High Quality Study Data Standards for Submission



Tuesday, February 26th 2019

Majdoub Haloui / Ellen Asam

Statistical Programming



Agenda

Regulatory Agency Requirements

 CDSIC Submission Data Package Components

 Detailed Steps for CDISC Submission Data Packages Quality Review

Conclusion





Regulatory Agencies Requirements: FDA

Binding Guidance Documents

Providing Regulatory Submissions in Electronic Format — Standardized Study Data

Guidance for Industry

Additional copies are available from:

Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
1000 New Hampshire Ave, Hillandale Bdg, 4th Floor
Silver Spring, MD 20093
Phone: 855-543-3784 or 301-796-4009. Fax: 301-431-6353
Email: drugsinfof@fda.hts.gov
ts.gov.Drug Cindance Compliance Regulatory Information Guidancecideduit has

anaw or Office of Communication, Outreach and Development Center for Biologics Evaluation and Research Food and Drug Administration 10903 New Hampshire Ave., Bidg. 71, Room 3128 Silver Spring, MD 20993

Phone: 800.828.4790 or 240.00.7800

Email: ocod@fda.hhs.gov http://www.fda.gov/RiologicsRloodVaccines/Guidances/amblanceRegulators/information/Guidances/default.htm

> U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologies Evaluation and Research (CBER)

> > December 2014 Electronic Submissions

Providing Regulatory Submissions in Electronic Format — Submissions Under Section 745A(a) of the Federal Food, Drug, and Cosmetic Act

Guidance for Industry

Additional copies are available from:

Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
10001 New Hampshire Ave., Hillandale Bidg, 4th Floor
Silver Spring, MD 20993
Phone: 855-543-3784 or 301-796-3400; Fax: 301-431-6353
Email: drugstip/Gdida.hts. 200

ov Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.ht and/or

Office of Communication, Outreach and Development Center for Biologie: Evaluation and Research Theory of the Communication of the Communication 10903 New Hampshire Ave., Bilg. 71, Room 3128 Sher Spring, MD 20993 Phone: 800-835-4709 or 240-4027-800 Email: occodi@da.hts.gov

http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER)

> December 2014 Electronic Submissions

Providing Regulatory Submissions in Electronic Format — Certain Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications Guidance for Industry

Additional copies are available from:

Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10001 New Hampshire Ave., Hillandale Bidg., 4th Floor
Silver Spring, MD 20993-0002
Phone: 855-543-3784 or 301-796-3400; Fax: 301-431-6353
Email: drugtin/Giglida hhs 1gov

v.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.ht and/or Office of Communication, Outreach and Development

Office of Communication, Outreach and Developme Center for Biologics Evaluation and Research Food and Drug Administration 3093 New Hampshire Ave., Bidg. 71, Room 3128 Silver Spring, MD 20993-0002 Phone: 800-835-4709 or 240-402-8010 Email: coedifilds hhs.gov

tp://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER)

> January 2019 Electronic Submissions Revision 6





Regulatory Agencies Requirements: FDA

Supporting Documents: TCG, Data Standard Catalog and Validation Rules

STUDY DATA TECHNICAL CONFORMANCE GUIDE

Technical Specifications Document

This Document is incorporated by reference into the following Guidance Document(s):

Guidance for Industry Providing Regulatory Submissions in Electronic Format – Standardized Study Data

For questions regarding this technical specifications document, contact CDER at cder-edata@fda.hhs.gov or CBER at cber.cdisc@fda.hhs.gov

U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER)

