

SDTM ARTs - a CRF-Driven SDTM Automation Tool for Generating aCRF, SDTM Mapping Specification and Programs

Haiqiang Luo, Hengrui Pharma, Guangzhou, China
Kaiying Li, Hengrui Pharma, Guangzhou, China
Xiaoyi Niu, Hengrui Pharma, Guangzhou, China
Xiaoxia Zhou, Hengrui Pharma, Guangzhou, China

ABSTRACT

Study Data Tabulation Model (SDTM) dataset package is one of the New Drug Application (NDA) Submission components required by U.S. Food and Drug Administration (FDA) and Pharmaceuticals and Medical Devices Agency (PMDA) in Japan. Its preparation process mainly includes annotating Case Report Form (CRF), writing SDTM mapping specifications, SAS programming, and generating SDTM datasets, which is time-consuming, laborious, and error-prone. Albeit the preparation work across different projects is highly repetitive, maintaining the same data mapping criteria is challenging. The timely preparation of SDTM datasets inevitably affects progress of analysis work. Our purpose of this paper is to introduce an innovative tool named the SDTM Automation Relevant Toolkits (ARTs) and its CRF-driven mechanism regarding SDTM annotated CRF (aCRF) preparation procedures. Developed through JavaScript, the ARTs reads in information from CRF, completes annotations, generates core SDTM mapping specifications and SDTM programs via Visual Basic for Applications (VBA) with no use of SAS. The idea behind this effective tool is readily accessible and is not hard to be implemented.

INTRODUCTION

While conducting analyses in clinical trials, SDTM dataset package preparation is acting as a pivotal and prerequisite step, hinge on which derivation of Analysis Data Model (ADaM) and Tables, Figures, and Listings (TFL) can be carried out smoothly. Thus, automation of SDTM dataset preparation has always been the pursued goal in the pharmaceutical industry. Nevertheless, most programmers put more emphasis on aCRF by following a semi-automatic routine: they export the annotation information into xfdf or fdf files from prepared aCRF of other pervious project and re-import them after manual editing. This, unfortunately, will introduce unsolved issues. For example, vague page number and coordinate information may cause imprecise positioning of annotations, redundant annotation notes, etc., leading to intricate manual adjustments. Moreover, some programmers might even put the cart before the horse. They create aCRF based on prepared SDTM mapping specifications, which does not conform to the conventional workflow.

This paper will introduce SDTM ARTs, an innovative tool that provides a CRF-driven solution to automatically generate aCRF. Using the tool requires little knowledge on annotation position parameters (e.g., page numbers and coordinates) and involves no manual adjustments. The procedure does not adhere to the semi-automatic routine mentioned above, instead, it fulfills the automation process centered on blank CRF question text as per user-customized annotation style to produce aCRF. In the meanwhile, automatic creation of core SDTM mapping specifications and SDTM programs is also feasible to some degree. Detailed concept and steps will be discussed in the corresponding sections; sample code and user-interface example are attached as Appendix I & II, respectively. Although beyond the scope of this paper, the tool is far more powerful than what we have discovered and always worth discussion.

WHY CRF DRIVEN?

The information collected by CRF is the most fundamental data source for SDTM and will be stored as raw database. Based on Clinical Data Interchange Standards Consortium (CDISC) standard, programmers re-organize and map these raw data to create SDTM domains and variables. Therefore, regardless of format, raw database and SDTM database are both CRF information carriers. Moving on to data processing, SDTM mapping specification and SDTM

program are acting as a bridge to help convert raw data into SDTM data through arranging and translating the CDISC mapping rules displayed on aCRF into domains. In other words, the essence of SDTM work is information exchange.

Now let's go further in depth with a real-world example. The bulk of participants' information, which is also the core of clinical studies, originates from CRF. The ultimate purpose of SDTM work is to map that information onto the SDTM database in an accurate manner. As shown in Figure 1 (demographic page), to realize automatic generation of SDTM files, targeting on CRF will be the most effective solution because raw database and its derived SDTM mapping specification and SDTM programs are not required in NDA submission and should not be the main focus.

