SDTM SUBMISSION PACKAGE: WHAT LIES BENEATH?

Puja Kawle

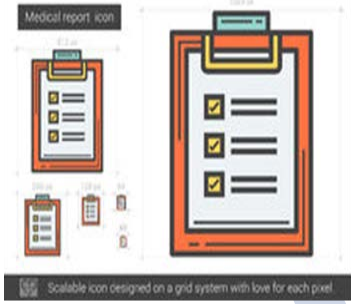
Cytel Statistical Software & Services Pvt. Ltd.

(Subsidiary of Cytel Inc.)

Agenda

- Data Standards for FDA Submissions – Background and Process Overview
- SDTM Submission Package
- Overview of SDTM Metadata documents
- Issues in define.xml, acrf.pdf and sdrg.pdf
- Challenges and Recommendation

Submission Data Review Before Standards

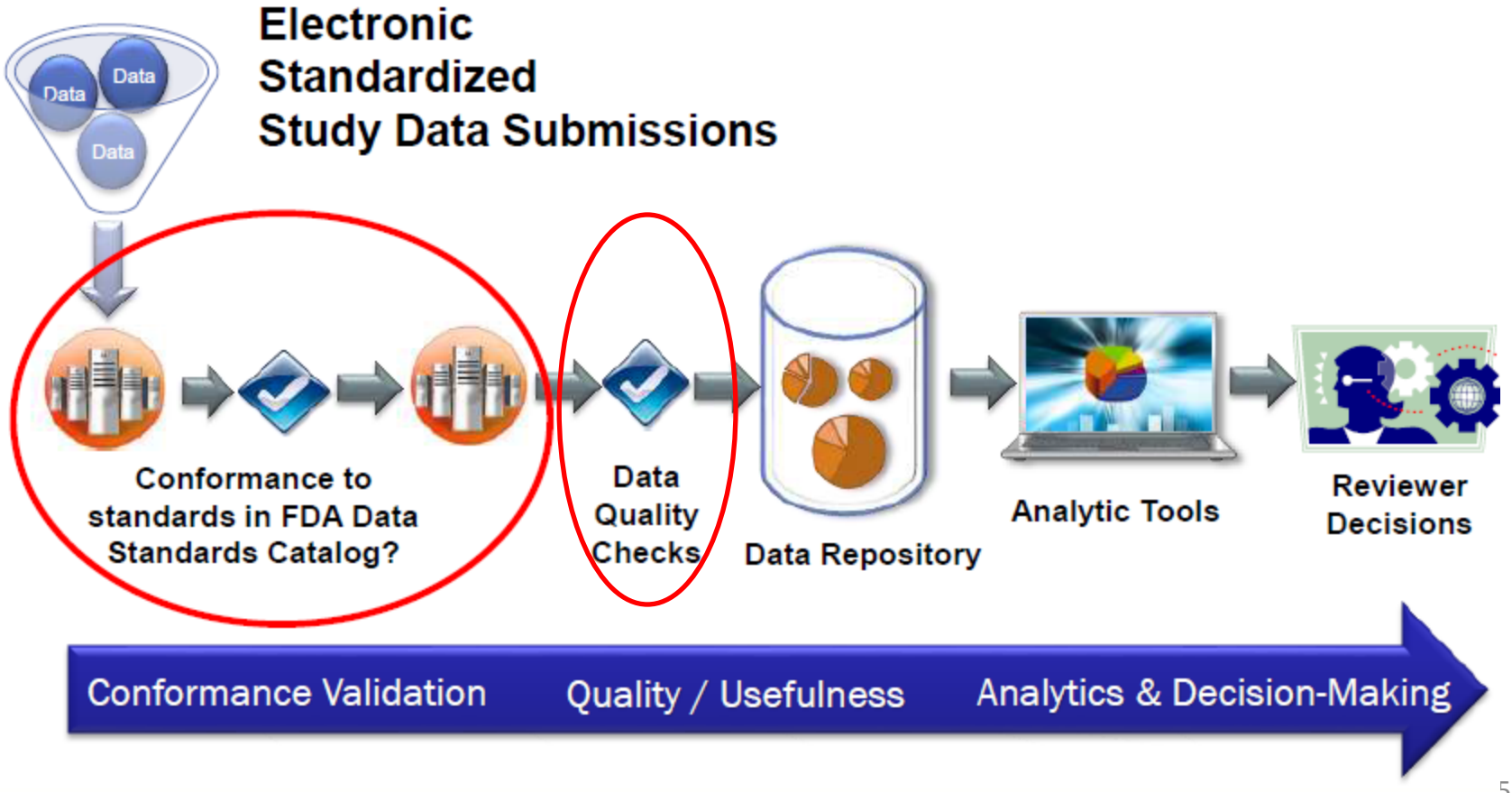


Non Standardized Clinical Data

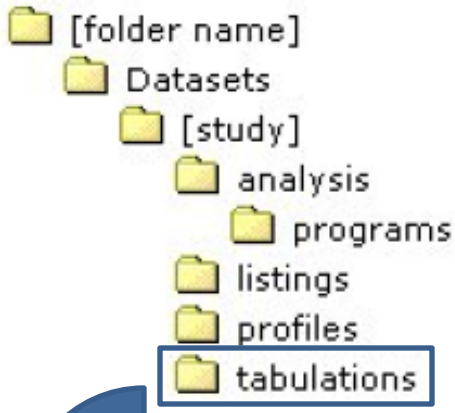
FDA's 21st Century Review Initiative

FDA Jumpstart Service & DataFit Program

KEY: High quality standardized data



Submitting SDTM



Replace with folder name, e.g., m5

Replace with study identifier, e.g., 123-070

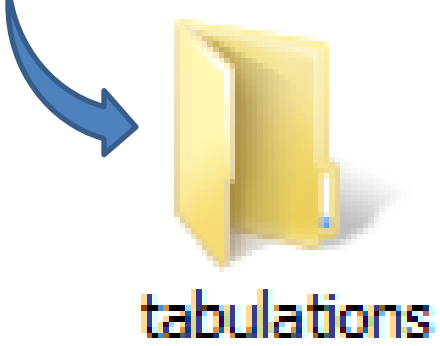
Contains analysis datasets and associated files

Contains program files

Contains data listing datasets and associated files

Contains subject profiles

