

WITH STANDARDS – UNLOCK THE POWER OF DATA



2022

CHINA

INTERCHANGE

29-30 JULY | VIRTUAL EVENT

CRF Question Text-Driven SDTM aCRF, Mapping Specification and Programs Automation

Presented by Haiqiang Luo, Director, Clinical Programming, Kelun-Biotech (科伦)



Meet the Speaker

Haiqiang Luo

Title: Director

Organization: Kelun-Biotech (科伦)

Haiqiang Luo is currently director of clinical programming at Kelun-Biotech.

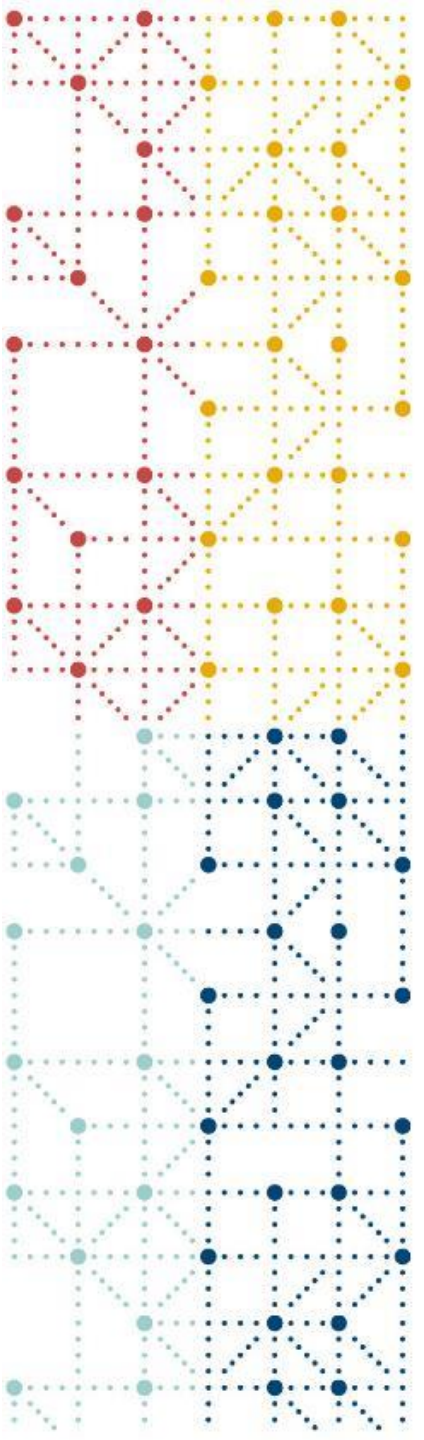
He has been working as a SAS programmer in the pharmaceutical industry for more than 12 years (including 7.5 years at global CRO and 2.5 years at local pharma in China). He has significant experience in CDISC and various therapeutic areas.

Haiqiang is interested in tools development and new technology implementation in daily work.



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 and Kelun-Biotech.*
- *The author(s) have no real or apparent conflicts of interest to report.*



Agenda

1. Steps of SDTM Work and CRF Question Text-Driven Automation Solution
2. Automatic generation of SDTM aCRF
3. Automatic generation of SDTM Core Mapping Spec and Programs
4. Summary

Current “Automation” Strategy in Industry?

➤ SDTM aCRF:

1. **Exporting** annotation information from prepared aCRF in pervious projects into **xdfd or fdf** file.
2. **Manually** preparing (partial of) new xdfd files for the current study with **SAS** and **Excel** software, using annotation information retrieved in Step 1 (might require **manual** input of **pages** and **coordinates** information)
3. **Re-importing xdfd** file (annotations) created in Step 2 into the new study (might require **manual adjustment** of annotation boxes)
Note: The accuracy of annotations depends on similarity in CRF contents between the new and previous projects

➤ SDTM Mapping Specification:

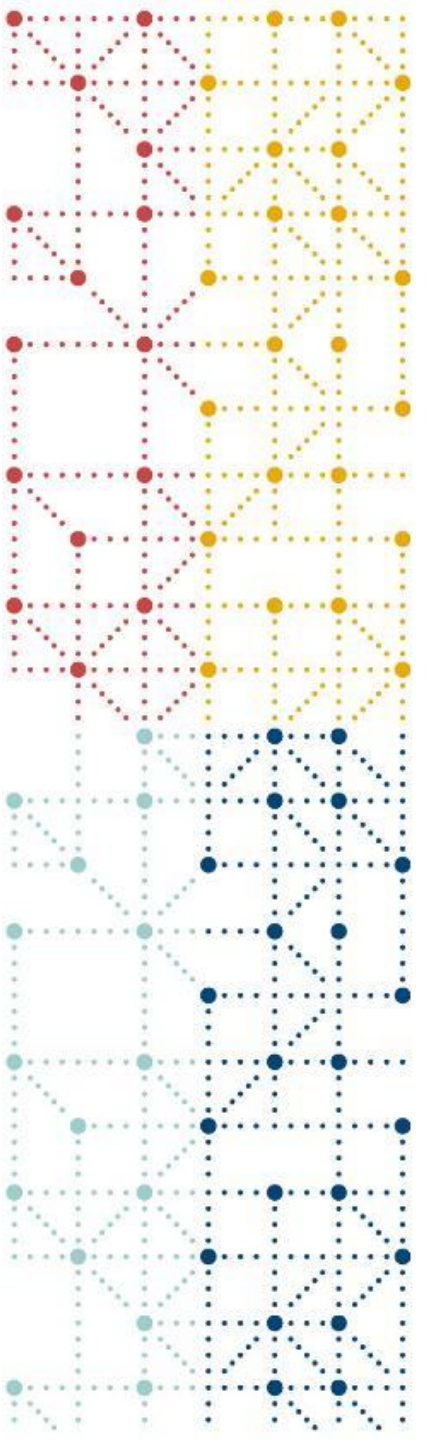
Does automation process take place?

If yes, how does the final product – Spec look? Is manual work involved?

➤ SDTM Programs:

Does automation process take place?