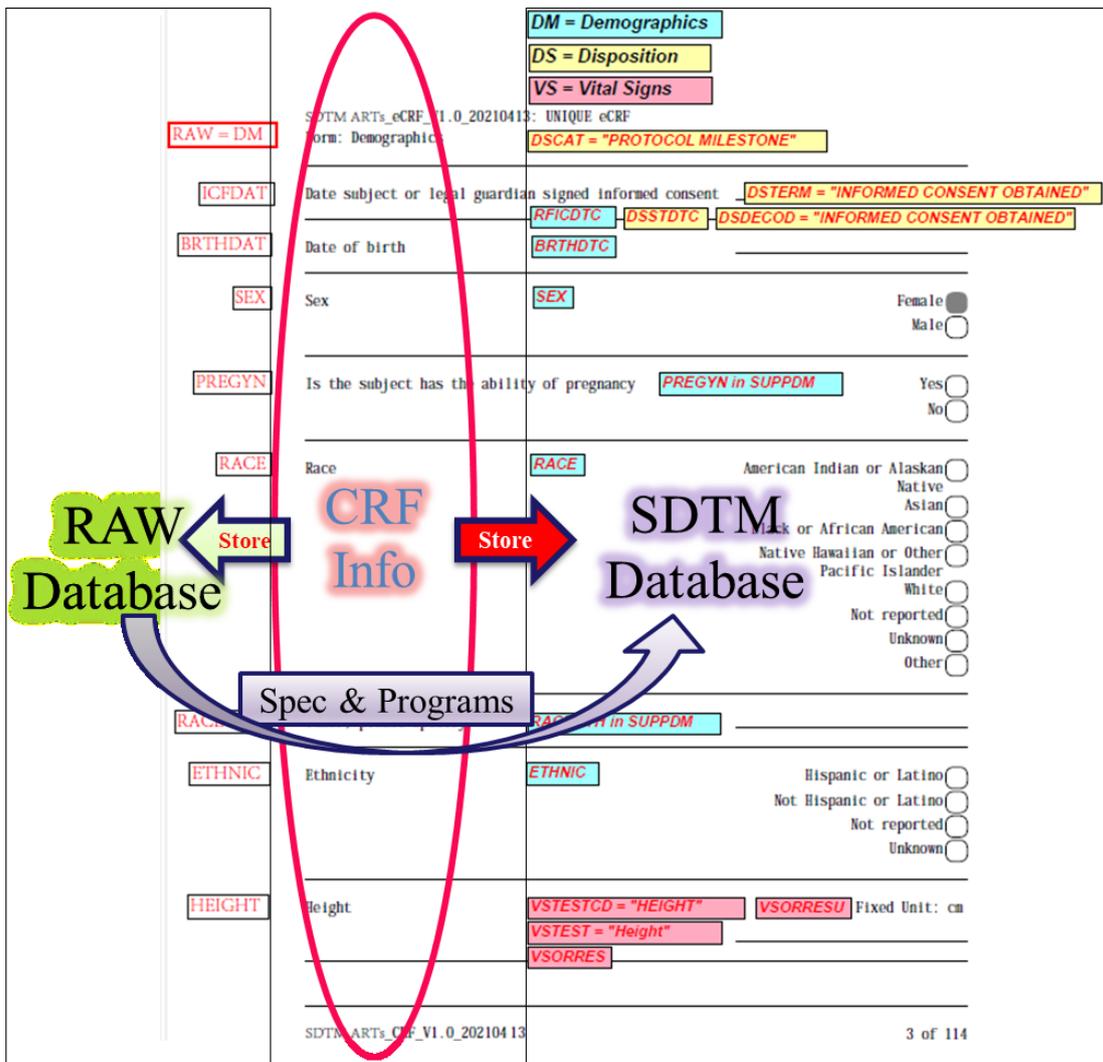


Figure 1: The essence of SDTM work is information exchange

A REAL CRF DRIVEN AUTOMATION TOOL: SDTM ARTs

STEPS OF SDTM WORK AND SDTM ARTS SOLUTIONS

The following table summarizes the tasks for preparing SDTM datasets, along with the difficulties and challenges one may encounter during each step, the corresponding degree of automation for the steps, and CRF-driven automation solutions proposed by the ARTs.

Task [Step] (Percentage of time spent in this task)	Difficulties & Challenges	Degree of Automation	SDTM ARTs Solution (Percentage of time saved)
1. SDTM aCRF (10 ~15%)	<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%+)

HOW SDTM ARTs WORKS - SDTM aCRF AUTOMATION

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

Above all, the indispensable elements in automation are page numbers, annotation box coordinates, and annotation text. We can keep other setting to default based on annotation text.

TWO FUNCTIONS TO EXTRACT CRF QUESTION TEXTS, PAGE NUMBERS AND COORDINATES IN JavaScript: getPageNthWord AND getPageNthWordQuad

By using the functions of getPageNthWord and getPageNthWordQuad in JavaScript (Appendix I), we can obtain each separate word and its coordinates from CRF pages (Figure 2). These words will be combined back into original question texts (Figure 3) based on the relatedness of coordinates. After grouping the words, we have the information to precisely locate the SDTM annotation box.

1	Word	UpperLeftX	UpperLeftY	UpperRightX	UpperRightY	LowerLeftX	LowerLeftY	LowerRightX	LowerRightY
2	Form:	90.00001526	687.9998779	115.0000153	687.9998779	90.00001526	676.6600342	115.0000153	676.6600342
3	Demographics	120.0000153	687.9998779	180.0000153	687.9998779	120.0000153	676.6600342	180.0000153	676.6600342
4	Date	90	650.9998779	110	650.9998779	90	639.6600342	110	639.6600342
5	subject	115	650.9998779	150	650.9998779	115	639.6600342	150	639.6600342
6	or	155	650.9998779	165	650.9998779	155	639.6600342	165	639.6600342
7	legal	170	650.9998779	195	650.9998779	170	639.6600342	195	639.6600342
8	guardian	200	650.9998779	240	650.9998779	200	639.6600342	240	639.6600342
9	signed	245	650.9998779	275	650.9998779	245	639.6600342	275	639.6600342
10	informed	280	650.9998779	320	650.9998779	280	639.6600342	320	639.6600342
11	consent	325	650.9998779	360	650.9998779	325	639.6600342	360	639.6600342
12	Date	90	615.9998779	110	615.9998779	90	604.6600342	110	604.6600342
13	of	115	615.9998779	125	615.9998779	115	604.6600342	125	604.6600342
14	birth	130	615.9998779	155	615.9998779	130	604.6600342	155	604.6600342
15	Sex	90	580.9998779	105	580.9998779	90	569.6600342	105	569.6600342
16	is	90	525.9998779	100	525.9998779	90	514.6600342	100	514.6600342
17	the	105	525.9998779	120	525.9998779	105	514.6600342	120	514.6600342
18	subject	125	525.9998779	160	525.9998779	125	514.6600342	160	514.6600342
19	has	165	525.9998779	180	525.9998779	165	514.6600342	180	514.6600342
20	the	185	525.9998779	200	525.9998779	185	514.6600342	200	514.6600342
21	ability	205	525.9998779	240	525.9998779	205	514.6600342	240	514.6600342
22	of	245	525.9998779	255	525.9998779	245	514.6600342	255	514.6600342
23	pregnancy	260	525.9998779	305	525.9998779	260	514.6600342	305	514.6600342
24	Race	90	470.9999084	110	470.9999084	90	459.6600342	110	459.6600342
25	Other	90	303.9999084	120	303.9999084	90	292.6600342	120	292.6600342
26	please	125	303.9999084	155	303.9999084	125	292.6600342	155	292.6600342
27	specify	160	303.9999084	195	303.9999084	160	292.6600342	195	292.6600342
28	Ethnicity	90	268.9999084	135	268.9999084	90	257.6600342	135	257.6600342
29	Height	90	181.9999084	120	181.9999084	90	170.6600342	120	170.6600342