October 2018

FDA Data Standards Catalog v5.2 (12-19-2018) - Supported and Required Standards												
For full description of column headings, see Instr.& Column Descriptions tab												
Use	Data Exchange Standard	Exchange Format	Standards Development Organization (SDO)	Supported Version	Supported Implementation Guide Version	FDA Center(s)	Date Support Begins (MM/DD/YYYY)	Date Support Ends (MM/DD/YYYY)	Date Requirement Begins (MM/DD/YYYY)	Date Requirement Ends	Statutory, Regulatory, or Guidance Authority	Information Sources
Udidatia	(SDTM)		Standards Consortium								Study Data Technical Conformance Guide	
Clinical study datasets	SDTM	ХРТ	CDISC	1.2	3.1.2	CDER, CBER	10/30/2009	03/15/2019 [1] 03/15/2020 [2]	12/17/2016 [1] 12/17/2017 [2]		Standardized Study Data	CDISC.org - SDTM
Clinical study datasets	SDTM	ХРТ	CDISC	1.2	Version 3.1.2 Amendment 1	CDER, CBER	08/07/2013	03/15/2019 [1] 03/15/2020 [2]	12/17/2016 [1] 12/17/2017 [2]		Standardized Study Data	CDISC.org-SDTM
Clinical study datasets	SDTM	XPT	CDISC	1.3	3.1.3	CDER, CBER	12/01/2012		12/17/2016 [1] 12/17/2017 [2]		Standardized Study Data	CDISC.org - SDTM
Clinical study datasets	SDTM	XPT	CDISC	1.4	3.2	CDER, CBER	08/17/2015		03/15/2018 [1] 03/15/2019 [2]		Standardized Study Data	CDISC.org-SDTM
Clinical study datasets	Analysis Data Model (ADaM)	ХРТ	CDISC	21	1.0	CDER, CBER	Ongoing	03/15/2019 [1] 03/15/2020 [2]	12/17/2016 [1] 12/17/2017 [2]		Standardized Study Data	CDISC.org - ADaM
Clinical study datasets	ADaM	ХРТ	CDISC	21	1.1	CDER, CBER	03/15/2018		03/15/2019 [1] 03/15/2020 [2]		Standardized Study Data	CDISC.org - ADaM
Animal study datasets	Standard for Exchange of Nonclinical Data (SEND)	XPT	CDISC	1.2	3.0	CDER	06/13/2011	03/15/2019 [1] 03/15/2020 [2]		03/15/2019 [1] 03/15/2020 [2]	Standardized Study Data	CDISC orq - SEND
→ Instr. & Column Descriptions Submission Data Exchange Stds SubmissionTerminology Stds Change History ⊕ : •												





Regulatory Agencies Requirements: PMDA

PMDA is currently in a transitional period. After this transition period, starting April 1, 2020, all required study data will need to be submitted in CDISC format. No waivers will be allowed after this date.

Provisional Translation (as of July 2015) *

Notification No. 0427001 April 27, 2015

To: Prefectural Health Department (Bureau)

Director of the Advanced Review with Electronic Data Promotion Group,
Pharmaceuticals and Medical Devices Agency

Technical Conformance Guide on Electronic Study Data Submissions

Basic principles on the submission of electronic study data for new drug applications have been notified in "Basic Principles on Electronic Submission of Study Data for New Drug Applications" (PFSB/ELD Notification No. 0620-6, by the Director of Evaluation and Licensing Division, Pharmaceutical and Food Safety Bureau, Ministry of Health, Labour and Welfare, dated June 20, 2014) (hereinafter referred to as notification of basic principles), and "Question and Answer Guide Regarding [Basic Principles on Electronic Submission of Study Data for New Drug Applications]" (Administrative Notice by the Evaluation and Licensing Division, Pharmaceutical and Food Safety Bureau, Ministry of Health, Labour and Welfare, dated June 20, 2014), and practical matters on the electronic study data submission have been notified in "Notification on Practical Operations of Electronic Study Data Submissions" (PFSB/ELD Notification No. 0427-1, by the Director of Evaluation and Licensing Division, Pharmaceutical and Food Safety Bureau, Ministry of Health, Labour and Welfare, dated April 27, 2015) (hereinafter referred to as notification on practical operations) and "Question and Answer Guide Regarding [Notification on Practical Operations of Electronic Study Data Submissions]" (Administrative Notice by the Evaluation and Licensing Division, Pharmaceutical and Food Safety Bureau, Ministry of Health, Labour and Welfare, dated April 27, 2015). More detailed matters and precautions regarding submission of electronic study data for new drug applications have been compiled as shown in the appendix; therefore, we ask you to inform manufacturers and sellers placed under your administration.

