Lifting the Language Barrier: Choosing not to Choose

Presented by John McDade, Associate Director, Operational Excellence, PHASTAR

Meet the Speaker

John McDade

Title: Associate Director, Operational Excellence

Organization: PHASTAR

I have been in the industry programming for around 16 years and been with PHASTAR for the last 8 years as part of the Operational Excellence team. Here I enjoy working on a variety of process and tool development initiatives and have started the open-source journey.

Outside of work I love playing golf and the guitar and I'm a long-suffering fan of Scottish football! Most of my time is spent chasing my 6 year old son around and love nothing more than getting away to the coast with my family.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author and do not necessarily reflect the official policy or position of CDISC.*

- I have no real or apparent conflicts of interest to report

Agenda

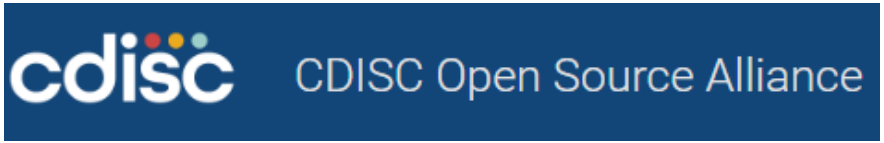
1. Open-Source Movement
2. CDISC 360 SDTM Automation
3. Mapping Library
4. Study Metadata
5. Final Automation Components
6. Summary
7. Q&A

Open-Source Movement

CDISC Open Source Alliance

- **Admiral**
  ADaM in R Asset Library.
- **CDISC Rules Engine (CORE)**
  Deliver and execute a governed set of executable Conformance Rules for each Foundational Standard
- **CDISC-ODM-XML-CRF-SDTM-Annotations**
  An XMLMAP for ODM-XML and for Define-XML along with a small set of SAS macros for each, converting the XML documents to SAS datasets following familiar data models shared by MDR and validation tools.
- **CORE - Rule Editor**
  Creating additional Conformance Rules in a common specification for CORE
- **Dataset-JSON Hackathon Projects**
  Projects developed as part of the Dataset-JSON hackathon
- **Define-XML XSL Stylesheets**
  This projects provides a Define-XML v2.0 and v2.1 XSL stylesheet
- **defineR**
  An open-source R package capable of generating the Define-XML.
- **Digital Data Flow**
  The DDF initiative aims to modernize clinical trials by enabling a digital workflow that allows for automated creation of study content and configuration of study systems to support clinical trial execution.
- **ODM XML Stylesheet**
  Apply a style sheet to ODM-XML, exactly as you apply a style sheet to define-xml to display it in a browser.
- **odmlib**
  odmlib is a Python library that simplifies creating and processing ODM and its extensions, such as Define-XML.
- **Open Study Builder**
  The OpenStudyBuilder is a new approach to working with studies that once fully implemented will drive end-to-end consistency and more efficient processes.
- **openCST**
  The open-source release of the SAS Clinical Standards Toolkit (openCST) is a framework that allows for the registration of CDISC standards metadata to enable automation, such as working in SAS with XML based standards (Define-XML, ODM, ...) and validation of clinical data, such as SDTM.
- **R4DSXML**
  R4DSXML is R package for import both CDISC Dataset-XML and Define-XML as R data frame.
- **Smart Submission Dataset Viewer**
  Dataset viewer allowing to inspect CDISC SDTM, SEND and ADaM submission files.
- **TFL Designer**
  An open-source TFL designer to create study-specific analysis output display and in parallel generate machine-readable metadata.
- **tfrmt**
  The {tfrmt} R package is a table formatting framework that provides the means to flexibly design and build mock results summaries.
- **tidyCDISC**
  tidyCDISC is a shiny app to easily create custom tables and figures from ADaM-ish data sets.
- **Tplyr**
  Tplyr is a grammar of data format and summary, designed to simplify the creation of clinical safety summaries.
- **Visual Define-XML Editor**
  Visual Editor for Define-XML 2.0 and ARM standards.

