Paper SM01

Automating SDTM Annotated CRF (aCRF) Creation Leveraging Upstream Metadata

David Postma, Pfizer, Inc., New York, USA
Christine Rossin, Pfizer, Inc., New York, USA

ABSTRACT

This paper will cover details for the "SDTM aCRF Annotator Tool" used at Pfizer to create submission-ready SDTM annotated CRFs (aCRF), leveraging study documents from data management (DM). Within the tool, an Annotated Study Book (ASB) containing database metadata is used to identify the exact location for the SDTM annotations based on the form and field reference identifiers. A metadata repository (MDR) extract file, which is ultimately uploaded into the tool, links the database field to SDTM variables as annotations. The ASB is first annotated with the SDTM variables, then the Blank CRF is annotated with a mirror of the ASB annotations. Finally, the tool uses study visit metadata stored in a spreadsheet from DM outlining the Schedule of Assessments, to create the table of contents (TOC) and bookmarks for the final aCRF. This tool has proven to take the laborious and time-consuming process of annotating the Blank CRF and make it a quicker and more efficient experience.

INTRODUCTION

Annotating a Blank Case Report Form (CRF) with SDTM variables is a critical industry-wide expectation for submitting data to regulatory agencies such as the FDA, PMDA, or NMPA. An annotated CRF (aCRF) is a PDF document that maps the clinical data collection fields to the corresponding variables within the SDTM datasets. This document is essential for understanding how study data were collected and tracing back from study analysis results to their origin.

BACKGROUND AND CONTEXT

The majority of aCRFs have traditionally been created manually, a process that is both time-consuming, tedious and subject to human error. A survey by Pinnacle 21 in 2020 revealed that over 80% of respondents still manually annotate Blank CRFs. This labor-intensive process underscores the need for automation to improve efficiency and reduce the time and costs associated with aCRF creation.

GUIDANCE AND BEST PRACTICES

Guidance for creating aCRFs can be found in:

- SDTM aCRF Guideline v1.0 (wiki.cdisc.org)
- Study Data Tabulation Model Metadata Submission Guidelines (SDTM-MSG) v1.0 (cdisc.org)
- Study Data Tabulation Model Metadata Submission Guidelines (SDTM-MSG): Human Clinical Trials v2.0 (cdisc.org)

DEVELOPMENT OF THE SDTM ACRF ANNOTATOR TOOL

ORIGINAL WORKING MODEL

Initially at Pfizer, the creation of submission-ready aCRFs involved considerable manual effort, supported by a team of eSubmission (eSub) specialists. Annotations were placed on Blank CRFs using SAS programming but placed in a column along one side of each form, with each annotation manually moved to the correct location. Study teams were involved by reviewing annotations and confirming study-specific details. The working model to create the aCRF relied only on deliverables available late in study conduct, such as a stable set of SDTM data and a stable define.xml file.

AUTOMATION OPPORTUNITY

A team of three statistical programmers set out to develop a web-based tool to automate the placement of SDTM variables, creating a hyperlinked Table of Contents (TOC) and bookmarks by Visit and Domain. Utilizing Python scripts behind a downloaded interface, SDTM variable annotations were attached to database source variable fields on a blank Annotated Study Book (ASB) (Figure 1). Coordinates from field reference IDs on the ASB were employed to accurately place SDTM variables onto corresponding blank CRF pages (Figure 2). Utilizing CDASH naming for data collection standards facilitates linkage between collection fields and their corresponding SDTM variables.

Figure 1 Annotated Study Book, SDTM variables placed next to field reference IDs.

