



ADaM at Bayer – metadata and automation



Elena Glathe

06. Junel 2019 // Copenhagen SDE





Agenda

- // Intro
- // ADaM Metadata
- // ADaM Mapping
- // QC
- // Docu
- // Summary



Intro: MADAM I'M ADAM

Semi-automatic Adam Mapping solution released in Version 1 in Aug. 2011

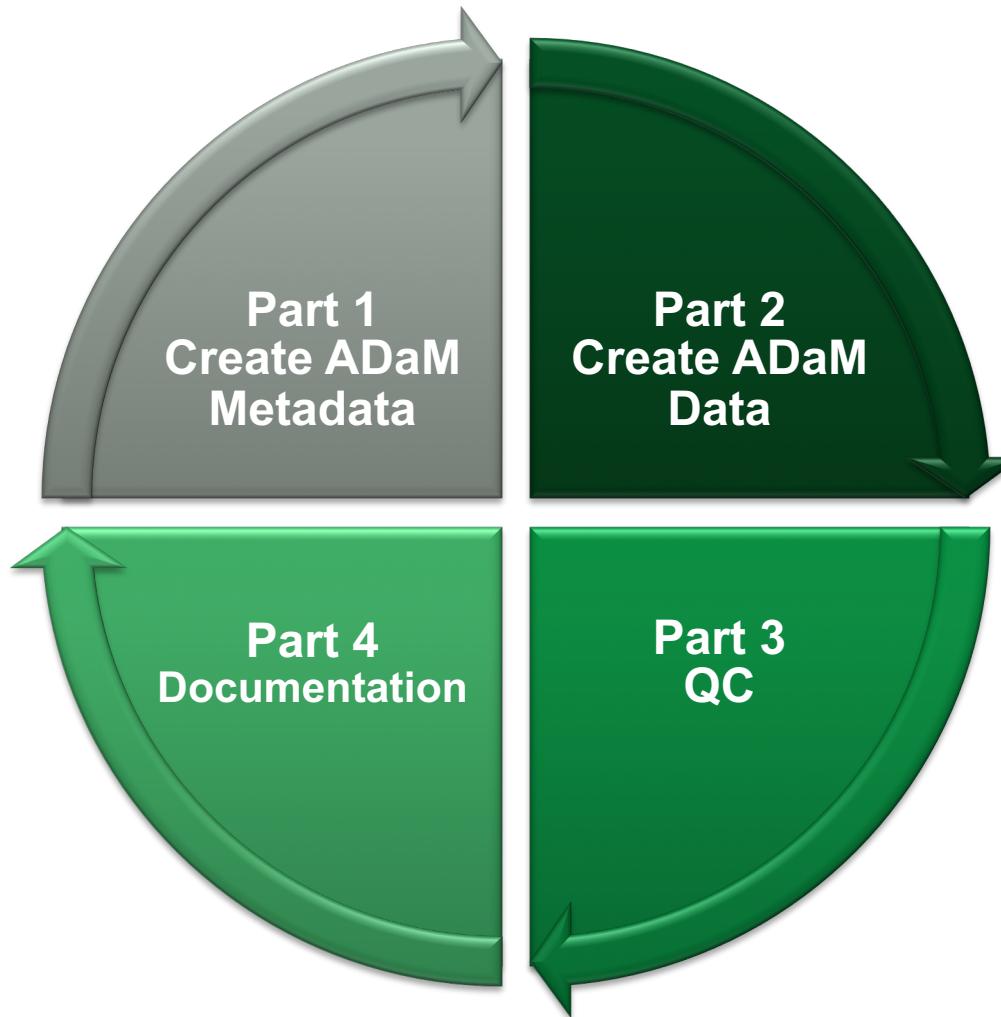
Metadata and SAS Macro framework for mapping from SDTM+ to ADaM

- Automatically collect all variables needed from SDTM+
- Automatically derive variable if algorithm is fix
- Automatically add decode variables
- Automatically apply formats, labels, type per Metadata, and more

Prerequisites

- Follow well defined standard
- SDTM+ data and metadata
- ADaM metadata concept is strictly used

AdaMap - Components



...big red button?



AdaMap is **NOT** to completely map a study to
ADaM automatically

BUT automate common derivations especially on
global level

AdaMap is **NOT** the reason, why ADaM Metadata
have to be created
BUT AdaMap uses Metadata