Contains data tabulation datasets and associated files



SDTM xpt files

Define

- define.xml
- define.pdf
- define.xls

METADATA

Annotated CRF (acrf.pdf)

Study Data Reviewer's Guide (sdrg.pdf)

Data Definition File (define.xml) – Overview

Detailed Specification

- Datasets
- Variables
- Codelist
- Derivations / Comments
- Data Origin
- Data Type

Evaluate Data Faster

Critical for FDA systems like JANUS

- Require define before loading study data

High Quality Define is MUST

DEFINE.XML v2.0 suggested

Data Definition File (define.xml) – Components

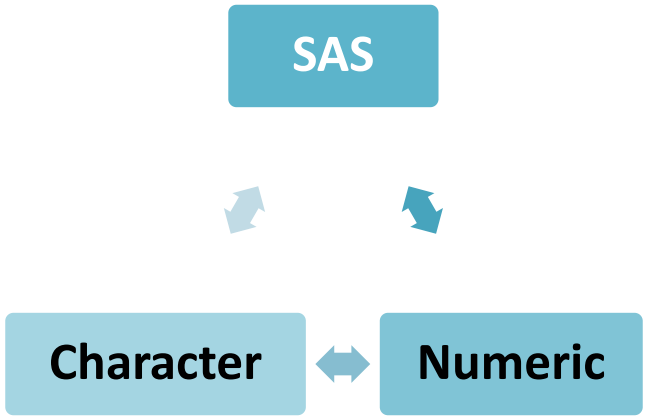
Dataset	Description	Class	Structure	Keys	Location
TE	Trial Elements	Trial Design	One record per element	STUDYID, ETCD	te.xpt
TA	Trial Arms	Trial Design	One record per planned element per arm	STUDYID, ARMCD, ETCD	ta.xpt
TV	Trial Vis				
TI	Trial Inclusio Criteria				
TS	Trial Su				
SE	Subject				
SV	Subject				
DM	Demogr				
CO	Comme				