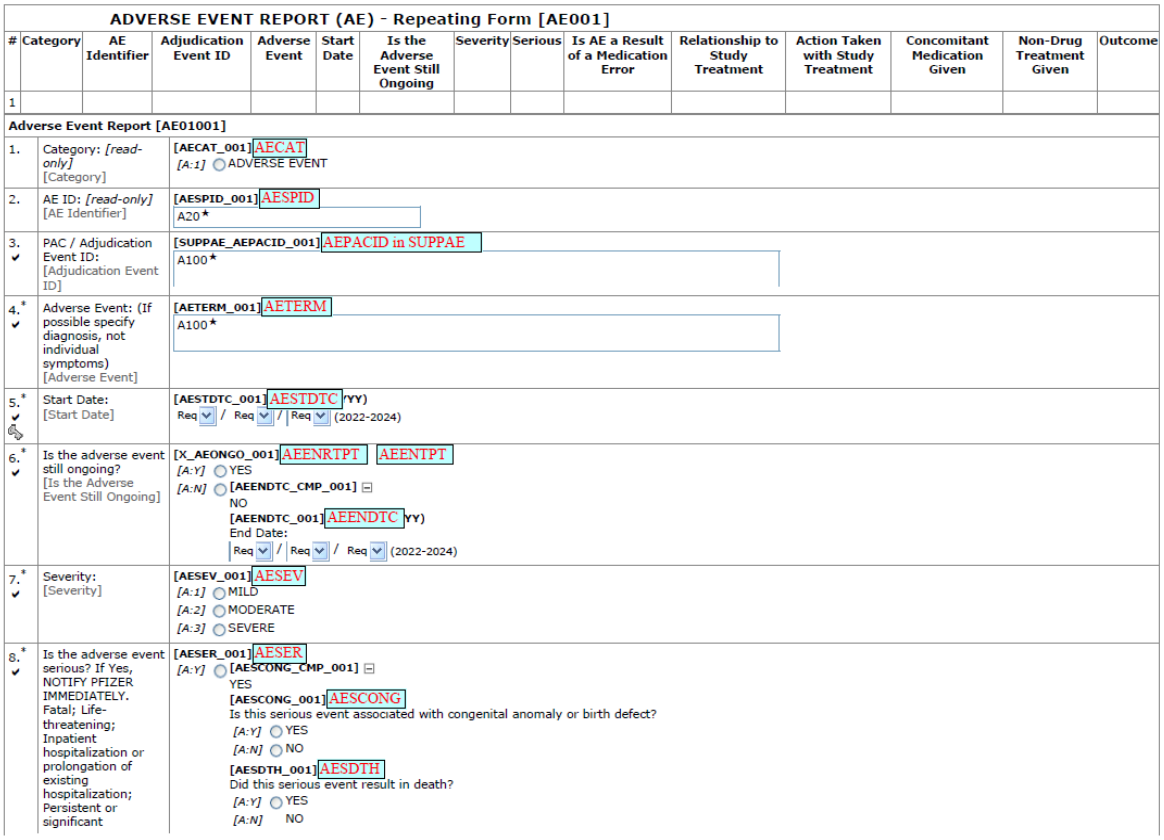
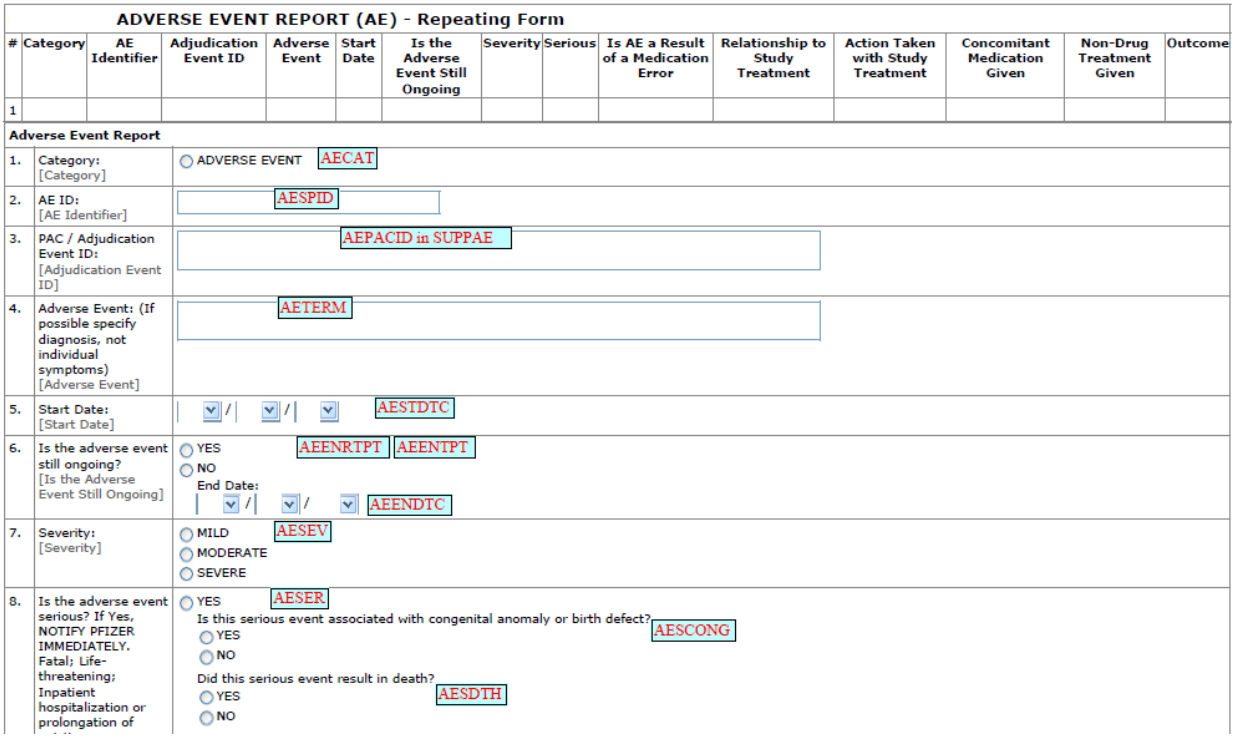


