PharmaSUG China 2017 – Paper 33

Extract Information from aCRF by using EXCEL VBA

Kai Zhou, PAREXEL Corporation, Shanghai
Annie Xu, PAREXEL Corporation, Shanghai

ABSTRACT

We need to extract information from annotations of aCRF to state the origin of SDTM variables to build the metadata. Such as CRF page, value list and code list. We usually search such information in the PDF file directly and that makes the task very tedious, time consuming and inaccurate. This paper introduces a method to automate and accelerate the process by using EXCEL VBA.

INTRODUCTION

HOW WE COLLECT METADATA INFORMATION?

As per the FDA statement that "sponsors should make certain that every data variable's codelist, origin, and derivation is clearly and easily accessible from the define file". That is what we called SDTM metadata. And this also makes metadata creation very important. Usually we collect such metadata information as below,

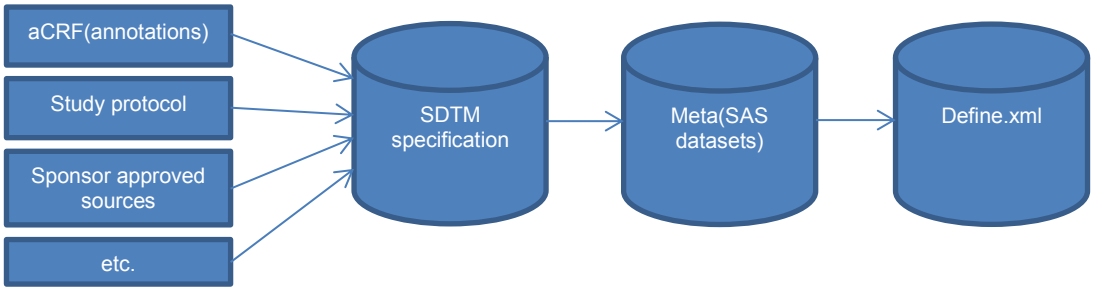


Figure 1. Overview of metadata creation

SDTM annotations in CRF is the mostly origin/source of SDTM variables. In annotations, usually we need to extract variable's CRF page, variable's code list and value definition (value list). Normally we search every variable's corresponding origin information manually. This makes this task very tedious and time consuming. I would like to introduce a small VBA tool to extract such information more quickly.

EXTRACT METADATA FROM ANNOTATIONS

Annotations within a STDM annotated CRF could be extracted as XML Form Data Format (XFDF) using Adobe Acrobat. The steps to export the annotation data vary with different Acrobat version. The Acrobat version 7.0 has a very simple process.

1. Select "Comments → Export Comments →To File" from the main menu.
2. A new box with the title "Export Comments" appears. Save this resulting PDF as XFDF file at your location. Select save as type to be "Acrobat XFDF Files (*.xfdf)".

When we get the xfdf file, we could read and process it by using EXCEL VBA to EXCEL file.