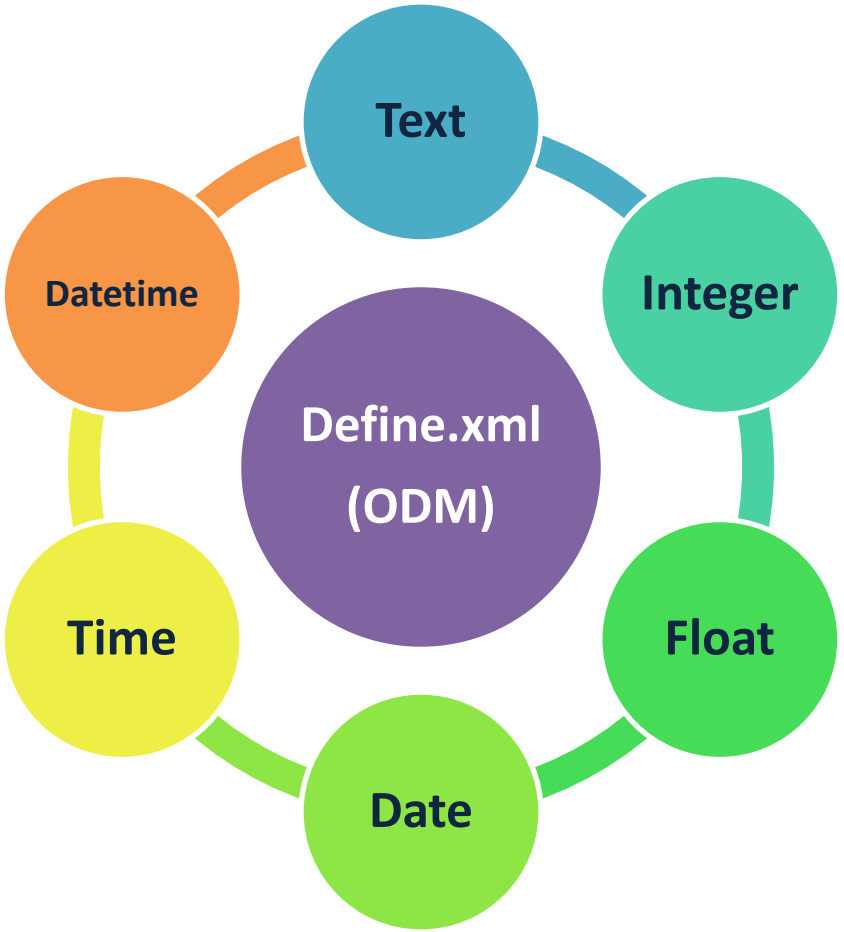
Table of Contents (TOC)

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
STUDYID	Study Identifier	1	text	8		Protocol	
DOMAIN	Domain Abbreviation		text	2		Assigned	
USUBJID	Unique Subject Identifier	2	text	20		Derived	Concatenation of STUDYID and SUBJID (hyphen separated) from the first study the subject participated in.
CESEQ	Sequence Number		integer	8		Derived	Sequential number ensuring uniqueness of records within each USUBJID in the domain.
CESPID	Sponsor-Defined Identifier	5	text	40		Assigned	Identifier created in the SDTM programming to facilitate traceability back to the source operational database.
CETERM	Reported Term for the Clinical Event	3	text	200	CETERM	CRF Page 28	See Study Data Reviewer's Guide for details regarding clinical events.
CELLT	Lowest Level Term		text	200		Assigned	
CELLTCD	Lowest Level Term Code		integer	8		Assigned	
CEDECOD	Dictionary-Derived Term		text	200	CEDECOD	Assigned	
CEPTCD	Preferred Term Code		integer	8		Assigned	
CEHLT	High Level Term		text	200		Assigned	