Open-Source Adoption



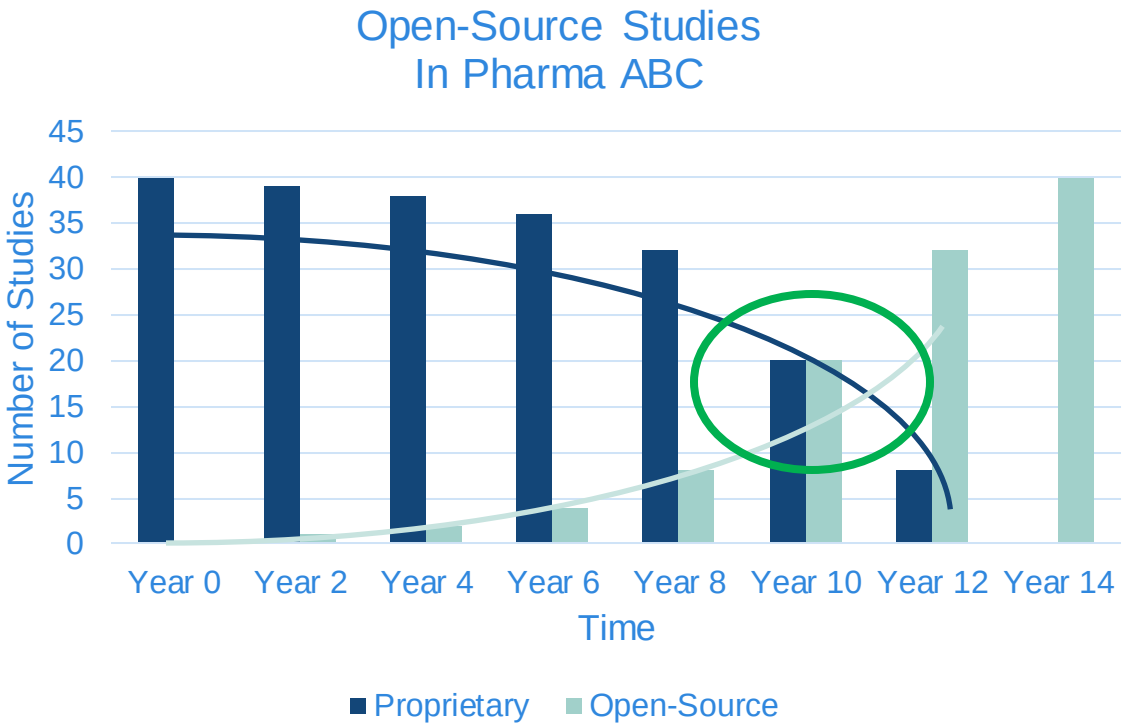
Open-Source Languages

Open-Source Transition Rate

Size of Organisation

Number of Ongoing Studies

Duration of Ongoing Studies



