Metadata Driven Automation

For Analysis and Reporting

SANOFI

Beilei Xu
B&P Data Standards and Operations

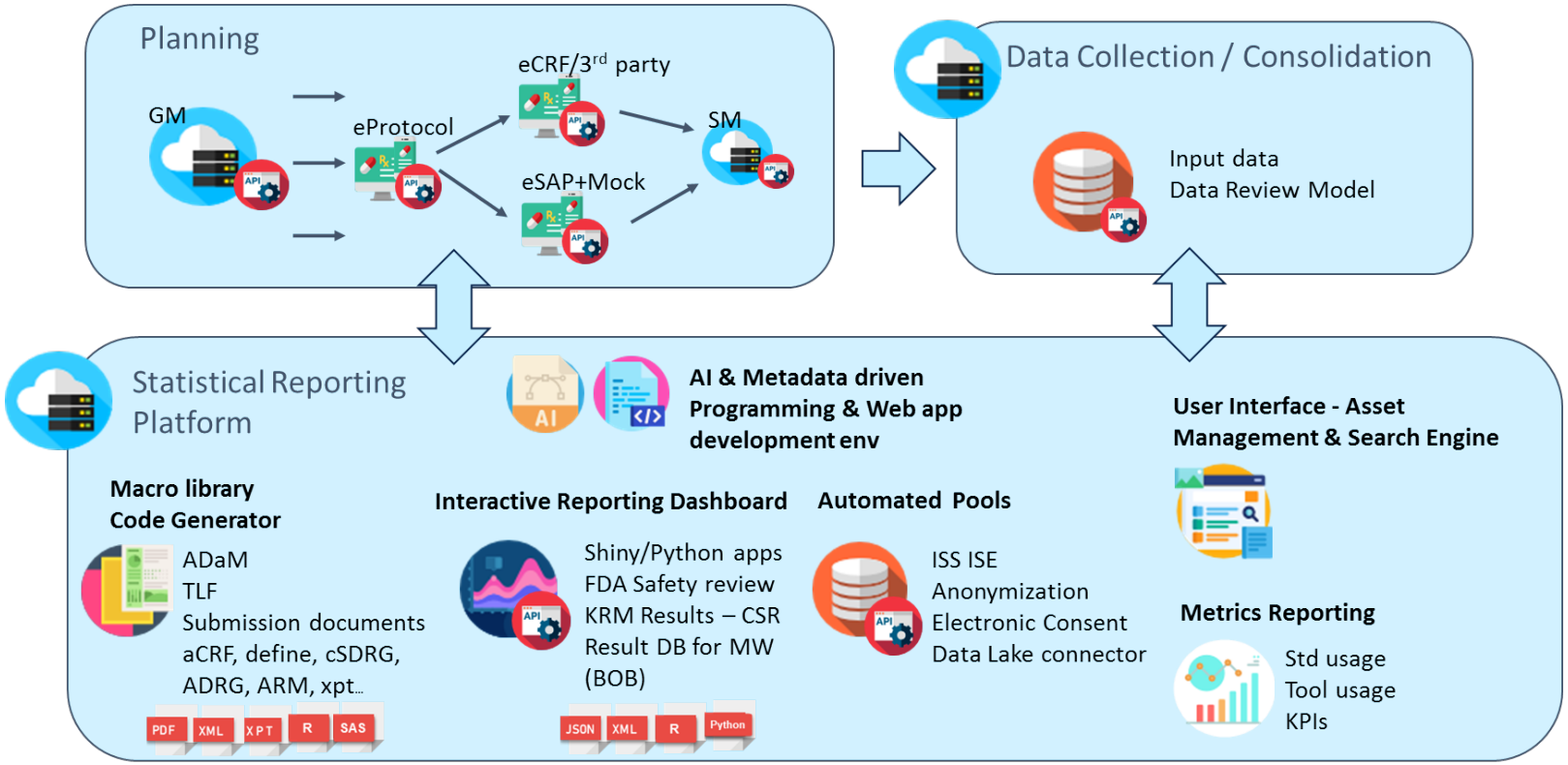
Outline

- **Background and Overview**
- **Machine Readable Metadata Review**
  - Study Design Metadata
  - TLF Metadata
  - SDTM/ADaM Metadata
- **Automation Tools**
  - Clinical Trial Designer/Mock Shell Generator
  - Metadata Generator
  - Code generator
- **Considerations**