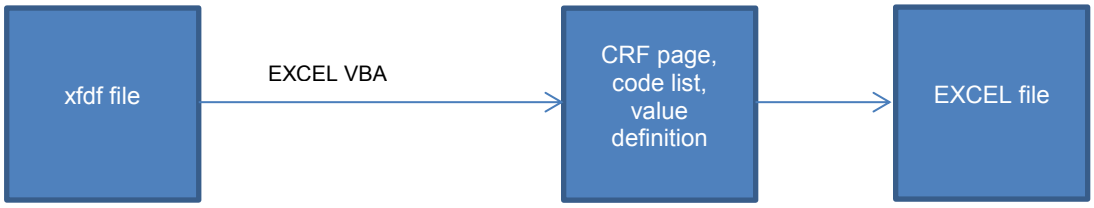
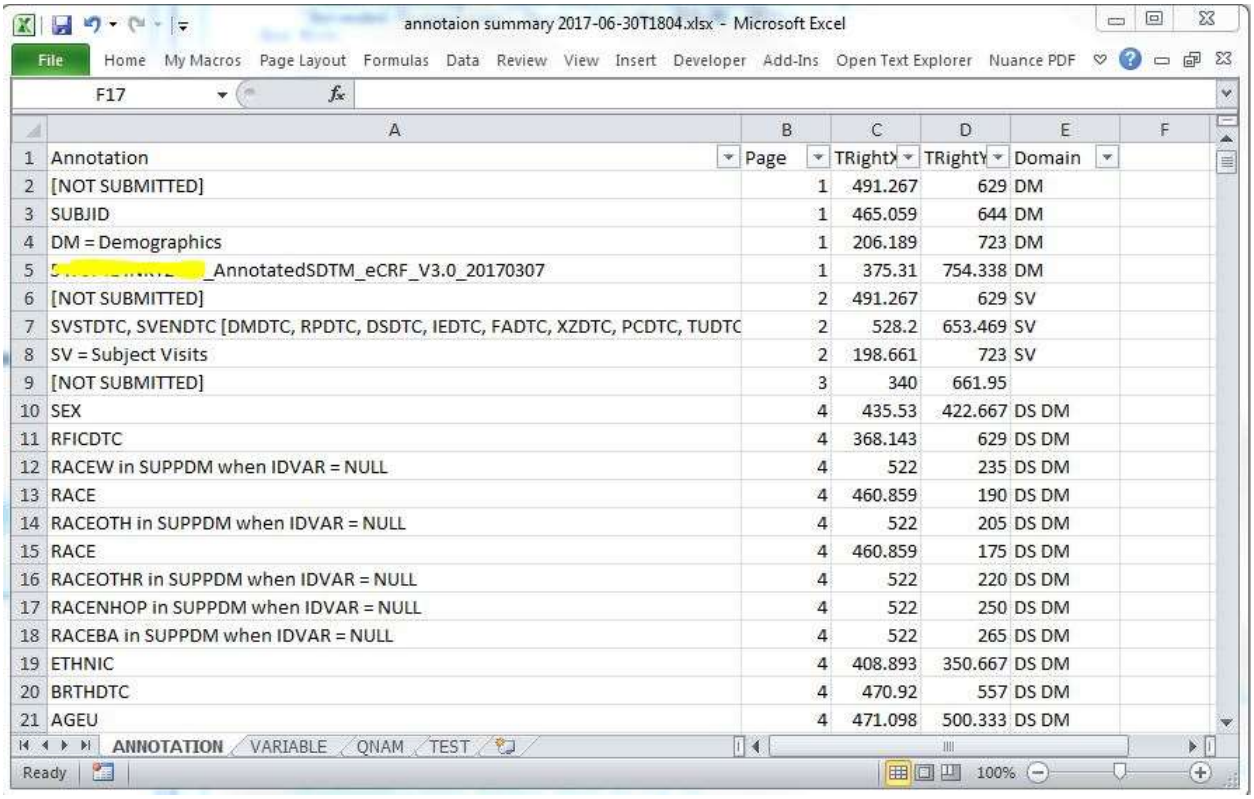


Figure 2. Overview of process of extract metadata by EXCEL VBA



annotaion summary 2017-06-30T1804.xlsx - Microsoft Excel

File Home My Macros Page Layout Formulas Data Review View Insert Developer Add-Ins Open Text Explorer Nuance PDF