Figure 2: Extracted words from CRF questions and their corresponding coordinates

1	Form Label	CRF Question Text	Page	UpperLeftX	UpperLeftY	UpperRightX	UpperRightY	LowerLeftX	LowerLeftY	LowerRightX	LowerRightY
2	Demographics	Form_Demographics	3	120	687.9998779	180	687.9998779	120	676.6600342	180	676.6600342
3	Demographics	Date subject or legal guardian signed informed consent	3	325	650.9998779	360	650.9998779	325	639.6600342	360	639.6600342
4	Demographics	Date of birth	3	130	615.9998779	155	615.9998779	130	604.6600342	155	604.6600342
5	Demographics	Sex	3	90	580.9998779	105	580.9998779	90	569.6600342	105	569.6600342
6	Demographics	Is the subject has the ability of pregnancy	3	260	525.9998779	305	525.9998779	260	514.6600342	305	514.6600342
7	Demographics	Race	3	90	470.9999084	110	470.9999084	90	459.6600342	110	459.6600342
8	Demographics	Other, please specify	3	160	303.9999084	195	303.9999084	160	292.6600342	195	292.6600342
9	Demographics	Ethnicity	3	90	268.9999084	135	268.9999084	90	257.6600342	135	257.6600342
10	Demographics	Height	3	90	181.9999084	120	181.9999084	90	170.6600342	120	170.6600342
11	Demographics	Form_Demographics	4	120	687.9998779	180	687.9998779	120	676.6600342	180	676.6600342
12	Demographics	Height Unit	4	125	650.9998779	145	650.9998779	125	639.6600342	145	639.6600342
13	Demographics	Is the subject rescreened?	4	165	611.9998779	220	611.9998779	165	600.6600342	220	600.6600342
14	Demographics	Previous Subject ID	4	175	556.9998779	185	556.9998779	175	545.6600342	185	545.6600342

Figure 3: Concatenated CRF question texts and their page numbers and coordinates after processing based on relatedness of coordinates

MAPPING CRF QUESTIONS TO SDTM ANNOTATION TEXTS USING SDTM MAPPING WAREHOUSE

The ARTs we built embedded a central file called SDTM Mapping Warehouse which maps extracted CRF question texts to SDTM annotation texts for each specific study. As shown in Figure 4, the Warehouse contains only 3 columns: form label, CRF question text and annotation text.

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
Breast Cancer History	Histological grade	HISTGR in SUPPMH

Figure 4: Example of SDTM Mapping Warehouse

Note: SDTM Mapping Warehouse is a central file, and it is not study-specific. It could be established based on the company's standard CRF library or accumulated meta-database. The Warehouse can be designed as company level or therapeutic area level based on needs.

Annotation Process: JavaScript Doc Method addAnnot

Once we successfully extracted CRF question texts and their coordinates and prepared SDTM Mapping Warehouse which holds information of annotation texts, automation of annotating CRF found upon SDTM standard becomes possible.

Recall that providing page numbers and text coordinates to the ARTs are not necessary, we are actually referring to an automatic read-in and output process conducted by the tool using blank CRF and information from SDTM Mapping Warehouse. Users can also customize properties of the annotation text and box on the settings menu to receive their preferred output format (Appendix II).

It is worth mentioning that as a fixed central file, "Annotation Text" in the third column, with contents such as QNAM in SUPPXX, is hard-coded. One common option programmers may choose to alter the contents in that column is that when using the ARTs on specific studies, they can utilize the raw dataset definition file (e.g. ALS in Medidata Rave Study) as input to replace those hard-coded texts. Introducing raw dataset definition file is an essential step in automatic generation of core mapping specification and programs, and it will be covered in detail later.

Below are the flowchart of automation for generating SDTM aCRF and an aCRF example (Figure 5 and Figure 6).