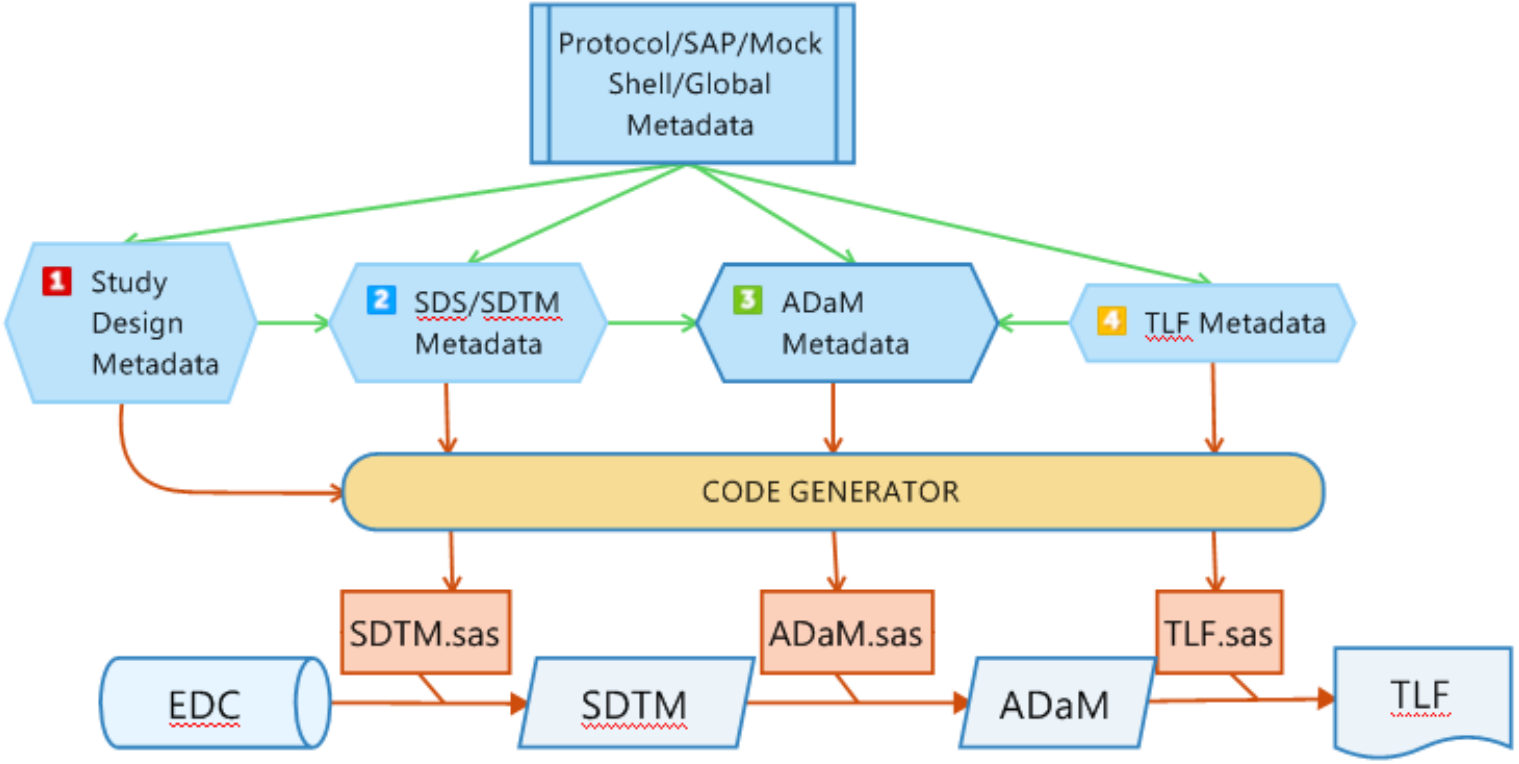
Background

- **Standards are required by HA – and drives the automation naturally**
- **CDISC data standards enable automation**
- **Metadata driven automation tools can be built for**
  - Metadata creation
  - Program creation
  - Dataset and TLF creation
  - Submission package creation