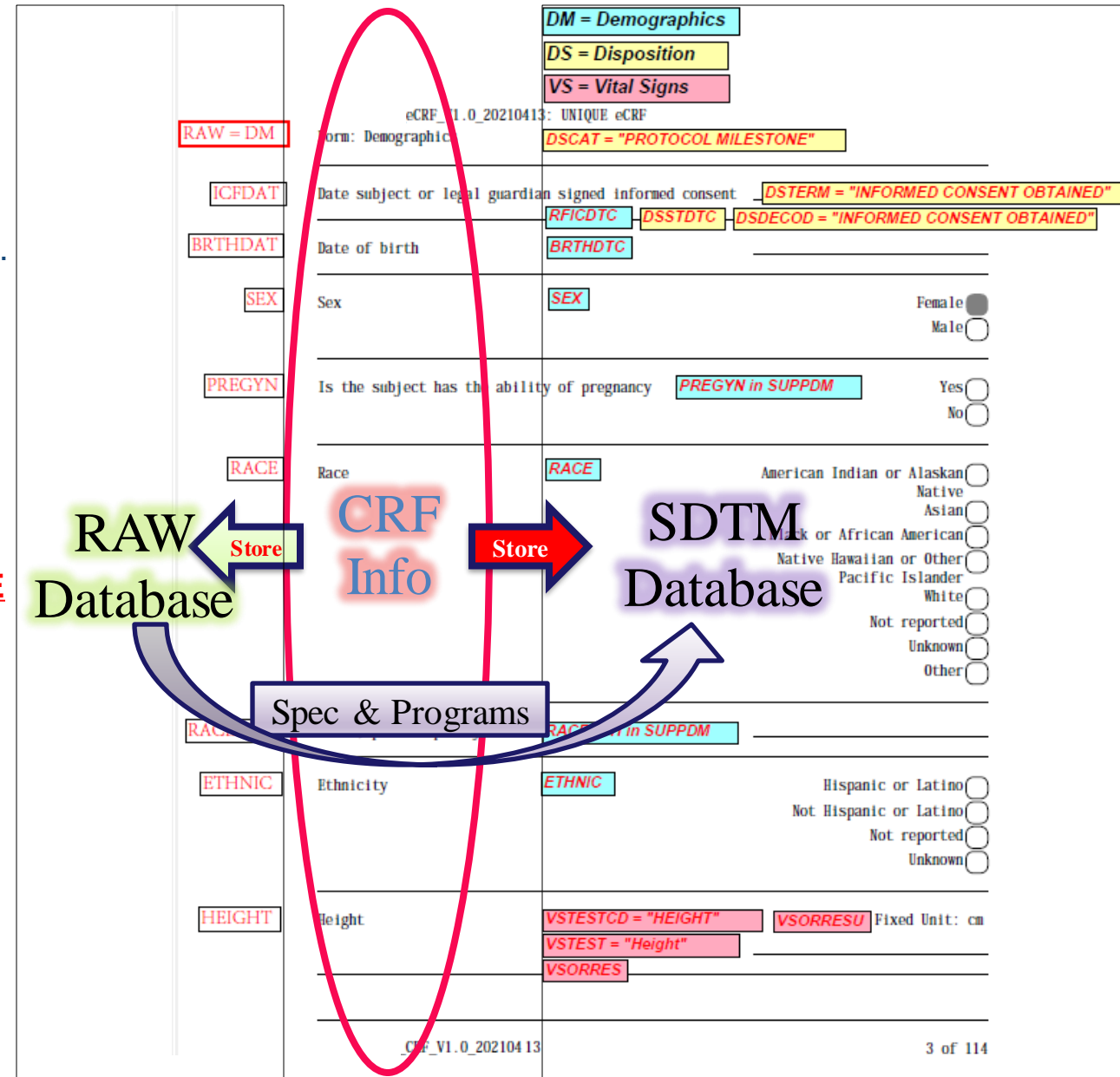
If yes, how does the final product – Program look? Is manual work involved?



Steps of SDTM Work and CRF Question Text-Driven Automation Solution

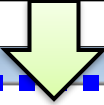
Dissect the SDTM Work

- Raw database and SDTM database are both **CRF information CARRIERS**, regardless of their formatting differences
 - **Raw database:** original information collected on CRF, with no naming criteria applied on datasets and variables.
 - **SDTM:** standardized (usually following CDISC) datasets which re-organize raw database variables into domains.
 - **Specification and Programs:** a bridge to help convert raw data into SDTM data through arranging and translating mapping rules displayed on aCRF into domains.
- The essence of SDTM work is **INFORMATION EXCHANGE**
 - Information on CRF is the origin of all datasets and the core of clinical studies
 - ◆ Source from CRF, reflect back to CRF! CRF Driven!
- Raw database, SDTM mapping specification and SDTM programs shall **NOT** be the main focus (They are **NOT required in NDA submission**)



Dissect the SDTM Work

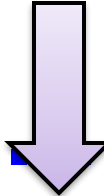
RAW Database



```
dm.sas
1 Data DM_DM;
2 Set RAW_DM;
3 RFICDTC = strip(put(ICFDAT, yymmdd10.));
4 BRTHDTC = strip(put(BRTHDAT, yymmdd10.));
5 SEX = SEX;
6 RACE = RACE;
7 ETHNIC = ETHNIC;
8 PREGYN = PREGYN; *** Supplemental variable;
9 RACEOTH = RACEOTH; *** Supplemental variable;
10 ;run;
```

SDTM Programs

```
ds.sas
23 Data DS_DM;
24 Set RAW_DM;
25 DSCAT = 'PROTOCOL MILESTONE';
26 if `missing(ICFDAT)` then DSDECOD = 'INFORMED CONSENT OBTAINED';
27 if `missing(ICFDAT)` then DSTERM = 'INFORMED CONSENT OBTAINED';
28 DSSTDTC = strip(put(ICFDAT, yymmdd10.));
29 ;run;
```



SDTM Database

DM = Demographics DS = Disposition VS = Vital Signs

CRF_V1.0 20210413: UNIQUE eCRF
Form: Demographics **DSCAT = "PROTOCOL MILESTONE"**

Date subject or legal guardian signed informed consent **DSTERM = "INFORMED CONSENT OBTAINED"**
RFICDTC DSSTDTC **DSDECOD = "INFORMED CONSENT OBTAINED"**

Date of birth **BRTHDTC**

Sex **SEX** Female Male

Is the subject has the ability of pregnancy **PREGYN in SUPPDM** Yes No

Race **RACE** Alaskan Native Asian Black or African American Native Hawaiian or Other Pacific Islander

Other, please specify **RACEOTH in SUPPDM**

Ethnicity **ETHNIC** Hispanic or Latino Not Hispanic or Latino Not reported Unknown

Height **VSTESTCD = "HEIGHT"** Fixed Unit: cm
VSTEST = "Height"
VSORRES

SDTM aCRF



Variable Name	Variable Label	Study Derivation
1 STUDYID	Study Identifier	
3 DOMAIN	Domain Abbreviation	DM
4 USUBJID	Unique Subject Identifier	concatenation of STUDYID, STUDYENVSITENUMBER, SUBJECT, separated by "-".
5 SUBJID	Subject Identifier for the Study	RAW.DM.SUBJID
10 RFICDTC	Date/Time of Informed Consent	RAW.DM.ICFDAT in ISO8601 format
17 BRTHDTC	Date/Time of Birth	RAW.DM.BRTHDAT in ISO8601 format
20 SEX	Sex	RAW.DM.SEX
21 RACE	Race	RAW.DM.RACE
22 ETHNIC	Ethnicity	RAW.DM.ETHNIC
32 PREGYN	Protocol Version	RAW.DM.PREGYN
33 RACEOTH	Race - Other	RAW.DM.RACE

SDTM Spec



Variable Name	Variable Label	Study Derivation
1 STUDYID	Study Identifier	
3 DOMAIN	Domain Abbreviation	DS
4 USUBJID	Unique Subject Identifier	concatenation of STUDYID, STUDYENVSITENUMBER, SUBJECT separated by "-".
DSSEQ	Sequence Number	Sequential number ensuring uniqueness of records within each USUBJID in the domain sorted by Key variables
DSTERM	Reported Term for the Disposition Event	RAW.DM: if ICFDAT is not missing then set to 'INFORMED CONSENT OBTAINED'
DSCAT	Category for Disposition Event	RAW.DM: set to 'PROTOCOL MILESTONE'
9 DSSTDTC	Start Date/Time of Disposition Event	RAW.DM.ICFDAT in ISO8601 format
12 DSSTDY	Study Day of Start of Disposition Event	=DSSTDTC - DM.RFSTDTC + 1 if DSSTDTC >= DM.RFSTDTC;

What the heck!
我嘍! 重复在搞事情啊!!!





