phuse·share
connect·share·advance

Single Day Event

Common/Best Practices of eCRT-Define.xml (Submission Documents)

Balakrishna Lagisetti,
Thermofisher Scientific

Disclaimer and Disclosures

The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of Thermofisher Scientific.

Agenda

Define-xml Introduction

Define-xml Components

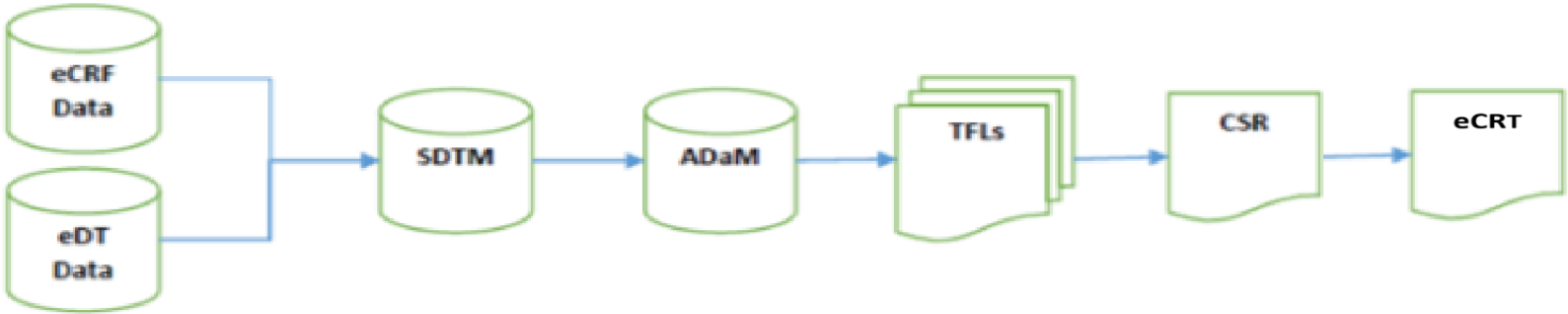
Define-xml Package

Best/Common Practices

Submission Ready PDFs

Define.xml Introduction

The Define-XML package, which provides a truthful representation of clinical trial metadata, is a critical component of clinical submissions to regulatory agencies such as the FDA and PMDA, in accordance with the electronic Common Technical Document (eCTD).



Define.xml Components

Study Level Metadata

Study Name	
Study Description	Dose Ranging Study to Evaluate the Immunogenicity, Safet ...t Influenza Virus Vaccine in Healthy Subjects 18 Years and Above
Protocol Name	V204_01
Metadata Name	CDISC ADaM Data Definitions
Metadata Description	Study Data Definitions

Dataset Level Metadata

Dataset	Description	Class	Structure	Purpose	Keys	Documentation	Location
ADSL [ADaMIG 1.3]	Subject-Level Analysis Dataset	SUBJECT LEVEL ANALYSIS DATASET	One record per subject.	Analysis	USUBJID		adsl.xpt
ADAE [ADaMIG 1.3]	Adverse Event Analysis Dataset	OCCURRENCE DATA STRUCTURE	One record per category per adverse event per term per start date per end date per subject.	Analysis	USUBJID, AECAT, AEDECOD, AETERM, ASTDT, AENDT		adae.xpt

Variable Level Metadata

Variable	Label / Description	Type	Role	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
STUDYID	Study Identifier	text	Analysis	7		Predecessor: DM.STUDYID
USUBJID	Unique Subject Identifier	text	Analysis	17		Predecessor: DM.USUBJID

Define.xml Components

Controlled Terminology (CT)

AGEU	Age Units	text	Analysis	5	Age Unit • "YEARS"
AGEGR1	Pooled Age Group 1	text	Analysis	16	Pooled Age Group 1 • "Born After 1968" • "Born Before 1968"

Value Level Metadata

BASE VLM		Baseline Value	Integer	Analysis	5		Source / Derivations are described per parameter in
	PARAMCD = "H2N3HI" (A/H2N3, Haemagglutination Inhibition (HI) (titer))	A/H2N3, Haemagglutination Inhibition (HI) (titer)	Integer		4		Derived (Source: Sponsor) Assigned to AVAL when ABLFL="Y" per BASETYPE.
	PARAMCD IN ("H2N3ELLA" (A/H2N3, Enzyme-Linked Lectin Assay (ELLA) (titer)), "H2N3MN" (A/H2N3, MicroNeutralization Assay (MN) (titer)))	Result or Finding in Orig Units - Set 3	Integer		5		Derived (Source: Sponsor) Assigned to AVAL when ABLFL="Y".