Regulator Feedback

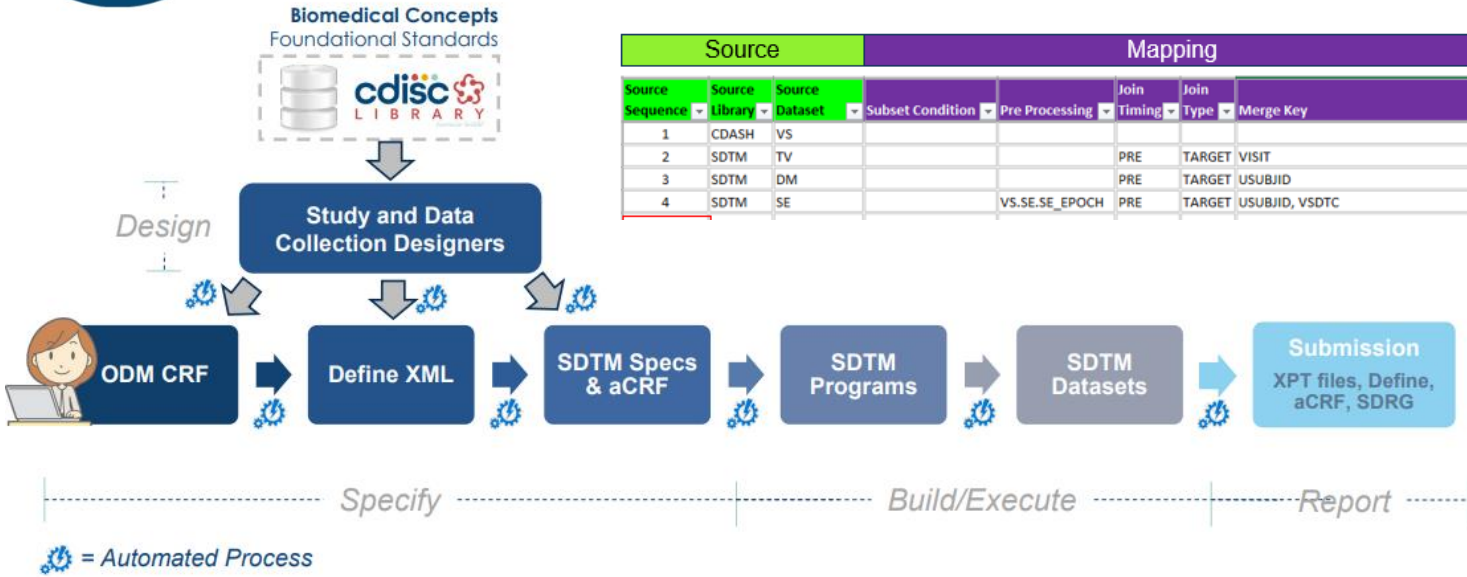
Infrastructure IT / SCE

Progamming Capability

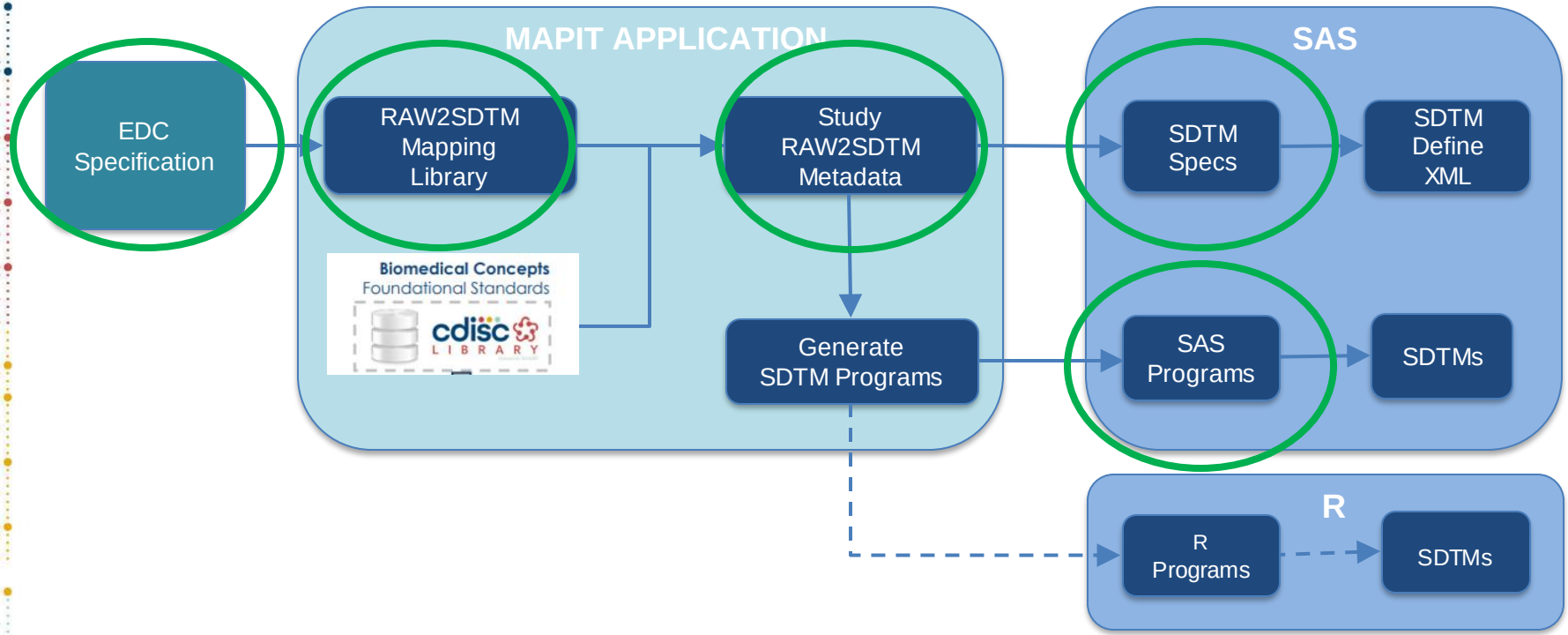
Future SDTM Programming Approach (from 2020 perspective)