Part 1: ADaM Metadata

ADaM Metadata

ADaM Metadata

memname	label	sasname	type	outform	codist	methtyp	deri_wag	method	vartype	var_link	deri_txt
		Dataset Label									
ADAE	Dataset Name	ADSNAME	C	8	X_ADSNM	derived	simple	=%_adamap_create_adsname			Name of the analysis data set. Must be available in X_ADSNM-codelist
ADAE	Sponsor Defined Analysis Sequence Number	SEQ	N	8		derived	complex	=%_adamap_create_seq(by=USUBJI D, order=AECATN AEST_Y)			Calculated sequence number starting with 1 and increasing by 1 within a sort order of ADAE.AECATN, ADAE.AEST_Y, ADAE.AEST_M, ADAE.AEST_D, ADAE.AEST_TC, ADAE.AEEN_Y, ADAE.AEEN_M, ADAE.AEEN_D, ADAE.AEEN_TC, ADAE.AETERM,
ADAE	SDTM+	AETERM	SDTM+	S							
ADAE	SDTM+	AELLTCD	SDTM+	S							AELLT
ADAE	SDTM+	AELOC	SDTM+	S							AELOCN
ADAE	SDTM+	AELOCN	SDTM+	S							AELOC
ADAE	Body System or Organ Class Second Path	AEBDS#CD	C	8							AEBDSYS #
ADAE	All Reasons for Seriousness	AESREAS	C	2							Derived; Body system or organ class used by the sponsor from the codelist (e.g. MedDRA). Secondary path(s). Add as many as needed. For integrated analysis, Derived; Attach SOC_CD format to AFMLLT in Med
ADAE	SDTM+	AEREL	SDTM+	S							AERELN
ADAE	SDTM+	AERELN	SDTM+	S							AEREL
ADAE	SDTM+	AEOUT	SDTM+	SDTM+	SDTM+	.	AE.AEOUT	decode	AEOUTN		
ADAE	SDTM+	AEOUTN	SDTM+	SDTM+	SDTM+	.	AE.AEOUTN	code	AEOUT		
ADAE	Analysis Start Date	ASTDT	N	DATE9.		derived	simple	=%_adamap_create_adt(rule=NONE, createFlagVar=Y)			Derived using SAS function MDY(AE.AEST_M, AE.AEST_D, AE.AEST_Y). If date parts are missing, imputation rules as defined in SAP are to be applied.
ADAE	Analysis Start Date Imputation Flag	ASTDTF	C	1	DATEFL	derived	simple	=%_adamap_ref(refvar=ASTDT)			If ASTDT is imputed, the following rules apply: If year is imputed set to Y, otherwise if month is imputed set to M, otherwise if day is imputed set to nothing was imputed keep empty.
ADAE	Duration of Adverse Event	AEDUR	N	8		derived	simple	=AENDT - ASTDT + 1			Number of days from ASTDT to AENDT, rule is AENDT - ASTDT + 1. Different rules or units have to be defined in SAP.
ADAE	Treatment Emergent Analysis Flag	TRTEMFL	C	14		decode	.				
ADAE	Treatment Emergent										

Typical Items:

- Variable Label
- Variable Name
- Type
- Length
- Codelist (Controlled Terms)
- Core

ADaM Metadata - Automation

Automation:

→ Method type

SDTM+, derived, decode

→ Method

Algorithm as SAS code (if algorithm is fix for all studies within a scope)

Empty in case it's a custom derivation

→ Derivation way

Define whether its a inline variable mapping or whether the mapping requires data step and merging and so on

→ Derivation text

Derivation description used in ADaM Mapping Spec and also for define.xml

sasname	meth type	deri_way	method
USUBJID	SDTM+		AE.USUBJID
AEBOCD	DERIVED	Simple	=PUT(AEMLLT, SOC_CD.)
AESREAS	DERIVED	Simple	=%_adamap_create_adae_aesreas
AENTM	DERIVED	Complex	=%_adamap_create_atm(rule=NONE,imputeMissing=N,createFlagVar=Y)
ANL##FN	DERIVED	Manual	
TRTEMFL	DECODE		

ADAE	Analysis Start Date Imputation Flag	ASTDTF	C	1	DATEFL	derived	simple	=%_adamap_ref(refvar=ASTDT)	decode	AEOUTN	
ADAE	Duration of Adverse Event	AEDUR	N	8		derived	simple	=AENDT - ASTDT + 1		AEOUT	Derived using SAS function MDY(AE.AEST_M, AE.AEST_D, AE.AEST_Y). If date parts are missing, imputation rules as defined in SAP are to be applied.
ADAE	Treatment Emergent Analysis Flag	TRTEMFL	C	14		decode	.		decode	TRTEMFL	If ASTDT is imputed, the following rules apply: If year is imputed set to Y, otherwise if month is imputed set to M, otherwise if day is imputed set to nothing was imputed keep empty.

ADaM Metadata - Decodes

ADaM Metadata – Re-use Attributes

memname	sasname	Label	Method	type	outform	codlst	methtyp	vartype	var_link
ADAE	STUDYID	SDTM+	AE.STUDYID	SDTM+	SDTM+	SDTM+	SDTM+		
ADAE	USUBJID	SDTM+	AE.USUBJID	SDTM+	SDTM+	SDTM+	SDTM+		
ADAE	AEDECOD	SDTM+	AE.AEDECOD	SDTM+	SDTM+	SDTM+	SDTM+	decode	AEPTCD
ADAE	AELLT	SDTM+	AE.AELLT	SDTM+	SDTM+	SDTM+	SDTM+	decode	AELLTCD

Be lazy!
Have trust!
Take what you get!

