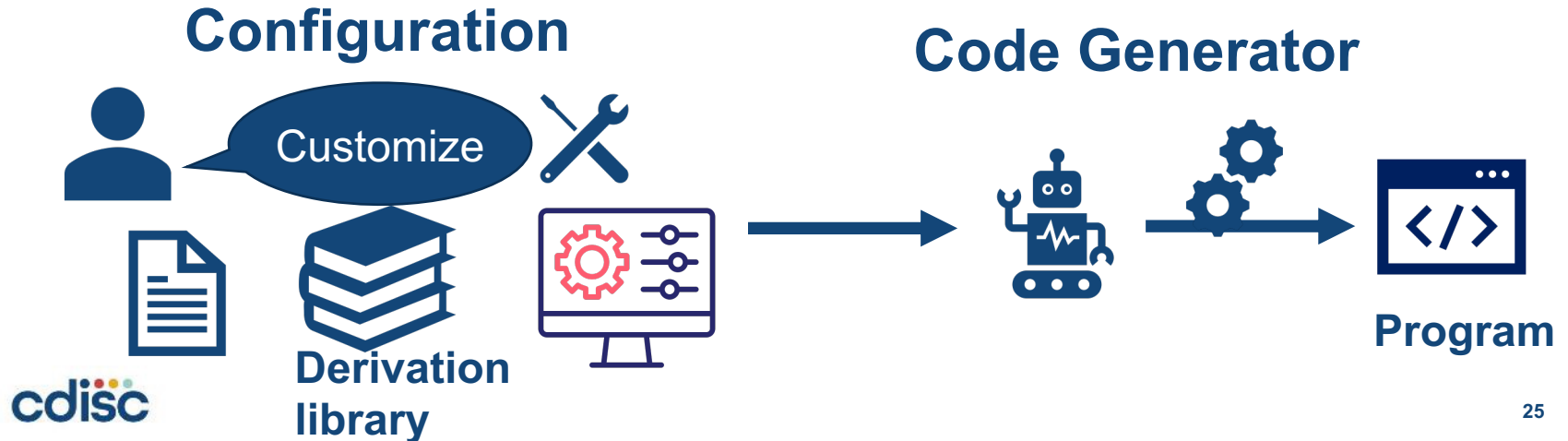


# Automation of programs

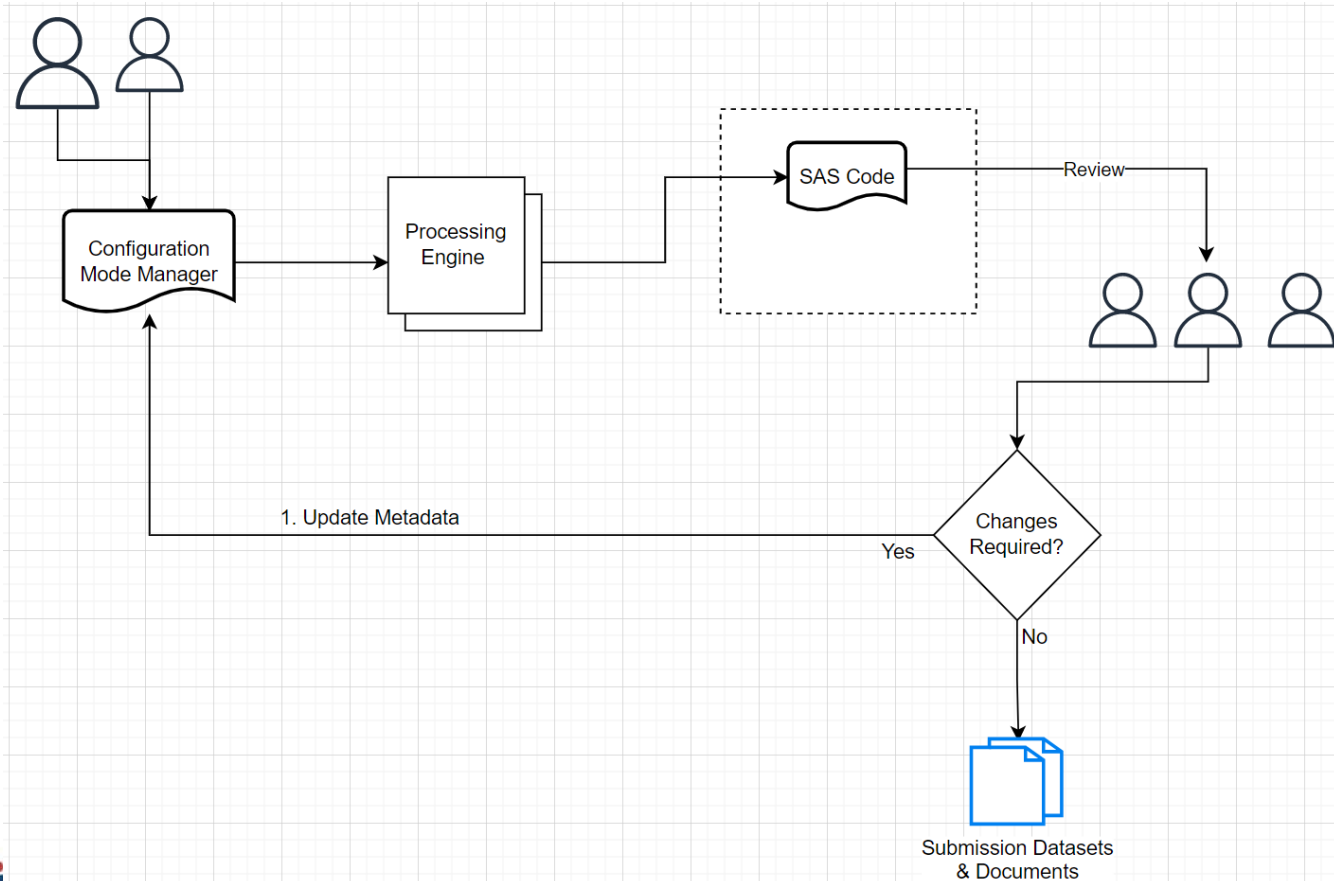


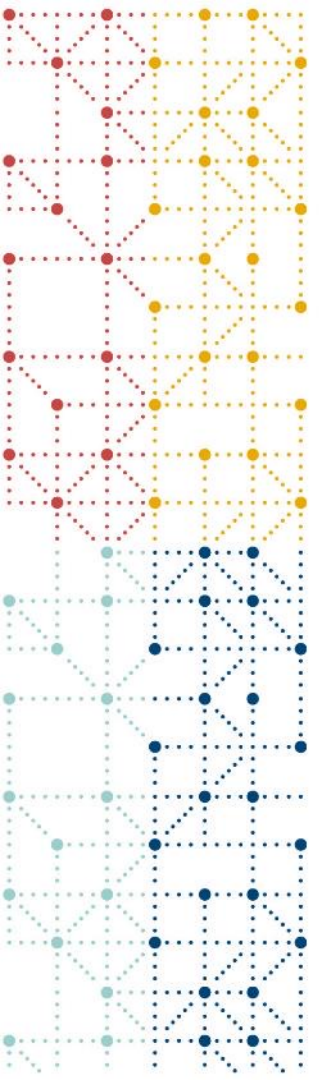
# AUTOMATION OF PROGRAMS

- To achieve automation, we are taking a **configuration-based approach**
- To derive a variable, we created a generic configuration model, that captures all necessary information. The **code generator** is utilizing this configuration along with other parameters (program formats etc.) to generate output code.
- A **code generator** is a software tool that automatically generates code based on input specifications or parameters.