SDTM aCRF

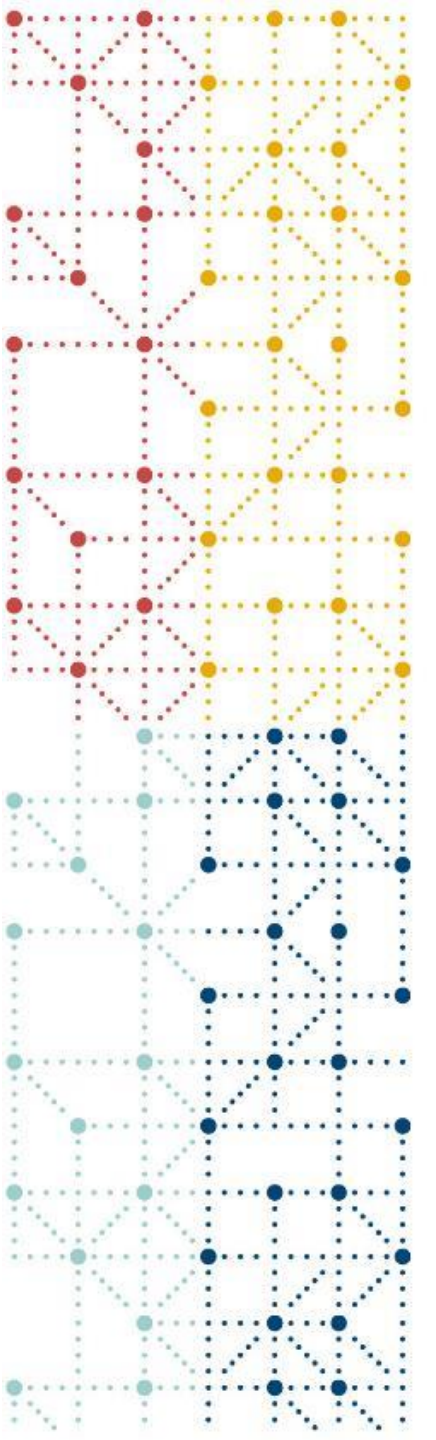
SDTM Programs

SDTM Spec



Steps of SDTM Work and CRF Question Text-Driven Solution

Task [Step] (Percentage of time spent in this task)	Difficulties & Challenges	Degree of Automation	CRF Question Text-Driven Solution (Percentage of time saved)
1. SDTM aCRF (10 ~15%) - Core	<ul style="list-style-type: none"> ● Ways to annotate specific information ● Performing repetitive work when starting new projects ● Consistency in data processing criteria among all projects 	Automation is possible throughout the step.	JavaScript Mechanism: By reading in CRF pdf document and processing contained information (100%)
2. SDTM Mapping Specification (40 ~45%)	<ul style="list-style-type: none"> ● Looking for relationship between SDTM database and raw database <p>Note: Most mapping rules are those transferred from aCRF.pdf to Excel file.</p>	Automation can be implemented on some parts of SDTM mapping specification if the data is obtained from CRF. i.e. except: <ul style="list-style-type: none"> • Core variables in the DM domain • Special domains. E.g. SV, SE, RELREC, etc. • Data from external source • Coding information • Special variables, E.g. EPOCH, VISITNUM, etc. • Some required CTs • Data from Protocol 	JavaScript + VBA Mechanism: By reading in information from aCRF.pdf and displaying dataflow from raw to SDTM database as an excel file following mapping specification template (75%+)
3. SDTM Programs (40 ~50%)	<ul style="list-style-type: none"> ● Raw datasets + SDTM domains and variables input <p>Note: This step is to translate mapping rules into SAS Code.</p>	Some basic SAS codes and some default derived codes such as --DY, --EPOCH, VISITNUM, can be automatically generated.	JavaScript + VBA Mechanism: By reading in information from aCRF.pdf and displaying dataflow from raw to SDTM database as SAS code following SAS program template (60%+)



Automatic generation of SDTM aCRF

Automatic generation of SDTM aCRF

To create an **annotation text box** in CRF, two parts of information are required:

➤ **Properties of the annotation box:**

- 1) Filling color
- 2) CRF page and coordinates**
- 3) The width and height of the box

➤ **Properties of the annotation text:**

- 1) Font (Italic/Bold) and size, alignment and color
- 2) Contents of the text**

Automatic generation of SDTM aCRF

To create an **annotation text box** in CRF, two parts of information are required:

➤ **Properties of the annotation box:**

1) Filling color - Default ✓

2) CRF page and coordinates - JavaScript ✓

3) The width and height of the box - Annotation Text Driven ✓

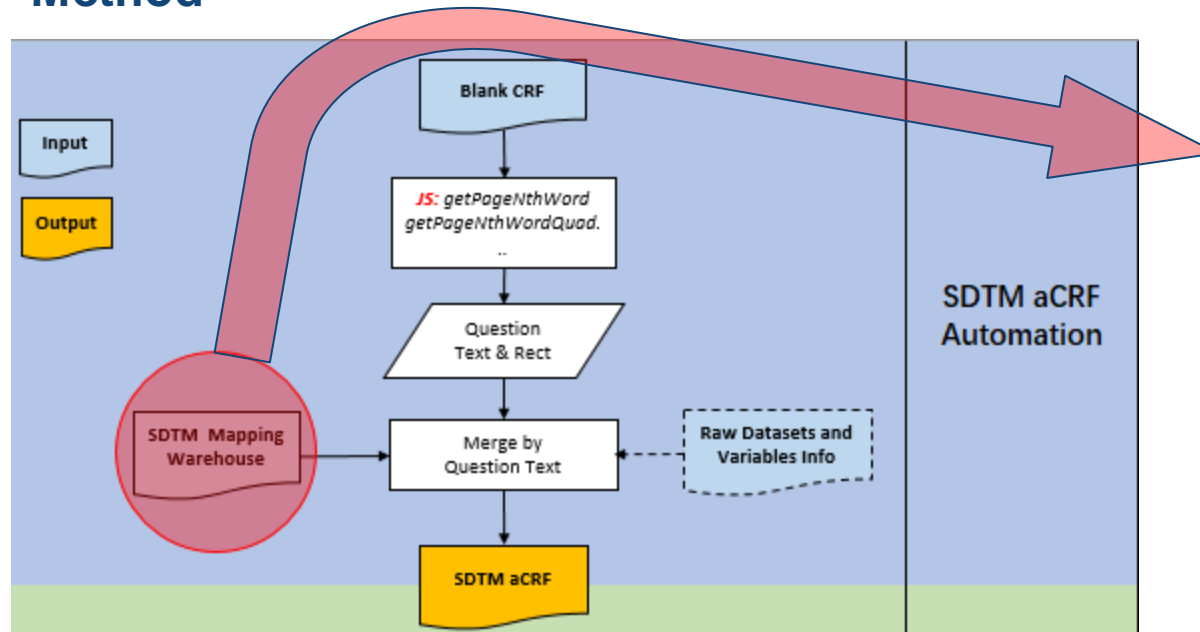
➤ **Properties of the annotation text:**

1) Font (Italic/Bold) and size, alignment and color - Default ✓

2) Contents of the text – SDTM Mapping Warehouse ✓

Automatic generation of SDTM aCRF

1. Obtaining **CRF question texts** and their corresponding **page numbers and coordinates** on pdf <- JavaScript App Method
2. Mapping extracted CRF question texts to SDTM **annotation texts** <- SDTM Mapping Warehouse (Central File) + JavaScript Merge
3. Annotating **on the correct CRF page and to the right location** next to the text <- JavaScript Doc Method



Form Label	CRF Question Text	Annotation Text
Subject	Subject	SUBJID
Visit Date	Was the visit done?	[NOT SUBMITTED]
Visit Date	If not done, provide the reason	[NOT SUBMITTED]
Visit Date	Visit date	SVSTDTC
Visit Date	Visit date	SVENDTC
Demographics	Form: Demographics	DSCAT = "PROTOCOL MILESTONE"
Demographics	Date subject or legal guardian signed informed consent	DSTERM = "INFORMED CONSENT OBTAINED"
Demographics	Date subject or legal guardian signed informed consent	DSDECOD = "INFORMED CONSENT OBTAINED"
Demographics	Date subject or legal guardian signed informed consent	DSSTDTC
Demographics	Date subject or legal guardian signed informed consent	RFICDTC
Demographics	Date of birth	BRTHDTC
Demographics	Sex	SEX
Demographics	Race	RACE
Demographics	Other, please specify	RACEOTH in SUPPDM
Demographics	Ethnicity	ETHNIC
Demographics	Height	VSTESTCD = "HEIGHT"
Demographics	Height	VSTEST = "Height"
Demographics	Height	VSORRES
Demographics	Height Unit	VSORRESU
Demographics	Is the subject has the ability of pregnancy	PREGYN in SUPPDM
Demographics	Is the subject rescreened?	SUBSCRYN in SUPPDM
Demographics	Previous Subject ID	SUBJIDP in SUPPDM
Breast Cancer History	Form: Breast Cancer History	MHCAT = "Breast Cancer History"
Breast Cancer History	Form: Breast Cancer History	MHTERM = "Breast Cancer"
Breast Cancer History	Date of initial diagnosis	MHSTDTC