Source				Mapping					Target						
Source Sequence	Source Library	Source Dataset	Source Variable	Map Sequence	Origin	Method	Comment	Code List	Target Library	Target Dataset	Target Variable	Target Description	Target Data Type	Target Length	Target Sorting Order
1	CDASH	VS			Assigned		CDISC360-2		SDTM	VS	STUDYID	Study Identifier	text	10	1
1	CDASH	VS			Assigned		VS	DOMAIN	SDTM	VS	DOMAIN	Domain Abbreviation	text	2	2
1	CDASH	VS	SUBJID		Assigned	ALL.USUBJID			SDTM	VS	USUBJID	Unique Subject Identifier	text	14	3
1	CDASH	VS			Assigned	VS.VSSPID			SDTM	VS	VSSPID	Sponsor-Defined Identifier	text	4	5
1	CDASH	VS	VISIT		Convert			VISITNUM	SDTM	VS	VISITNUM	Visit Number	integer	8	16
1	CDASH	VS	VISIT		Predecessor			VISIT	SDTM	VS	VISIT	Visit Name	text	18	17
1	CDASH	VS	VSDAT		Assigned	VS.VSDTC			SDTM	VS	VSDTC	Date/Time of Measurements	date	10	19

Source			Mapping					Target		
Source Sequence	Source Library	Source Dataset	Subset Condition	Pre Processing	Join Timing	Join Type	Merge Key	Target Sequence	Target Library	Target Dataset
1	CDASH	VS						5	SDTM	VS
2	SDTM	TV			PRE	TARGET	VISIT	5	SDTM	VS
3	SDTM	DM			PRE	TARGET	USUBJID	5	SDTM	VS
4	SDTM	SE		VS.SE.SE_EPOCH	PRE	TARGET	USUBJID, VSDTC	5	SDTM	VS

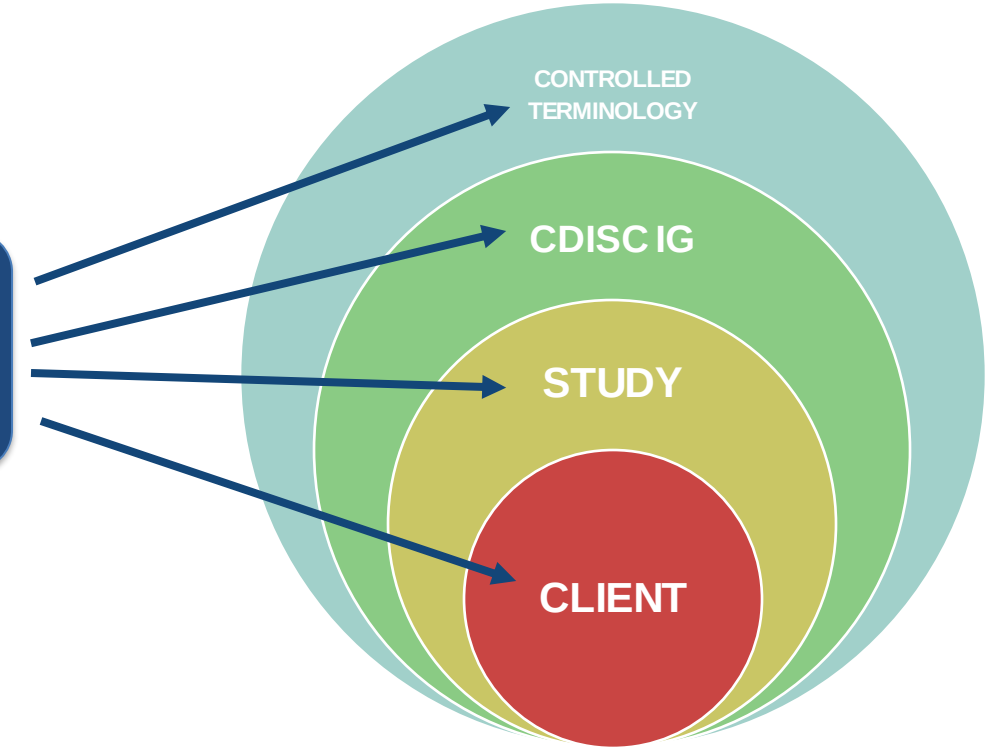


MAPIT SDTM Programming Approach



MAPPING LIBRARY

RAW2SDTM Mapping Library



Building Study Metadata

EDC SPEC VARIABLES

A	B	I
FormOID	FieldOID	DataDictionaryName
AE	AEOUT	$OUT
AE	AEACN	$AEACN

EDC SPEC CODELISTS

DataDictionaryName	CodedData	Ordinal	UserDataString
$AEACN	0	1	Dose not changed
$AEACN	1	2	Dose increased
$AEACN	2	3	Dose reduced
$AEACN	3	4	Drug interrupted
$AEACN	4	5	Drug permanently discontinued
$AEACN	98	6	Not applicable
$AEOUT	0	1	Recovered/resolved
$AEOUT	1	2	Recovering/resolving
$AEOUT	2	3	Recovered/resolved with sequelae
$AEOUT	3	4	Not recovered/not resolved
$AEOUT	4	5	Fatal