# AUTOMATION OF PROGRAMS





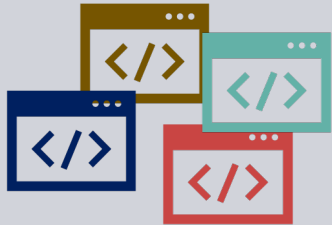
## Benefits and challenges of pseudo-code

# BENEFITS AND CHALLENGES OF PSEUDO-CODE

- Even with the project still in development, these are the proposed benefits of the pseudo-codes:

## Agnostic

- Pseudo-code is language agnostic

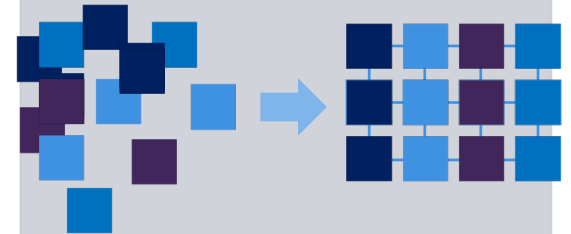


## Consistent

- Pseudo-code and additional user input in corresponding columns and tabs serve as the source for code generator to generate the program as well as the derivation specifications for the programmers

## Structured

- Pseudo-code provides a structure to the definition of derivation algorithms



# BENEFITS AND CHALLENGES OF PSEUDO-CODE

- We tested out code generator on several pilot studies, these were the challenges we faced:
  - Pseudo-code and code generator functions are new concepts and need user training and support.
  - Could be challenging at times to define correct/complicated pseudo-codes at the study level.
  - This process adds a non-trivial layer of metadata which requires modern tools to be able to preview future SAS codes and enter the pseudo-codes.



## ACKNOWLEDGEMENT

This presentation is based on the paper written by Lyubov Sushchenko (Lead author) and Rie Ichihashi (Co-author):

[Automating SDTM and ADaM Creation in Clinical Trials](#)

# ACKNOWLEDGEMENT

The authors of the paper would like to express their gratitude to the Sanofi ART team, for their support and ideas, especially Faisal Mohammed. ADaM MAP team for their dedication and effort. B&P programmers for providing feedback and volunteering in testing out new tools. And Beilei Xu for inspiring to write this paper.

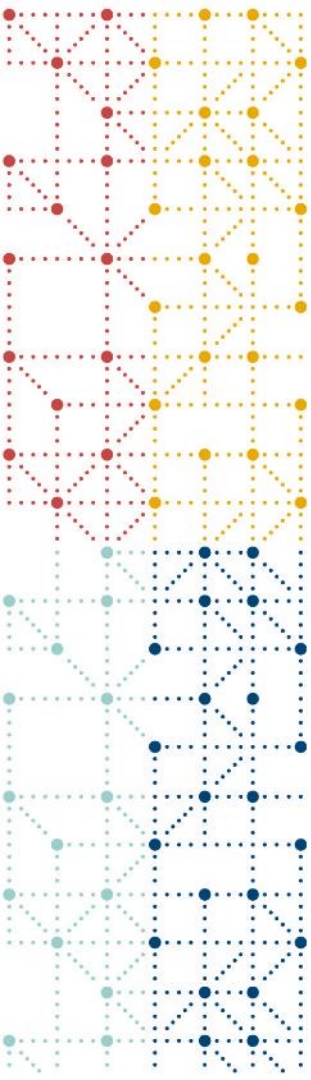
## CONTACT INFORMATION

Your comments and questions are valued and encouraged.

### •Contact the authors at:

Author Name: Lyubov Sushchenko  
Company: Sanofi US  
Email: [Lyubov.Sushchenko@sanofi.com](mailto:Lyubov.Sushchenko@sanofi.com)  
Web: <https://www.sanofi.com>

Author Name: Rie Ichihashi  
Company: Sanofi K.K.  
Email: [Rie.Ichihashi@sanofi.com](mailto:Rie.Ichihashi@sanofi.com)  
Web: <https://www.sanofi.com>



**Thank You!**

**cdisc**