Analysis Results Metadata (for ADaM Define only):

It is an optional component of ADaM metadata, as specified in the ADaM version 2.1 document. However, the PMDA Technical Conformance Guide states that including ARM in the ADaM define.xml is highly recommended. Although the FDA submission guidelines do not explicitly require ARM, it is increasingly valued by reviewers for facilitating the review of results, datasets, variables, and methods used in the analysis. As a result, submissions that include ARM are often recommended.

Define.xml Package

Annotated aCRF (SDTM Define) and it should be submission ready

SDTM/ADaM Datasets (.xpt)

Define.xml and Style sheet

Reviewer's Guide and it should be submission ready

Programs (optional for ADaM Define)

Analysis Results Metadata (for ADaM Define only)

Additional Definition and Supplemental file and it should be submission ready

Best Practices-DDT/Derivation Comment

- DDT/Derivation comments need to begin with Capital letter.
- Use right framing sentences for all the DDT/Derivation comments and correctly spelled.
- DDT comment shouldn't have reference of raw data names or external data files variables as we can use the label of the variables in correct sentence.
- DDT comment need to be in layman's terms. DO not use SAS languages.
- Ensure to have consistent comment for similar variables across domains (Variables like XXDY, XXSEQ, EPOCH) unless any different derivation used.
- DDT/Derivation comments need to be in Perfect past tenses.
- Ensure to use consistent separator for two statements in DDT/Derivation comment (e.g. using comma or semicolon).
- The DDT/Derivation comment must be present all Derived and Assigned origin variables.

Best Practices-CT/VLM

- Every CT name and CT description should be unique.
- We should add all the data and CRF values in CT for respective variables.
- We should add all the supp variables into QNAM and QLABEL CT even though there is no data.
- We should create separate CT for each variables if the values are different ex: Domain, DOSU, DOSFRM etc..
- We can create one global CT (not mandatory but best practice) if the same values are present in different domains and variables.
- Create CT for all the character result variables for which the values are pre-defined may be in CRF or derivation. Free text fields should not be use for CT.
- Make sure to check NCI codes for standard CTs
- Decoded values should be displayed only when the coded values are unclear. For ex: ARMCD, --TESTCD, --TPTNUM, AGEGRPN, RACEN, etc..
- If there is a codelist for the variable coming from the SDTM dataset, then applicable values from same codelist must be carry forward to ADaM.

Best Practices-CT/VLM

- It is always recommended to create VLM for all the PARAMCDs defined in particular dataset if the derivation is different.

Parameter Value List – adfia [ASTDTM]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Source/Derivation/Comment
ASTDTM	PARAMCD IN ("FIBLD" (Therapy-Blood) , "FIIV" (Therapy-IV Fluids) , "FIPLAT" (Therapy-Platelets))	integer	DATETIME20.		Derived: Time is collected for each blood transfusion or IV replacement therapy start and stop date, also for blood sample collection date for central lab. Set to FAFI.FASTDTC. Convert to numeric.
ASTDTM	PARAMCD IN ("FICACU" (Blood Sample Collected Acute Phase) , "FICCONV" (Blood Sample Collected Convalescent Ph.))	integer	DATETIME20.		Derived: Time is collected for each blood transfusion or IV replacement therapy start and stop date, also for blood sample collection date for central lab. Set to FAFI.FADTC. Convert to numeric.
ASTDTM	PARAMCD IN ("FIACON" (Subject experience fever (>/= 38 degree C) for any 2 out of 3 consecutive days) , "FIAINF" (Febrile Illness) , "FIBLEED" (Evidence of Bleeding) , "FICDF" (Dengue Fever) , "FICDHF" (Dengue Haemorrhagic Fever) , "FICENT" (Blood Sample Collected for Local Lab) ,	integer	DATETIME20.		Assigned: Set to missing.