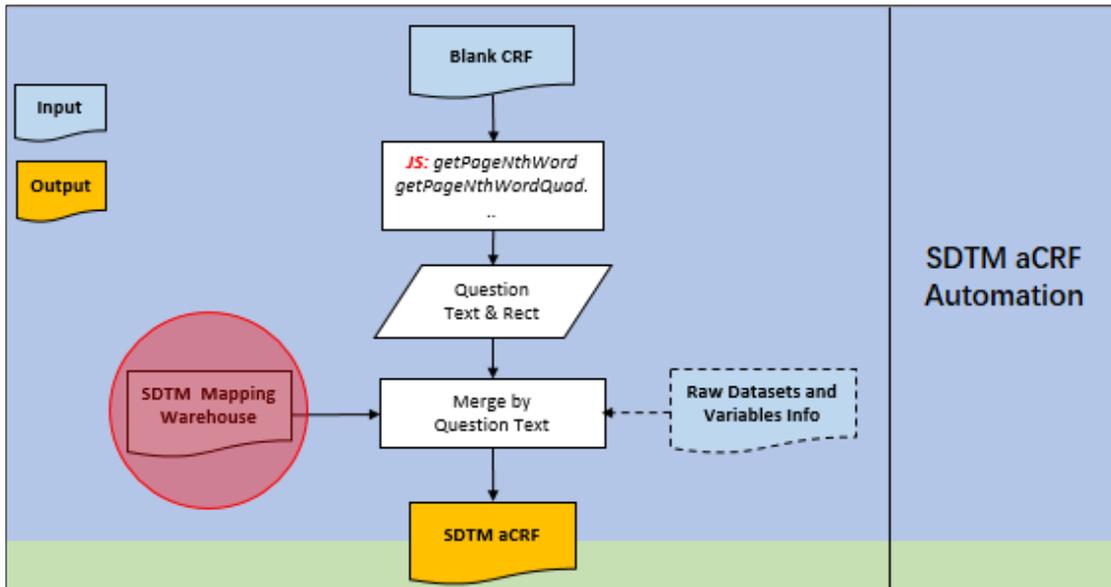


Figure 5: SDTM aCRF Automation Flowchart

Form Label	CRF Question Text	Annotation Text
Subject	Subject	SUBID
Visit Data	Was the visit done?	(NOT SUBMITTED)
Visit Data	If not done, provide the reason	(NOT SUBMITTED)
Visit Data	Visit date	SVSTDT
Visit Data	Visit date	SVENDT
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	DSSTDC
Demographics	Date subject or legal guardian signed informed consent	RFICDTC
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 Unit	VSORRESU
Demographics	Is the subject rescreened?	SUBSCRYN in SUPPDM
Demographics	Previous Subject ID	SUBIDP 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	MHSTDT
Breast Cancer History	TNM staging (TNM)	BNVMSH in SUPPDM
Breast Cancer History	TNM staging (TNM)	TNM
Breast Cancer History	Clinical staging	CLM
Breast Cancer History	Clinical staging at screening	IXSTG in SUPPDM
Breast Cancer History	Date of recurrence or metastasis	MHRDAT in SUPPDM
Breast Cancer History	Most recent progression/relapse date	RCULD in SUPPDM

DM = Demographics	DS = Disposition	VS = Vital Signs
SDTM ARTs_eCRF_V1.0_20210413: UNIQUE eCRF		
Form: Demographics DSCAT = "PROTOCOL MILESTONE"		
Date subject or legal guardian signed informed consent DSTERM = "INFORMED CONSENT OBTAINED"		
Date of birth BRTHDTC		
Sex SEX		
Is the subject has the ability of pregnancy PREGYN in SUPPDM		
Race RACE		
Other, please specify RACEOTH in SUPPDM		
Ethnicity ETHNIC		
Height VSTESTCD = "HEIGHT"		
VSTEST = "Height"		
VSORRESU Fixed Unit: cm		

Blank CRF -> SDTM aCRF

Figure 6: Example of aCRF

HOW SDTM ARTS WORKS - AUTOMATION FOR GENERATING SDTM CORE MAPPING SPECIFICATION AND SDTM PROGRAMS

With the concept introduced above, we now move our focus from generating aCRF to generating SDTM core mapping specification and SDTM program.

A JavaScript Doc Method: getAnnots

The getAnnots method in JavaScript (Appendix I) can be employed to extract all annotation texts and their coordinates from aCRF, based on which domains, variables and certain values are derived.

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	458.160004	151	473.160004
22	3 RACEOTH in SUPPDM	DM	RACEOTH		Y	200	291.160004	327.5	306.160004
23	3 ETHNIC	DM	ETHNIC			140	256.160004	188	271.160004
24	3 VSTESTCD = "HEIGHT"	VS	VSTESTCD	HEIGHT		125	169.160004	267.5	184.160004
25	3 VSTEST = "Height"	VS	VSTEST	Height		125	152.660004	252.5	167.660004
26	3 VSORRES	VS	VSORRES			125	136.160004	181	151.160004
27	3 VSORRESU	VS	VSORRESU			380.984558	167.461456	444.9845581	182.461456

Figure 7: Using JavaScript Doc Method to obtain information to produce SDTM Mapping Spec and Programs

As discussed in the above sections, we obtained information for mapping the question texts to SDTM texts. CRF question texts are associated with raw datasets, therefore, we can link SDTM variables and raw dataset variables together. These relationships will allow us to use VBA tool to facilitate generating SDTM core mapping specification and SDTM programs, as long as the mapping information is available on aCRF.

Note: The VBA script used for producing the SDTM core mapping specification and SDTM program is not included in this paper because each company has its own templates for those two items, however, examples are demonstrated. Above all, the critical step of these automations is to relate SDTM variables to raw dataset variables sourced from aCRF.

In summary, to automate production of SDTM core mapping specification and SDTM programs, we need the following components: the coordinates of both CRF question texts and annotation texts extracted by JavaScript, and the linkage between CRF question texts and raw dataset variables.

Figure 8 to 11 are the flowchart and the examples of automation for generating SDTM core mapping specification and programs.