→ Label, Type, Length (Outform) and codelist take over from SDTM+

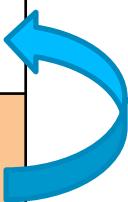
ADaM Metadata - # Variables

MEM NAME	label	method	sasname	varType	var_link	deri_txt
	Dataset Label	Input Definition	Variable Name	Code-Decod	Variable Pair	Derivation Description
ADAE	Treatment Emergent Flag: Treatment #		AETRTT#	decode	AETRTT#N	
ADAE	Treatment Emergent Flag: Treatment # (N)		AETRTT#N	code	AETRTT#	Set to "Y" if adverse event is treatment emergent with regard to treatment #.
ADAE	SMQ ## Name		SMQ##NAME			The standardized MedDRA query's name. Blank for terms that are not in the SMQ. The number of the SMQ variables must be constant for a query within a project.
ADAE	SMQ ## Code		SMQ##CD			The standardized MedDRA query's number code. Blank for terms that are not in the SMQ. The number of the SMQ variables must be constant within a query and within a project.
ADBM	Analysis Criterion #		CRIT#	decode	CRIT#N	
ADBM	Analysis Criterion # (N)		CRIT#N	code	CRIT#	A code for identifying a criterion of interest within a project as defined in SAP.
ADBM	Analysis Record Flag ##		ANALREC#	decode	ANALREC#N	
ADBM	Analysis variables with a number in the name → collection of several variables of the same type → Use # wildcard for easier maintenance TRT##A, CRIT#, CRIT#N					
ADSL	Age Group ##		AGEGR##	decode	AGEGR##N	

ADaM Metadata – Several Variables on a Stroke



Memname	Sasname	Label	Method
ADAE	ASTDT	Analysis Start Date	=%_adamat_create_adt(rule=NONE, createFlagVar=Y)
ADAE	ASTDTF	Analysis Start Date Imputation Flag	=%_adamat_ref(refvar=ASTDT)



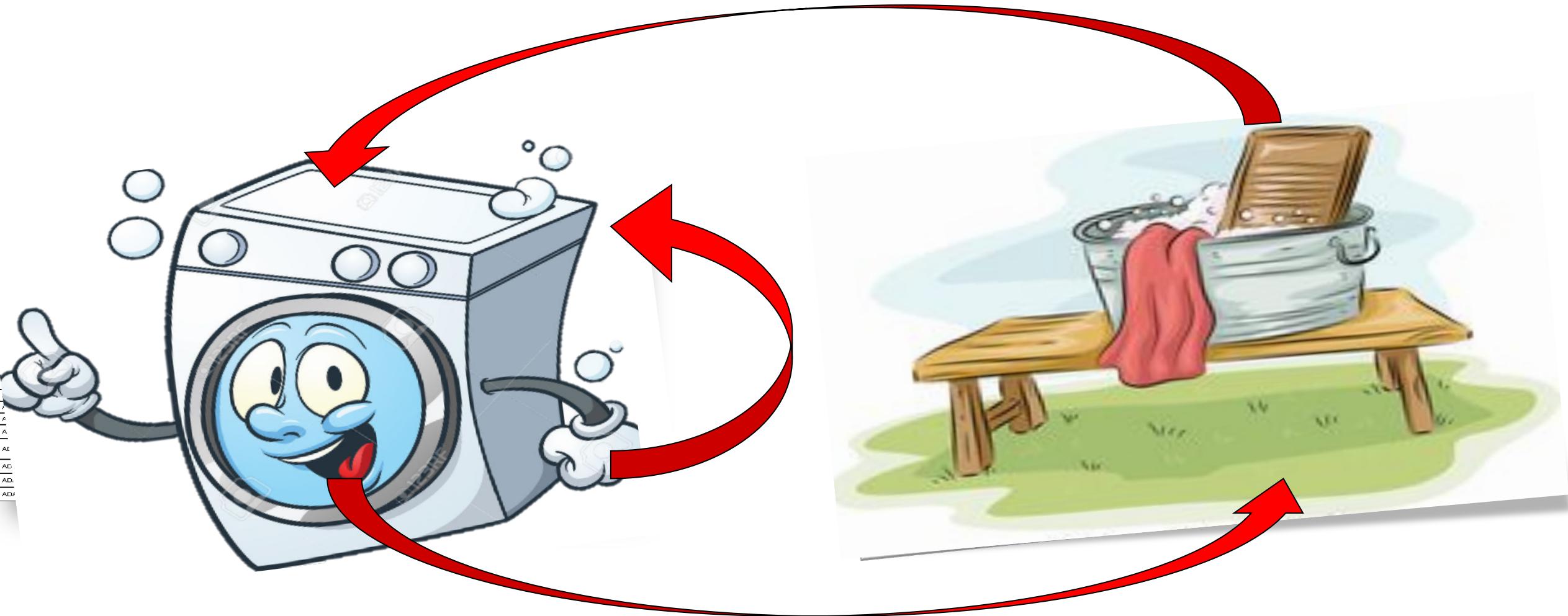
Controlled ADaM Metadata Management

- Update of METHOD requires update of derivation description; and vice versa
- Add ADaM variables automatically **if** SDTM+ variable(s) available
- Disable ADaM variables automatically **if** input SDTM+ variable(s) **not** available



Part 2: ADaM Mapping

Mapping Cycles



Mapping Process: Very Verbose Log

NOTE: CREATE_INITIAL_ADS_DAT - Will create ADS.ADSL from SDTMP.DM!

INFO: CREATE_INITIAL_ADS_DAT - Add SDTMP variables from DM to ADS.ADSL

INFO: CREATE_INITIAL_ADS_DAT - Will rename DM --> PPROT to PPROTFL

INFO: CREATE_INITIAL_ADS_DAT - Will rename DM --> PPROTN to PPROTFN

NOTE: CREATE_INITIAL_ADS_DAT - Will create ADS.ADAE from SDTMP.AE (where = (AENYN=1))!