RAW2SDTM METADATA

RAW_TABLE	RAW_COLUM	RAW_TYP	RAW_LENGT	RAW_LABEL	RAW_FORMAT	RAW_COLUMN_WHRVALUE
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"0"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"1"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"2"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"3"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"4"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"98"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"0"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"1"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"2"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"3"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"4"

Building Study Metadata

RAW_TABLE	RAW_COLUMN	RAW_TYPE	RAW_LENGTH	RAW_LABEL	RAW_FORMAT	RAW_COLUMN_WHRVALUE	SDTM_TAB	SDTM_COLUM	SDTM_WHRCLAUSE	SDTM_ASSIGNED_VALUE
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"0"	AE	AEACN		"DOSE NOT CHANGED"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"1"	AE	AEACN		"DOSE INCREASED"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"2"	AE	AEACN		"DOSE REDUCED"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"3"	AE	AEACN		"DRUG INTERRUPTED"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"4"	AE	AEACN		"DRUG WITHDRAWN"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"98"	AE	AEACN		"NOT APPLICABLE"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"0"	AE	AEOUT		"RECOVERED/RESOLVED"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"1"	AE	AEOUT		"RECOVERING/RESOLVING"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"2"	AE	AEOUT		"RECOVERED/RESOLVED WITH SEQUELAE"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"3"	AE	AEOUT		"NOT RECOVERED/NOT RESOLVED"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"4"	AE	AEOUT		"FATAL"

```
*** RAW column =AEACN ***;
if AEACN= "0" then S__AEACN = "DOSE NOT CHANGED";
else if AEACN= "1" then S__AEACN = "DOSE INCREASED";
else if AEACN= "2" then S__AEACN = "DOSE REDUCED";
else if AEACN= "3" then S__AEACN = "DRUG INTERRUPTED";
else if AEACN= "4" then S__AEACN = "DRUG WITHDRAWN";
else if AEACN= "98" then S__AEACN = "NOT APPLICABLE";
```

```
# RAW column = AEACN
mutate(S__AEACN =
       case_when(
         AEACN=="0" ~ "DOSE NOT CHANGED",
         AEACN=="1" ~ "DOSE INCREASED",
         AEACN=="2" ~ "DOSE REDUCED",
         AEACN=="3" ~ "DRUG INTERRUPTED",
         AEACN=="4" ~ "DRUG WITHDRAWN",
         AEACN=="98" ~ "NOT APPLICABLE"
       ))  %>%
```

```
# Map the "AEACN" column to a new column called "S__AEACN"
mapping = {
    "0": "DOSE NOT CHANGED",
    "1": "DOSE INCREASED",
    "2": "DOSE REDUCED",
    "3": "DRUG INTERRUPTED",
    "4": "DRUG WITHDRAWN",
   "98": "NOT APPLICABLE",
}

# Apply mapping to create a new column "s__aeacn"
df['s__aeacn'] = df['aeacn'].map(mapping)
```

```
*** RAW column =AEOUT ***;
if AEOUT= "0" then S__AEOUT = "RECOVERED/RESOLVED";
else if AEOUT= "1" then S__AEOUT = "RECOVERING/RESOLVING";
else if AEOUT= "2" then S__AEOUT = "RECOVERED/RESOLVED WITH SEQUEL
else if AEOUT= "3" then S__AEOUT = "NOT RECOVERED/NOT RESOLVED";
else if AEOUT= "4" then S__AEOUT = "FATAL";
```

```
# RAW column = AEOUT
mutate(S__AEOUT =
       case_when(
         AEOUT=="0" ~ "RECOVERED/RESOLVED",
         AEOUT=="1" ~ "RECOVERING/RESOLVING",
         AEOUT=="2" ~ "RECOVERED/RESOLVED WITH SEQUELAE",
         AEOUT=="3" ~ "NOT RECOVERED/NOT RESOLVED",
         AEOUT=="4" ~ "FATAL"
       ))  %>%
```

```
# Map the "AEOUT" column to a new column called "S__AEOUT"
mapping = {
    "0": "RECOVERED/RESOLVED",
    "1": "RECOVERING/RESOLVING",
    "2": "RECOVERED/RESOLVED WITH SEQUELAE",
    "3": "NOT RECOVERED/NOT RESOLVED",
    "4": "DRUG WITHDRAWN",
}

# Apply mapping to create a new column "s__aeout"
df['s__aeout'] = df['aeout'].map(mapping)
```