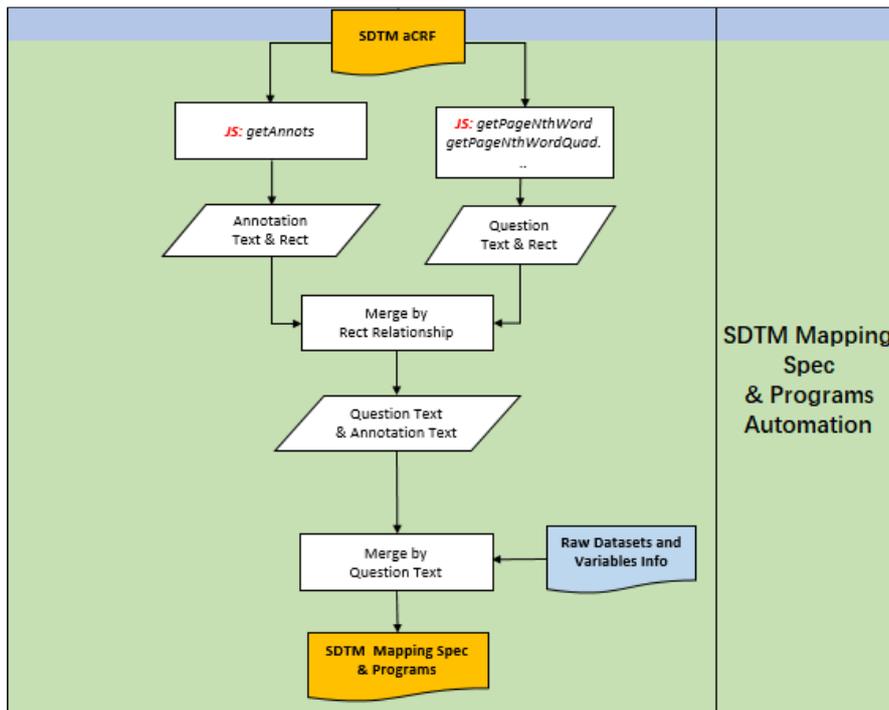


Figure 8: Flowchart of automation for generating SDTM core mapping specification and programs

The image shows the SDTM aCRF tool interface. On the left is a CRF form for 'Demographics' with fields like 'Date subject or legal guardian signed informed consent', 'Date of birth', 'Sex', 'Is the subject has the ability of pregnancy', 'Race', 'Ethnicity', and 'Height'. On the right is a table titled 'RAW Information' with columns: FormOID, FieldOID, PreText, and FixedUnit. Blue arrows map CRF fields to table entries: 'Date subject or legal guardian signed informed consent' to 'ICFDAT', 'Date of birth' to 'BRTHDAT', 'Sex' to 'SEX', 'Is the subject has the ability of pregnancy' to 'PREGYN', 'Race' to 'RACE', 'Ethnicity' to 'ETHNIC', and 'Height' to 'HEIGHT'. A red dashed box highlights the 'ICFDAT', 'BRTHDAT', 'SEX', 'PREGYN', 'RACE', 'ETHNIC', 'HEIGHT', and 'PREVIOUS SUBJECT ID' rows in the table.

Figure 9: Automation process for generating SDTM core mapping specification and programs

A	B	C	D	E	F	G	H	I	J
Variable	Variable Label	Key	Type	Length	Format	Controlled	Origin	CRF Page	Study Derivation
STUDYID	Study Identifier	1	Char	200			Protocol		'SDTM ARTs'
DOMAIN	Domain Abbreviation		Char	2	DS		Assigned		DS
USUBJID	Unique Subject Identifier	2	Char	200			Derived		concatenation of STUDYID, STUDYENV/SITENUMBER, SUBJECT separated by "-".
DSSEQ	Sequence Number		Num	8			Derived		Sequential number ensuring uniqueness of records within each USUBJID in the domain sorted by Key variables
DSSPID	Sponsor-Defined Identifier		Char	200			Assigned		concatenation of Form Name(lowercase) and Recordid by "-". E.g. "xx-3870634"
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) ...
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
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' ...
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' ...
EPOCH	Epoch		Char	200		EPOCH	Derived		Based on DSSTDTC, merge with SE domain.
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.CWDAT 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' ...
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.
DTHREAS	Death Reason		Char	200			CRF	CRF Page 111	RAW.DS.DTHREAS
DTHOTH	Other		Char	200			CRF	CRF Page 111	RAW.DS.DTHOTH

