

Every Click Counts: A Smarter Way to Annotate CRFs

Prasanth Prabhakaran, Catalyst Flex, Thiruvananthapuram, India

Bimal Thomas, Catalyst Flex, Thiruvananthapuram, India

ABSTRACT

Manual annotation of Case Report Forms (CRFs) is time-consuming and error-prone, especially when CRF layouts frequently change. This paper presents an automated, scalable solution that integrates SAS® macros, R® functions, and Forms Data Format (FDF) files to streamline the annotation process. Annotations are defined in an auto-generated Excel template from CRF that captures key metadata, including fields, page numbers, coordinates, annotation types, annotation text, and color codes. A SAS macro processes this input to generate an FDF file, which is then applied to the CRF in PDF format. This approach ensures consistent and accurate annotations with precise formatting aligned with metadata submission guidelines. It significantly reduces manual effort, improves adaptability to layout changes, and enhances the overall efficiency of clinical documentation workflows.

INTRODUCTION

Creating annotated Case Report Forms (aCRF) is an essential part of clinical data submissions, but the process is often manual, repetitive, and vulnerable to errors. To simplify and standardize this workflow, an automated solution was developed that integrates SAS macros, R-based PDF extraction, and structured Excel templates. This system enables consistent and accurate annotations while significantly reducing manual effort.

The workflow consists of four steps. First, key properties of the CRF, such as page dimensions and margin settings, are extracted from the PDF using the *pdftools* R package and stored in an Excel file for review. Next, the `%autoanno(usage=ref)` macro uses these properties to generate a reference aCRF with line numbers, along with two supporting worksheets: one for line-based positions and another for annotation mapping.

Users then populate the Mapping sheet with annotation text, position, alignment, type, and optional width details. Finally, running `%autoanno` produces an FDF file containing all annotations, which can be imported into the original CRF to generate the fully annotated aCRF.

FLOWCHART OF ACRF ANNOTATION PROCESS

The flowchart below provides a simple overview of the automated aCRF annotation process. It highlights the main steps involved, which are explained in more detail in the following sections.

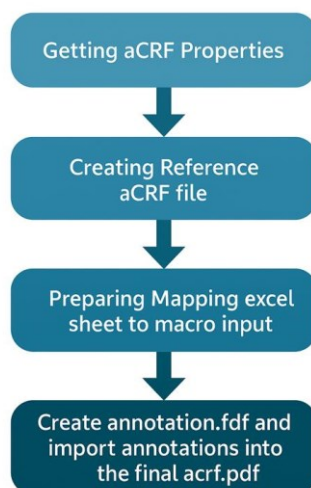


Figure 1. Flowchart of aCRF annotation process

ACRF AUTO ANNOTATION STEPS

To streamline and standardize aCRF annotation process, an automated macro-driven workflow is employed, supported by R-based PDF extraction and structured Excel mapping templates.

The overall process consists of four systematic steps. First, the structural properties of the CRF such as page

dimensions and margin settings are extracted from the PDF using the *pdftools* R package. These properties are stored in a standardized spreadsheet, allowing users to review and adjust any parameters before annotation. Next, a reference aCRF file is created by running the `%autoanno(usage=ref)` macro, which uses the extracted page properties to produce line-numbered reference pages. This step also generates two essential Excel sheets: Position, containing vertical coordinates for each line number, and Mapping, which serves as the template for entering annotation details.

In the third step, users populate the Mapping sheet with all required annotations, specifying page numbers, annotation text, type, position, alignment, domain order, and optionally annotation box width. These values provide the macro with precise instructions for annotation of placement on the final aCRF. Finally, executing `%autoanno` produces an FDF file that can be imported into the original CRF to generate the fully annotated aCRF with accurately positioned annotation boxes.

Additional logic within the macro supports automated width calculation based on character lengths and determines annotation coordinates using the reference file and margin settings. Together, these components create a robust and repeatable framework for generating high-quality aCRF documents with minimal manual effort.