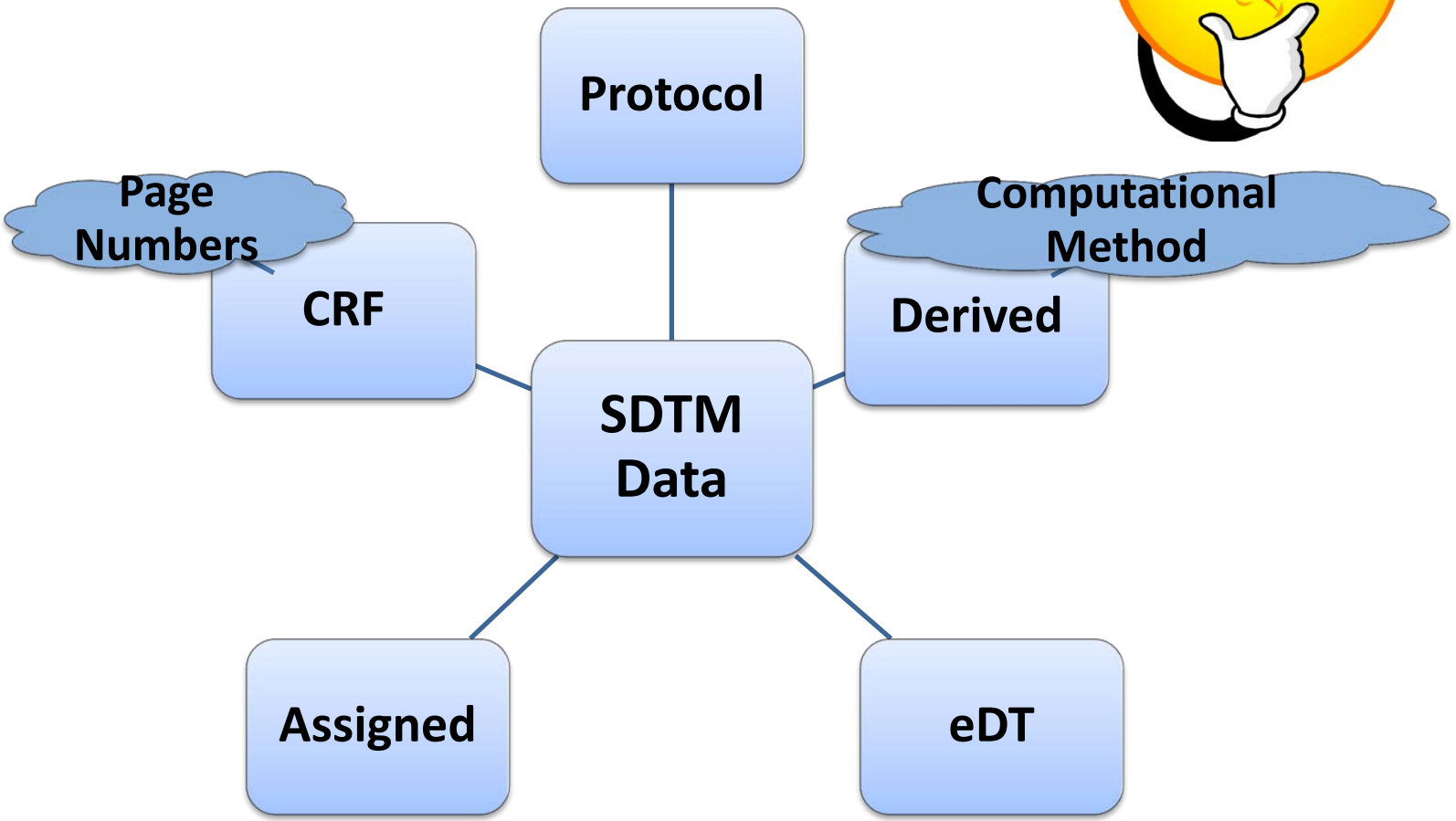
Data Definition Table

Define.xml
Data Type





Define.xml
SDTM Data Origin



Common Issues in Define.xml

Type	Issue
Data Type	• Missing Data Type • Mismatch Data Type • Mismatch Data Type in VLM
Length / Format	• Length not matching actual data • Mismatch: Data Precision • Missing / Incorrect format
Origin	• Missing Origin • Incorrect Origin • CRF Page Number issues when Origin = 'CRF' • Computational Method Issues when Origin = 'Derived'