SDTM aCRF Automation



Date subject or legal guardian signed informed consent

Date of birth

Sex

Is the subject has the ability of pregnancy

Race

Other, please specify

Ethnicity

Height

Form Label	CRF Question Text	Annotation Text
Subject	Subject	SUBJID
Visit Date	Was the visit done?	[NOT SUBMITTED]
Visit Date	If not done, provide the reason	[NOT SUBMITTED]
Visit Date	Visit date	SVSTDTC
Visit Date	Visit date	SVENDTC
Demographics	Form: Demographics	DSCAT = "PROTOCOL MILESTONE"
Demographics	Date subject or legal guardian signed informed consent	DSTERM = "INFORMED CONSENT OBTAINED"
Demographics	Date subject or legal guardian signed informed consent	DSDECOD = "INFORMED CONSENT OBTAINED"
Demographics	Date subject or legal guardian signed informed consent	DSSTDTC
Demographics	Date of birth	RFICDTC
Demographics	Sex	SEX
Demographics	Date of birth	BRTHDTC
Demographics	Sex	SEX
Demographics	Is the subject has the ability of pregnancy	PREGYN in SUPPDM
Demographics	Race	RACE
Demographics	Other, please specify	RACEOTH in SUPPDM
Demographics	Ethnicity	ETHNIC
Demographics	Height	VSTESTCD = "HEIGHT"
Demographics	Height	VSTEST = "Height"
Demographics	Height	VSORRES
Demographics	Height Unit	VSORRESU
Demographics	Is the subject rescreened?	SUBSCRYN in SUPPDM
Demographics	Previous Subject ID	SUBJIDP in SUPPDM
Breast Cancer Histor	Form: Breast Cancer History	MHCAT = "Breast Cancer History"
Breast Cancer Histor	Form: Breast Cancer History	MHTERM = "Breast Cancer"
Breast Cancer Histor	Date of initial diagnosis	MHSTDTC
Breast Cancer Histor	Histological grade	HISTGR in SUPPMH
Breast Cancer Histor	Primary site	PRIMSITE in SUPPMH
Breast Cancer Histor	TNM staging (N)	RLYMPH in SUPPMH
Breast Cancer Histor	TNM staging (M)	MEFDIS in SUPPMH
Breast Cancer Histor	Clinical staging (N)	CLNCLN in SUPPMH
Breast Cancer Histor	Clinical staging (M)	CLNCLM in SUPPMH
Breast Cancer Histor	Date of recurrence or metastasis	MHRDAT in SUPPMH
Breast Cancer Histor	Most recent progression/relapse date	RCULD in SUPPMH

Form Label	CRF Question Text	Page	UpperLeftX	UpperLeftY
Demographics	Form: Demographics	3	687.999	
Demographics	Date subject or legal guardian signed informed consent	3	687.999	
Demographics	Date of birth	3	687.999	
Demographics	Sex	3	687.999	
Demographics	Is the subject has the ability of pregnancy	260	25.999	
Demographics	Race	90	470.999	
Demographics	Other, please specify	3	160	303.999
Demographics	Ethnicity	3	90	268.999
Demographics	Height	3	90	181.999
Demographics	Form: Demographics	4	120	687.999
Demographics	Height Unit	4	125	650.999
Demographics	Is the subject rescreened?	4	165	611.999
Demographics	Previous Subject ID	4	175	556.999

Blank CRF

SDTM Warehouse (Central File)

SDTM aCRF

eCRF_V1.0_20210413: UNIQUE eCRF
Form: Demographics

Date subject or legal guardian signed informed consent

Date of birth

Sex
Female
Male

Is the subject has the ability of pregnancy
Yes
No

Race
American Indian or Alaskan Native
Asian
Black or African American
Native Hawaiian or Other Pacific Islander
White
Not reported
Unknown
Other

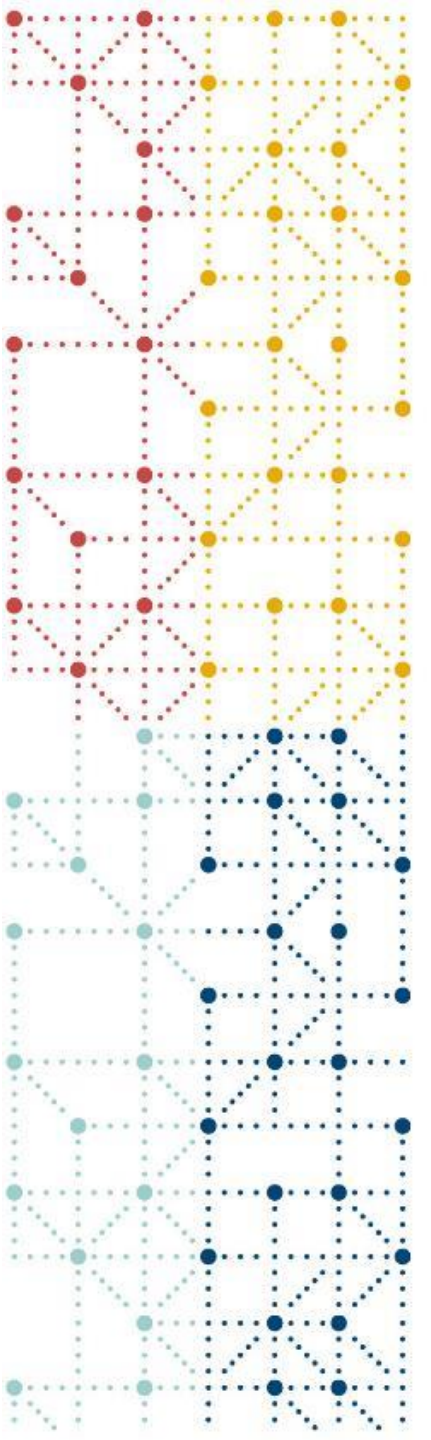
Other, please specify

Ethnicity
Hispanic or Latino
Not Hispanic or Latino
Not reported
Unknown

Height
Fixed Unit: cm

Automatic generation of SDTM aCRF

LET'S
DEMO!



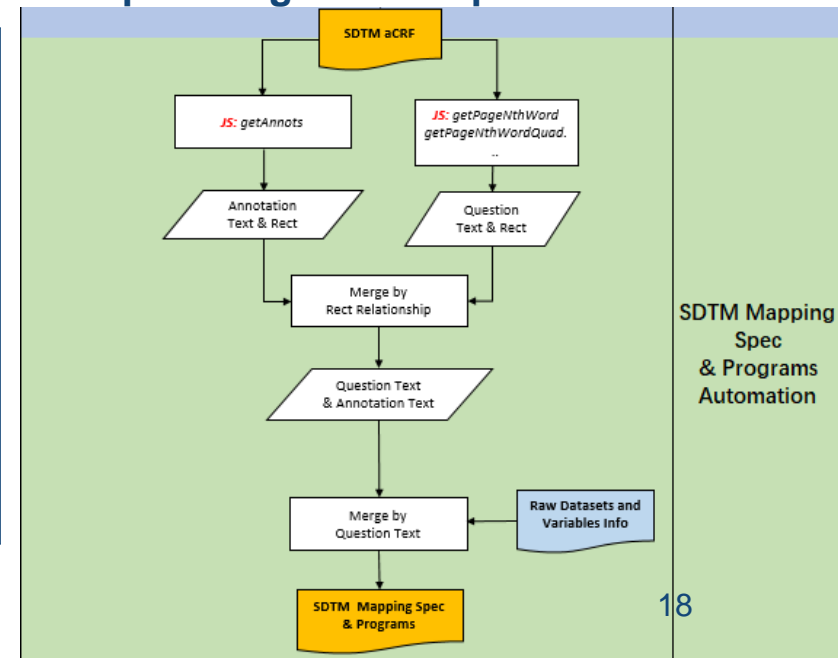
Automatic generation of SDTM Core Mapping Spec and Programs