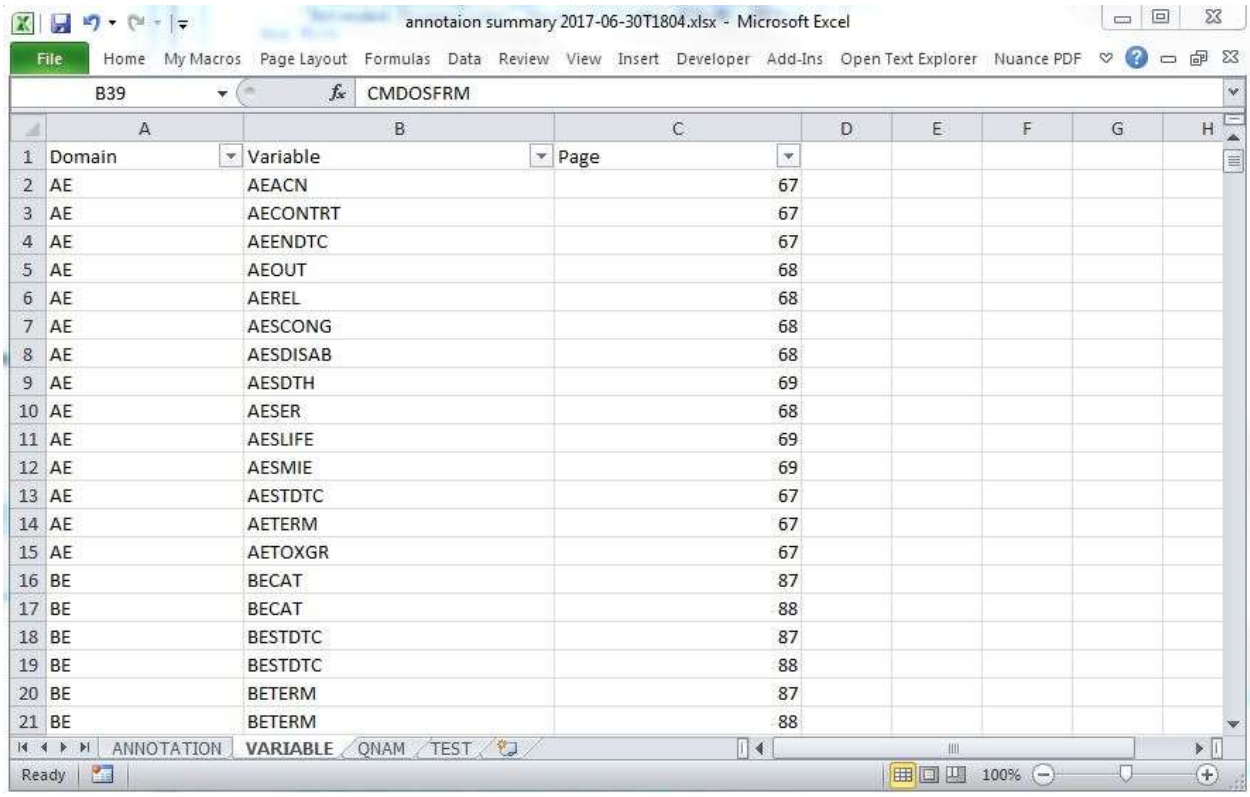
F17

	A	B	C	D	E	F
1	Annotation	Page	TRightX	TRightY	Domain	
2	[NOT SUBMITTED]	1	491.267	629	DM	
3	SUBJID	1	465.059	644	DM	
4	DM = Demographics	1	206.189	723	DM	
5	[redacted]_AnnotatedSDTM_eCRF_V3.0_20170307	1	375.31	754.338	DM	
6	[NOT SUBMITTED]	2	491.267	629	SV	
7	SVSTDTC, SVENDTC [DMDTC, RPDTC, DSDTC, IEDTC, FADTC, XZDTC, PCDTC, TUDTC	2	528.2	653.469	SV	
8	SV = Subject Visits	2	198.661	723	SV	
9	[NOT SUBMITTED]	3	340	661.95		
10	SEX	4	435.53	422.667	DS DM	
11	RFICDTC	4	368.143	629	DS DM	
12	RACEW in SUPPDM when IDVAR = NULL	4	522	235	DS DM	
13	RACE	4	460.859	190	DS DM	
14	RACEOTH in SUPPDM when IDVAR = NULL	4	522	205	DS DM	
15	RACE	4	460.859	175	DS DM	
16	RACEOTHR in SUPPDM when IDVAR = NULL	4	522	220	DS DM	
17	RACENHOP in SUPPDM when IDVAR = NULL	4	522	250	DS DM	
18	RACEBA in SUPPDM when IDVAR = NULL	4	522	265	DS DM	
19	ETHNIC	4	408.893	350.667	DS DM	
20	BRTHDTC	4	470.92	557	DS DM	
21	AGEU	4	471.098	500.333	DS DM	

ANNOTATION VARIABLE QNAM TEST

Ready 100%

Display 1. Result of annotations



annotaion summary 2017-06-30T1804.xlsx - Microsoft Excel

File Home My Macros Page Layout Formulas Data Review View Insert Developer Add-Ins Open Text Explorer Nuance PDF