Figure 2 Annotated CRF, Mirrors annotations from ASB and places onto Blank CRF



Study-level annotations were managed manually through a spreadsheet uploaded into the utility as an interfile (noted as “Intermediate file” in Figure 6). Over time, a comprehensive library of annotations (Figure 3 and noted as “DataBase” in Figure 6) was developed, leveraging each study interfile of annotations, expanding to include the new annotations needed from each study. The library of annotations served as a set of standard and repeating

annotations that is applied to new studies being uploaded into the utility. The review and approval of the annotation library was conducted by the repository management team, which was a labor-intensive task.

Figure 3 CRF annotation library spreadsheet

Id	Form Title	Form Ref Name	Source Variable	Annotation Text
fgkAAA==	DEMOGRAPHY (DEMOG)	[DM001]	[BRTHDTC_001]	BRTHDTC
fwkAAA==	DEMOGRAPHY (DEMOG)	[DM001]	[DMPYN_001]	USUBJID
gAkAAA==	DEMOGRAPHY (DEMOG)	[DM001]	[D_SUBJID_001]	SUBJID
gQkAAA==	DEMOGRAPHY (DEMOG)	[DM001]	[ETHNIC_001]	ETHNIC
ggkAAA==	DEMOGRAPHY (DEMOG)	[DM001]	[RACE_001]	RACE, when more than one selected, RACE=MULTIPLE and individual responses are RACE1, RACE2, etc. in SUPPDM
gwkAAA==	DEMOGRAPHY (DEMOG)	[DM001]	[SEX_001]	SEX
hAkAAA==	DEMOGRAPHY (DEMOG)	[DM001]	[SUBJID_001]	SUBJID
hQkAAA==	DEMOGRAPHY (DEMOG)	[DM001]	[SUPPDM_ETHDES_001]	ETHDES in SUPPDM
hgkAAA==	DEMOGRAPHY (DEMOG)	[DM001]	[SUPPDM_RACIALD_001]	RACIALD in SUPPDM
hwkAAA==	DEMOGRAPHY (DEMOG)	[DM001]	[SUPPDM_VACSTAT_001]	VACSTAT in SUPPDM
iAkAAA==	DEMOGRAPHY (DEMOG)	[DM001]		DM=Demographics
Xw4AAA==	DEMOGRAPHY (DEMOGRAPHY)	[DM301]	[AGEDRVU]	AGEU
Xg4AAA==	DEMOGRAPHY (DEMOGRAPHY)	[DM301]	[AGEDRV]	AGE
iQkAAA==	DEMOGRAPHY (DEMOGRAPHY)	[DM301]	[BRTHDAT]	BRTHDTC
igkAAA==	DEMOGRAPHY (DEMOGRAPHY)	[DM301]	[ETHNIC]	ETHNIC
iwkAAA==	DEMOGRAPHY (DEMOGRAPHY)	[DM301]	[RACE]	RACE
jAkAAA==	DEMOGRAPHY (DEMOGRAPHY)	[DM301]	[SEX]	SEX
jQkAAA==	DEMOGRAPHY (DEMOGRAPHY)	[DM301]	[SUBJID]	SUBJID
jgkAAA==	DEMOGRAPHY (DEMOGRAPHY)	[DM301]		DM=Demographics
zQkAAA==	DISPOSITION - FOLLOW-UP (DI	[DS001_6]	[DSDECOD_001_6]	DSDECOD
zgkAAA==	DISPOSITION - FOLLOW-UP (DI	[DS001_6]	[DSSCAT_001_6]	DSSCAT

The Study Design Specifications (SDS), including a spreadsheet for time and events schedules (Figure 4), was used to create a hyperlinked Table of Contents (TOC), with bookmarks categorized by Visit and Domain (Figure 5). Additionally, footers were created to reflect the new electronic file page numbers.