Figure 10: Example of automation for generating SDTM core mapping specification

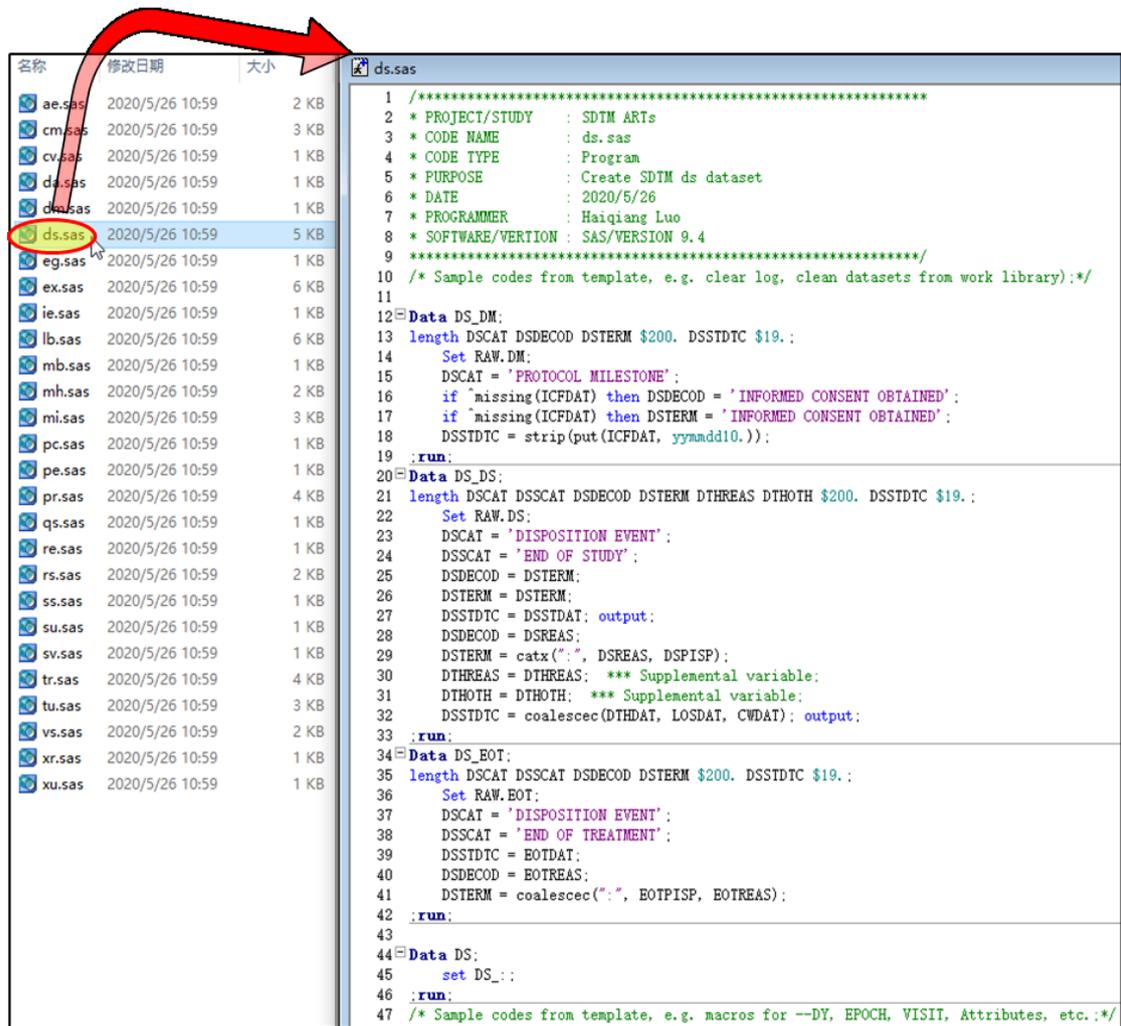


Figure 11: Example of automation for generating SDTM core programs

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”.

THE ARTs ACHIEVES SELF-LEARNING AND IMPROVES SDTM MAPPING WAREHOUSE

In the process of mapping CRF information to SDTM variables and building SDTM Mapping Warehouse, the ARTs gathers data into the Warehouse to satisfy self-learning and upgrades to attain fully-automation in the studies in succession (Figure 12). In other words, given a black CRF, the tool automatically creates SDTM aCRF, SDTM mapping specification and programs, which is illustrated by Figure 13 below.

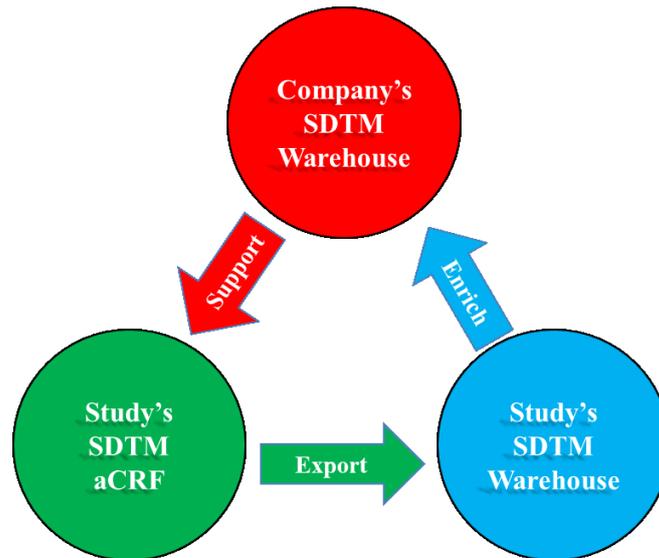


Figure 12: Self-Learning and Improving

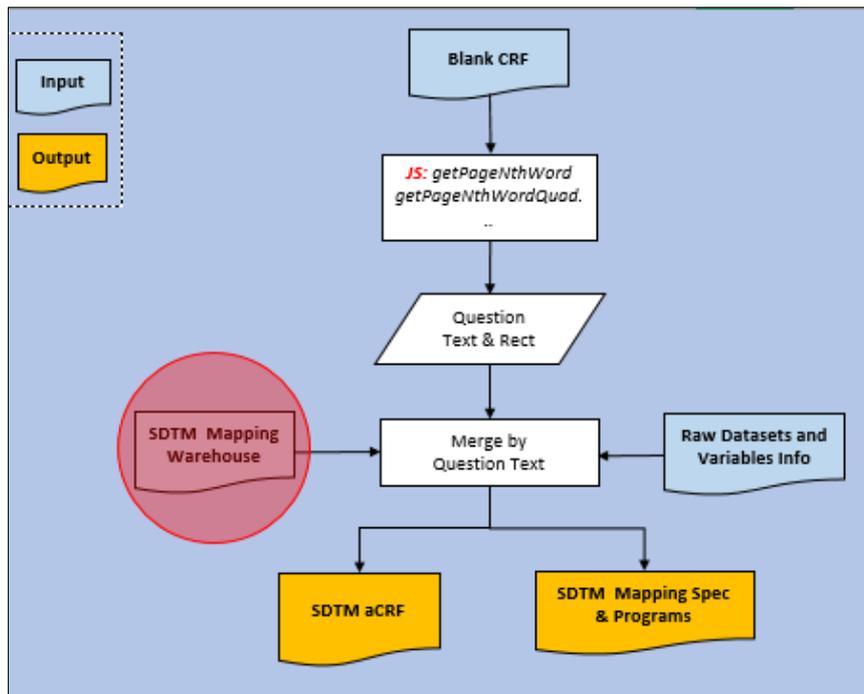


Figure 13: One-step automation for generating SDTM aCRF, Core Mapping Spec and Programs

CONCLUSION

The SDTM ARTs provides a full set of CRF-driven SDTM automation solutions. The challenge is extracting CRF question texts and their corresponding coordinates. With SDTM Mapping Warehouse as input file and metadata from raw database, SDTM aCRF, core mapping specification and core program can be automatically generated, which saves a considerable amount of time in SDTM works. With the clear idea, this tool is easy to understand, use, and reproduce. The idea behind this highly innovative tool is readily accessible and is not hard to be implemented. Although beyond the scope of this paper, the tool is far more powerful than what we have discovered and always worth discussion.