Metadata Driven Automation Process Overview



Analysis and Reporting



Study Design Metadata

A	B	C	E	F	G
Item #	**Item Name**	**Definition**	**Possible Values**	**Notes**	**Study_SOA Value**
1	sdtm_type	Full SDTM or Submission Ready SDTM	Full (default), null	The SDTM dataset format: Full= full SDTM (Suppquals are horizontal in parent domain), null = submission ready SDTM	Full
2	rand_study	Randomized Study or not?	Yes (default), No		Yes
3	nb_study_period	Number of periods in the study	1 (default), >1...	Use this to determine the xx in TRTxxP, TRTxxA etc. variables	1
4	epoch_treat	Treatment EPOCH	TREATMENT		TREATMENT
5	data_cut_off	Data Cut-Off Date for interim analysis			
6	age_group	Age Group	See Format Tab		f_ageqr1

TLF Metadata

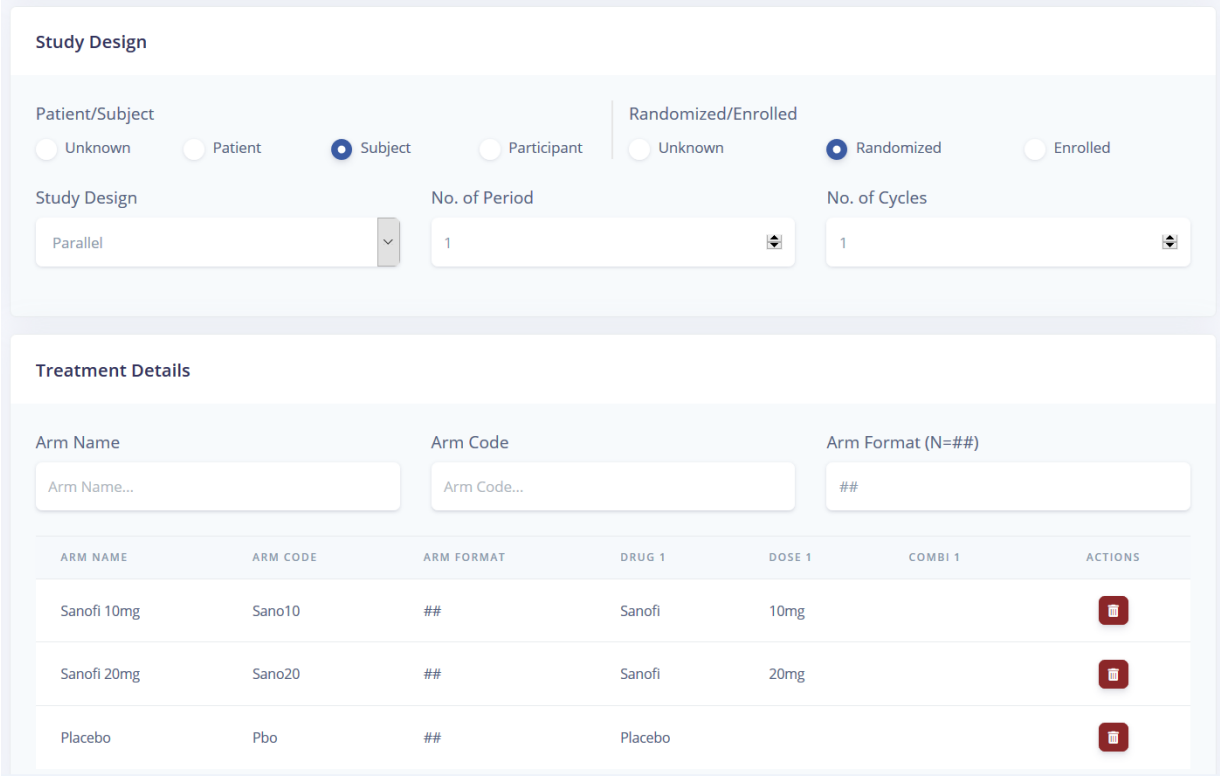
TA	filename	domains	
All	ae_selae_allae_s_l	ADAE |ADSL	USUBJID AEBODSYS AEDECOD AETERM AESTDTC ASTDY PREFL TRTEMFL FUPFL AEACN AEREL A
All	ae_death_s_l	ADAE |ADSL	USUBJID AEBODSYS AEDECOD AETERM AESTDTC ASTDY PREFL TRTEMFL FUPFL ASEV AEACN AE
All	ae_socptfreq_byage_s_t	ADAE |ADSL	AEDICTV AEBODSYS AEDECOD BDSYORD PSOCFL PTORD TRTEMFL USUBJID |AGEGR1 AGEGR1N
All	ae_overview_cntrysite_s_t	ADAE |ADSL	AEACN AEOUT AEREL AESER ASEVN PSOCFL TRTEMFL USUBJID |CNTRYNAM COUNTRY EXPOSE
Oncology	ae_death_studyperiod_onco_s_t	ADSL	DTHCAUS DTHDT DTHFL DTHONSTD DTHTRTEM SAFFL TRTxxAN TRTSDT USUBJID
Oncology	ae_socpt_ltdeath_othpd_onco_s_t	ADAE |ADSL	AEDICTV AEBODSYS AEDECOD AEOUT BDSYORD PSOCFL TRTEMFL USUBJID |DTHCAUS DTHETD