Building Study Metadata

RAW_TABLE	RAW_COLUMN	RAW_TYPE	RAW_LENGTH	RAW_LABEL	RAW_FORMAT	RAW_COLUMN_WHRVALUE	SDTM_TAB	SDTM_COLUMN	SDTM_WHRCLAUSE	SDTM_ASSIGNED_VALUE
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"0"	AE	AEACN		"DOSE NOT CHANGED"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"1"	AE	AEACN		"DOSE INCREASED"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"2"	AE	AEACN		"DOSE REDUCED"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"3"	AE	AEACN		"DRUG INTERRUPTED"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"4"	AE	AEACN		"DRUG WITHDRAWN"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"98"	AE	AEACN		"NOT APPLICABLE"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"0"	AE	AEOUT		"RECOVERED/RESOLVED"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"1"	AE	AEOUT		"RECOVERING/RESOLVING"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"2"	AE	AEOUT		"RECOVERED/RESOLVED WITH SEQUELAE"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"3"	AE	AEOUT		"NOT RECOVERED/NOT RESOLVED"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"4"	AE	AEOUT		"FATAL"

ID	Name	NCI Codelist Code	Data Type	Order	Term	NCI Term Code
AEACN	Action Taken with Study Treatment	C66767	text	1	DOSE NOT CHANGED	C49504
AEACN	Action Taken with Study Treatment	C66767	text	2	DOSE INCREASED	C49503
AEACN	Action Taken with Study Treatment	C66767	text	3	DOSE REDUCED	C49505
AEACN	Action Taken with Study Treatment	C66767	text	4	DRUG INTERRUPTED	C49501
AEACN	Action Taken with Study Treatment	C66767	text	5	DRUG WITHDRAWN	C49502
AEACN	Action Taken with Study Treatment	C66767	text	98	NOT APPLICABLE	C48660
AEOUT	Outcome of Event	C66768	text	1	RECOVERED/RESOLVED	C49498
AEOUT	Outcome of Event	C66768	text	2	RECOVERING/RESOLVING	C49496
AEOUT	Outcome of Event	C66768	text	3	RECOVERED/RESOLVED WITH SEQUELAE	C49495
AEOUT	Outcome of Event	C66768	text	4	NOT RECOVERED/NOT RESOLVED	C49494
AEOUT	Outcome of Event	C66768	text	5	FATAL	C48275

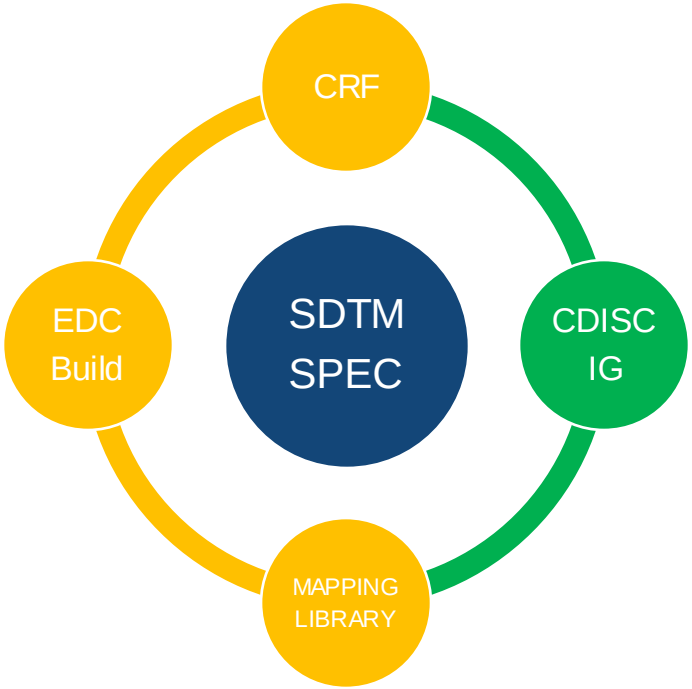
Action Taken with Study Treatment [*C66767*]