NOTE: CREATE_INITIAL_ADS_DAT - Will create ADS.ADEGF from SDTMP.EG (where = (EGCAT="FINDING"))!

INFO: CREATE_INITIAL_ADS_DAT - Add SDTMP variables from EG to ADS.ADEGF

INFO: PERFORM_DERIVATIONS_DAT - Will Derive UASR method = COMPRESS(CATX('/', USUBJID, AGE, SEX, PUT(RACEALLN, X_RACES.)))

INFO: PERFORM_DERIVATIONS_DAT - Will Derive BRTHDT method = %_ADAMAP_CREATE_DTT

INFO: PERFORM_DERIVATIONS_DAT - Will Derive BRTHDTL method = %_ADAMAP_CREATE_DTL

INFO: PERFORM_DERIVATIONS_DAT - Will Derive SMOKHXN method = %_ADAMAP_CREATE_ADSL_VAR

INFO: ADDADSMETADATA2DAT - ATTRIB STUDYID LENGTH = \$10 FORMAT = \$10. LABEL = "Study Identifier"

INFO: ADDADSMETADATA2DAT - ATTRIB USUBJID LENGTH = \$20 FORMAT = \$20. LABEL = "Unique Subject Identifier"

INFO: ADDADSMETADATA2DAT - ATTRIB SUBJIDN FORMAT = 9. LABEL = "Subject Identifier for Study(N)"

NOTE: PERFORM_DERIVATIONS_DAT - Postponed derivation of 1 variable(s) in ADS.ADDA: TRTPN

INFO: PERFORM_DERIVATIONS_DAT - Will Derive PARAMTYP method = ""

INFO: PERFORM_DERIVATIONS_DAT - Will Derive DTTYPE method = ""

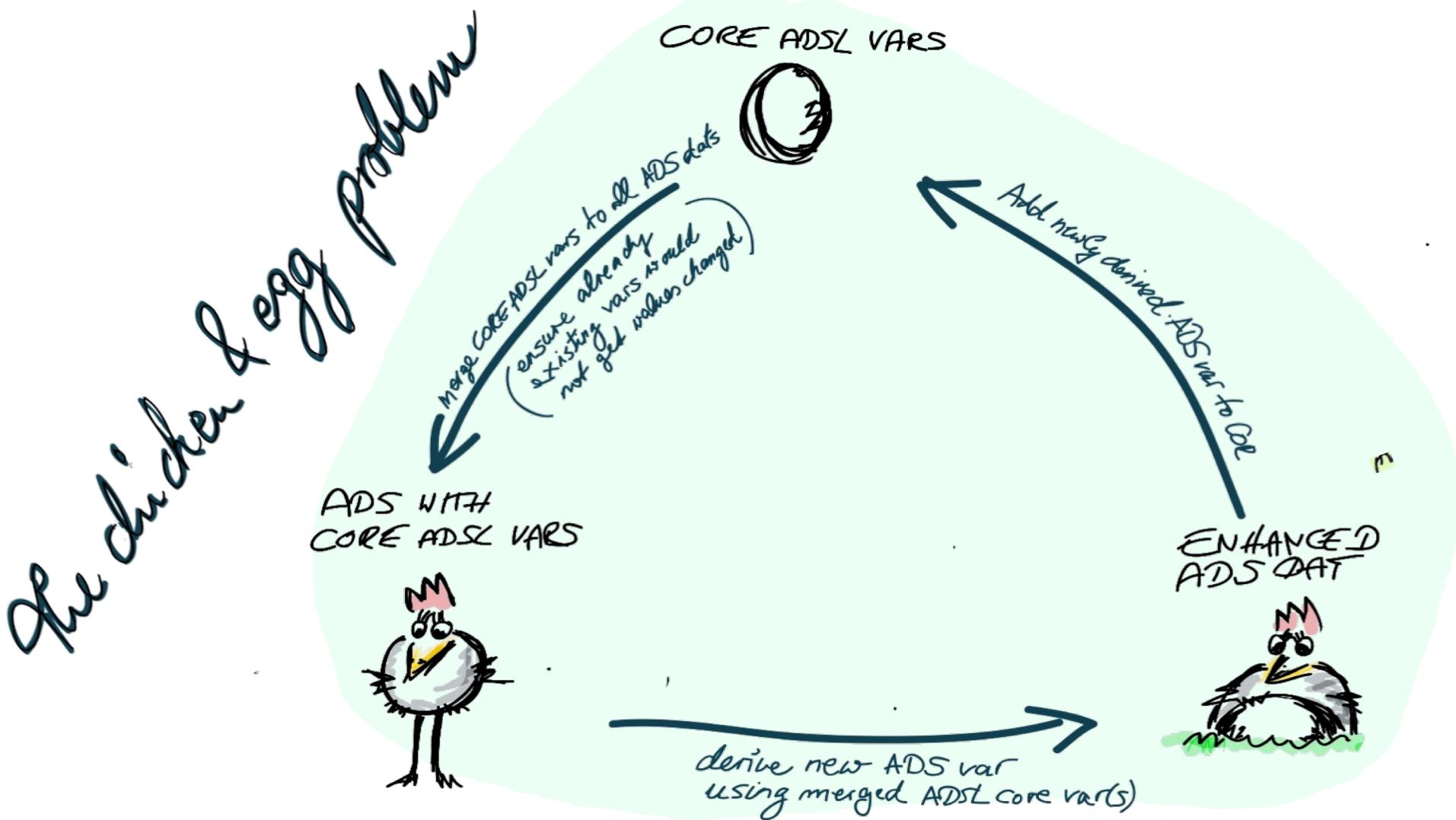
INFO: ADDDECODES2DAT - Code variable APERIOD is not available. Will not add decode APERIODC!