- CT should be added for each VLM group wherever possible.

Parameter Value List – adfia [PARCAT1]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Source/Derivation/Comment
PARCAT1	PARAMCD = "FIHOSP" (Febrile Illness Require Hosp.)	text	15	["Hospitalization"] <Parameter Category 1 for Febrile Illness Assessments [HOSP]>	Assigned: Set to 'Hospitalization'.
PARCAT1	PARAMCD IN ("FIACON" (Subject experience fever (>/= 38 degree C) for any 2 out of 3 consecutive days) , "FIAINF" (Febrile Illness) , "FIMANI" (Any CNS Manifestations) , "FIORGAN" (Other Signs of Organ Damage))	text	1		Assigned: Set to missing.
PARCAT1	PARAMCD IN ("FIBLD" (Therapy-Blood) , "FIIV" (Therapy-IV Fluids) , "FIPLAT" (Therapy-Platelets) , "FITRANS" (Blood Transfusion or IV Replacement))	text	35	["Blood Transfusion or IV Replacement"] <Parameter Category 1 for Febrile Illness Assessments [BT/IVR]>	Assigned: Set to 'Blood Transfusion or IV Replacement'.
PARCAT1	PARAMCD IN ("FIBLEED" (Evidence of Bleeding) , "FICDF" (Dengue Fever) , "FICDHF" (Dengue Haemorrhagic Fever) , "FIDENG" (Clinically Dengue Case) , "FIECCHY" (Bleeding-Ecchymoses or Purpura) ,	text	22	["CLINICALLY DENGUE CASE", "EVIDENCE OF BLEEDING"] <Parameter Category 1 for Febrile Illness Assessments [CDC/EOB]>	Predecessor: CEFI.CSCAT

Best Practices-Specification

- If the variable is subject CT, then prefix the CT name with CT: in ADaM specification.

  Example: CT: DATEFL

- If only a subset of the CT is applicable, then prefix the CT name with CT_S: in ADaM specification.

  Example CT_S: NY

- If there is no existing CDISC ADaM CT name, then you need to create the name for CT ***(with no prefix added)***.

  Example: Region

- If the variable is directly mapping from SDTM and not modifying values, then source column should be updated with <Domain Name>.<Variable Name>

- If new variable is created with 1-1 mapping from SDTM variable values then source column should be updated with 'Assigned'

SEXN	Sex (N)		integer	1	[1 = "M", 2 = "F"] <Sex (N)>	Assigned: Set to 1, if DM.SEX='M'; Set to 2, if DM.SEX='F'.
RACE	Race		text	41	Race	Predecessor: DM.RACE
RACEC	Race (C)		text	2	Race (C)	Assigned: Set abbreviated race (RACEC) from RACE as follows: 'AI'='AMERICAN INDIAN OR ALASKA NATIVE'; 'As'='ASIAN'; 'B'='BLACK OR AFRICAN AMERICAN'; 'NH'='NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER'; 'W'='WHITE'; 'MR'='MULTIPLE'.
RACEN	Race (N)		integer	1	Race (N)	Assigned: Set to 1, if DM.RACE='AMERICAN INDIAN OR ALASKA NATIVE'; Set to 2, if DM.RACE='ASIAN'; Set to 3, if DM.RACE='BLACK OR AFRICAN AMERICAN'; Set to 4, if DM.RACE='NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER'; Set to 5, if DM.RACE='WHITE'; Set to 6, if DM.RACE='MULTIPLE'.

Best Practices-Supplemental file

- Comment/Documentation section should be updated if any programming notes applicable at dataset level

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
adsl	Subject Level Analysis Dataset	SUBJECT LEVEL ANALYSIS DATASET	One record per subject	Analysis	USUBJID	adsl.xpt	Note: Please exclude records affected by the legal age ICF issue in all source SDTM datasets used in this domain. E.g. delete SDTM records if ADSL.EXCL18FL='Y' and [DS.DSSTDTC / VS.VSDTC / LB.LBDTC / IS.ISDTC / SV.SVSTDTC / SV.SVENDTC / XP.XPDTC / XR.XRDTC] > BD18DT.
adae	Adverse Event Analysis Dataset	OCCURRENCE DATA STRUCTURE	One record per subject per adverse event	Analysis	USUBJID, AEDECOD, ASTDT, AENDT	adae.xpt	Remove records where ADSL.EXCL18FL='Y' and ASTDT > BD18DT.