Figure 4 Study Design Specifications, metadata source for creating the Table of Contents and bookmarks

		Visit Title	System Screening	System Enrollment	COHORT_SELECTION	SCR	V1_DAY1_VAX1_S
		Visit Short Title / Mnemonic	(SCR)	(ENR)	(COHSEL)	(SCR)	(V1D1VX1S)
		Visit RefName	SYSSCR	SYSENR	COHORT_SELECTION	SCR	V1_DAY1_VAX1_S
		Scheduled/Dynamic/Unplanned (indicate using 'S', 'D' and/or 'U')	[S]	[S]	[S]	[S/D]	[S/D]
		Protocol Visit Window	0 hours	0 hours	1 hour	0 hours	0 to 28 days after Screening
		Start Hour Increment	0	0	0	1	673
Form Title /Assessment	Form Short Title / C	Form RefName					
INFORM SCREENING	SCREEN	D_INFSCR001	1				
INFORM ENROLLMENT	ENROLL	D_INFENR001		1			
COHORT SELECTION	COHORT SEL	R_TRIG001			1		
MAIN INFORMED CONSENT	CONSENT	DS001			2		
DEMOGRAPHY	DEMOG	DM001			3		
DATE OF VISIT	DOV	SV001				1	1
INCLUSION/EXCLUSION CRITERIA	INC EXC S	IE001				2-DF	
INCLUSION/EXCLUSION CRITERIA	INC EXC S	IE001_3				3-DF	
INCLUSION/EXCLUSION CRITERIA	INC EXC S	IE001_6				4-DF	
DISPOSITION - SCREENING	DISP SCR	DS001_4				6	
GENERAL MEDICAL HISTORY	MEDHX	MH001				7-DF	
CONCOMITANT MEDICATIONS - BASELINE	CONMED BSL	CM001				8-DF-RF	
PHYSICAL EXAMINATION	PHYS EXAM	PE001				9-DF	2
VITAL SIGNS - BASELINE	VITALS BSL	VS001				10-DF	
ELECTRONIC SAMPLE TRACKING - PRIOR COVID-19 INFECTION	PRIORCOV19	BEETRK001_3				11	
MICROBIOLOGY SPECIMEN	COV19 SITE	MBMIB002_2				12-DF-RF	

Figure 5 aCRF Table of Contents

LOCAL UTILITY DEVELOPED INTO GLOBALLY ACCESSIBILITY

The first version of the tool was a locally executed utility (Figure 6). For it to be used by study teams, a validated version was required. A GUI interface was developed to be hosted on an internal server, with a formal support ticketing process established for addressing issues and enhancement requests. The second version of the tool, a globally accessible GUI, required a shared server for global access (Figure 7).

Figure 6 First version of aCRF File Generator, local utility interface using Python scripts

