

How to Prepare High-quality Metadata for Submission

Varun Debbeti, Cytel Inc., Waltham, MA, USA

Paper SI12

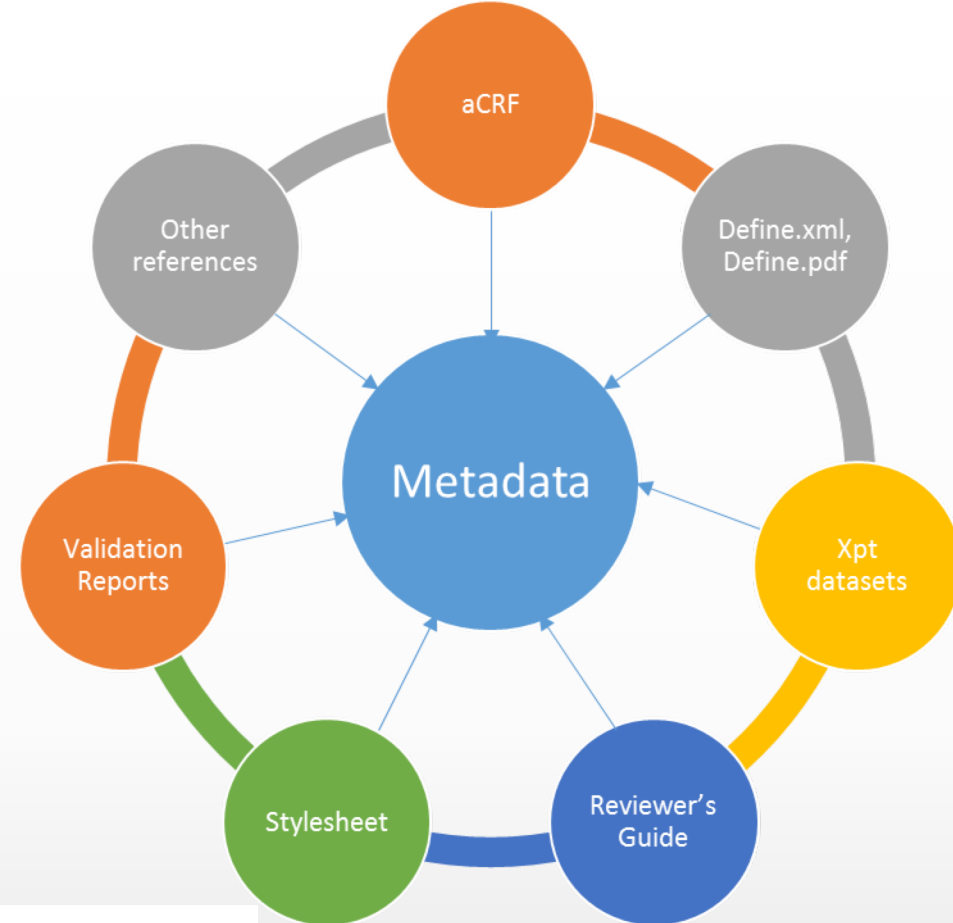


Contents

- What is Metadata?
- How to Prepare High-quality Metadata
- Conclusion
- References

What is Metadata?

- Definition: “structured data about data”
- “Metadata is descriptive information about an object or resource, whether it is physical or electronic”.
- Description could be about content, origin etc.



Metadata Constituents

	STUDYID	DOMAIN	USUBJID	SUBJID	RFSTDTC	RFENDTC	RFXSTDTC	RFXENDTC
1	XXX-XX-101	DM	XXX-XX-101-001-004	101-001-004	2016-03-07T11:27	2016-03-07T11:58	2016-03-02T15:58	2016-03-07T11:58
2	XXX-XX-101	DM	XXX-XX-101-001-005	101-001-005	2016-05-16T09:47	2016-05-16T10:19	2016-05-11T14:59	2016-05-16T10:19
3	XXX-XX-101	DM	XXX-XX-101-001-006	101-001-006	2016-08-03T10:50	2016-08-03T11:16	2016-07-29T21:24	2016-08-03T11:16
4	XXX-XX-101	DM	XXX-XX-101-002-001	101-002-001	2015-05-19T11:18	2015-05-19T11:41	2015-05-14T10:24	2015-05-19T11:41
5	XXX-XX-101	DM	XXX-XX-101-002-003	101-002-003	2015-06-01T11:27	2015-10-13T12:41	2015-05-27T10:59	2015-10-13T12:41
6	XXX-XX-101	DM	XXX-XX-101-002-004	101-002-004	2015-07-01T12:30	2015-07-01T13:00	2015-06-26T09:38	2015-07-01T13:00
7	XXX-XX-101	DM	XXX-XX-101-002-006	101-002-006	2015-12-23T12:50	2016-05-18T11:35	2015-12-18T11:31	2016-05-18T11:35
8	XXX-XX-101	DM	XXX-XX-101-002-007	101-002-007	2016-02-01T13:25	2016-09-22T12:26	2016-01-27T10:00	2016-09-22T12:26