INFO: ADDDECODES2DAT - Add variable ANRLO with decode of ANRLON

INFO: ADDDECODES2DAT - Add variable ANRHI with decode of ANRHIN

INFO: MERGE_ADSL - Will merge ADSL core variables from ADS.ADSL to ADS.ADAE!

Complexity of Mapping Process



Part 3: QC

```
ERROR: CHECK_ADS_META - Sequence number 4
NOTE: CHECK_ADS_META - Will check structure
INFO: Character variables have defaulted
(Line):(Column). Truncation may result. 4
ERROR: CHECK_ADS_META - DERI_TXT is not a
variables!(ADCE.STATUS METHTYP=derived DE
WARNING: CHECK_ADS_META - METHOD is not a
4!(ADCE.ANL01FN deri_way=4 )
ERROR: CHECK_ADS_META - OUTFORM contains
ERROR: CHECK_ADS_META - DERI_TXT is not a
!(ADDS.EXTENNY METHTYP=decode
DERI_TXT=Decode of EXTENNYN )
ERROR: CHECK_ADS_META - OUTFORM contains
WARNING: CHECK_ADS_META - OUTFORM should
OUTFORM=
WARNING: CHECK_ADS_META - TYPE should be
ERROR: CHECK_ADS_META - VAR_LINK cannot b
```

ADaM Metadata

%check_adam_meta

```
ERROR: CHECK_ADS_DAT - Codelist UNIT is incom
XLSTRESU=/m1
OAD values in 2559 not derived observations have c
he following variables: AMENDNO DSCAT DSCATN DSCNSI
WARNING: CHECK_ADS_DAT - No variable for metadata
*****
6050 issues in data set ADVS
*****
AD0018 - METADATA - ADaM dataset variable label
AD0042 - FORMAT - * TM does not have the ADaM re
AD0059 - FORMAT - * TM is not a numeric variable
AD0116 - CONSISTENCY - ATPTREF is populated and
AD0124 - CONSISTENCY - Inconsistent value for PA
AD1001 - PRESENCE - Required variable is not pre
: AD1002 - PRESENCE - Expected variable is not p
```

ADaM Data

%check_adam
%pinnacle_validator

Check AdAM – Great Job

INFO: CHECK_ADS - Will check ADAE!

INFO: CHECK_ADS_DAT - Will merge ADSL core variable metadata to ADSMETA.ADAE

INFO: CHECK_ADS_DAT - Will check metadata compliance

INFO: CHECK_ADS_DAT - No metadata compliance error found. Great Job!

INFO: CHECK_ADS_DAT - Will perform decode / codelist checks!

INFO: CHECK_ADS_DAT - No wrong decode values found. Great Job!

INFO: CHECK_ADS_DAT - Will need to perform extra checks on following variables: ADURU

INFO: CHECK_ADS_DAT - Will check variable ADURU against codelist UNIT!

INFO: CHECK_ADS_DAT - Will check number of non-derived observations between ADS.ADAE and SDTMRP.AE(WHERE=(AETERM

INFO: CHECK_ADS_DAT - Found 1358 non-derived observations in both data sets. Great Job!

INFO: CHECK_ADS_DAT - Will check if SDTMRP variables are unchanged

INFO: CHECK_ADS_DAT - Found no changed SDTMRP values. Great Job!

INFO: CHECK_ADS_DAT - Will check keys variables as defined in ADS.AD

INFO: CHECK_ADS_DAT - ADAE is defined in ADS.AD. Great Job!

INFO: CHECK_ADS_DAT - Will check if keys variables is defined in ADS.AD

INFO: CHECK_ADS_DAT - Keys variables are defined in ADS.AD as STUDYID,USUBJID,SEQ. Great Job!