STEP-1 GETTING ACRF PROPERTIES

In this step, the properties of the CRF PDF file are extracted using the R function `acrf_properties`. Provide the CRF PDF as input and run the function:

```
acrf_properties(acrf_file = "acrf.pdf")
```

This will generate an Excel file named *annotation_map.xlsx* containing a worksheet titled "Properties" (Figure 2). The sheet includes the following columns:

- page - total number of pages in the CRF.
- height - height of each CRF page
- width - width of each CRF page
- lmargin - left margin position
- rmargin - right margin position

	A	B	C	D	E
1	page	height	width	lmargin	rmargin
2	52	792	612	90	522

Figure 2. "Properties" tab of the excel file 'annotation_map.xlsx'

Review these values, and if any adjustments are required, update them directly in the Excel file before proceeding to step 2. We are using the *r* package "*pdftools*" to extract the CRF contents from the pdf CRF file.

```
library(pdftools)
pdf_data <- pdf_data(acrf_file)
```

STEP-2 CREATING REFERENCE CRF FILE

After completing step 1, run the following macro to generate the *ref_acrf.fdf* file:

```
%autoanno(usage=ref)
```

This macro reads the inputs from the "Properties" sheet created in Step 1. Once executed, it produces the *ref_acrf.fdf* file.

Next, import the *ref_acrf.fdf* file into the original CRF PDF to create *acrf_ref.pdf*. This reference PDF includes line numbers added to the left and right margins of every page (Figure 3), enabling precise identification of annotation positions.

In addition to the reference PDF, the macro also updates *annotation_map.xlsx* by adding two new worksheets:

- Position - contains the coordinates corresponding to each line number (Figure 4).
- Mapping - serves as a template to enter annotation text and its properties in the next step (Figure 5).

The *acrf_ref.pdf* file will be used as the definitive reference for accurate annotation positioning.

1		1
2		2
3		3
4		4
5		5
6		6
7	PRD_000V1_01_21FEB2024: All CRFs	7
8	Form: Subject Identification	8
9	Generated On: 21 Feb 2024 20:02:24	9
10		10
11	Subject Number	11
12		12
13		13
14		14
15		15
16		16
17		17
18		18
19		19
20		20
21		21
22		22
23		23
24		24
25		25
26		26
27		27
28		28
29		29
30		30
31		31
32		32
33		33
34		34
35		35
36		36
37		37
38		38
39		39
40		40
41		41
42		42
43		43
44		44
45		45
46		46
47	PRD_000V1_01_21FEB2024 (3)	47
48		48
49		49
50		50
51		51
52		52

Figure 3. A page from *acrf_ref.pdf*

	A	B	C
1	y1	position	x2
2	780	1	520
3	765	2	520
4	750	3	520
5	735	4	520
6	720	5	520
7	705	6	520
8	690	7	520

Figure 4. “Position” sheet of the file *annotation_map.xlsx*.

	A	B	C	D	E	F	G
1	page	text	type	position	domain_order	align	width
2							

Figure 5. “Mapping” sheet of the file *annotation_map.xlsx*.

STEP-3 PREPARE THE EXCEL SHEET “MAPPING”

In this step, all required annotation text must be entered into the “Mapping” sheet of the *annotation_map.xlsx* file generated in step 2. This worksheet will serve as the input for creating final annotations.

It contains the following columns:

- *page* - Specifies the page number where the annotation should appear.
- *text* - Contains the annotation text to be displayed.
- *type* - Indicates the annotation type. Valid values are:
 - 1 - Domain annotation
 - 2 - Annotation in a solid box
 - 3 - Annotation in a dotted box
- *position* - Defines the line number at which the annotation should be placed. Line numbers can be referenced from the *acrf_ref.pdf* file.
- *domain_order* - Determines the order of domains. This value controls the color assignment when multiple domain annotations appear on a single CRF page.

- *align* -Specifies the alignment of the annotation text. Supported values are left, right, and center, chosen based on the desired visual appearance.
- *width* - Optional field. The macro automatically computes the annotation width, but it may be manually adjusted here if needed.

This completed sheet will be used in the next step to generate the final annotated aCRF.