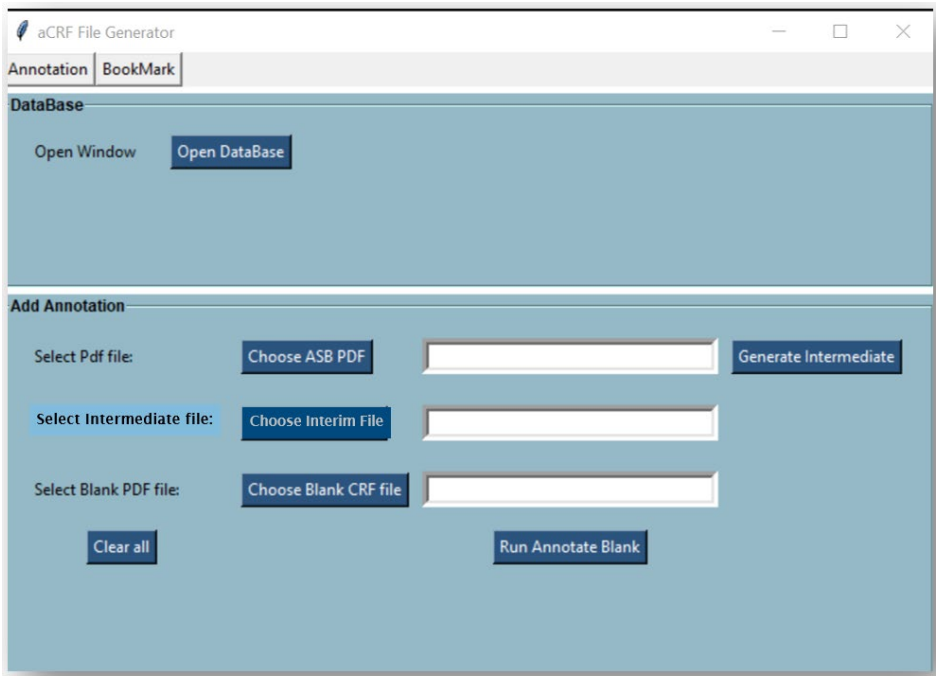


Figure 7 Updated version SDTM aCRF Annotator Tool, globally accessible validated GUI

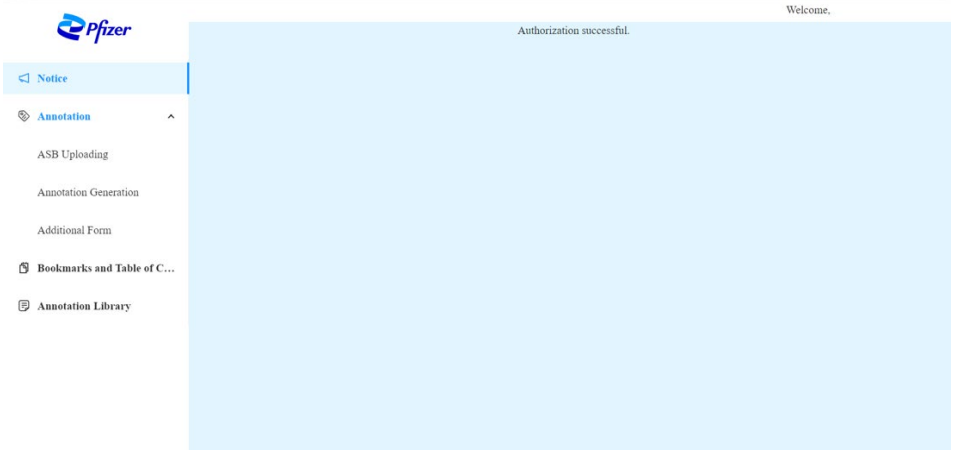


Figure 8 Reviewing and updating annotations metadata in the SDTM aCRF Annotator Tool

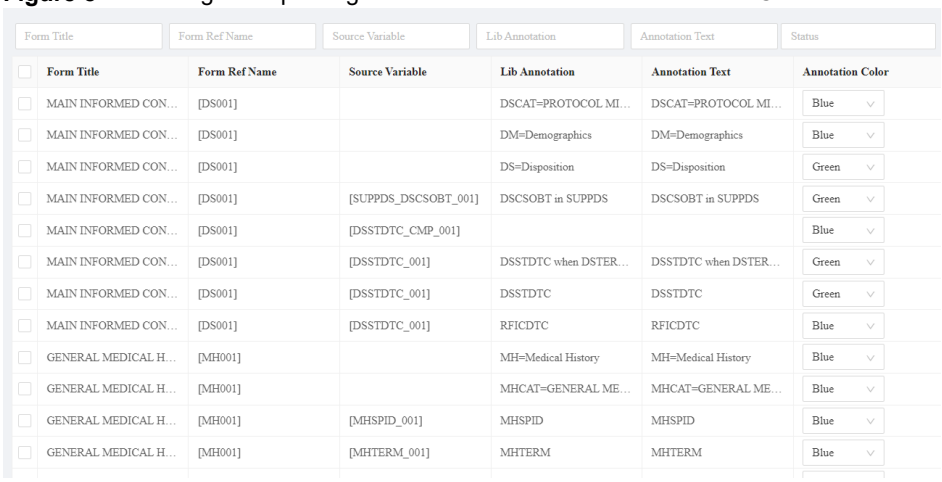
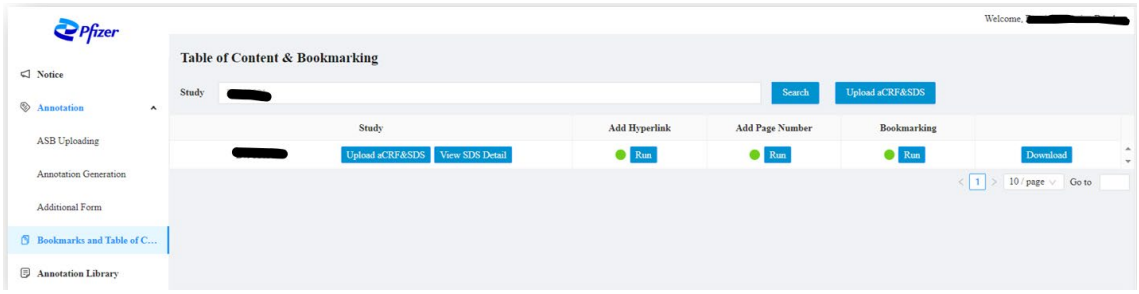