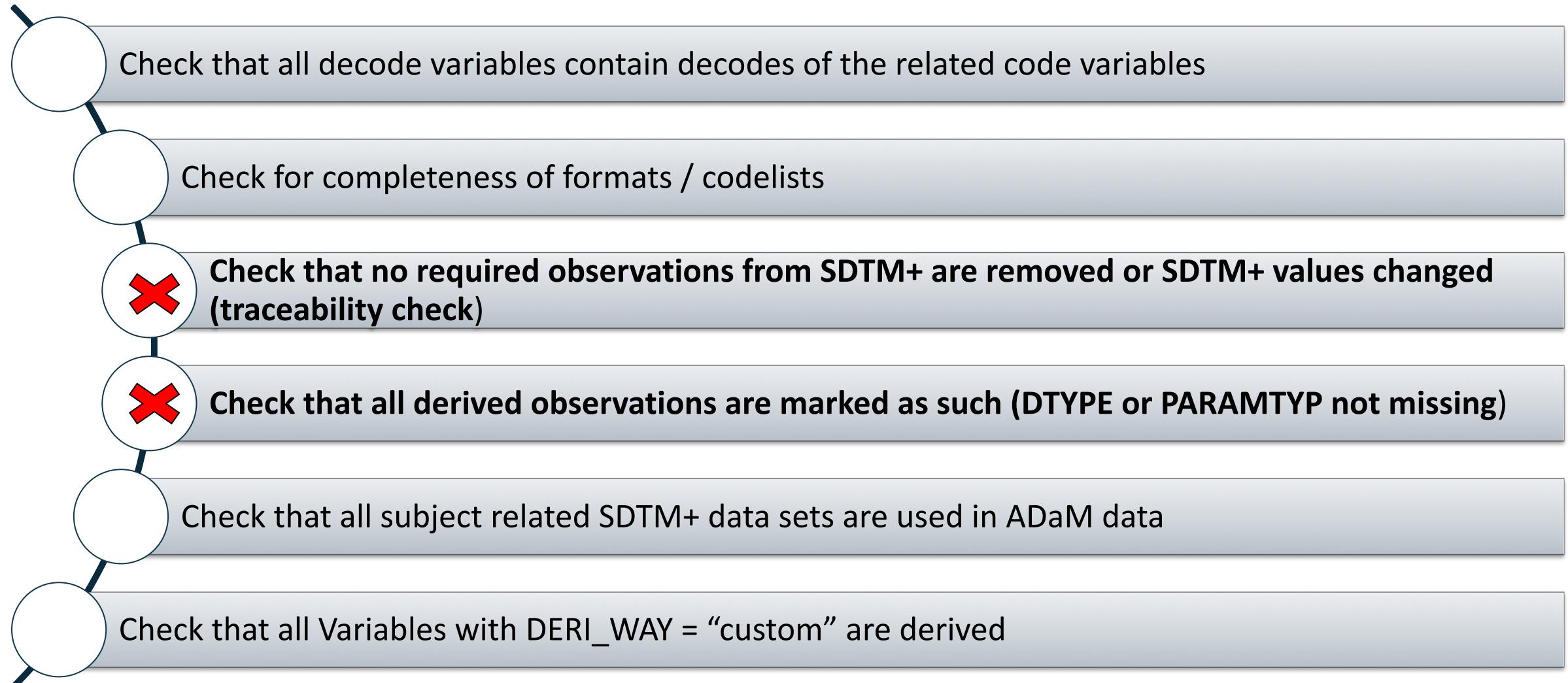
INFO: CHECK_ADS_DAT - Will check if keys variables exist in ADS.ADAE

INFO: CHECK_ADS_DAT - All keys variables are exist in ADS.ADAE. Great Job!

INFO: CHECK_ADS_DAT - Will check if any duplicate (by STUDYID,USUBJID,SEQ) exist in ADS.ADAE

INFO: CHECK_ADS_DAT - No duplicated found in ADS.ADAE. Great Job!

QC – Data Traceability Checks



Part 4: Documentation

Documentation

%Meta_docu	ADaM specification for SAP
%Comp_metadata	see changes between 2 Metadata versions
%findderivedrecs	New records derivation documentation including derivation rule in deri_txt → used for define.xml
Define.xml	Serves as input for define.xml creation for submissions

Derivation comment in SAS code

.... added based upon the metadata

```
*< AVISITN;
/** LABEL = Analysis Visit (N) - Format: _AVISIT - Outform: 8.3
** A numeric representation of AVISIT. AVISITN may contain the
** visit number as observed (i.e., from SDTM VISITNUM), derived
** visit numbers, time window numbers, conceptual description
** numbers (such as Average, Endpoint, etc.) or a combination
** of any of these */
Data ADS.ADLB;
  SET ...;
```