A	B	C	D	E	F	G
page	text	type	position	domain_order align		width
5	DM (Demographics)	1	5	1 LEFT		
5	DS (Disposition)	1	5	2 Center		
5	DSDTC / DSSTDTC	2	14	2 RIGHT		
5	DSREFID	2	16	2 RIGHT		
5	DSTERM = INFORMED CONSENT OBTAINED	3	8	2 RIGHT		
5	RESCRN in SUPPDS	2	19	2 RIGHT		
5	RFICDTC	2	13	1 RIGHT		
5	[NOT SUBMITTED]	3	11	1 RIGHT		
6	DM (Demographics)	1	5	1 LEFT		
6	RP (Reproductive System Findings)	1	5	2 Center		
6	AGE	2	15	1 RIGHT		
6	AGEU = YEARS	3	14	1 Center		
6	BRTHDTC	2	11	1 RIGHT		
6	ETHNIC	2	24	1 RIGHT		
6	RACE	2	27	1 RIGHT		
6	RPORRES when RPTESTCD = CHILDPOT	2	19	2 RIGHT		
6	RPREASND when RPSTAT=NOT DONE	2	22	2 RIGHT		
6	SEX	2	17	1 RIGHT		
6	When multiple values are selected then RACE	3	38	1 RIGHT		

Figure 6. "Mapping" sheet after adding annotation text.

STEP-4 CREATING FDF FILE AND UPLOADING INTO CRF

Once the *annotation_map.xlsx* file is fully populated and finalized, run the macro:

%autoanno;

This macro reads the annotation details from the Mapping sheet and generates an FDF file named *annotation.fdf*. After the macro completes, import the *annotation.fdf* file into the original CRF PDF using your preferred PDF editor. This will apply all annotations to the CRF and produce the final annotated aCRF (Figure 7 & 8).

DM (Demographics)	DS (Disposition)
Form: Informed Consent	DSTERM = INFORMED CONSENT OBTAINED
Generated On: [REDACTED]	
Was Informed Consent provided?	[NOT SUBMITTED]
Date Informed Consent Form Signed	RFICDTC
	DSDTC / DSSTDTC
Version Number of protocol consented to	DSREFID
Is the subject being re-screened?	RESCRN in SUPPDS

Figure 7. Annotated CRF page after importing the FDF file.

values stored in *width.xlsx*. This ensures that each annotation box is sized appropriately to fit the text. If a custom width is needed, it can be specified manually in the 'width' column of the Mapping sheet.

A	B
letter	len
#	7
[	4
]	4
0	6
1	6
2	6
3	6

Figure 10. Reference file named *width.xlsx* for handling width.

HANDLING POSITION OF ANNOTATIONS IN THE MACRO

The macro determines the placement of each annotation using the positional information generated earlier in step 2. In that step, a reference aCRF containing line numbers was created. These line numbers must be entered into the "position" column of the Mapping sheet for every annotation.

During macro execution:

- Vertical placement (Y-coordinates) is derived from the line numbers specified in the Mapping sheet, using the positional values available in the reference aCRF file.
- Horizontal placement (X-coordinates) is calculated using the left and right margin settings defined in the annotation_map.xlsx file created in step 1.
- The Align field allows adjustment of the annotation alignment left, right, or center ensuring precise and visually consistent placement.

This ensures that each annotation is positioned accurately and consistently across all CRF pages.

CONCLUSION

The automated aCRF annotation workflow offers a reliable, scalable solution that replaces time-consuming and error-prone manual processes. By combining SAS macros, R-based PDF extraction, and structured Excel templates, the system ensures accurate, consistent, and well-formatted annotations aligned with regulatory expectations. Automated width handling, coordinate-based positioning, and metadata-driven annotation rules enable precise placement even when CRF layouts change. Overall, this approach lowers manual effort, reduces errors, and improves the speed and quality of aCRF development.

REFERENCES

SDTM Metadata Submission Guidelines v2.0 <https://www.cdisc.org/standards/foundational/sdtm/sdtm-metadata-submission-guidelines-v2-0>

ACKNOWLEDGMENTS

I would like to extend my appreciation to the Catalyst leadership and organization for providing the opportunity to develop and present this work. I am also grateful to Aromal Krishna for his valuable support throughout the macro testing activities. In addition, I acknowledge the use of AI-based tools, including Copilot, which assisted in structuring concepts and refining the clarity of this document.

CONTACT INFORMATION

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

Author Name: Prasanth Prabhakaran

Company: Catalyst Flex

Address: Second Floor, Nila Building, Technopark Campus, Thiruvananthapuram, Kerala 695581

Email: prasanth.prabhakaran@catalystcr.com

Website: <https://catalystcr.com/>

Co-Author Name: Bimal Thomas

Company: Catalyst Flex

Address: Second Floor, Nila Building, Technopark Campus, Thiruvananthapuram, Kerala, 695581

Email: bimal.thomas@catalystcr.com

Website: <https://catalystcr.com/>

Brand and product names are trademarks of their respective companies.