Best Practices-Specification

- Derivation column should be cleaned by following below rules
  - Should be descriptive (Avoid SAS syntax)
  - Sentence formation using symbols appropriately
  - Variable names should be in UPCASE
  - Variable values should be in quotation marks and per data
  - Use domain name as reference when a variable in one domain using in other domain derivations.

DRYRUNFL	Dry-Run Flag		text	1	["N" = "No", "Y" = "Yes"] <No Yes Response, Subset 1>	Derived: Set to 'Y', if XP.XPORRES = 'YES' where XP.XPTESTCD = 'CONTYN'; Set to 'N', if XP.XPORRES = 'NO' where XP.XPTESTCD = 'CONTYN'.

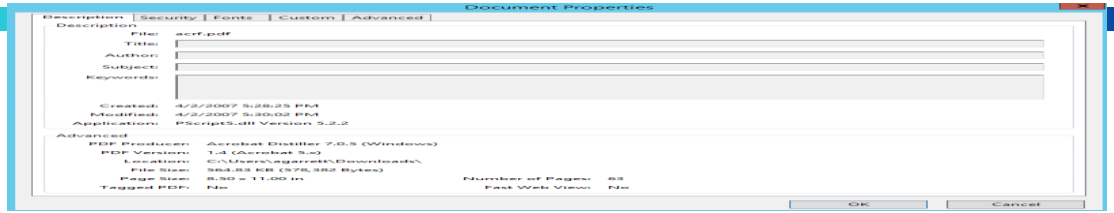
Best Practices-Supplemental file

- When complex algorithms (Supplemental file) is required?
  - Derivation exceed 2000 characters length
  - Same derivation across all datasets (for ex: APHASE, EPOCH)
  - Derivation need bullet point presentation
  - Derivation need tabular format
  - Complex derivations at Value Level Metadata (VLM)

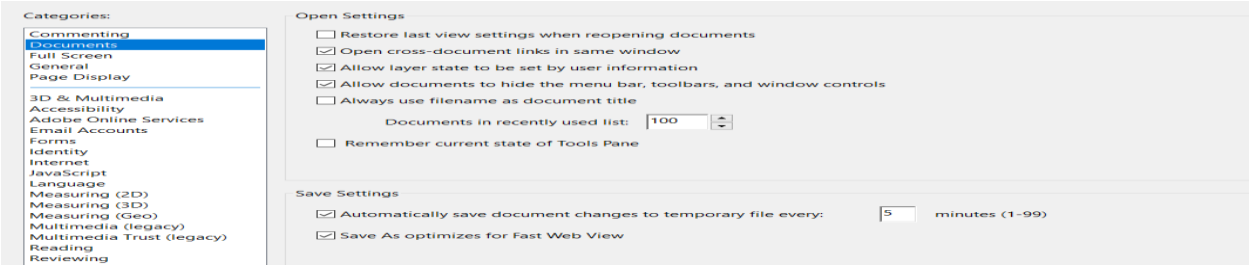
PARAM	Parameter		text	200	Parameter for Febrile Illness Assessments	Derived: Please refer to ADFIA section of the complex algorithms below. Complex Algorithms
PARAMCD	Parameter Code		text	10	Parameter Code for Febrile Illness Assessments	Assigned: Please refer to ADFIA section of the complex algorithms below. Complex Algorithms
PARAMN	Parameter (N)	4	integer	2	Parameter (N) for Febrile Illness Assessments	Assigned: Please refer to ADFIA section of the complex algorithms below. Complex Algorithms