Oncology	ae_socpt_ltdeath_pd_onco_s_t	ADAE |ADSL	AEDICTV AEBODSYS AEDECOD AEOUT BDSYORD PSOCFL TRTEMFL USUBJID |DTHCAUS DTHETD
Oncology	ae_death_othpop_onco_a_t	ADSL	DTHCAUS DTHFL EXPOSEFL RANDFL USUBJID
All	ae_death_studyperiod_s_t	ADSL	DTHFL DTHONSTD DTHTRTEM SAFFL TRTxxA TRTxxAN USUBJID
All	ae_death_othpop_a_t	ADSL	DTHFL EXPOSEFL RANDFL USUBJID
Oncology	ae_overview_onco_s_t	ADAE |ADSL	AEACN AEREL AESER ATOXGRN TRTEMFL USUBJID |EXPOSEFL RANDFL SAFFL USUBJID
Oncology	ae_overview_posttrt_onco_s_t	ADAE |ADSL	AEACN AEREL AESER ATOXGRN FUPFL USUBJID |EXPOSEFL RANDFL SAFFL USUBJID
Oncology	ae_overview_pretrt_onco_s_t	ADAE |ADSL	AEACN AEREL AESER ATOXGRN PREFL USUBJID |EXPOSEFL RANDFL SAFFL USUBJID
All	ae_socptfreq_byrace_s_t	ADAE |ADSL	AEDICTV AEBODSYS AEDECOD BDSYORD PSOCFL PTORD TRTEMFL USUBJID |RACE RACEN SAFF
All	ae_socptfreq_bysex_s_t	ADAE |ADSL	AEDICTV AEBODSYS AEDECOD BDSYORD PSOCFL PTORD TRTEMFL USUBJID |SAFFL SEX TRTxxA
All	ae_correctivetrt_cqzz_s_t	ADAE |ADSL	AECONTRT CQzzCD PSOCFL TRTEMFL USUBJID |SAFFL TRTxxA TRTxxAN USUBJID
Oncology	ae_correctivetrt_onco_cqzz_s_t	ADAE |ADSL	AECONTRT ATOXGRN CQzzCD PSOCFL TRTEMFL USUBJID |SAFFL TRTxxA TRTxxAN USUBJID
Oncology	ae_correctivetrt_onco_smqzz_s_t	ADAE |ADSL	AECONTRT ATOXGRN PSOCFL SMQzzCD TRTEMFL USUBJID |SAFFL TRTxxA TRTxxAN USUBJID
All	ae_correctivetrt_smqzz_s_t	ADAE |ADSL	AECONTRT PSOCFL SMQzzCD TRTEMFL USUBJID |SAFFL TRTxxA TRTxxAN USUBJID

