

Developing annotated CRF: SAS, Excel and patience as your friends

Ilias Pynokokis, OCS Consulting B.V., 's-Hertogenbosch, The Netherlands

ABSTRACT

As part of the mandatory submission of electronic data to the authorities, annotated Case Report Forms (aCRF) are included. Clinical Data Interchange Standards Consortium (CDISC) has determined the requirements imposed by the regulatory body to guarantee that variable annotations are placed in their corresponding fields of CRF. This will create an efficient way for the reviewer to check the metadata definitions by tracing data originating from the CRF. However, there is a challenge to achieve the desired outcome in an efficient way, since there are many sources of data to be combined and many features to be dealt with, such as the font, color and placement of the annotation boxes. In this paper we conduct such an analysis by using both SAS base and Excel, showing the advantages and disadvantages of such an approach.

INTRODUCTION

Electronic datasets of study data analyzed are required to be submitted to the authorities. The documentation of those datasets is also mandatory and consisted of data definition tables and annotated case report forms (aCRF).

An annotated case report form is a blank CRF with annotations that document the location of the data with the corresponding names of the datasets and the names of those variables included in the submitted datasets. This blank CRF aims to map each of the items on the CRF to the corresponding variables in the dataset. Nothing should be left behind, all the pages and items need to be present and are annotated by mentioning the variables names as available in the data tabulation datasets. Even items that have not been submitted in datasets are to be annotated as such. The annotated CRF should be provided as a PDF file and named "blankcrf.pdf".

It is crucial to meet the above specifications in order for the submission to the regulatory bodies to be successful. However, such a task is demanding as many CRFs can include hundreds of pages and even more items within those pages. Hence, it can be understandable that it is important to achieve efficiency by saving time but not decreasing or threatening quality and standards.

In this paper, a method of creating annotations for CRF is illustrated. It is described the method and tools used for such an assignment.

REQUIREMENTS AND CHALLENGES

Apart from the above mentioned general requirements for annotated case report forms, Clinical Data Interchange Standards Consortium (CDISC) has determined more specific technical requirements to guarantee and achieve two important goals. Firstly, to create an efficient way for the reviewer to check the metadata definitions by tracing data that was originated from the CRF and secondly to promote standardization of the data submitted across and within the industry. Some of these specifications are listed below

- Annotated CRF is a blank CRF
- All the pages should be included regardless the use or lack of use of data
- Annotated CRF should be provided as PDF file named "blankcrf.pdf"

PhUSE 2015

- Tool used by FDA is Adobe Acrobat Professional
- Annotations should reflect the data submitted with SDTM
- Operational or derived data should not be included as annotations on the blank CRF
- Annotations should be text-based and searchable using standard PDF viewers
- Annotations should appear as simple and clean as possible
- Annotations should be in the form of digital boxes and never handwritten
- Annotations should reflect specific appearance and format recommendations, such as font and color
- Every occurrence of the data which is unique should not be discarded, but always annotated
- In case of repetitive pages, annotation of items can be excluded but it is mandatory to create one annotation box referring to the original page. For example, if in page 15 adverse events are present which have been initially observed in page 10 and they are totally the same—that is crucial as it might be the same sector but with additional items—then this page can be annotated as “For Annotations see CRF page 10”
- Data that has been recorded but not submitted in SDTM datasets, should be annotated as “[NOT SUBMITTED]”
- For Value Level Metadata elements, such as VSTESTCD where different attributes are combined, there should be clarification of which exactly this annotation is for. It is well known that VSTESTCD may contain weight, height and temperature as different attributes. VSORRES and VSORRESU variables include their values and units. Hence, if we need to annotate an item as VSORRES, it is required to mention for which attribute of VSTESTCD this value belongs. If the item refers to height of the subject filling in the questionnaire, then the annotation should be “VS.VSORRES when VSTESTCD = “HEIGHT”.
- Most of the times, each page reflects one specific domain, such as adverse events, concomitant medications etc. It is recommended to put the initial of the domain following a dot and then the variable name when we annotate items. For example, DM.AGE for the item which refers to the age of the subject in demographics section. However, this is not always the case as one page might include items which refer to more than one domain. If these items are applicable to those domains, then the annotation can be __.__.AAA. For example, if on one page an item of date and time is applicable for DM and VS domain, it is allowed to annotate it as __.__.DTC than to annotate two times VS.VSDTC and DM.DMDTC.
- For Supplemental qualifier variables which are considered as non-standard, the annotation should be SUPP__.QVAL when QNAM = “”.