PMDA Data Standards Catalog (2017-03-03) - Data Exchange Standards									
Use	Data Exchange Standard	Supported Implementatio Version(s) Guide Version		Exchange Format	Date Support Begins (YYYY-MM-DD)	Date Support Ends (YYYY-MM-DD)	Notes		
Clinical study datasets - Transport	SAS Transport (XPORT)	5	-	XPT	2016-10-01				
Clinical study datasets	SDTM	1.4	3.2	XPT	2016-10-01				
Clinical study datasets	SDTM	1.3	3.1.3	XPT	2016-10-01				
Clinical study datasets	SDTM	1.2	3.1.2 Amendment1	XPT	2016-10-01				
Clinical study datasets	SDTM	1.2	3.1.2	XPT	2016-10-01				
Clinical study datasets	ADaM	2.1	1.0	XPT	2016-10-01				
Clinical study data definition files	Define	2.0	-	XML	2016-10-01				
Clinical study data definition files	Define	1.0	-	XML	2016-10-01				
Documents	PDF	1.4-1.7	-	PDF	2016-10-01		In principle, eCTD PDF specification should be referenced for details.		
Data Exchange Standards		Terminology S	tandards Change	History	(+)		for details.		





Current Regulatory Agencies Requirements: Others

European Medicines Agency (EMA)

 The EMA does not require individual patient data to be submitted electronically

China Food and Drug Administration (

 cFDA has recently started to accept the submission of electronic data to support drug applications but does not yet mandate it.

Health Canada (Canada)

 Canada does not require submission of electronic data but have asked for data for selected submissions.

Ministry of Food and Drug Safety (Korea)

 Korea has asked for data and programs for select submissions





CDSIC Submission Data Packages Components



SDTM

ADaM





Study Data Standardization Plan (SDSP)

SDSP Based on PhUSE template

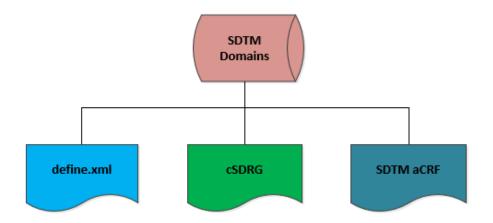
Plan describing the submission of standardized study data to FDA

- Planned Studies (Clinical & Non Clinical)
- Study Type and design
- Data standards, Formats, Terminologies & their versions
- Justification for studies that may not conform to the currently supported standards
- Is located in module 1.13.9 general investigational plan
- Assists FDA in identifying potential data standardization issues early in the development program
- Sponsors may also initiate discussions at the pre-IND stage
- Should be updated in subsequent communications with FDA as the development program expands and additional studies are planned
- For clinical studies that will be submitted to CBER, the SDSP appendix should be provided no later than the end-of-phase 2 meeting
- The CBER SDSP appendix should include tables of proposed SDTM domain/variable usage, supplemental domain usage and proposed analysis



SDTM

- aCRF
- Datasets
- Define.xml
- Clinical Study Data Reviewer's Guide (cSDRG)

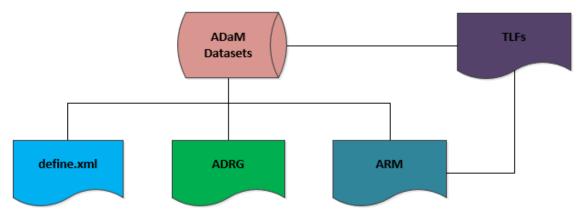






ADaM

- Datasets
- Define.xml
- Analysis Data Reviewer's Guide (ADRG)
- Software programs
- Analysis Results Metadata (ARM)





Detailed Steps for CDISC Submission Data Packages Quality Review



www.fda.gov



High quality data is
the key to enabling
regulatory reviewers
to fully utilize
the Computational
Science Center's tools
and services
to support decision
making





aCRF

aCRF is a main reference for reviewers to understand how data was collected

- Regardless of whether the clinical database is legacy or SDTM compliant, an aCRF should be submitted and the file name should be 'acrf.pdf'
- Only unique CRF forms should be annotated
- Annotation of SDTM variables, not raw data variables
- aCRF should not be a scanned document and should be searchable
- The text "NOT SUBMITTED" should be annotated on the CRF where data is recorded on the CRF but not submitted, and an explanation should be provided in the SDRG
- · All collected variables in the domain datasets (xpt files) should be accounted for in the aCRF





Datasets (1 of 2)

The study data should conform to CDISC SDTM, ADaM, SEND standards and sponsor's custom standards