No REFERENCES

ACKNOWLEDGMENTS

The authors sincerely thank the management teams for their valuable comments and suggestions on this paper/presentation.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Haiqiang Luo
Jiangsu Hengrui Pharmaceuticals Co.,Ltd.
+86 13811139137
Haiqiang.Luo@hengrui.com
Haiqiang911@126.com

Kaiying Li
Jiangsu Hengrui Pharmaceuticals Co.,Ltd.
+86 17721287680
Kaiying.Li@hengrui.com

Xiaoyi Niu
Jiangsu Hengrui Pharmaceuticals Co.,Ltd.
+86 13580467517
Xiaoyi.Niu@hengrui.com

Xiaoxia Zhou
Jiangsu Hengrui Pharmaceuticals Co.,Ltd.
+86 13710009381
Xiaoxia.Zhou@hengrui.com

APPENDIX I: SAMPLE CODES AND OUTPUTS

JavaScript App Method getPageNthWord and getPageNthWordQuad

```

var outWords = "";
for (var p = 1; p < this.numPages; p++) {
  {
    for (var i = 0, totalNumWords = this.getPageNumWords(p); i < totalNumWords; i++) {
      var currentWord = this.getPageNthWord(p,i, false).replace(/\r\n/g, "");

      var currentWordQuads = this.getPageNthWordQuads(p,i);
      currentWordQuads = currentWordQuads.toString().replace(/,/g, "\t");

      outWords += (currentWord + "\t" + p + "\t" + currentWordQuads + "\n");
    }
  }
}
this.createDataObject({cName: "myWordList.txt", cValue: outWords});

```

	Word	Page	UpperLeftX	UpperLeftY	UpperRightX	UpperRightY	LowerLeftX	LowerLeftY	LowerRightX	LowerRightY
2	Form:	3	90.00001526	687.9998779	115.0000153	687.9998779	90.00001526	676.6600342	115.0000153	676.6600342
3	Demographics	3	120.0000153	687.9998779	180.0000153	687.9998779	120.0000153	676.6600342	180.0000153	676.6600342
4	Date	3	90	650.9998779	110	650.9998779	90	639.6600342	110	639.6600342
5	subject	3	115	650.9998779	150	650.9998779	115	639.6600342	150	639.6600342
6	or	3	155	650.9998779	165	650.9998779	155	639.6600342	165	639.6600342
7	legal	3	170	650.9998779	195	650.9998779	170	639.6600342	195	639.6600342
8	guardian	3	200	650.9998779	240	650.9998779	200	639.6600342	240	639.6600342
9	signed	3	245	650.9998779	275	650.9998779	245	639.6600342	275	639.6600342
10	informed	3	280	650.9998779	320	650.9998779	280	639.6600342	320	639.6600342
11	consent	3	325	650.9998779	360	650.9998779	325	639.6600342	360	639.6600342
12	Date	3	90	615.9998779	110	615.9998779	90	604.6600342	110	604.6600342
13	of	3	115	615.9998779	125	615.9998779	115	604.6600342	125	604.6600342
14	birth	3	130	615.9998779	155	615.9998779	130	604.6600342	155	604.6600342
15	Sex	3	90	580.9998779	105	580.9998779	90	569.6600342	105	569.6600342
16	Is	3	90	525.9998779	100	525.9998779	90	514.6600342	100	514.6600342
17	the	3	105	525.9998779	120	525.9998779	105	514.6600342	120	514.6600342
18	subject	3	125	525.9998779	160	525.9998779	125	514.6600342	160	514.6600342
19	has	3	165	525.9998779	180	525.9998779	165	514.6600342	180	514.6600342
20	the	3	185	525.9998779	200	525.9998779	185	514.6600342	200	514.6600342
21	ability	3	205	525.9998779	240	525.9998779	205	514.6600342	240	514.6600342
22	of	3	245	525.9998779	255	525.9998779	245	514.6600342	255	514.6600342
23	pregnancy	3	260	525.9998779	305	525.9998779	260	514.6600342	305	514.6600342
24	Race	3	90	470.9999084	110	470.9999084	90	459.6600342	110	459.6600342
25	Other	3	90	303.9999084	120	303.9999084	90	292.6600342	120	292.6600342
26	please	3	125	303.9999084	155	303.9999084	125	292.6600342	155	292.6600342
27	specify	3	160	303.9999084	195	303.9999084	160	292.6600342	195	292.6600342
28	Ethnicity	3	90	268.9999084	135	268.9999084	90	257.6600342	135	257.6600342
29	Height	3	90	181.9999084	120	181.9999084	90	170.6600342	120	170.6600342

JavaScript Doc Method addAnnot

```

var annot = this.addAnnot({
  author : "Haiqiang",
  contents : "SDTM ARTs Annotation",
  page : 0,
  rect : [100, 100, 260, 120],
  strokeColor : ["RGB",0,0,0],
  style : "S",
  textFont : "Times-Roman",
  type : "FreeText",
});

```

SDTM ARTs Annotation

JavaScript Doc Method getAnnots

```

var annotsInfo = "Page\tAnnotation\tLowerLeftX\tLowerLeftY\tUpperRightX\tUpperRightY";
var annots = this.getAnnots({nSortBy: ANSB_Page});
for (var i = 0; i < annots.length; i++) {
    var annotPage = annots[i].page;
    var annotContents = annots[i].contents;
    var annotRect = annots[i].rect;
    annotsInfo += ("\n" + annotPage + "\t" +
        annotContents.replace(/\r\n/g, " ") + "\t" + annotRect.toString().replace(/,/g, "\t"));
}
this.createDataObject({cName: "My annotations.txt", cValue: annotsInfo});