Permitted Value (Code)
DOSE NOT CHANGED [*C49504*]
DOSE INCREASED [*C49503*]
DOSE REDUCED [*C49505*]
DRUG INTERRUPTED [*C49501*]
DRUG WITHDRAWN [*C49502*]
NOT APPLICABLE [*C48660*]

Outcome of Event [*C66768*]

Permitted Value (Code)
RECOVERED/RESOLVED [*C49498*]
RECOVERING/RESOLVING [*C49496*]
RECOVERED/RESOLVED WITH SEQUELAE [*C49495*]
NOT RECOVERED/NOT RESOLVED [*C49494*]
FATAL [*C48275*]

Specification Puzzle



AE	Adverse Events	STUDYID,USUBJID,AESTDTC,AEDECOD
Variable	**Label**	**Data Type**
STUDYID	Study Identifier	text
DOMAIN	Domain Abbreviation	text
USUBJID	Unique Subject Identifier	text
AESEQ	Sequence Number	integer
AESPID	Sponsor-Defined Identifier	text
AETERM	Reported Term for the Adverse Event	text
AELLT	Lowest Level Term	text
AELLTCD	Lowest Level Term Code	integer
AEDECOD	Dictionary-Derived Term	text
AEPTCD	Preferred Term Code	integer
AEHLT	High Level Term	text
AEHLTCD	High Level Term Code	integer
AEHLGT	High Level Group Term	text

CDISC IG Metadata to Code

```
**********************************;
** code to create SEQ variable. **;
**********************************;

proc sort data=ae_addvars1;
  by studyid usubjid aestdtc aedecod;
run;

data ae_seq;
  set ae_addvars1;
  by studyid usubjid aestdtc aedecod;
  if first.usubjid then aeseq=1;
  else aeseq+1;
run;
```

```
##############################################
# code to create SEQ variable and final dataset
##############################################

ae_seq <-
  ae_addv  1 %>%
  arrange(STUDYID, USUBJID, AESTDTC, AEDECOD) %>%
  ungroup() %>%
  group_by(USUBJID) %>%
  mutate(AESEQ = 1:n()) %>%
  ungroup()

final_ae <-
  ae_seq %>%
  select(STUDYID, DOMAIN, USUBJID, AESEQ, AESPID, AETERM, AELLT, AELLTCD, AEDECOD, AEPTCD,
         AEHLT, AEHLTCD, AEHLGT, AEHLGTCD, AECAT, AEBODSYS, AEBDSYCD, AESOC, AESOCCD, AESER,
         AEACN, AEREL, AERELNST, AEOUT, AESCONG, AESDISAB, AESDTH, AESHOSP, AESLIFE, AESMIE,
         AECONTRT, AETOXGR, EPOCH, AESTDTC, AEENDTC, AESTDY, AEENDY)
```

```
**************************************;
** code to create final main dataset. **;
**************************************;

data sdtm.ae (label="Adverse Events");
  set templ_ae_seq;
  keep STUDYID DOMAIN USUBJID AESEQ AESPID AETERM AELLT AELLTCD AEDECOD AEPTCD
       AEHLT AEHLTCD AEHLGT AEHLGTCD AECAT AEBODSYS AEBDSYCD AESOC AESOCCD AESER
       AEACN AEREL AERELNST AEOUT AESCONG AESDISAB AESDTH AESHOSP AESLIFE AESMIE
       AECONTRT AETOXGR EPOCH AESTDTC AEENDTC AESTDY AEENDY;
run;
```

Final Automation Components

Combining Raw Data

SDTM_TABLE	rawn
AE	2
CE	2
CM	3
CO	5
CV	1
DA	1
DD	1
DM	7
DS	4
EC	4
EG	1
FA	7
HO	1
IE	1
LB	2
MH	2

JOIN_ORDER	SDTM_TABLE	RAW_TABLE	JOIN_TYPE	JOIN_KEY
1	AE	AE, SERAE	LEFT JOIN	"USUBJID", "AESPID", "AETERM"
1	CE		SET	
1	CM		SET	
1	CO		SET	

```
*********************************************
** Join raw datasets as per join metadata **;
*********************************************

proc sort data=ae_ae_2;
  by usubjid aespid aeterm;
run;

proc sort data=ae_serae_2;
  by usubjid aespid aeterm;
run;

data ae_addvars;
  merge ae_ae_2 (in=a)
        ae_serae_2 (in=b);
  by usubjid aespid aeterm;
  if a;
  ;
run;
```

```
######################################################################
# Join raw datasets as per join metadata
######################################################################

ae_addvars <-
  ae_ae_2 %>%
  left_join(ae_serae_2, by = c("USUBJID", "AESPID", "AETERM"))
```