Microsoft Excel Worksheet

Common Issues in Define.xml

Type	Issue
Codelist / Controlled Terminologies	• Missing codelist • Incorrect / incomplete codelist • Single codelist for multiple variables
Value Level Metadata	• Missing VLM for multiple origin variables • Mismatch: Origin with VLM • Where clause not defined correctly
Derivation Logic / Comments	• Missing Comments • Inappropriate and Vague comments • Incorrect / incomplete derivation logic

Common Issues in Annotated CRF

EDC Bookmarks not removed

Unclean Annotation

Incorrect Annotation

Bookmarks not set to 'Inherit Zoom'

Missing Notes, if any

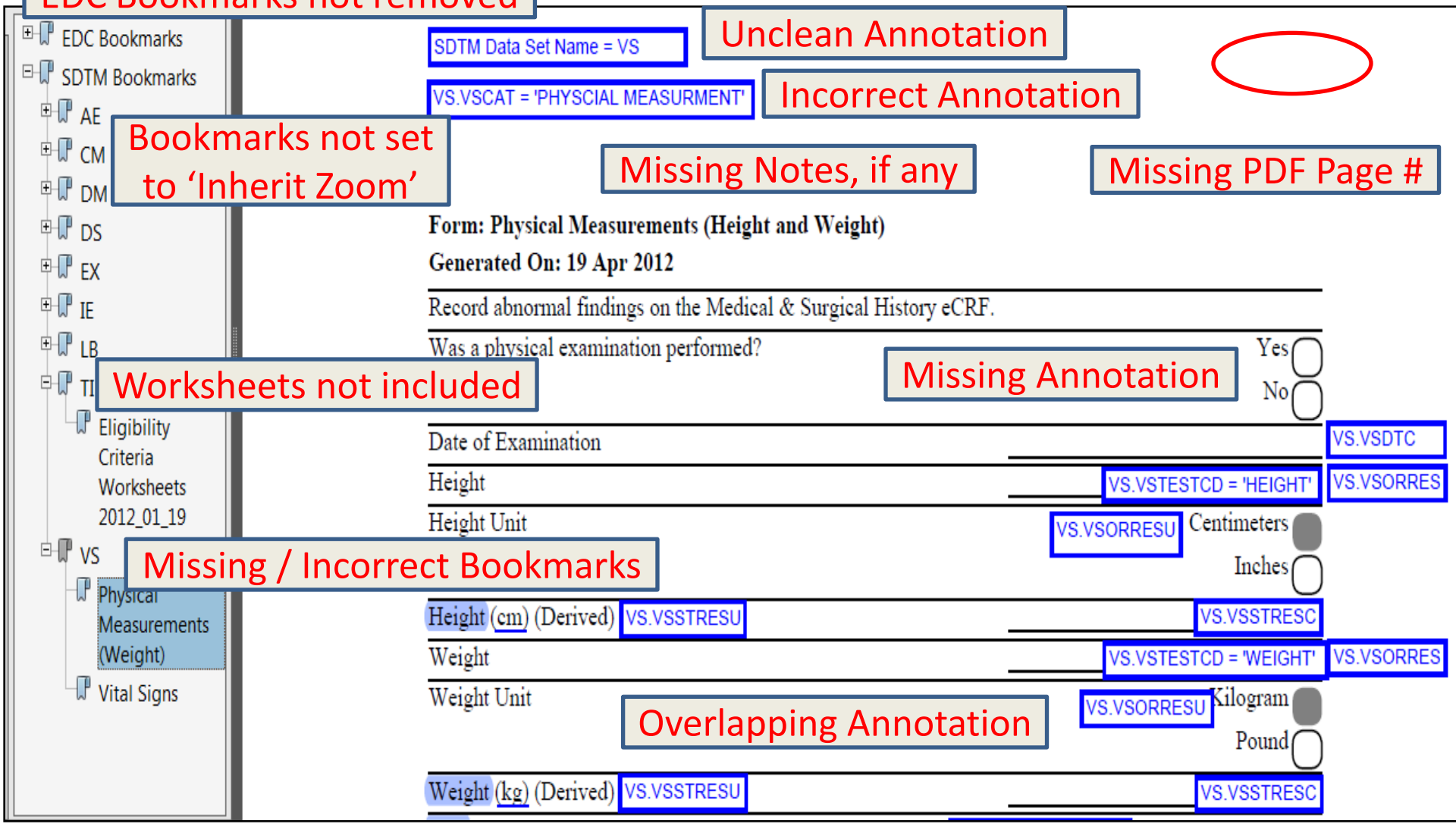
Missing PDF Page #

Worksheets not included

Missing Annotation

Missing / Incorrect Bookmarks

Overlapping Annotation



Study Data Reviewer's Guide (SDRG)