Data

Metadata

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
DM	Demographics	SPECIAL PURPOSE	One record per subject	Tabulation	STUDYID, USUBJID	dm.xpt	

How to Prepare High-quality Metadata



1. Knowledge on standards and latest requirements
2. Efficient preparation and Extensive Validation
3. Interpretation of issues
4. Evaluation of Impact of Issues
5. Fixing Critical Issues
6. Explaining non-fixable issues

Knowledge on Standards and Latest Requirements



FDA Data Standards Catalog v4.10 (10-24-2017) - Supported and Required Standards

This table contains a listing of the data exchange, file formats and terminology standards supported at FDA. These standards have gone through all the steps necessary to make this part of the regulatory review process, including posting of regulatory guidance documents and associated implementation guidelines and technical specifications. The submission of standardized data using any standard not listed, or to an FDA Center not listed, should be discussed with the Agency in advance. This catalog is incorporated by reference in the guidance to industry, *Providing Regulatory Submissions in Electronic format-Standardized Study Data* (<http://www.fda.gov/downloads/Drugs/Guidances/UCM292334.pdf>).

Use	Data Exchange Standard	Exchange Format	Standards Development Organization (SDO)	Supported Version	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	Regulatory Reference and Information Sources
Regulatory Applications (IND, NDA, ANDA, BLA, master files)	Electronic Common Technical Document (eCTD)	Extensible Markup Language (XML)	International Council for Harmonisation (ICH)	3.2.2	M2 eCTD: Electronic Common Technical Document Specifications	CDER, CBER	06/01/2008		05/05/2017 [5] 05/05/2018 [6]		Electronic Submissions-Electronic Common Technical Document (eCTD)
Clinical and Non-Clinical study data sets - Transport	SAS Transport (XPORT)	XPT	SAS	5	SAS Technical Support TS-140	CDER, CBER	Ongoing		12/17/2016 [1] 12/17/2017 [2]		For CDER and CBER only: Technical Conformance Guide
Clinical study datasets	Study Data Tabulation Model (SDTM)	XPT	Clinical Data Interchange Standards Consortium (CDISC)	1.1	3.1.1	CDER, CBER	Ongoing	01/28/2015		01/28/2015	CDISC.org - SDTM See Technical Conformance Guide
Clinical study datasets	SDTM	XPT	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]	03/15/2019 [1] 03/15/2020 [2]	CDISC.org - SDTM
Clinical study datasets	SDTM	XPT	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]	03/15/2019 [1] 03/15/2020 [2]	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]		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]		CDISC.org - SDTM
Clinical study datasets	Analysis Data Model (ADaM)	XPT	CDISC	2.1	1.0	CDER, CBER	Ongoing	03/15/2019 [1] 03/15/2020 [2]	12/17/2016 [1] 12/17/2017 [2]	03/15/2019 [1] 03/15/2020 [2]	CDISC.org - ADaM
Clinical study datasets	Analysis Data Model (ADaM)	XPT	CDISC	2.1	1.1	CDER, CBER	03/15/2018		03/15/2019 [1] 03/15/2020 [2]		CDISC.org - ADaM
Study data definition	Define	XML	CDISC	1.0	N/A	CDER, CBER, CDRH	Ongoing	03/15/2018 [1] 03/15/2019 [2]	12/17/2016 [1] 12/17/2017 [2]	03/15/2018 [1] 03/15/2019 [2]	CDISC.org - Define-XML
Study data definition	Define	XML	CDISC	2.0	N/A	CDER, CBER, CDRH	08/07/2013		12/17/2016 [1] 12/17/2017 [2]		CDISC.org - Define-XML

Knowledge on Standards and Latest Requirements



The Catalog lists

- currently supported and/or required standards
- the FDA support begin/end date
- the requirement begin/end date

Any study starting after 17DEC2016 for NDA, BLA and ANDAs should follow the set standards in order to avoid refuse to file (RTF).