The above specifications create quite a few challenges, hence a developer should take them into consideration in order to successfully submit aCRF and assist the reviewer to clearly investigate all the elements and match them with SDTM domains and variables. Moreover, the real challenge is to annotate CRF automatically and be effective on such a demanding task

APPROACH

The process is divided into three parts. Firstly, to determine which annotations are to be available in which pages and store them into an Excel file. Secondly, the development of a SAS application which generates an XFDF file containing the annotation boxes. Last part of the process is to import the XFDF file into the PDF file and the manual placement of the annotations/boxes to their exact location.

DETERMINE ANNOTATIONS

EXCEL FILE

As the first step, the sources used to determine the annotations are the following:

- Original CRF to investigate all the items
- Study Data (SDTM) and metadata to match items with their corresponding variables

After determining the annotations, they are stored in an Excel spreadsheet manually.

PhUSE 2015

	A	B	C
1	CRFPAGE	VARIABLE	ORIENTATION
2	3	__STUDYID	4
3	3	DM.SITEID	4
4	3	__VISITNUM	4
5	3	DM.SUBJID	4
6	3	[NOT SUBMITTED]	4
7	3	__DTC	4
8	3	DM.BRTHDTC	4
9	3	DM.SEX	4
10	3	DM.RACE	4
11	3	DM.ETHNIC	4
12	3	SUPPDM.QVAL when QNAM = "RACEOTH"	4

The first column, named CRFPAGE, includes the page of the CRF in which the annotations should belong to. This is an important input in order to control that all annotations will be placed to their corresponding page. The second column, named VARIABLE, includes the exact text to be used as annotation. The third column, named ORIENTATION, is the orientation of the annotation box and it can have 4 discrete values, ranking from 1 to 4 depending on the type of the PDF. More details on the use and purpose of the latter column will be discussed in the section "SAS APPLICATION".

PROCESS AND TOOLS

SAS APPLICATION

Second part of the process is to develop a SAS program which is going to generate an XFDF file. This file will contain all the annotation boxes in digital format and with the support of Acrobat Reader Professional it will be imported to the blank CRF. The initial part of the program is to read the excel file in SAS and convert it into a SAS dataset. Before reading the excel file, it is crucial to ensure that the format of the excel file must be .XLSX. Below it is figured the code of the standard macro to read the excel file.

```
%infile_xls(inxls=&xlsfile,
            sheetn=ANNOT,
            varlist= CRFPAGE * $200 @
                   LABEL   * $200 @
                   ORIENT  * $200 @,
            outdd=anno);

proc sort data=work.anno;
  by crfpage;
run;
```

Having the SAS dataset with all the variables, next step of the application is to develop code to create the annotation boxes which will include the variables names. After extended research, it was determined that the user must have two options for the size of the boxes depending the size of the PDF which is going to be annotated. Hence, the code illustrated below shows these two options

PhUSE 2015

```
%if &boxsize = large %then %do;
  %let boxheight=15;
  %let fontsize=9;
  %let charwidth=7;
%end;
%else %do;
  %let boxheight=10;
  %let fontsize=6;
  %let charwidth=4;
%end;
```