- Introduced by FDA in 2013 as supplemental document to define.xml
- Accommodate additional information regarding sponsor defined domains, mapping decisions, sponsor defined extensions to CDISC controlled terminology etc.
- Act as a **single point study reference** for the FDA
- **Capture unresolved data validation issues with sponsor's justification**

Common Issues in SDRG

All sections of SDRG not completed making it redundant

Comprehensive study summary along with mapping decisions, deviation details etc. not provided

Missing or unclear justification to data conformance issues

Example:

- *'Codelist is extensible'*
- *'Sponsor wished to keep as is'*

Incorrect interpretation of CDISC standards

Example;

Issue: '*Date is after RFPENDTC*'

Sponsor Explanation: '*RFPENDTC is set to End of Study Date in DS domain; which is last study date*'

Incorrect / missing list of variables with no data

Maintain Quality – HOW?

- Adhere to Data Standards
- Perform thorough data validation
- **Execute Pinnacle reports (data checks and define checks) – MUST**
- Address validation findings
- Thorough review of submission documents
  - Links are working
  - Correct details provided etc.
- Develop internal data validation tools

Challenges and Recommendations

Challenges	Recommendation
Creating VLM • Values not present in data • Study Specific values	• Create repository for standard values • Obtain comprehensive spec for vendor data
Identify data type for VLM values • Particularly challenging for large domains like LB	• Develop automation tools to handle these scenarios
Entering CRF page numbers manually • Time consuming • Increase chance of error	• Tools present providing page numbers for variables • Tools can be developed for VLM ???

References

CDER 21st Century Review Process

https://www.google.co.in/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwj_reOIh_TRAhXErI8KHZacDAcQFggZMAA&url=http%3A%2F%2Fwww.fda.gov%2Fdownloads%2FAboutFDA%2FCentersOffices%2FCDER%2FManualofPoliciesProcedures%2FUCM218757.htm&usg=AFQjCNG7hBc4TdlRPUU4LBDG2HbBEbFiOA&bvm=bv.146094739,d.c2I

Providing Regulatory Submission in Electronic Format

https://www.google.co.in/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&uact=8&ved=0ahUKEwj2-tK9h_TRAhUfR48KHWobBzQQFggiMAI&url=http%3A%2F%2Fwww.fda.gov%2Fdownloads%2Fdrugs%2Fguidancecomplianceregulatoryinformation%2Fguidances%2Fucm333969.pdf&usg=AFQjCNHzOOUx73N7V2QAGZubzZRRjqZ6-A&bvm=bv.146094739,d.c2I

Case Report Tabulation Data Definition Specification – CDISC

https://www.google.co.in/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=0ahUKEwi54Kzkh_TRAhUFuY8KHZSvCdAQFggfMAE&url=https%3A%2F%2Fwww.cdisc.org%2Fsystem%2Ffiles%2Fall%2Fstandard_category%2Fapplication%2Fpdf%2Fcrt_ddspecification1_0_0.pdf&usg=AFQjCNGIzUIYBlvZ-P_ZCuOvhnKpvy_OOw&bvm=bv.146094739,d.c2I

The Most Common Issues in Submission Data

https://www.google.co.in/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwiqg6iJiPTRAhVItY8KHeD4BAQQFggZMAA&url=http%3A%2F%2Fwww.pharmasug.org%2Fproceedings%2F2015%2FSS%2FPharmaSUG-2015-SS06.pdf&usg=AFQjCNHXisxwBZ5YIghqYJDPLagNGeHE_g&bvm=bv.146094739,d.c2I

Contact Details

Puja Kawle

Cytel Statistical Software & Services Pvt. Ltd.

(Subsidiary of Cytel Inc.)

puja.kawle@cytel.com


Thank you!