Automatic generation of SDTM Core Mapping Spec and Programs

1. Obtaining **CRF question texts** and their **corresponding coordinates** <- JavaScript App Method
2. Obtaining **annotation texts** (restructured into columns of “Domain”, “Variable” and “Value”) and their **coordinates** <- JavaScript Doc Method
3. Correlating information retrieved from Step 1 and Step 2 based on the **relatedness of coordinates** <- JavaScript Merge
4. **Relating** CRF question texts to the **raw** datasets and variables <- JavaScript Merge
i.e. Obtaining relationship between SDTM Domain/Variables and Raw Datasets/Variables
5. Producing SDTM **Core** Mapping Specification and SDTM **Core** Programs <- VBA + Spec/Programs template

1	Page	Annotation	Domain	Variable	Value	SUPP?	LowerLeftX	LowerLeftY	UpperRightX	UpperRightY
10	3	DM = Demographics	DM		Demographics		90	707	217.5	725
11	3	DS = Disposition	DS		Disposition		222.5	707	342.5	725
12	3	VS = Vital Signs	VS		Vital Signs		347.5	707	467.5	725
13	3	DSCAT = "PROTOCOL MILESTONE"	DS	DSCAT	PROTOCOL MILESTONE		185	675.159973	381	690.159973
14	3	DSTERM = "INFORMED CONSENT OBTAINED"	DS	DSTERM	INFORMED CONSENT OBTAINED		365	638.835999	599.565979	653.159973
15	3	DSDECOD = "INFORMED CONSENT OBTAINED"	DS	DSDECOD	INFORMED CONSENT OBTAINED		365	621.817017	600.2210083	636.523987
16	3	DSSTDTC	DS	DSSTDTC			306.744995	622.041992	362.7449951	637.041992
17	3	RFICDTC	DM	RFICDTC			248.145004	622.005005	304.144989	637.005005
18	3	BRTHDTC	DM	BRTHDTC			160	603.159973	216	618.159973
19	3	SEX	DM	SEX			110	568.159973	137	583.159973
20	3	PREGYN in SUPPDM	DM	PREGYN		Y	310	513.159973	430	528.159973
21	3	RACE	DM	RACE			115	458.159973	151	473.160004
22	3	RACEOTH in SUPPDM	DM	RACEOTH		Y	90	458.159973	151	306.160004
23	3	ETHNIC	DM	ETHNIC			90	458.159973	188	271.160004
24	3	VSTESTCD = "HEIGHT"	VS	VSTESTCD	HEIGHT		90	458.159973	188	271.160004
25	3	VSTEST = "Height"	VS	VSTEST	Height		90	458.159973	188	167.660004
26	3	VSORRES	VS	VSORRES			90	458.159973	188	151.160004
27	3	VSORRESU	VS	VSORRESU			15581	182.461456		

NO SAS involved



SDTM Core Mapping Spec & Programs Automation

DM = Demographics DS = Disposition VS = Vital Signs

eCRF_V1.0_20210413: UNIQUE eCRF

Form: Demographics **DSCAT = "PROTOCOL MILESTONE"**

Date subject or legal guardian signed informed consent **DSTERM = "INFORMED CONSENT OBTAINED"**

RFICDTC DSSTDTC DSDECOD = "INFORMED CONSENT OBTAINED"

Date of birth **BRTHDTC**

Sex **SEX** Female Male

Is the subject has the ability of pregnancy **PREGYN in SUPPDM**

Race **RACE**
 American Indian or Alaskan Native
 Asian
 Black or African American
 Native Hawaiian or Other Pacific Islander
 White
 Not reported
 Unknown
 Other

Other, please specify **RACEOTH in SUPPDM**

Ethnicity **ETHNIC**
 Hispanic or Latino
 Not Hispanic or Latino
 Not reported
 Unknown

Height **VSTESTCD = "HEIGHT"** **VSORRESU** Fixed Unit: cm
VSTEST = "Height"
VSORRES

	A	B	O	P
	FormOID	FieldOID	PreText	FixedUnit
9	DM	ICFDAT	Date subject or legal guardian signed informed consent	
10	DM	BRTHDAT	Date of birth	
	DM	SEX	Sex	
11	DM	PREGYN	Is the subject has the ability of pregnancy	
12	DM	RACE	Race	
13	DM	RACEOTH	Other, please specify	
14	DM	ETHNIC	Ethnicity	
15	DM	HEIGHT	Height	cm
16	DM	HEIGHTU	Height Unit	
17	DM	RERANGD	Is the subject rescreened?	
18	DM	PSUBJID	Previous Subject ID	
19	DM			

RAW Information

Link

Page	Annotation	Domain	Variable	Value	SUPP?	lowerLeftX	lowerLeftY	upperRightX	upperRightY
3	DM = Demographics	DM	Demographics			90	707	317.5	725
3	DS = Disposition	DS	Disposition			222.5	707	342.5	725
3	VS = Vital Signs	VS	Vital Signs			247.5	707	467.5	725
3	DSCAT = "PROTOCOL MILESTONE"	DS	DSCAT	PROTOCOL MILESTONE		185	675.15997	381	690.1599731
3	DSTERM = "INFORMED CONSENT OBTAINED"	DS	DSTERM	INFORMED CONSENT OBTAINED		365	638.836	599.565979	653.1599731
3	DSDECOD = "INFORMED CONSENT OBTAINED"	DS	DSDECOD	INFORMED CONSENT OBTAINED		365	621.81702	600.221008	636.523986
3	DSSTDTC	DS	DSSTDTC			306.745	622.04199	362.744995	637.0419922
3	RFICDTC	DM	RFICDTC			248.145	622.005	304.144989	637.0050049
3	BRTHDTC	DM	BRTHDTC			160	603.15997	216	618.1599731
3	SEX	DM	SEX			110	568.15997	137	583.1599731
3	PREGYN in SUPPDM	DM	PREGYN		Y	310	513.15997	430	528.1599731
3	RACE	DM	RACE			115	458.16	151	473.1600037
3	RACEOTH in SUPPDM	DM	RACEOTH		Y	200	291.16	327.5	306.1600037
3	ETHNIC	DM	ETHNIC			140	256.16	188	271.1600037
3	VSTESTCD = "HEIGHT"	VS	VSTESTCD	HEIGHT		125	169.16	267.5	184.1600037
3	VSTEST = "Height"	VS	VSTEST	Height		125	152.66	252.5	167.6600037
3	VSORRES	VS	VSORRESU			125	136.16	181	151.1600037
3	VSORRESU	VS	VSORRESU			380.98456	167.46146	444.984558	182.4614563

SDTM Annotation Information

Form Label	CRF Question Text	Page	UpperLeftX	UpperLeftY	UpperRightX	UpperRightY
Demographics	Form: Demographics	3	120	687.9998779	180	687.9
Demographics	Date subject or legal guardian signed informed consent	3	320	650.9998779	360	650.9
Demographics	Date of birth	3	130	615.9998779	155	615.9
Demographics	Sex	3	90	580.9998779	105	580.9
Demographics	Is the subject has the ability of pregnancy	3	260	525.9998779	305	525.9
Demographics	Race	3	90	470.9999084	110	470.9
Demographics	Ethnicity	3	160	303.9999084	195	303.9
Demographics	Height	3	90	268.9999084	135	268.9
Demographics	Height Unit	3	90	181.9999084	120	181.9
Demographics	Form: Demographics	4	120	687.9998779	180	687.9
Demographics	Height Unit	4	125	650.9998779	145	650.9
Demographics	Is the subject has the ability of pregnancy	4	165	611.9998779	220	611.9
Demographics	Previous Subject ID	4	175	556.9998779	185	556.9