SDTM/ADaM Metadata

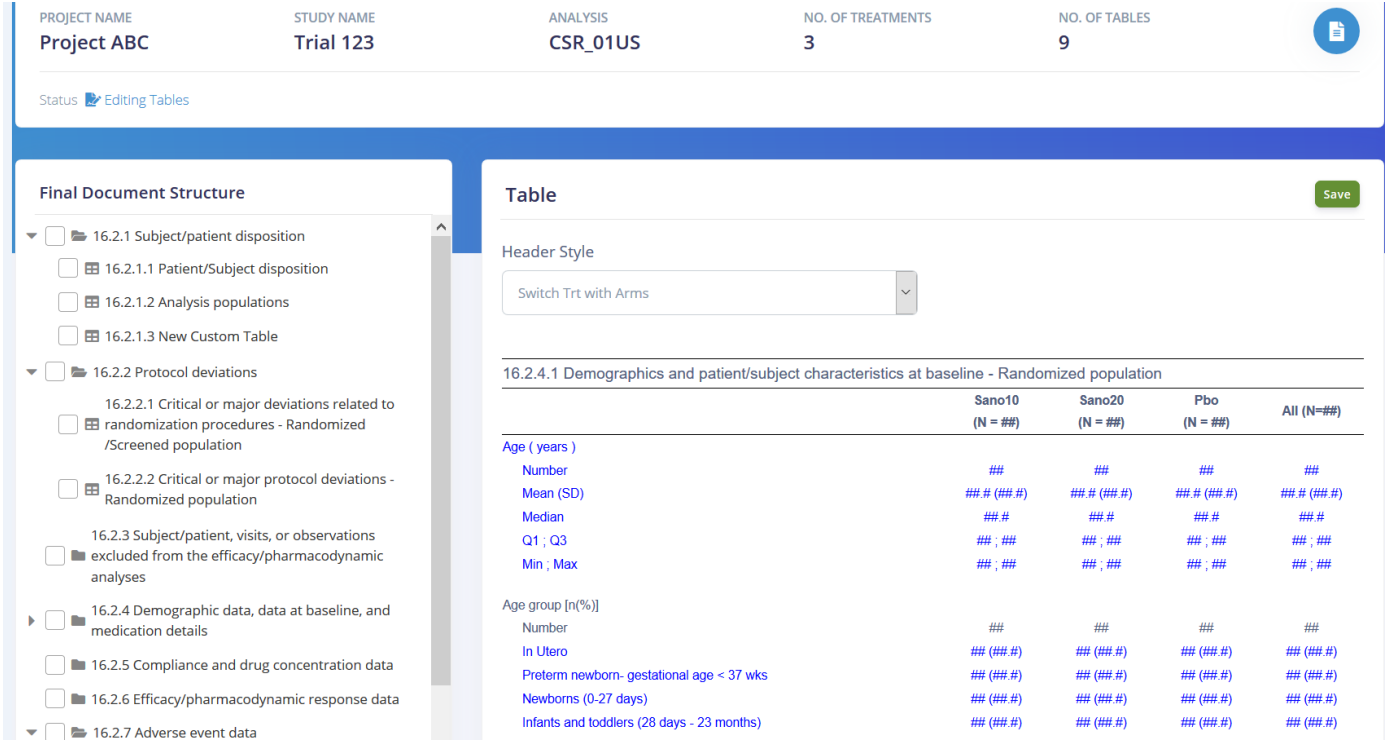
Dataset Name	Relative Order	Variable Name	Variable Label	Type	Standard Length	Significant Digits	Display Format	Controlled Terminolog	Origin	Derivation Type	Pseudo Code	Core
ADSL	1	STUDYID	Study Identifier	text	15				Predecessor		source	Req
ADSL	2	USUBJID	Unique Subject Identifier	text	25				Predecessor		source	Req
ADSL	100	SUBJID	Subject Identifier for the Study	text	12				Predecessor		source	Req
ADSL	110	SITEID	Study Site Identifier	text	7				Predecessor		source	Req
ADSL	130	COUNTRY	Country	text	3			ISO3166	Predecessor		source	Req
ADSL	140	AGE	Age	integer	8				Predecessor		source	Req
ADSL	150	AGEU	Age Units	text	6			AGEU	Predecessor		source	Req
ADSL	180	AGEGR1	Pooled Age Group 1	text	16			[AGEGR1]	Derived	Computation	recode	Cond
ADSL	210	SEX	Sex	text	1			SEX	Predecessor		source	Req
ADSL	220	RACE	Race	text	50			RACE	Predecessor		source	Req
ADSL	310	TRTSDT	Date of First Exposure to Treatment	integer	8		YYMMDD10.		Derived	[Imputation or Computation]	derivation	Cond
ADSL			Date of Last Exposure							[Imputation or		