Submission Ready PDFs

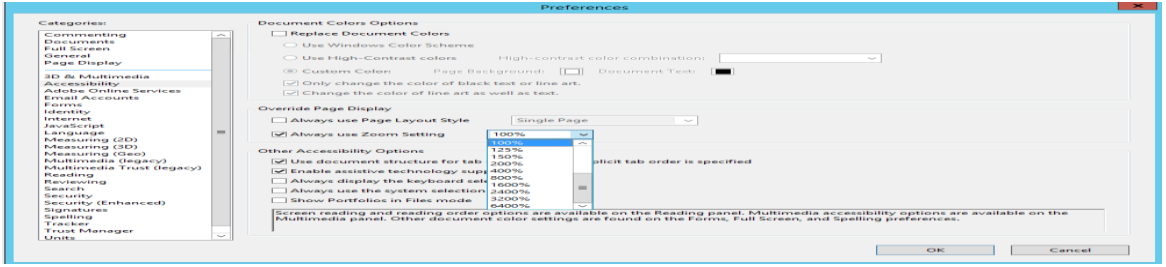
- Update the proper description



- Optimize the fast web view



- Always zoom setting 100%



Submission Ready PDFs

- Optimize pdf

Challenges and Solutions

- Challenge: Challenges in uploading large files.
- Solution: Utilization of the authentic tools for compressing files that cannot compromise the quality
- divide large files into parts so that it can comply with the authoritative requirements.

- Challenge: Challenges in Meeting Tight Deadlines and Resource Limitations
- Solution: We should have proper submission timelines from the beginning of the study going for submission, also allocate the proper resource allocation through out the life cycle of the study.

- Challenges: Regarding Post Submission (missing documents or inappropriate formatting and delayed responses to authorities queries).
- Solution: Using your respective tools or process to identify and provide the quick response. We should also maintain detailed updates or versions for audit trail. We should have exp staff to handle the regulatory requirements.

FDA vs PMDA

- The FDA and PMDA have started accepting electronic study data (e-study data) submissions from 2016. Most of the requirements in e-submissions of PMDA are similar to the FDA's requirement. However, there are some key differences in the requirements between both health authorities
- Below table focuses on focus on key differences between e-submission requirements for the FDA and PMDA

FDA vs PMDA

FDA	PMDA
SDSP: A specific template for a Study Data Standardization Plan is not specified.	A specific template should be used for SDSP.
Conformance checks with severity 'ERROR', 'WARNING' and 'NOTE' Conformance checks with severity 'REJECT', 'ERROR' and 'WARNING' Discrepancies between study data and standard/business rules should be corrected or explained in the Reviewer's Guide.	REJECT: the review will be suspended until corrections have been made. ERROR': without any prior explanations, the review will be suspended until corrections have been made REJECT errors must be corrected Explain un-correctable errors in Reviewer's Guide and discuss at e-data consultation
SAS programs to be submitted with ".txt" file extension in place of ". sas" file extension. Therefore, it is recommended that sponsors change all ". sas" file extensions to ".txt" before loading SAS programs to the eCTD folder.	The PMDA explicitly requests that "The file name should include the extension attached by the analysis software". So, sponsors should not change ". sas" file extensions for SAS programs for submissions to the PMDA.
Programs to create the ADaM datasets	Programs to create the ADaM datasets
Programs to Intermediate datasets	Analyses performed to obtain the important results on efficacy and safety and clinical study results that provide rationales for setting the dosage and administration
define.xml v2.0, v2.1 and/or pdf files, Style sheet (define-2-0-0.xsl and define-2-1.xsl)	Both define.xml v1.0 and v2.0 are accepted
SDRG (filename: csdrg.pdf) ADRG (filename: adrg.pdf)	sdrg.pdf / study-data-reviewers-guide.pdf adrg.pdf /analysis-data-reviewers-guide.pdf
ARM data are optional	ARM is required to support main results of efficacy and safety that provide rationales for setting of the dosage and administration
Dataset > 5 GB to be split in a split folder	To be consulted if sponsors have datasets >= 5 GB
Both split datasets and non-split datasets must be submitted	No requirement to split datasets
The FDA could accept legacy format submission for studies starting before December 16, 2016 and with additional datasets DM, TS and ADSL.	Legacy data conversion is one of the biggest challenges for PMDA submission.

Questions

phuse

The Global Healthcare Data Science Community

Contact Channels
UK +44 1843 609600
office@phuse.global
phuse.global

Social Media
@phusetwitta
/phusebook
/phusetube
/company/phuse