CRF Question Information

Link by relatedness of coordinates

Automatic generation of SDTM Core Mapping Spec and Programs

	A	B	C	D	E	F	G	H	I	J
1	Variable	Variable Label	Key	Type	Length	Format	Controlled	Origin	CRF Page	Study Derivation
2	STUDYID	Study Identifier	1	Char	200			Protocol		
3	DOMAIN	Domain Abbreviation		Char	2	DS		Assigned		DS
4	USUBJID	Unique Subject Identifier	2	Char	200			Derived		concatenation of STUDYID, STUDYENVSITENUMBER, SUBJECT separated by "-".
5	DSSEQ	Sequence Number		Num	8			Derived		Sequential number ensuring uniqueness of records within each USUBJID in the domain sorted by Key variables
6	DSSPID	Sponsor-Defined Identifier		Char	200			Assigned		concatenation of Form Name(lowercase) and RecordId by "-". E.g. "xx-3870634"
7	DSTERM	Reported Term for the Disposition Event		Char	200			CRF	CRF Page 3 72 111	RAW.DM: if ICFDAT is not missing then set to 'INFORMED CONSENT OBTAINED' RAW.EOT.EOTREAS (set to EOTPISP when value is 'Other') RAW.DS.DSPISP (set to DSREAS when value is missing) ...
8	DSDECOD	Standardized Disposition Term	3	Char	200			CRF	CRF Page 3 72 111	RAW.DM: if ICFDAT is not missing then set to 'INFORMED CONSENT OBTAINED' RAW.EOT.EOTREAS RAW.DS.DSREAS
9	DSCAT	Category for Disposition Event		Char	200		DSCAT	CRF	CRF Page 3 72 111	RAW.DM: set to 'PROTOCOL MILESTONE' RAW.EOT: set to 'DISPOSITION EVENT' RAW.DS: set to 'DISPOSITION EVENT' ...
10	DSSCAT	Subcategory for Disposition Event		Char	200		*	CRF	CRF Page 72 111	RAW.EOT: set to 'END OF TREATMENT' RAW.DS: set to 'END OF STUDY' ...
11	EPOCH	Epoch		Char	200		EPOCH	Derived		Based on DSSTDTC, merge with SE domain.
12	DSSTDTC	Start Date/Time of Disposition Event	4	Char	19	ISO 8601		CRF	CRF Page 3 72 111	RAW.DM.ICFDAT RAW.EOT.EOTDAT RAW.DS.CW DAT when DSSCAT = 'END OF STUDY' and DSDECOD = 'WITHDRAWAL BY SUBJECT' RAW.DS.LOSDAT when DSSCAT = 'END OF STUDY' and DSDECOD = 'LOST TO FOLLOW-UP' RAW.DS.DTHDAT when DSSCAT = 'END OF STUDY' and DSDECOD = 'DEATH' ...
13	DSSTDY	Study Day of Start of Disposition Event		Num	8			Derived		=DSSTDTC - DM.RFSTDTC + 1 if DSSTDTC>=DM.RFSTDTC; =DSSTDTC - DM.RFSTDTC if DSSTDTC<DM.RFSTDTC.
14	DTHREAS	Death Reason		Char	200			CRF	CRF Page 111	RAW.DS.DTHREAS
15	DTHOTH	Other		Char	200			CRF	CRF Page 111	RAW.DS.DTHOTH

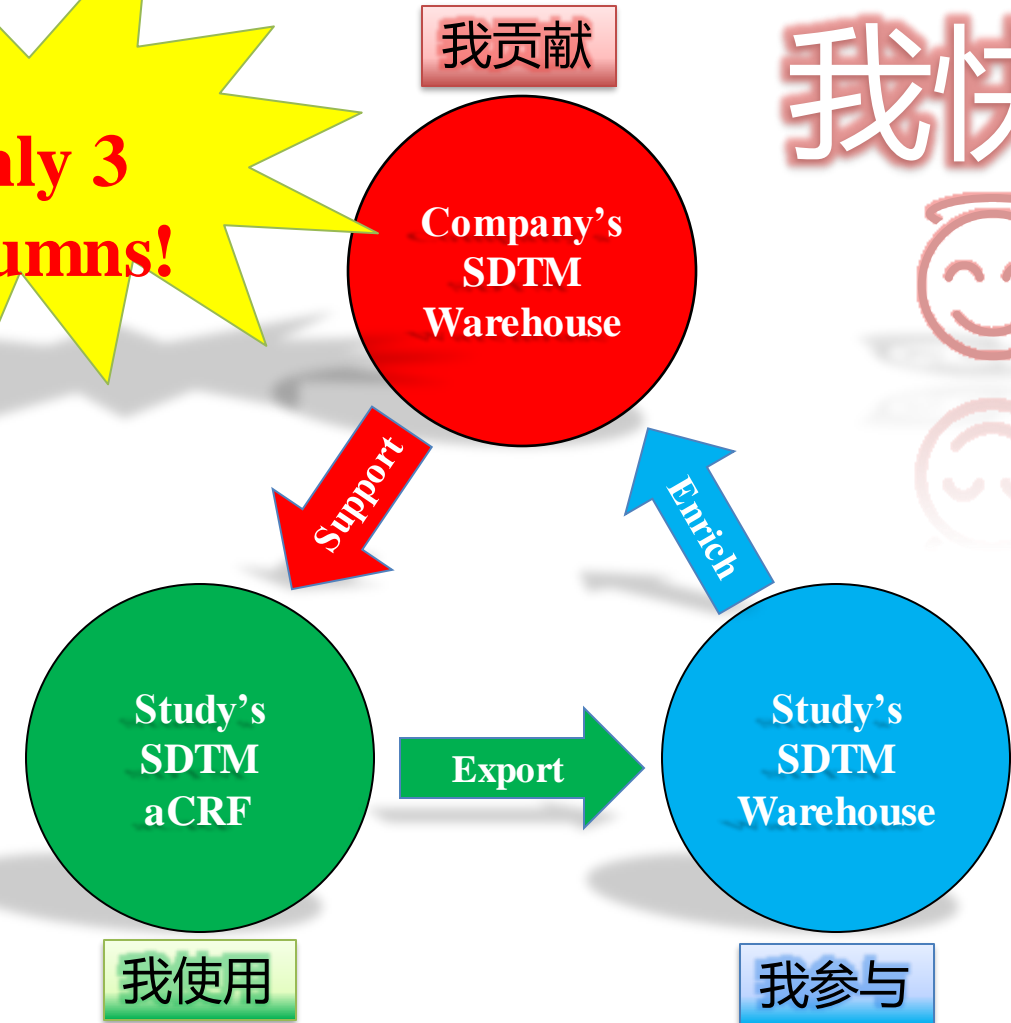
名称	修改日期	大小
ae.sas	2020/5/26 10:59	2 KB
cm.sas	2020/5/26 10:59	3 KB
cv.sas	2020/5/26 10:59	1 KB
da.sas	2020/5/26 10:59	1 KB
dm.sas	2020/5/26 10:59	1 KB
ds.sas	2020/5/26 10:59	5 KB
eg.sas	2020/5/26 10:59	1 KB
ex.sas	2020/5/26 10:59	6 KB
ie.sas	2020/5/26 10:59	1 KB
lb.sas	2020/5/26 10:59	6 KB
mb.sas	2020/5/26 10:59	1 KB
mh.sas	2020/5/26 10:59	2 KB
mi.sas	2020/5/26 10:59	3 KB
pc.sas	2020/5/26 10:59	1 KB
pe.sas	2020/5/26 10:59	1 KB
pr.sas	2020/5/26 10:59	4 KB
qs.sas	2020/5/26 10:59	1 KB
re.sas	2020/5/26 10:59	1 KB
rs.sas	2020/5/26 10:59	2 KB
ss.sas	2020/5/26 10:59	1 KB
su.sas	2020/5/26 10:59	1 KB
sv.sas	2020/5/26 10:59	1 KB
tr.sas	2020/5/26 10:59	4 KB
tu.sas	2020/5/26 10:59	3 KB
vs.sas	2020/5/26 10:59	2 KB
xr.sas	2020/5/26 10:59	1 KB
xu.sas	2020/5/26 10:59	1 KB

Note: The relationship between the SDTM variables and raw dataset variables is an essential input to generate SDTM core mapping specification and SDTM program, with the former document displaying “per SDTM domain per SDTM variable” and the latter one displaying “per SDTM domain per raw dataset per SDTM variable”.