Automatically document derived Records

Manually filled

MEMNAME	paramcd Parameter Code	param Parameter	pcd_codl Codelist of PARAMCD	pn_codl Codelist of PARAMN	dtype Derivati Type	paramty Paramete Type	deri_txt Derivation Description
ADTTE	COMP07	Composite: Cardiov...	\$X_TTE.	X_TTEN.		DERIVED	Time (in days) to Composite of Cardiovascula
ADTTE	COMP06	Composite: Death o...	\$X_TTE.	X_TTEN.		DERIVED	Time (in days) to Composite of Death of any
ADTTE	COMP08	Composite: Heart Fa...	\$X_TTE.	X_TTEN.		DERIVED	Time (in days) to Composite of Heart Failure
ADTTE	COMP09	Composite: Heart Fa...	\$X_TTE.	X_TTEN.		DERIVED	Time (in days) to Composite of Heart Failure
ADTTE	MACE	Major Adverse Cardi...	\$X_TTE.	X_TTEN.		DERIVED	Time (in days) to Major Adverse Cardiovascul
ADTTE	HOSPCV	Time to Cardiovascu...	\$X_TTE.	X_TTEN.		DERIVED	Time (in days) to Cardiovascular Hospitaliza
ADTTE	DEATH	Time to Death (days)	\$X_TTE.	X_TTEN.		DERIVED	Time (in days) to Death
ADTTE	WCHFWORS	Time to Emergency P...	\$X_TTE.	X_TTEN.		DERIVED	Time (in days) to Emergency Presentation due
ADTTE	HOSPHF	Time to Heart Failu...	\$X_TTE.	X_TTEN.		DERIVED	Time (in days) to Heart Failure Hospitalizat
ADVS	BMI	Body Mass Index (kg...)	\$X_VSPAR.	X_VSPARN.	LVPD		The records are created from the last non-mi
ADVS	DIABP	Diastolic Blood Pre...	\$X_VSPAR.	X_VSPARN.	AVERAGE		Average of three non-missing measurements. I
ADVS	DIABP	Diastolic Blood Pre...	\$X_VSPAR.	X_VSPARN.	LVPD		The records are created from the last non-mi
ADVS	HR	Heart Rate (BEATS/MIN)	\$X_VSPAR.	X_VSPARN.	AVERAGE		Average of three non-missing measurements. I
ADVS	HR	Heart Rate (BEATS/MIN)	\$X_VSPAR.	X_VSPARN.	LVPD		The records are created from the last non-mi
ADVS	HEIGHT	Height (cm)	\$X_VSPAR.	X_VSPARN.	LVPD		The records are created from the last non-mi
ADVS	MAP	Mean Arterial Press...	\$X_VSPAR.	X_VSPARN.	AVERAGE		Average of three non-missing measurements. I
ADVS	MAP	Mean Arterial Press...	\$X_VSPAR.	X_VSPARN.	LVPD		The records are created from the last non-mi
ADVS	PULSEPR	Pulse Pressure (mmHg)	\$X_VSPAR.	X_VSPARN.	AVERAGE		Average of three non-missing measurements. I
ADVS	PULSEPR	Pulse Pressure (mmHg)	\$X_VSPAR.	X_VSPARN.	LVPD		The records are created from the last non-mi
ADVS	SYSBP	Systolic Blood Pres...	\$X_VSPAR.	X_VSPARN.	AVERAGE		Average of three non-missing measurements. I

Summary

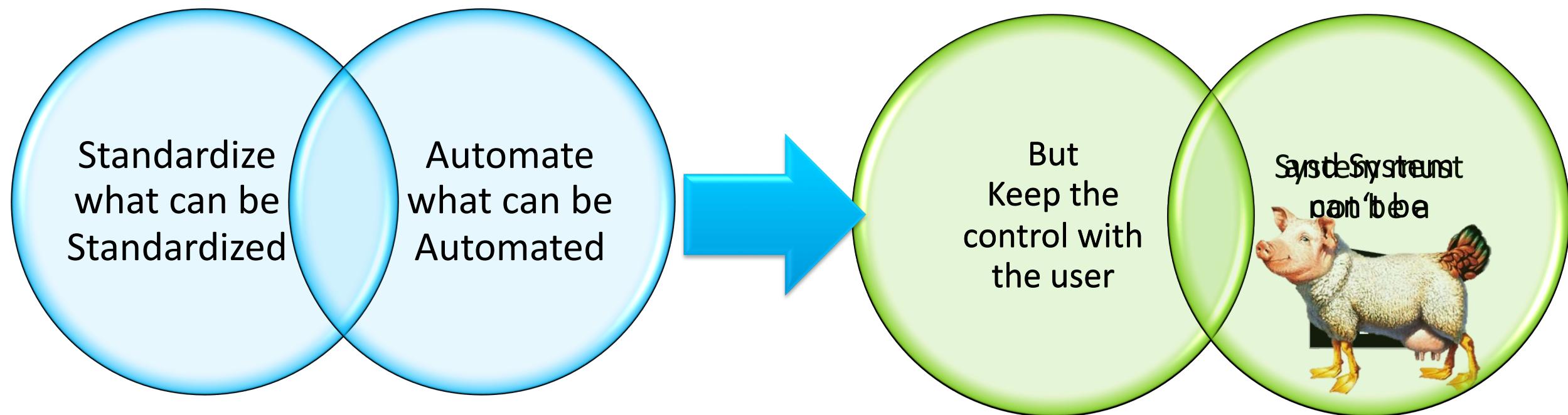
Summary

- // Embedded environment. Mapping macros depend on underlying AdaMap macro environment and on underlying ADaM metadata concept
 - // Ensures **traceability** from SDTM+ to ADaM. Including QC
- ➔ **Default mapping works right away for every new study**

Features:

- // Inherit SDTM+ attributes
- // Support of # variables
- // Create multiple derived variables at once with clear documentation in metadata
- // Automatic handling of decode creation
- // Automatically adds labels and codelist to ads data sets ➔ apply how often you wish
- // Merge of Core ADSL in multiple stages possible, with consistency check

Conclusion – Metadata driven automation



Future...

AdaMap Version 2 – what comes next?

- New ADaM IG**
- Go over pure SDTM**



How is SDTM+ different from (pure) SDTM?