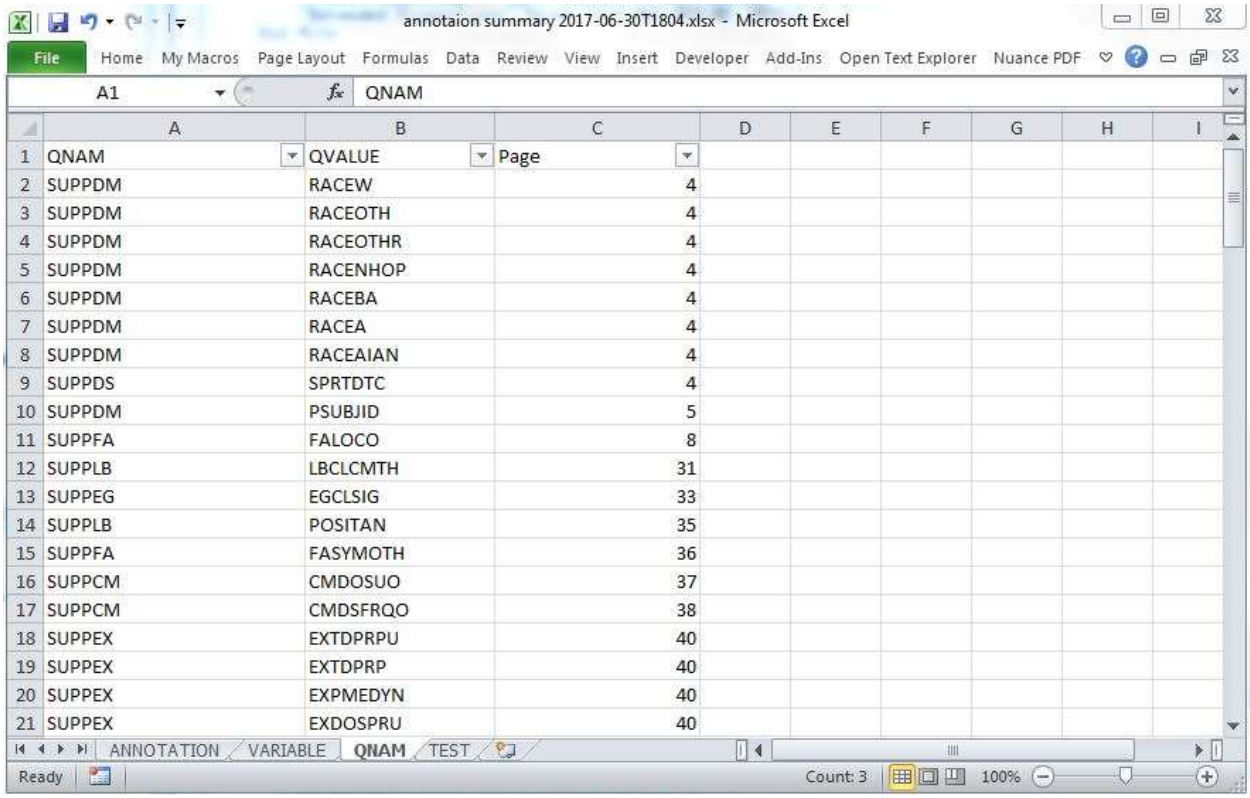
B39 CMDOSFRM

	A	B	C	D	E	F	G	H
1	Domain	Variable	Page					
2	AE	AEACN	67					
3	AE	AECONTRT	67					
4	AE	AEENDTC	67					
5	AE	AEOUT	68					
6	AE	AEREL	68					
7	AE	AESCONG	68					
8	AE	AESDISAB	68					
9	AE	AESDTH	69					
10	AE	AESER	68					
11	AE	AESLIFE	69					
12	AE	AESMIE	69					
13	AE	AESTDTC	67					
14	AE	AETERM	67					
15	AE	AETOXGR	67					
16	BE	BECAT	87					
17	BE	BECAT	88					
18	BE	BESTDTC	87					
19	BE	BESTDTC	88					
20	BE	BETERM	87					
21	BE	BETERM	88					