After research on various PDF files with totally different format, it was decided to give to the user the option to determine the orientation of the boxes based on the format of the PDF of the CRF. Considering the orientation of the file, if it is a landscape or a portrait, if it has a vertical or a horizontal structure, the user will determine the optimal solution for each case in order the annotation boxes and their texts to be displayed correctly based on each situation. Our code takes into consideration those four types. This gives the user the flexibility to choose what is best for his case.

```
data work.items2;
  retain counter column;
  length coords $ 200 coord1-coord4 counter column len rotation 8;
  set work.anno;
  by crfpage;

  if first.crfpage then do;
    counter = 1;
    column = 1;
  end;
  else if counter = 40 then do;
    counter = 1;
    column = column + 1;
  end;
  else counter = counter + 1;

  /* COORD1/COORD2 = upper left corner
     COORD3/COORD4 = lower right corder */
  coord1 = 10 + ((counter-1) * (&boxheight + 5));
  coord2 = 10 + ((column - 1) * (100 * &charwidth.));
  coord3 = coord1 + &boxheight;
  len = max(11, length(label));
  coord4 = coord2 + 10 + (len * &charwidth.);

  %htmlspecialchars(label);

  /* if length(label) <= 8 then coord4 = */
  /* if length(label) <= 11 then coord4 = &boxwidth+10;*/
  /* else coord4 = &boxwidth_xl+10;*/

  /* coords = catx(",", of coord:);*/

  page = strip(put(input(strip(crfpage), 8.) - 1, 8.));

  if orient = "1" or missing(orient) then do;
```

PhUSE 2015

```
rotation = 90;
coords = catx(",", of coord:);
end;
else if orient = "2" then do;
rotation = 180;
coords = catx(",", of coord:);
end;
else if orient = "3" then do;
rotation = 270;
coords = catx(",", of coord:);
end;
else if orient = "4" then do;
rotation = 0;
coords = catx(",", coord2, coord1, coord4, coord3);
end;

drop crfpage counter len;
run;
```

Finally, the last part of the code is to create a data step to write the XML language needed in order to create the XFDF file which was the initial goal. Below is part of the code.

```
data work.output;
length line $ 500;
set work.items2 end=last;

if _n_ = 1 then do;
line = "<?xml version='1.0' encoding='UTF-8'?">"; output;
line = "<xfdf xmlns='http://ns.adobe.com/xfdf/' xml:space='preserve'>";
output;
line = "  <annots>"; output;
end;

line = "    <freetext rect='\" || strip(coords) || \"\"' rotation='\" ||
strip(rotation) || \"\"'>"; output;
line = "      opacity='1' color='\"#cccccc\"' flags='\"print\"'"; output;
line = "      subject='\" || strip(label) || \"\"' page='\" || strip(page) ||
\"\">"; output;
line = "        <contents-richtext>"; output;
line = "          <body xmlns='\"http://www.w3.org/1999/xhtml\"'"; output;
line = "            xmlns:xfa='\"http://www.xfa.org/schema/xfadata/1.0/\""
"; output;
line = "              xfa:APIVersion='\"Acrobat:7.0.0\"'"; output;
line = "              xfa:spec='\"2.0.2\"' style='\"font-size:&fontsize..0pt;font-
family:Helvetica,sans-serif\"><p><span style='\"font-family:Helvetica\"><b>\" ||
strip(label) || \"</b></span></p></body>"; output;
line = "            </contents-richtext>"; output;
line = "          <defaultappearance>0 0 0 rg /Helv &fontsize.
Tf</defaultappearance>"; output;
line = "        <defaultstyle>font: Helvetica,sans-serif &fontsize..0pt; text-
align:left; color:#0000ff</defaultstyle>"; output;
line = "      </freetext>"; output;

if last then do;
```

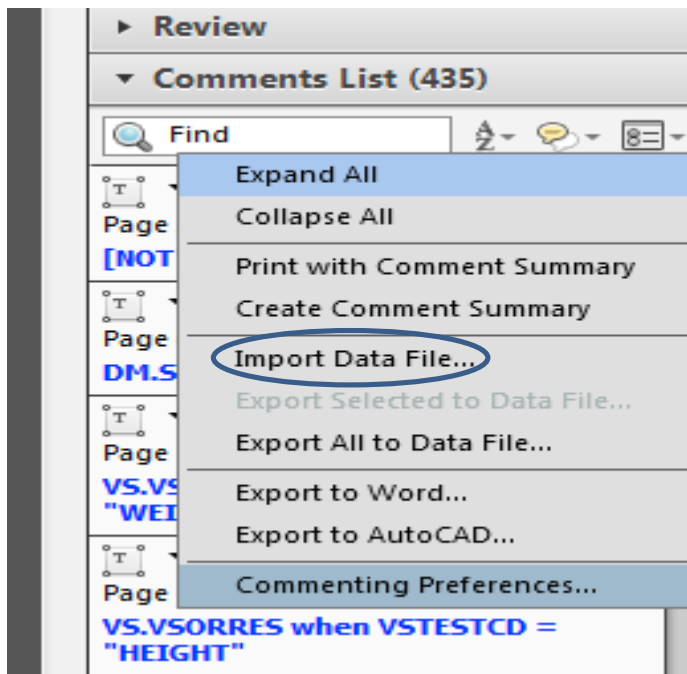
PhUSE 2015

```
line = " </annots>"; output;  
line = "</xpdf>"; output;  
end;  
run;
```