- SAS Transport Format (XPORT) version 5 should be the datasets file format
- All datasets and dataset variables should have labels
- Dataset labels should be unique across SDTM & ADaM
- For all datasets, the character variable length should be set to the maximum length
- All dataset and variables should have the appropriate length (Name 8 and label 40)
- Variable names should contain only uppercase letters, numbers, and must start with a letter
- Dataset names should contain only lowercase letters, numbers, and must start with a letter.
- No other symbols or special characters should be included in these names



Datasets (2 of 2)

- Variable names, as well as dataset and variable labels should include American Standard
 Code for Information Interchange (ASCII) text codes only
- Dataset and variable labels can include punctuation characters but no special characters
- All required and expected variables should be present in the dataset
- All permissible variables for which data were collected (or necessary for derivations) should be present in the dataset
- All permissible variables that have NULL or missing values should be dropped from the dataset unless they were intended to be collected on the CRF
- Datasets greater than 5 gigabytes (GB) in size should be split into smaller datasets
- File names should only contain letters, numbers or hyphens no underscores or spaces
- All file names should be in lower case





Define.xml (1 of 3)

The define file is the most important part of the electronic dataset submission for regulatory review. The structure and content of the define.xml for both tabulations and analysis should be checked for accuracy, clarity and traceability

- Define.xml should be created as per CDISC Define-XML Specification Version 2.0
- Datasets should be listed in alphabetical order by name attribute within each class in the define.xml file
- All datasets submitted are listed in the define.xml and vice versa
- All dataset attributes (description, structure, class, keys etc.) should be populated in the defne.xml
- Class attributes should have a value of 'SPECIAL PURPOSE', 'FINDINGS', 'EVENTS',
 'INTERVENTIONS', 'TRIAL DESIGN', or 'RELATIONSHIP' for SDTM and SEND data and
 'SUBJECT LEVEL ANALYSIS DATASET', 'BASIC DATA STRUCTURE', 'OCCURRENCE
 DATA STRUCTURE' or 'ADAM OTHER' for ADaM data
- Structure should match the mapping specification for SDTM datasets and ADaM specification for ADaM datasets



Define.xml (2 of 3)

- The dataset keys listed in the define.xml should be "natural keys" and should identify a unique record within a dataset
- Variable attributes should match with the dataset (xpt file) variable attributes
- Variable origin type attribute should have a value of 'CRF', 'Derived', 'Assigned', 'Protocol', 'eDT', or 'Predecessor'
- If a variable has multiple origins/datatypes, leave the origin/datatype blank at the variable level and provide origin/datatype at the value level metadata
- If the origin of a variable is 'CRF', CRF page numbers must be populated for that variable
- If the origin of a variable is 'Predecessor', predecessor value must be provided





Define.xml (3 of 3)

- If the origin of a variable is 'Derived', derivation must be populated for that variable
- The explanation of derivations in the "Source/Derivation/Comment" column (or in computational algorithm section) should be written in English instead of programming code.
 The derivations should be easily understood by any reviewer
- Derivations should be clear and concise to help replicate the results
- The controlled terminology should be provided in the define.xml for all variables that have controlled terminology associated with it as per the SDTM-IG/ADaM-IG version being used for the study
- The define.xml should have bookmarks for navigating to external documents such as aCRF,
 CSDRG, ADRG or Analysis results metadata





Datasets/define.xml Validation

Study data validation helps to ensure that the study data are compliant, useful, and will support meaningful review and analysis. Study data should be validated against CDISC standards, eCTD and FDA/PMDA validation rules

- Compliance with Technical Rejection Criteria for both FDA and PMDA
 - ✓ **FDA**: All Rejects must be resolved, Errors/Warnings that can not be resoled should be documented in the SDRG/ADRG with a clear rationale
 - ✓ PMDA: All Rejects must be resolved, Errors need to be discussed before submission and warnings do not require any explanation
- Validation of the datasets and define.xml must be performed using the latest Pinnacle21 version used by the agency
- The results of the validations should be reviewed carefully
- The define.xml file should open without any issues in the web browser and all bookmarks and links are working
- All dataset/variable attributes should be populated in the defne.xml and match with the actual dataset/variable attributes