ANNOTATION VARIABLE QNAM TEST

Ready 100%

Display 2. Result of Variable


annotaion summary 2017-06-30T1804.xlsx - Microsoft Excel

File Home My Macros Page Layout Formulas Data Review View Insert Developer Add-Ins Open Text Explorer Nuance PDF

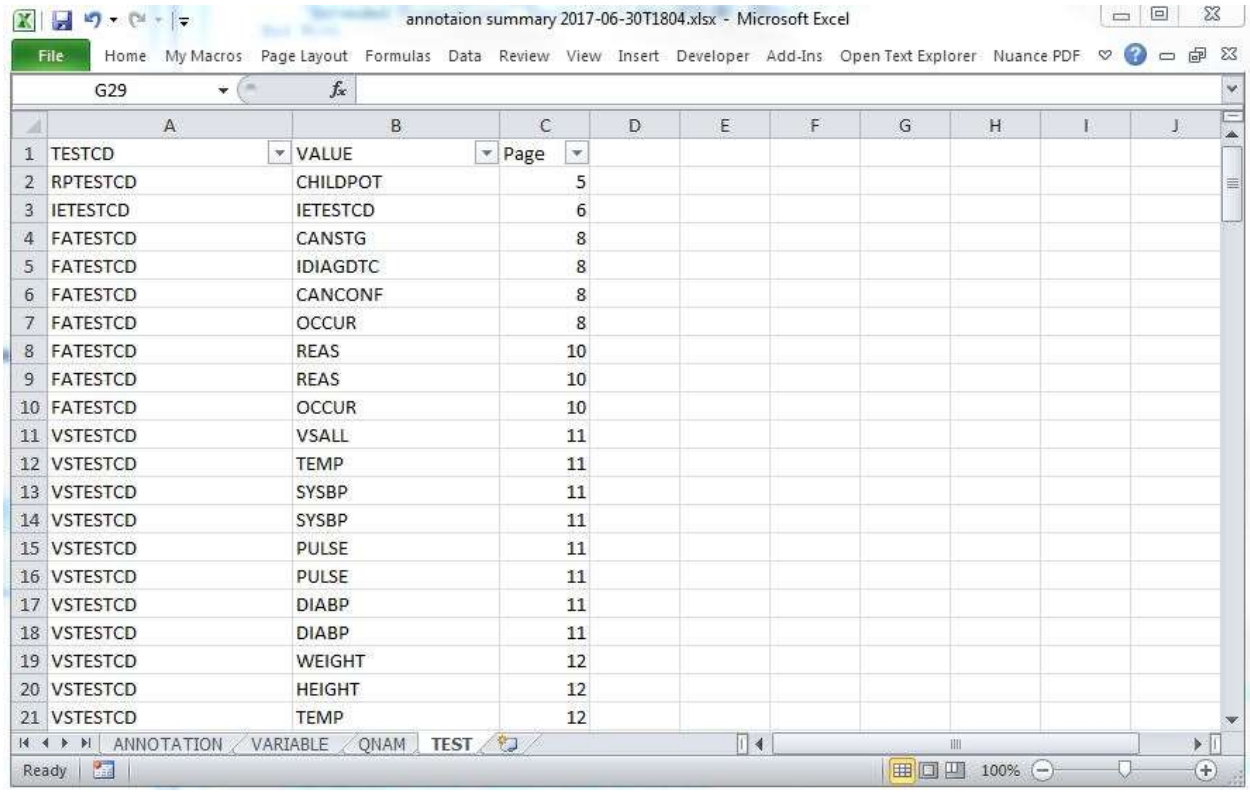
A1 QNAM

	A	B	C
1	QNAM	QVALUE	Page
2	SUPPDM	RACEW	4
3	SUPPDM	RACEOTH	4
4	SUPPDM	RACEOTHR	4
5	SUPPDM	RACENHOP	4
6	SUPPDM	RACEBA	4
7	SUPPDM	RACEA	4
8	SUPPDM	RACEAIAN	4
9	SUPPDS	SPRTDTC	4
10	SUPPDM	PSUBJID	5
11	SUPPFA	FALOCO	8
12	SUPPLB	LBCLCMTH	31
13	SUPPEG	EGCLSIG	33
14	SUPPLB	POSITAN	35
15	SUPPFA	FASYMOTH	36
16	SUPPCM	CMDOSUO	37
17	SUPPCM	CMDSFRQO	38
18	SUPPEX	EXTDPRPU	40
19	SUPPEX	EXTDPRP	40
20	SUPPEX	EXPMEDYN	40
21	SUPPEX	EXDOSPRU	40