Use	Data Exchange Standard	Exchange Format	Standards Development Organization (SDO)	Supported Version	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
Regulatory Applications (IND, NDA, ANDA, BLA, master files)	Electronic Common Technical Document (eCTD)	Extensible Markup Language (XML)	International Council for Harmonisation (ICH)	3.2.2	M2 eCTD: Electronic Common Technical Document Specifications	CDER, CBER	06/01/2008		05/05/2017 [5] 05/05/2018 [6]	
Clinical and Non-Clinical study data sets - Transport	SAS Transport (XPORT)	XPT	SAS	5	SAS Technical Support TS-140	CDER, CBER	Ongoing		12/17/2016 [1] 12/17/2017 [2]	
Clinical study datasets	Study Data Tabulation Model (SDTM)	XPT	Clinical Data Interchange Standards Consortium (CDISC)	1.1	3.1.1	CDER, CBER	Ongoing	01/28/2015		01/28/2015
Clinical study datasets	SDTM	XPT	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]	03/15/2019 [1] 03/15/2020 [2]
Clinical study datasets	SDTM	XPT	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]	03/15/2019 [1] 03/15/2020 [2]
Clinical study datasets	SDTM	XPT	CDISC	1.3	3.1.3	CDER, CBER	12/01/2012		12/17/2016 [1] 12/17/2017 [2]	
Clinical study datasets	(SDTM)	XPT	CDISC	1.4	3.2	CDER, CBER	08/17/2015		03/15/2018 [1] 03/15/2019 [2]	
Clinical study datasets	Analysis Data Model (ADaM)	XPT	CDISC	2.1	1.0	CDER, CBER	Ongoing	03/15/2019 [1] 03/15/2020 [2]	12/17/2016 [1] 12/17/2017 [2]	03/15/2019 [1] 03/15/2020 [2]
Clinical study datasets	Analysis Data Model (ADaM)	XPT	CDISC	2.1	1.1	CDER, CBER	03/15/2018		03/15/2019 [1] 03/15/2020 [2]	
Study data definition	Define	XML	CDISC	1.0	N/A	CDER, CBER, CDRH	Ongoing	03/15/2018 [1] 03/15/2019 [2]	12/17/2016 [1] 12/17/2017 [2]	03/15/2018 [1] 03/15/2019 [2]
Study data definition	Define	XML	CDISC	2.0	N/A	CDER, CBER, CDRH	08/07/2013		12/17/2016 [1] 12/17/2017 [2]	

<https://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM340684.xlsx>

How to Prepare High-quality Metadata



1. Knowledge on standards and latest requirements
2. Efficient preparation and Extensive Validation
3. Interpretation of issues
4. Evaluation of Impact of Issues
5. Fixing Critical Issues
6. Explaining non-fixable issues

Metadata Preparation/Validation

Guidance documents that help in preparation and validation

- Define.xml schema validation
- Define.xml Pinnacle 21 validation
- FDA Business run/Validator rules
- SDTM-Metadata Submission Guidelines
- Define-XML-2-0-Specification
- eCTD Technical Conformance Guide
- Technical Rejection Criteria
- Study Data - Technical Conformance guide (TCG)
- Other manual checks

Define.xml schema validation

- First step is to validate the structure of the define.xml
- CDISC provides schema files to be used by validator tools <http://www.cdisc.org/define-xml>
- External websites for some XML tools evaluated by CDISC E.g.,
 - DefineValidator <https://www.phaseforward.com/products/cdisc/>
 - Oxygen XML Editor <http://www.oxygenxml.com>
 - Eclipse <http://www.eclipse.org/>
 - Cladonia XMLEXchanger <http://www.exchangerxml.com>
 - Liquid XML <http://www.liquid-technologies.com/>
 - Validome <http://www.validome.org>

Define.xml Pinnacle 21 validation

- It is recommended to validate datasets in relation to define.xml
- Pinnacle 21 community can validate
 - Datasets alone
 - Define.xml alone
 - Datasets in relation to define.xml

4. Data Conformance Summary

4.1 Conformance Inputs

Was OpenCDISC used to evaluate conformance? Yes/No

If yes, specify the versions of OpenCDISC and the OpenCDISC validation rules:

[Text Here]

Were sponsor-defined validation rules used to evaluate conformance? Yes/No

If yes, describe any significant sponsor-defined validation rules:

[Text Here]

Were the SDTM datasets evaluated in relation to define.xml? Yes/No

Was define.xml evaluated? Yes/No

Provide any additional compliance evaluation information:

*Study-data-reviewers-guide template from
PhUSE (released as SDRG Package v1.2)*

Pinnacle 21 ID	Publisher ID	Message	Description	Category	Severity
DD0001		XML schema validation issue within Define.xml	There is an XML schema validation issue with either Define.xml structure or content that is not covered by one of the other rules. It is most likely an issue with element declarations.	Structure	Error

Define.xml Pinnacle 21 validation

Pinnacle 21 Community

File View Help

Home

Validator

Define.xml

Converter

ClinicalTrials.gov

Validator *check compliance with SDTM, SEND, ADaM, and Define.xml*

✓ Validate Data

Standard SDTM Configuration SDTM 3.2

Source Format SAS® Transport (XPORT) Report Format Excel

Source Data

Drop files here to select them
or
Browse...