SDRG & ADRG Sections

The SDRG/ADRG describe any special considerations or directions to quickly orient the reviewers to the submitted data and help them understand the relationship between the study report and the SDTM or ADaM data. Follow the appropriate template from PhUSE and fill out all sections

cSDRG Sections

- 1. Introduction
 - 1.1 Purpose
 - 1.2 Study Data Standards and Dictionary Inventory
- Protocol Description
 - 2.1 Protocol Number and Title
 - 2.2 Protocol Design
 - 2.3 Trial Design Datasets
- Subject Data Description
 - 3.1 Overview
 - 3.2 Annotated CRFs
 - 3.3 SDTM Subject Domains
- Data Conformance Summary
 - 4.1 Conformance Inputs
 - 4.2 Issues Summary
 - 4.2.1 SDTM Dataset and Define.xml Validation Issues
 - 4.2.2 Define xml Validation Issues

Appendix I: Inclusion/Exclusion Criteria

ADRG Sections

- Introduction
 - 1.1 Purpose
 - 1.2 Acronyms
 - 1.3 Study Data Standards and Dictionary Inventory
 - 1.4 Source Data Used for Analysis Dataset Creation
- Protocol Description
 - 2.1 Protocol Number and Title
 - 2.2 Protocol Design in Relation to ADaM Concepts
- 3. Analysis Considerations Related to Multiple Analysis Datasets
 - 3.1 Comparison of SDTM and ADaM Content
 - 3.2 Core Variables
 - 3.3 Treatment Variables
 - 3.4 Subject Issues that Require Special Analysis Rules
 - 3.5 Use of Visit Windowing, Unscheduled Visits, and

Record Selection

- 3.6 Imputation/Derivation Methods
- Analysis Data Creation and Processing Issues
 - 4.1 Split Datasets
 - 4.2 Data Dependencies
 - 4.3 Intermediate Datasets
 - 4.4 Variable Conventions
- 5. Analysis Dataset Descriptions
 - 5.1 Overview
 - 5.2 Analysis Datasets
 - 5.2.xDataset Dataset Label
- 6. Data Conformance Summary
 - 6.1 Conformance Inputs
 - 6.2 Issues Summary
- Submission of Programs
- Directory Structure



SDRG & ADRG Data Conformance

- Data Conformance Issues
 - All issues that can be fixed should be fixed
 - Provide meaningful explanations for data conformance issues that remain, including the reason why
 - Always use the latest version of Pinnacle 21 software used by the agency





SDRG & ADRG Quality Checklist

The reviewers guide should be well formatted, polished and professional. It should be concise, clear, easy to read, use and navigate

- ✓ Technical functionality: Bookmarks & hyperlinks function and point to correct location
- ✓ Adherence to technical specifications
- ✓ Accuracy and completeness of content
- ✓ Cross-document consistency: use of concepts, definitions and acronyms across documents
- ✓ References in documents are verified.
- ✓ Readability
- ✓ Quality of spelling, grammar etc.
- ✓ Is the tense, person and voice the same throughout the document?
- ✓ Amount of information in document is balanced
 - ✓ Document contains only information that is relevant or important
 - ✓ Document is not missing any information
 - ✓ No extraneous or editorial information
- ✓ Formatting is consistent
 - ✓ Are all fonts the same/correct size?
 - ✓ Are all fonts the same/correct type (bold, italic, etc.)?
 - ✓ Are all fonts the same/correct color?





Analysis Results Metadata (ARM)

The ARM assists reviewers by identifying critical analyses and supports understanding and reproducing analysis results. Therefore, the structure and content of the ARM should be checked for accuracy and clarity.

- All tables/figures listed in ARM should be present in the CSR
- All programs referenced in ARM should be present in the programs folder and linking correctly
- All dataset and variables referenced in ARM should be present in the dataset folder and linking correctly



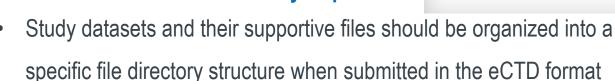


Electronic Common Technical Document (eCTD)

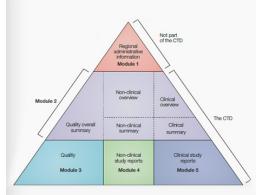
The eCTD is a standard format for electronic regulatory submission.