ANNOTATION VARIABLE QNAM TEST

Ready Count: 3 100%

Display 3. Result of TEST value list


annotaion summary 2017-06-30T1804.xlsx - Microsoft Excel

File Home My Macros Page Layout Formulas Data Review View Insert Developer Add-Ins Open Text Explorer Nuance PDF

G29

	A	B	C
1	TESTCD	VALUE	Page
2	RPTESTCD	CHILDPOT	5
3	IETESTCD	IETESTCD	6
4	FATESTCD	CANSTG	8
5	FATESTCD	IDIAGDTC	8
6	FATESTCD	CANCONF	8
7	FATESTCD	OCCUR	8
8	FATESTCD	REAS	10
9	FATESTCD	REAS	10
10	FATESTCD	OCCUR	10
11	VSTESTCD	VSALL	11
12	VSTESTCD	TEMP	11
13	VSTESTCD	SYSBP	11
14	VSTESTCD	SYSBP	11
15	VSTESTCD	PULSE	11
16	VSTESTCD	PULSE	11
17	VSTESTCD	DIABP	11
18	VSTESTCD	DIABP	11
19	VSTESTCD	WEIGHT	12
20	VSTESTCD	HEIGHT	12
21	VSTESTCD	TEMP	12

ANNOTATION VARIABLE QNAM TEST

Ready 100%

Display 4. Result of QNAM value list

As display 1, 2, 3 and 4 shows, through this VBA tool, the most part of the information like CRF page, xxTEST value list and QNAM value list could be extracted.

SOURCE CODE(EXCEL ADD-IN)

Automate metadata
.xlam

WHY EXCEL VBA?

First, we need to collect information from aCRF to STDM specification. Generally the specification is EXCEL file too. If we manually do that, that will cost too much time. If we use this tool, we could search this summary EXCEL file to get CRF page, xxTEST and QNAM from supplemental domain directly, which will accelerate this process.

Second, if we use SAS to analysis this XFDF file, we had to read it into SAS then output to EXCEL file. It is more tedious than VBA. Also we need to process special character like “new line” if we are in UNIX SAS environment.

Third, VBA is a very strong language built in MS EXCEL. So it is very easy to use and learn. This is add-in tool, it is also very easy to install and use.

CONCLUSION

The VBA tool described in this article is very useful tool to accelerate the task that collects metadata information from annotations. Let us avoid a part of manual work and improve the efficiency.

REFERENCES

FDA CDER, 2011. “Common Data Standards Issue Document.” Available at http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM254113.pdf

Allen Zeng,Shenglin Zhang. “Define.xml Content Validation – CRF Page Check”. PharmaSUG 2016. Available at

http://www.lexjansen.com/pharmasug-cn/2016/CD/PharmaSUG-China-2016-CD10.pdf

ACKNOWLEDGMENTS

I would like to thank my line manager Annie Xu for helpful discussion.

CONTACT INFORMATION

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

Name: Kai Zhou
Enterprise: PAREXEL International
Address: 9F & Unit A/B/C 10F, No.506, Shangcheng Road, Pudong District, Shanghai, China, 200120
City, State ZIP: Shanghai, 200120
Work Phone: +86 2120505341
E-mail: Kai.Zhou@PAREXEL.com
Web: http://www.parexel.com/

Name: Annie Xu
Enterprise: PAREXEL International
Address: 9F & Unit A/B/C 10F, No.506, Shangcheng Road, Pudong District, Shanghai, China, 200120
City, State ZIP: Shanghai, 200120
Work Phone: +86 2120505324
E-mail: Annie.Xu@PAREXEL.com
Web: http://www.parexel.com/