Figure 9 Uploading and reviewing the SDS, generating the TOC, Page Numbers and Bookmarking in the SDTM aCRF Annotator Tool



PILOT TO PREPARE FOR ROLLOUT TO STUDY TEAMS

PILOT SCOPE AND EXECUTION

In the pilot phase to prepare for the rollout to study teams, it was essential to conduct pressure testing within a controlled scope. Low-priority studies, handled by in-house programming teams, served as the ideal candidates, particularly those in the early stages of their conduct or even recently completed studies. The focus remained on studies utilizing the current database system. The pilot phase time commitment was an estimated 20% time over two weeks, equating to less than two days of total effort, though this varied depending on the complexity of the study. During execution, programmers received training and leveraged the pilot to assess the adequacy of the training materials. Teams independently created an aCRF, with Subject Matter Experts (SMEs) available for support. Any gaps identified during this process were addressed and incorporated into the materials for broader rollout.

PILOT PHASE REVIEW: PROS AND CONS

The average time to complete the first aCRF was 12.3 days, with 57% of cases categorized as high complexity, 29% as medium and 14% low complexity. The pilot phase presented several pros and cons. On the positive side, the

system was easy to use and significantly better than the current process, with a faster processing time. However, it faced challenges, including the fact that only 5% to 20% of the CRF annotations were automated, making the review and adjustment process time-consuming. Additionally, there was still a considerable amount of manual work needed by the participants to add new variables for annotations, which resulted in no time savings for new annotations.

REFINING SDTM ACRF ANNOTATOR TOOL

In response to the pilot feedback, several updates have been made to enhance user experience. The tool has been refined to better align with user needs, ensuring a smoother experience.

GUIDANCE MATERIALS UPDATED

The user guide has been revised for clarity, to guide users through the creation of aCRFs from start to finish. Additionally, the software requirements for finalizing aCRFs have been clarified, and the required Adobe version has been confirmed for study team users to ensure compatibility.

ANNOTATION GAP ADDRESSED

To address annotation gaps, a library of annotations has been created using direct extraction from the metadata repository (MDR), significantly enhancing automation from the initial 5-20% to an impressive 80-90%. These updates aim to bridge the gaps identified during the pilot phase, offering a more robust and efficient solution for CRF automation.

CURRENT STATE OF THE TOOL

The final version of the SDTM aCRF Annotator Tool is a validated, globally accessible tool leveraging upstream deliverables and metadata. It provides full traceability back to collection and SDTM standards, offering significant time savings, with increased quality and consistency in aCRF creation. The creation process for an aCRF has been reduced from several weeks to just a few days.

CONCLUSION

The SDTM aCRF Annotator Tool represents a significant advancement in the automation of aCRF creation. By leveraging upstream metadata from Data Management, the aCRF can be created earlier in the Clinical Reporting timeline without the need for a define.xml. Also, by reducing manual efforts, it ensures high-quality, consistent, and submission-ready aCRFs, ultimately facilitating more efficient regulatory submissions.

RECOMMENDED READING

FDA Study Data Technical Conformance Guide, Technical Specifications Document, v5.7, March 2024

SDTM aCRF Guideline, Guideline for SDTM annotation in Case Report Forms, Summary and Recommendations for Best Practice, v10 2020-11-20

ACKNOWLEDGMENTS

Gauravbir Dogra, Smriti Anand, Suvathi Veerapandian, Abdallah Kinani, Mercer Thorp, and Yao Huang for their contributions to the development of the tool.

CONTACT INFORMATION

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

Dave Postma, Pfizer, Inc., david.f.postma@pfizer.com
Christine Rossin, Pfizer, Inc., christine.rossin@pfizer.com