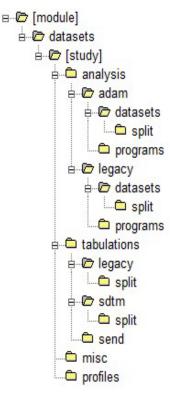
File folder structure for study datasets

- eCTD consists of 5 modules:
 - Module1: Region specific Information
 - Module 2: Summary Documents
 - Module 3: Quality
 - Module 4: Non-clinical Study Reports
 - Module 5: Clinical Study Reports



- Empty file/folders should not be submitted
- Files that exceeds file size limit (5GB) should be split into smaller files, these split files should be submitted in the "split" sub-folder in addition to the larger non-split file in the original data folder







The Need for Additional Checks Beyond P21 E

- Submission of study data standards involves many components and has become a complex process
- It is a significant challenge to ensure all the pieces are accurate, complete and consistent
- We have implemented additional automated quality checks beyond what is available in the Pinnacle 21 Enterprise tool, to ensure consistency between the define.xml document and the submission data (See eSubmission - Are you really Compliant? PharmaSUG paper for detailed information)







Conclusion

- High quality study metadata is extremely important for regulatory review process
 - ✓ Helps reviewers better understand study data
 - ✓ Helps automate review and analysis
 - ✓ Minimize the number of agency's information requests
 - ✓ Accelerate approval







THANKYOU





BACKUP SLIDES





Regulatory Agencies Requirements: FDA

Guidance

- Providing Regulatory Submissions in Electronic Format Standardized Study Data: Guidance for Industry (PDF - 136KB) (Dec. 2014)
- Providing Regulatory Submissions in Electronic Format Submissions Under Section 745A(a) of the FD&C Act: Guidance for Industry (PDF - 81KB) (Dec. 2014)

Supporting Docs

- FDA Data Standards Catalog
- Technical Rejection Criteria for study data
- FDA Technical Conformance Guides

Rules

- The <u>Business Rules</u> help ensure that the study data are compliant, useful, and will support meaningful review and analysis.
- The <u>Validator Rules</u> are used by the FDA to ensure data are standards compliant and support meaningful review and analysis



Regulatory Agencies Requirements: PMDA

PMDA is currently in a transitional period. After this transition period, starting April 1, 2020, all required study data will need to be submitted in CDISC format. No waivers will be allowed after this date.

Data Standards Catalog

Contains a list of acceptable versions of Data Exchange Standards and Terminology Standards that PMDA supports.

<u>Technical Conformance Guide on Electronic Study Data Submissions</u>

The Technical Conformance Guide provides technical details for electronic study data submission.

Basic Principles on Electronic Submission of Study Data for New Drug Applications

Document and Q&A about basic principles for electronic submissions of study data to PMDA.

• Notification on Practical Operations of Electronic Study Data Submissions

Additional information about the electronic submission of study data to PMDA.

Study Data Validation Rules

PMDA has published a set of Study Data Validation Rules for SDTM, ADaM, and define.xml.

FAQs on Electronic Study Data Submission (Excerpt)

The FAQs on Electronic Study Data Submission summarizes inquiries on electronic study data submission received by the PMDA in a Q&A format.





aCRF

aCRF is a main reference for reviewers to understand how data was collected

- Regardless of whether the clinical database is legacy or SDTM compliant, an aCRF should be submitted
- No spelling errors in dataset labels and variable names
- The annotated CRF should include all available forms used to collect source data but only unique CRF forms should be annotated
- Annotation of SDTM variables, not raw data variables
- aCRF should not be a scanned document and should be searchable
- File name should be 'acrf.pdf' not 'blankcrf.pdf'





aCRF

aCRF is a main reference for reviewers to understand how data was collected

- The text "NOT SUBMITTED" should be annotated on the case report form where data are recorded on the CRF but not submitted, and an explanation for why data was not submitted should explained in the Study Data Reviewer's Guide
- All collected variables in the domain datasets (xpt files) should be accounted for in the aCRF (No missing annotations)
- If there is more than one domain per page, each set of domain variables should be annotated in its own color
- Supplemental domains should be the same color as the parent domain
- Every domain on the aCRF page should have the 2-letter domain code at the top left of the page, followed by the domain decode that matches the dataset label
- All text in the annotations that represent variable and domain names should be capitalized
- If possible, the annotations should not obstruct any text on the CRF page