Apart from the standard text which is required for an XFDF file to be created, the above code includes the variables from our dataset "work.items2" which will create the annotation boxes based on the desired preferences with respect to the coordinates of the boxes, text of annotations and pages in which the annotations are to be displayed when imported into the PDF file.

IMPORTING THE XFDF FILE IN CRF

Using Adobe Acrobat Pro the generated XFDF file is imported into the CRF



As it can be observed in the below figure, the initial outcome is not the desired one as we only specified the page number of the annotations.

Annotated Studybook

SV.SV8TDC

DM.STUDYID = "V70_28"

DM.SUBJID

[NOT SUBMITTED] Design: Date of Visit (Date of Visit) [frn

DM.BRTHDTC Date of Visit [frmDateOfVisit__3]

DM.SEX Date of Visit [itmDateOfVisit] Req / Req / Req (2009-2012)

DM.RACE Key: [*] = Item is required. [✓] = Source verification required.

DM.ETHNIC

StudyDesign: Demographics (Demographics)

V8.V8ORRES when V8TESTCD = "WEIGHT"

V8.V8ORRES when V8TESTCD = "HEIGHT"

DM.AGEU Subject Code

DM.AGE

[NOT SUBMITTED] Date of Birth

[NOT SUBMITTED] Age (years)

[NOT SUBMITTED] Gender

[NOT SUBMITTED] Ethnic Origin

[NOT SUBMITTED]

CMTRT = "INFLUENZA VACCINE"

CM.CMOCCUR

CM.CM8TDC

7.* Weight
✓ Should be given with 1 decimal

The boxes needed for each page are indeed present and their orientation is correct, however they are not automatically placed to each item they correspond. Hence, in order to complete the entire process successfully, a labor intensive process must be followed and that is the manual placement of the annotations/boxes. Then the final result is the desired one.

2.* ✓	Subject Code	[itmSubjectCode_Demog] A3 [NOT SUBMITTED]
3.* ✓	Date of Birth	[itmDateOfBirth] Req/Unk / Req/Unk / Req/Unk (1900-1994) DM.BRTHDTC
4.*	Age (years) DM.AGEU	[itmAge] N3 DM.AGE
5.* ✓	Gender	[itmGender] (A:1) ♂ Male (A:2) ♀ Female DM.SEX

CONCLUSION

This paper illustrated a process to create an annotated CRF as part of the mandatory submission to the authorities. Manual editing is inevitable due to the nature and the variety of CRFs with respect not only to their formats but also to their content as the items can vary from one study to another. However, the approach described above helps to automate some parts of the procedure. Not only it takes consideration the different formats of the PDF file, but also through the application developed, it generates automatically all the annotation boxes. In a case where the CRF consists of hundreds of pages and the items on it as well, the workload to generate them is reduced and the manual creation of them is avoided. Certainly there is space for improvements, both to the entire process and the SAS application itself; however it is believed that such an approach is a good step to begin with when dealing with such a demanding task.

ACKNOWLEDGMENTS

This paper is based on an actual project which was undertaken in order to create and deliver various aCRF. I would like to deeply thank Jules van der Zalm for his support and guidance throughout the process.

CONTACT INFORMATION

Ilias Pynokokis
OCS Consulting B.V.
Ruwkampweg 2G
5222 AT 's-Hertogenbosch
+31 (0) 73 523 6000
sasquestions@ocs-consulting.com
<http://www.ocs-consulting.nl/>

Brand and product names are trademarks of their respective companies.