Filtering Raw Data

RAW_TABLE_WHRCLAUSE
AEYN EQ "1"

RAW_TABLE	RAW_COLUM	SDTM_TABLE	SDTM_COLUMN	SDTM_WHRCLAUSE	SDTM_ASSIGNED_VALUE
VS	SBP	VS	VSTESTCD	SBP NE .	"SYSBP"
VS	SBP	VS	VSTEST	SBP NE .	"Systolic Blood Pressure"
VS	SBP	VS	VSCAT	SBP NE .	"PULSE AND BLOOD PRESSURE"
VS	SBP	VS	VSPOS	SBP NE .	"UNSPECIFIED"
VS	SBP	VS	VSORRES	SBP NE .	
VS	SBP	VS	VSORRESU	SBP NE .	"mmHg"
VS	SBP	VS	VSSTRESC	SBP NE .	
VS	SBP	VS	VSSTRESN	SBP NE .	
VS	SBP	VS	VSSTRESU	SBP NE .	"mmHg"

```
data AE_AE(keep=s__: );
  set AE_AE_set (where=(AEYN EQ "1"));
```

```
ae_ae <-
  ae %>%
  filter(AEYN == "1") %>%
  mutate(
```

```
*** RAW column =SBP ***;
if SBP NE . then do;
  S__VSCAT = "PULSE AND BLOOD PRESSURE";
  S__VSORRES = strip(put(SBP,best.));
  S__VSORRESU = "mmHg";
  S__VSPOS = "UNSPECIFIED";
  S__VSSTRESC = strip(put(SBP,best.));
  S__VSSTRESN = SBP;
  S__VSSTRESU = "mmHg";
  S__VSTEST = "Systolic Blood Pressure";
  S__VSTESTCD = "SYSBP";
  output;
end;
```

Code Snippets

DERIVATION_ORDER	RAW_TABLE	SDTM_TABLE	SDTM_COLUMN	COL_TYPE	DER_TYPE	SAS_DERIVATION
1		AE	AETRTEM	SUPP	GENERAL	aetrtem=ifc(aestdtcn ge rfstdtn,'Y','N');
2		AE	AETOXGR	PARENT	PRE-PROCESS	proc contents data=raw.ae out=ae_cont noprint; run;
3		AE	AETOXGR	PARENT	PRE-PROCESS	proc sql noprint; select name into :tox_vars separated by ' ' from ae_cont where prxmatch("/AEC\d{2}TOX/i",name); quit; %let tox_varn=&sqlobs;

4		AE	AETOXGR	PARENT	GENERAL	array toxc{*} $ aetoxgr &tox_vars;
						maxv=0;
						do i=1 to dim(toxc);
						if input(toxc{i},best.)>maxv then maxv=input(toxc{i},best.);
						end;
						if cats(of toxc{*}) ne '' then aetoxgr=strip(put(maxv,best.));

```
**************************************************
** Pre-Processing for Derivations  **;
**************************************************

** AETOXGR;
proc contents data=raw.ae out=ae_cont noprint;
run;

proc sql noprint;
  select name into :tox_vars separated by ' '
    from ae_cont
      where prxmatch("/AEC\d{2}TOX/i",name);
quit;
%let tox_varn=&sqlobs;
```

```
************************************************************;
** Derivation code **;
************************************************************;

data ae_addvars1;
  set ae_addvars;

  **EPOCH;
  EPOCH = "";

  **AETRTEM;
  aetrtem=ifc(aestdtcn ge rfstdtn,'Y','N');

  **AETOXGR;
  array toxc{*} $ aetoxgr &tox_vars;
  maxv=0;
  do i=1 to dim(toxc);
    if input(toxc{i},best.)>maxv then maxv=input(toxc{i},best.);
  end;
  if cats(of toxc{*}) ne '' then aetoxgr=strip(put(maxv,best.));

run;
```

SAS and R Automated Code Benefits

- Metadata Driven
- Mapping library not language dependent
- Double programming complete?
- Double programming required?

Next Steps

- Synthetic Data with AI
- R Package / OAK Project
- TFL Automation in R with Analysis Results Standards

Thank You!

John.mcdade@phastar.com