Variable Metadata | recode | derivation | macro function | custom | Pseudo Code Descriptions | Variable Category Descriptions | CompMethod | Whe

Clinical Study Designer



Mock Shell Generator



Metadata Generator

- **Automatically populate study level metadata based on global metadata**
  - Populate model required variables
  - Populate TLF required variables based on TLF metadata
  - Populate study specific variables based on Study Design Variables

Dataset Name	Relative Order	Variable Name	Variable Label	Type	Standard Length	L	Signi fica	Display Format	Controlled Terminolog	Origin	Derivation Type	Source	Derivation
ADSL	1490	TRTxxP	Planned Treatment for Period xx	text	200				[TRTxxP]	Derived	Computation		[Treatment in DM.ARM corresponding to Period xx] or [TRTSEQP corresponding to Period xx]

Dataset Name	Relative Order	Variable Name	Variable Label	Type	Standard Length	L	Signi fica	Display Format	Controlled Terminolog	Origin	Derivation Type	Source	Derivation
ADSL	1490	TRT01P	Planned Treatment for Period 01	text	200				[TRT01P]	Derived	Computation		[Treatment in DM.ARM corresponding to Period 01] or [TRTSEQP corresponding to Period 01]
ADSL	1490	TRT02P	Planned Treatment for Period 02	text	200				[TRT02P]	Derived	Computation		[Treatment in DM.ARM corresponding to Period 02] or [TRTSEQP corresponding to Period 02]

Code Generator

- **R or Python program**
  - Read in the metadata
  - Parse the metadata
  - Generate sas code
- **Pseudo code in Variable Metadata**

Pseudo Code	Description
source	Indicates the variable has an origin of "Predecessor" where source variable is defined in "Source" column
gen_xxx	Indicates a macro function name that links to the "Function Name" in "macro function" tab
derivation	Indicates that detail derivation specs are available in "derivation" tab
recode	Indicates that detail derivation specs are available in "recode" tab

Machine Readable Metadata

Dataset Name	Relative Order	Variable Name	Variable Label	Type	Origin	Derivation Type	Pseudo Code
ADSL	180	AGEGR1	Pooled Age Group 1	text	Derived	Computation	recode
ADSL	1150	AGEGR1N	Pooled Age Group 1 (N)	integer	Assigned		recode

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

```
if 60 <= AGE <= 70 then do;
    AGEGR1 = "60-70 years";
    AGEGR1N = 1;
end;
else if 70 < AGE then do;
    AGEGR1 = "over 70 years";
    AGEGR1N = 2;
end;
```

Considerations

- **Metadata Repository is essential to capture, store and maintain the metadata**
- **Machine readable metadata is the key to drive the end to end automation**
- **Code generator can be expanded to generate R codes**
- **Study specific metadata is a potential area for machine learning**
- **Submission Package Tools can be developed:**
  - resize, xpt, acrf, define, cSDRG, ADRG, etc.

Acknowledgement

- Thanks to Andre Couturier for sharing his ideas and his slide (Slide 4)
- Thanks to Mei Wu for her review and support

Thank You!


phuse·share
connect·share·advance
Around the world
throughout 2019
Single Day
Events