- Supplemental data are not separated but stay in parent domain
- Items with pure operational purpose in SDTM+
e.g. trigger questions or additional variables for coding (eg Clarified Verbatim)
- Code and Decodes were present

Use SDTM instead of SDTM+

- DON'T use any „plus“ variable for mapping to ADaM but only „real“ SDTM variables → comply with traceability. E.g. use the decode variables
- Re-introduce numerical/coded variables out of SDTM variables for calculations, derivations and analysis
- Re-merge supplementals to the parent domain
- Pre-requisite: Bi-directional codelists and perfect metadata



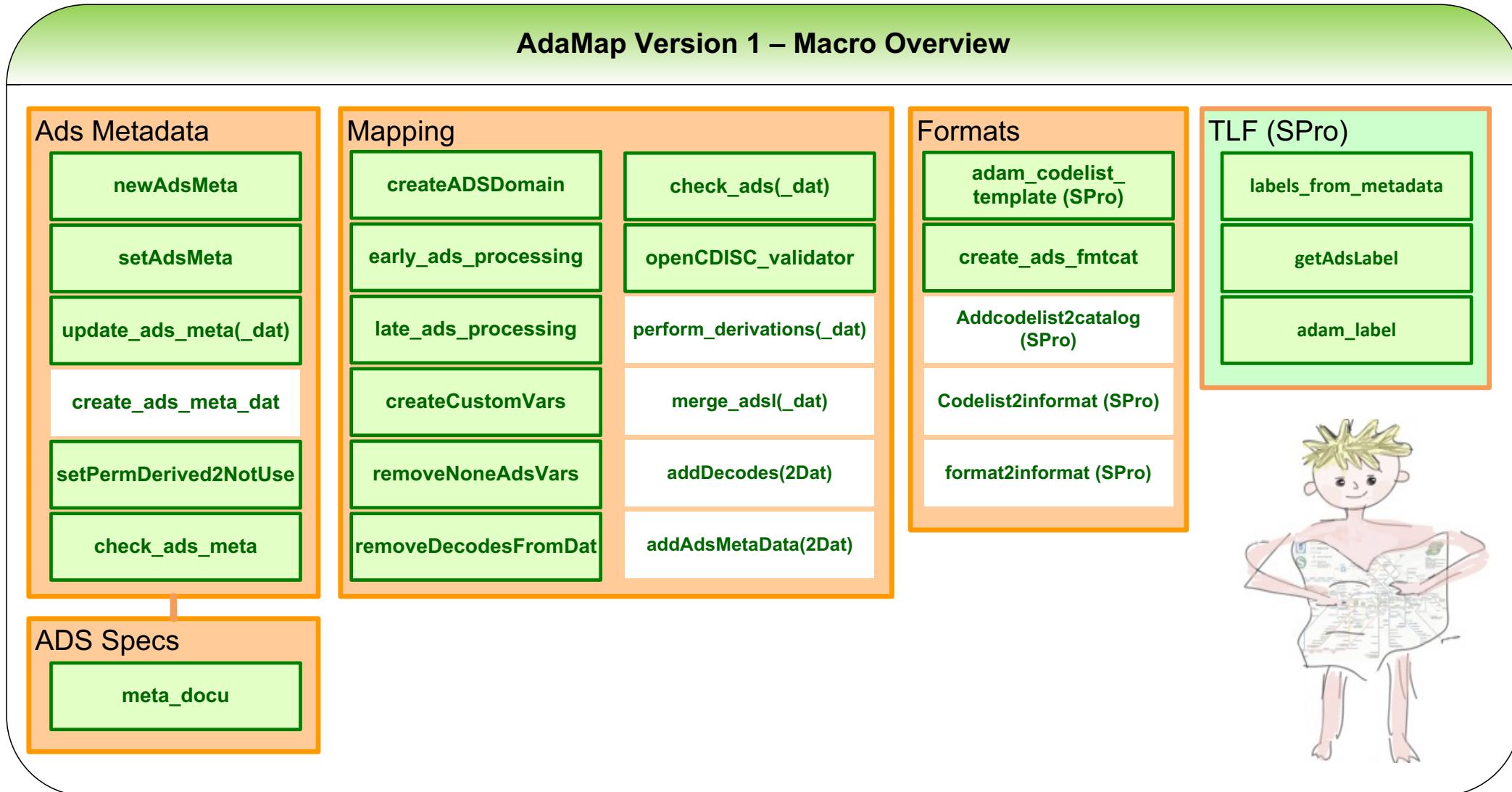
Thank you!



contact: elena.glathe@bayer.com



AdaMap – Macro Environment



What is SDTM+ at Bayer?

- All SDTM variables and controlled terminology
- Numeric representation SDTM characters, where applicable
- Code/decode pair for controlled terminology
- All Supplemental (SUPP) information together with respective primary domain (no separation)
- Leading/Indicator questions (ie. Adverse Event? Yes/No)
- Unique identifiers linked to data collection sources
(internal studies only)



The logo consists of the acronym "SDTM" in large, bold, dark blue letters. To the right of "SDTM", the word "Permissible" is stacked vertically above "Epoch". Above "SDTM", there is a vertical stack of terms: "Visits", "Elements", "Required", "Expected", and "Dataset". To the right of "Dataset", the word "Arms" is stacked vertically above "CDISC". Below "SDTM", the words "Domain" and "Rules" are stacked vertically.

What is SDTM(+) at Bayer?

- All SDTM variables and controlled terminology
- ~~Numeric representation SDTM characters, where applicable~~
- ~~Code/decode pair~~ for controlled terminology
- All Supplemental (SUPP) information together with respective primary domain (no separation)
- ~~Leading/Indicator questions (ie. Adverse Event? Yes/No)~~
- Unique identifiers linked to data collection sources
(internal studies only)

The plus from SDTM+ is being challenged!

Visits
Elements
Required
Expected
Arms
CDISC
SDTM
Permissible
Epoch
Domain
Rules
Dataset