Central File: SDTM Mapping Warehouse

Form Label	CRF Question Text	Annotation Text
Subject	Subject	SUBJID
Visit Date	Was the visit done?	[NOT SUBMITTED]
Visit Date	If not done, provide the reason	[NOT SUBMITTED]
Visit Date	Visit date	SVSTDTC
Visit Date	Visit date	SVENDTC
Demographics	Form: Demographics	DSCAT = "PROTOCOL MILEST"
Demographics	Date subject or legal guardian signed informed consent	DSTERM = "INFORMED"
Demographics	Date subject or legal guardian signed informed consent	DSDECOD = "INFORMED"
Demographics	Date subject or legal guardian signed informed consent	DSSTDTC
Demographics	Date subject or legal guardian signed informed consent	RFICDTC
Demographics	Date of birth	BRTHDTC
Demographics	Sex	SEX
Demographics	Race	RACE
Demographics	Other, please specify	RACEOTH in SUPPDM
Demographics	Ethnicity	ETHNIC
Demographics	Height	VSTESTCD = "HEIGHT"
Demographics	Height	VSTEST = "Height"
Demographics	Height	VSORRES
Demographics	Height Unit	VSORRESU
Demographics	Is the subject has the ability of pregnancy	PREGYN in SUPPDM
Demographics	Is the subject rescreened?	SUBSCRYN in SUPPDM
Demographics	Previous Subject ID	SUBJIDP in SUPPDM
Breast Cancer History	Form: Breast Cancer History	MHCAT = "Breast Cancer History"
Breast Cancer History	Form: Breast Cancer History	MHTERM="Breast Cancer"
Breast Cancer History	Date of initial diagnosis	MHSTDTC
Breast Cancer History	Histological grade	HISTGR in SUPPMH
Breast Cancer History	Primary site	MHILOC in SUPPMH
Breast Cancer History	TNM staging (T)	CAPRM in SUPPMH
Breast Cancer History	TNM staging (N)	RLYMPH in SUPPMH
Breast Cancer History	TNM staging (M)	METDIS in SUPPMH
Breast Cancer History	Clinical staging at initial diagnosis	IDXSTGL in SUPPMH
Breast Cancer History	Clinical staging at screening	IDXSTG in SUPPMH
Breast Cancer History	Date of recurrence or metastasis	MHRDAT in SUPPMH
Breast Cancer History	Most recent progression/relapse date	RCULD in SUPPMH

Only 3 Columns!

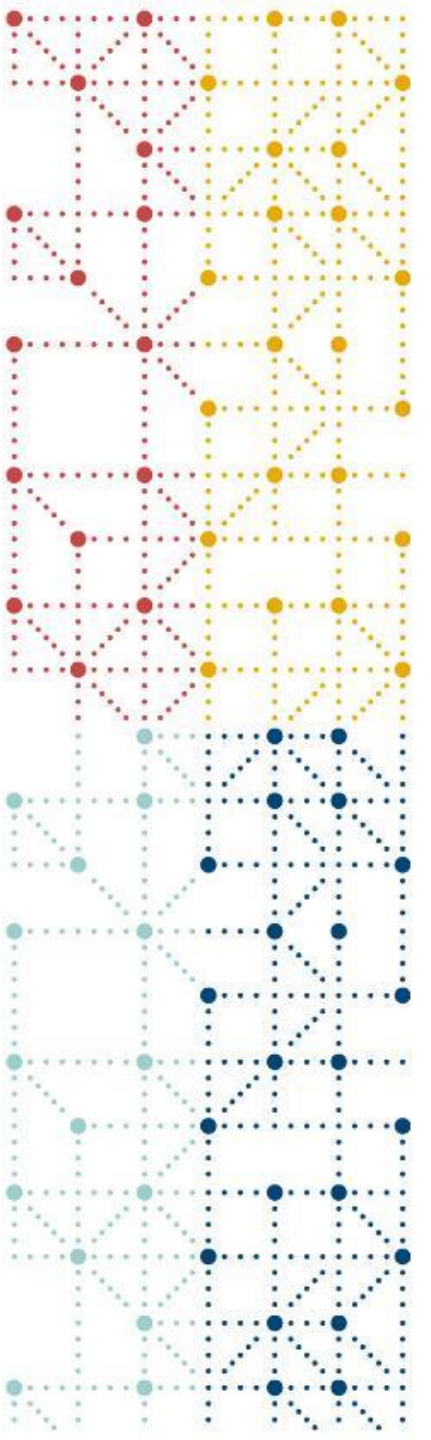


我快乐



Automatic generation of SDTM Core Mapping Spec and Programs

No Time
for
DEMO!



Summary

Summary

➤ The essence of SDTM work is information exchange

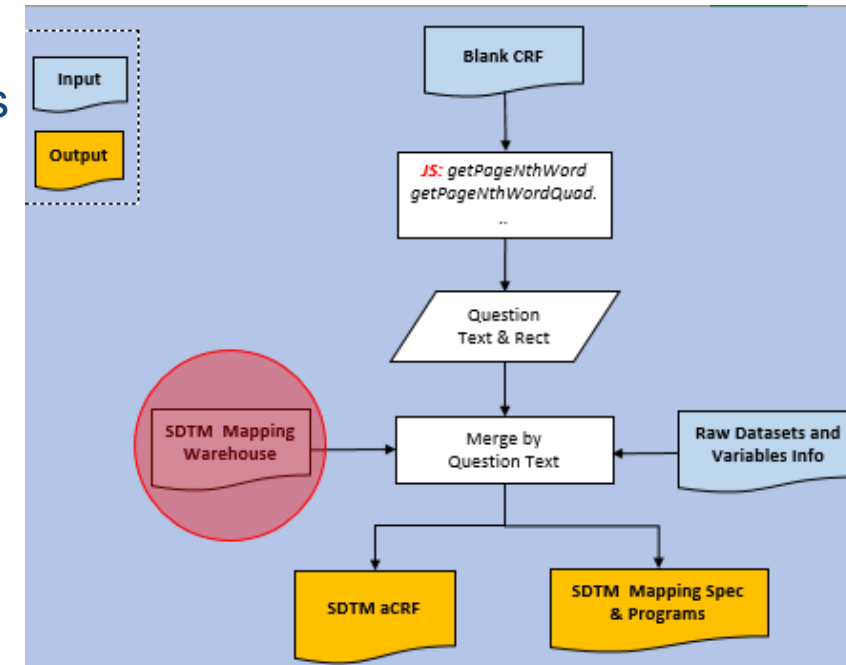
➤ CRF Question Text-Driven Solution covers:

1. Automatic Generation of SDTM aCRF - Blank CRF driven
2. Automatic Generation of SDTM Core Mapping Specification - SDTM aCRF driven
3. Automatic Generation of SDTM Core Programs - SDTM aCRF driven

Ideal Scenario: One-step automation driven by blank CRF for generating SDTM aCRF, Core Mapping Specification and Programs

➤ Hub: SDTM Mapping Warehouse

How efficient? A few minutes to process.