```

1	Page	Annotation	LowerLeftX	LowerLeftY	UpperRightX	UpperRightY
10	3	DM = Demographics	90	707	217.5	725
11	3	DS = Disposition	222.5	707	342.5	725
12	3	VS = Vital Signs	347.5	707	467.5	725
13	3	DSCAT = "PROTOCOL MILESTONE"	185	675.159973	381	690.159973
14	3	DSTERM = "INFORMED CONSENT OBTAINED"	365	638.835999	599.565979	653.159973
15	3	DSDECOD = "INFORMED CONSENT OBTAINED"	365	621.817017	600.2210083	636.523987
16	3	DSSTDTC	306.744995	622.041992	362.7449951	637.041992
17	3	RFICDTC	248.145004	622.005005	304.144989	637.005005
18	3	BRTHDTC	160	603.159973	216	618.159973
19	3	SEX	110	568.159973	137	583.159973
20	3	PREGYN in SUPPDM	310	513.159973	430	528.159973
21	3	RACE	115	458.160004	151	473.160004
22	3	RACEOTH in SUPPDM	200	291.160004	327.5	306.160004
23	3	ETHNIC	140	256.160004	188	271.160004
24	3	VSTESTCD = "HEIGHT"	125	169.160004	267.5	184.160004
25	3	VSTEST = "Height"	125	152.660004	252.5	167.660004
26	3	VSORRES	125	136.160004	181	151.160004
27	3	VSORRESU	380.984558	167.461456	444.9845581	182.461456

APPENDIX II: SDTM ARTs USER INTERFACE



SDTM ARTs

Automation
Relevant
Toolkits

Where is your CRF: ABC RAVE CN EN

/E/01 Project/_Tools/SDTM ARTs/for_PharmaSUG_2021/SDTM ARTs_eCRF_V1.0_20210413.pdf

What do you want:

<input type="checkbox"/> RAW Annotations <input checked="" type="checkbox"/> SDTM Annotations SDTMIG Version: <input type="text" value="3.3"/> <input type="checkbox"/> SDTM Mapping Warehouse <input checked="" type="checkbox"/> SDTM Mapping Spec <input checked="" type="checkbox"/> SDTM Programs <input type="checkbox"/> SDTM Datasets <input type="checkbox"/> Submission Bookmarks <input type="checkbox"/> RawValue Back to CRF	<p>Style: <input type="button" value="RAW = DM"/> <input type="button" value="ETHNIC"/> <input type="button" value="Raw Value"/></p> <p><input checked="" type="radio"/> CDISC <input type="radio"/> Company</p> <table border="0" style="width: 100%; font-size: small;"> <tr> <td><input type="button" value="DM = Demographics"/></td> <td><input type="button" value="DS = Disposition"/></td> <td><input type="button" value="EX = Exposure"/></td> </tr> <tr> <td><input type="button" value="XX in SUPPDM"/></td> <td><input type="button" value="ETHNIC"/></td> <td><input type="button" value="[NOT SUBMITTED]"/></td> </tr> <tr> <td><input type="button" value="DM = DEMOGRAPHICS"/></td> <td><input type="button" value="DS = DISPOSITION"/></td> <td><input type="button" value="EX = EXPOSURE"/></td> </tr> <tr> <td><input ethnico""="" type="button" value="SUPPDM.QVAL where QNAM = "/></td> <td><input type="button" value="DM.ETHNIC"/></td> <td></td> </tr> <tr> <td><input type="button" value="NOT SUBMITTED"/></td> <td></td> <td></td> </tr> </table> <p>Note: ① You can customize any other annotation styles from the selected style. ② You can customize the pattern of SUPP--, i.e. either ETHNICO in SUPPDM or SUPPDM.QVAL where QNAM = "ETHNICO" ③ Set XX as QNAM of SUPP-- annotation then you will get RAW variable name. ④ Set YY. as prefix of variable annotation then you will get a prefix domain name.</p> <p><input type="checkbox"/> NOT SUBMITTED Fields For cSDRG</p>	<input type="button" value="DM = Demographics"/>	<input type="button" value="DS = Disposition"/>	<input type="button" value="EX = Exposure"/>	<input type="button" value="XX in SUPPDM"/>	<input type="button" value="ETHNIC"/>	<input type="button" value="[NOT SUBMITTED]"/>	<input type="button" value="DM = DEMOGRAPHICS"/>	<input type="button" value="DS = DISPOSITION"/>	<input type="button" value="EX = EXPOSURE"/>	<input ethnico""="" type="button" value="SUPPDM.QVAL where QNAM = "/>	<input type="button" value="DM.ETHNIC"/>		<input type="button" value="NOT SUBMITTED"/>		
<input type="button" value="DM = Demographics"/>	<input type="button" value="DS = Disposition"/>	<input type="button" value="EX = Exposure"/>														
<input type="button" value="XX in SUPPDM"/>	<input type="button" value="ETHNIC"/>	<input type="button" value="[NOT SUBMITTED]"/>														
<input type="button" value="DM = DEMOGRAPHICS"/>	<input type="button" value="DS = DISPOSITION"/>	<input type="button" value="EX = EXPOSURE"/>														
<input ethnico""="" type="button" value="SUPPDM.QVAL where QNAM = "/>	<input type="button" value="DM.ETHNIC"/>															
<input type="button" value="NOT SUBMITTED"/>																

Press me to start >>>>

Other utilities

<input type="button" value="Delete RAW Annotations"/> <input type="button" value="Delete SDTM Annotations"/> <input type="button" value="Delete RawValue Annotations"/>		<input type="button" value="Get CRF Page xx for Variables"/> <input type="button" value="Reformat SDTM Annotations"/> <input type="button" value="Create TOC from Bookmarks"/> <input type="button" value="Add links to Dataset/Domain Annots"/>
---	--	---

Log

Developed by [Haiqiang Luo](#)
v1.0 [24May2020]
I'm DEMO