Steps of SDTM Work and CRF Question Text-Driven Solution

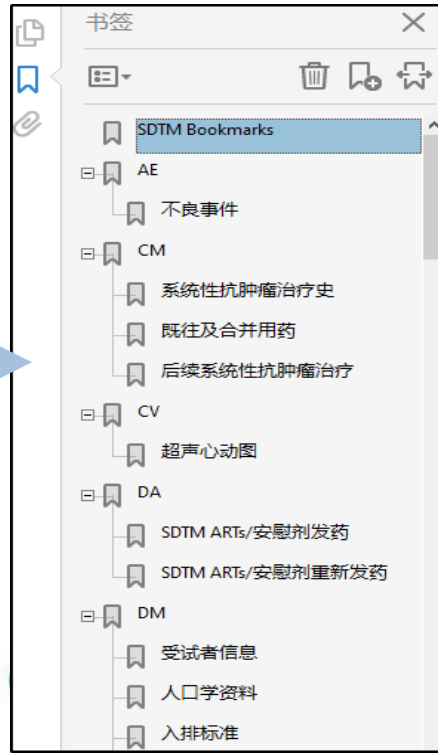
Task [Step] (Percentage of time spent in this task)	Difficulties & Challenges	Degree of Automation	CRF Question Text-Driven Solution (Percentage of time saved)
1. SDTM aCRF (10 ~15%) - Core	<ul style="list-style-type: none"> Ways to annotate specific information Performing repetitive work when starting new projects Consistent data processing among all projects 	Automation is possible throughout the step.	JavaScript Mechanism: By reading in information from aCRF.pdf and displaying dataflow from raw SDTM database as SAS code following SAS program template (75%+)
2. SDTM Mapping Specification (40 ~45%)	<ul style="list-style-type: none"> Looking for specific parts of SDTM obtained from aCRF.pdf to Excel Note: Most mapping rules are obtained from aCRF.pdf to Excel	Some parts of SDTM obtained from aCRF.pdf to Excel E.g. EPOCH, VISITNUM, etc.	JavaScript + VBA Mechanism: By reading in information from aCRF.pdf and displaying dataflow from raw SDTM database as SAS code following SAS program template (60%+)
3. SDTM Programs (40 ~50%)	<ul style="list-style-type: none"> Raw datasets + SDTM domains + variables input Note: This step is to translate mapping rules into SAS Code.	Some basic SAS codes and some default derived codes such as --Y, --EPOCH, VISITNUM, can be automatically generated.	JavaScript + VBA Mechanism: By reading in information from aCRF.pdf and displaying dataflow from raw SDTM database as SAS code following SAS program template (60%+)

Can save around
55%~75% >=
1 man-month /
project

Time saving,
Effort saving,
Stress-free

More solution by using CRF Question Text-Driven idea

- Annotating Raw dataset names and variable names on SDTM aCRF (helpful for Mapping Spec and Programs preparation and review)
- Annotating raw data collected by CRF
- Adding bookmarks for SDTM aCRF
- Get CRF Page number of all variables for define.xml



	A	B	C	D
1	Domain	Variable	supp?	Origin
75	DM	AGE		CRF Page 4
76	DM	AGEU		CRF Page 4
77	DM	BRTHDTC		CRF Page 4
78	DM	DM		CRF Page 2 4 40 69
79	DM	DTHDTC		CRF Page 69
80	DM	ETHNIC		CRF Page 4
81	DM	ETHNICOT	Y	CRF Page 4
82	DM	INVID		CRF Page 2
83	DM	INVNAM		CRF Page 2
84	DM	PROVER	Y	CRF Page 40
85	DM	RERAND	Y	CRF Page 4
86	DM	RESUBJID	Y	CRF Page 4
87	DM	RFICDTC		CRF Page 4
88	DM	SEX		CRF Page 4
89	DM	SITEID		CRF Page 2
90	DM	SUBJID		CRF Page 2
91	DM	SUBJINIT	Y	CRF Page 2
92	DS	DS		CRF Page 4 41 61 62 63 64 65 66 69
93	DS	DSCAT		CRF Page 4 41 61 62 63 64 65 66 69
94	DS	DSDECOD		CRF Page 4 41 61 62 63 64 65 66 69
95	DS	DSDTC		CRF Page 65 66
96	DS	DSSAT		CRF Page 4
97	DS	DSSCAT		CRF Page 41 61 62 63 64 65 66 69
98	DS	DSSPID		CRF Page 65 66
99	DS	DSSTDTC		CRF Page 4 41 61 62 63 64 69
100	DS	DSTERM		CRF Page 4 41 61 62 63 64 65 66 69
101	DS	DTHOTH	Y	CRF Page 69
102	DS	DTHREAS	Y	CRF Page 69
103	DS	EOTDTC	Y	CRF Page 62 63 64

Subject= 07001; Visit= 筛选期

CV = Cardiovascular System Findings

RAW = UCG

Unique blank eCRF_ 版本日期: 20200528

Form: 超声心动图

超声心动图 CVTEST = "Left Ventricular Ejection" [null] CVSTAT 未做 UCGPERF

CVTESTCD = "LVEF"

检查日期 CVDTC 2020/7/20 UCGDAT

LVEF CVORRES 68 Fixed Unit: % UCGORRES

单位 CVORRESU % UCGORESUS

临床意义 CVCLSIG in SUPPCV 异常, 无临床意义 正常 UCGCLSIG

异常描述 CVDESC in SUPPCV 左室舒张功能减退 UCGDESC

More solution by using CRF Question Text-Driven idea

- Get NOT SUBMITTED pages and Fields for cSDRG Section 3.3

3.3 Annotated CRFs

Collected fields that have not been tabulated have been annotated as “Not Submitted”. LDCP Inc. collects certain data elements to facilitate operational processes including data cleaning and dynamically creating additional forms in the electronic data capture system. All fields that have been annotated as “Not Submitted” meet this criterion.

Explanation of data fields [Not Submitted]

aCRF page Number(s)	Data Collection Field	Explanation of why [NOT SUBMITTED]
5	Were there any product complaints?	For internal use only.
30	PI Signature Date	Not needed for analysis.



With CRF Question Text-Driven Solution, I think you will

NOT FEAR (不害怕) PREPARING SDTM PACKAGE

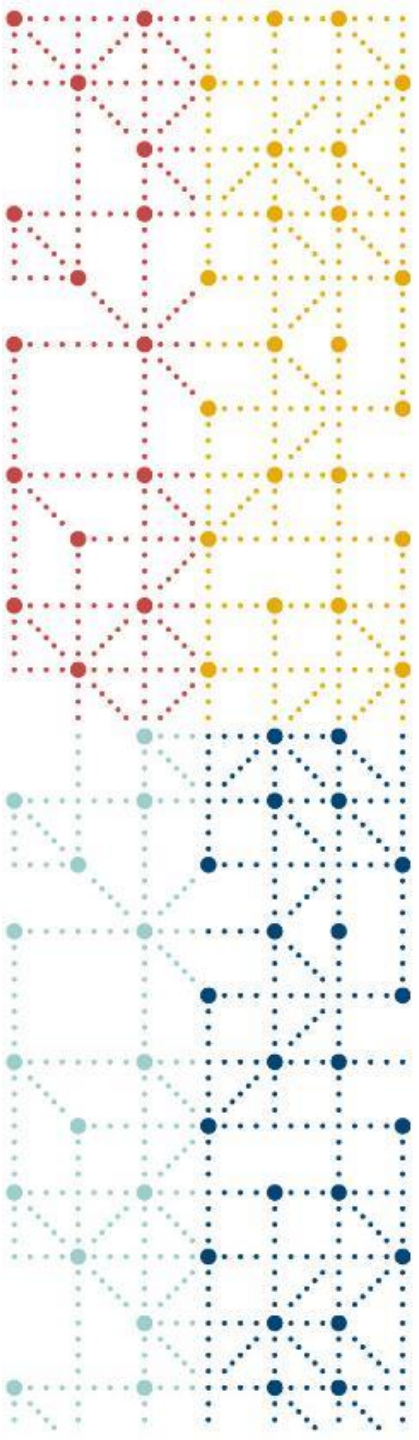
NOT REJECT (不排斥) PREPARING SDTM PACKAGE

FALL FOR (喜欢上) PREPARING SDTM PACKAGE



Reference

- **PharmaSUG China 2021:** SDTM ARTs - a CRF-Driven SDTM Automation Tool for Generating aCRF, SDTM Mapping Specification and Programs. Haiqiang Luo



Thank You!

Contact:

Email: haiqiang911@126.com

WeChat: haiqiang0911