Define.xml Optional Browse...

CDISC CT 2016-06-24 MedDRA 19.0 More dictionaries

Validate

© 2018 Pinnacle 21 LLC

Follow us: LinkedIn Twitter Facebook Email

Pinnacle 21
Community tool

FDA Business rules/Validator rules

- FDA Business rules: Published Mar, 2017
 - Developed to reinforce internal business processes.
 - Supersede previously published validation rules.
- FDA validation rules: Published Nov, 2014
 - The validator rules are technical version of business rules
 - The validator rules were published for transparency purposes.
- One may not have to read or remember all these rules. Tools like P21 cover these issues under its rules set.

FDA Business rules/Validator rules

FDA Business Rule ID	FDA Business Rule	FDA Validator Rule	Domains
FDAB035	The definition of datasets, variables, and codelists in define.xml must reflect the actual study data.	Datasets included in study data must be described in the data definition document (define.xml).	ALL
		Domains referenced in data definition document (define.xml) should be included in the submission.	ALL
		Variables listed in the data definition document (define.xml) should be included in the dataset.	ALL
		Variables included in the dataset must be described in the data definition document (define.xml).	ALL
		Variable Data Types in the dataset must match the variable data types described in the data definition document (define.xml).	ALL
		Variable values should be populated with terms found in the user-defined codelist associated with the variable in define.xml.	ALL

Pinnacle 21 ID	Publisher ID	Message	Description	Category	Severity
SD1063	FDAC023	Dataset is not present in define.xml	Datasets included in study data must be described in the data definition document (define.xml).	Metadata	Error
SD0061	FDAC024	Domain referenced in define.xml but dataset is missing	Domains referenced in data definition document (define.xml) should be included in the submission.	Metadata	Warning
SD0054	FDAC025	Variable in define.xml is not present in the dataset	Variables listed in the data definition document (define.xml) should be included in the dataset.	Metadata	Warning
SD0060	FDAC026	Variable in dataset is not present in define.xml	Variables included in the dataset must be described in the data definition document (define.xml).	Metadata	Error
SD0059	FDAC034	Define.xml/dataset variable type mismatch	Variable Data Types in the dataset must match the variable data types described in the data definition document (define.xml).	Metadata	Error
SD0037	FDAC037	Value for variable not found in user-defined codelist	Variable values should be populated with terms found in the user-defined codelist associated with the variable in define.xml.	Terminology	Error

FDA Business rules/Validator rules

- While most of the business rules/validator rules are covered by P21, there are a few issues that are not covered
- Issue: **Sequence of model permissible variables** in data which are **not matching** in define.xml are not found by P21 community
- e.g., When QSMETHOD is added to QS domain, and the order of QSMETHOD in define is different in define and dataset, it is not identified by P21 community tool.

SDTM-Metadata Submission Guidelines

- Purpose : To provide guidance for compiling the eCTD module 5 “sdtm” folder.
- Domain Name vs Dataset name
 - A dataset is just a collection of data with rows and columns.
 - A domain is defined as per the SDTM model.
- As per SDTM IG 3.2 “A domain may be comprised of more than one physical dataset” For e.g., QS domain = QS and SUPPQS datasets.
- Split domains
 - Annotated CRF: The annotated CRF refers to the domain name (QS) as opposed to the dataset name (QSCG, QSCS, or QSMM).
 - DOMAIN variable value should be QS (i.e., DOMAIN = “QS”) in all split datasets e.g, QSCG, QSCS or QSMM.
 - the define.xml controlled terminology-code list for Domains to include 4 letter split dataset names.
- Advise: Sponsor should check with their review agency regarding exactly **what needs to be included in the submission**, i.e. the split datasets or both the split datasets and the un-split datasets.

Define-XML-2-0-Specification



CDISC Define-XML Specification Version 2.0

Prepared by

CDISC Define-XML Team

- Purpose : The specification describes the Define-XML model that is used to describe SDTM/ADaM datasets for submission to FDA

Define-XML-2-0-Specification

- Dataset level: Natural keys describe the structure of each dataset. “A *natural key is a piece of data that uniquely identify and distinguish observation from any other observation in the dataset.*”
- Natural keys have business meaning, natural keys are often coupled to the business. Example:
 - Sponsor-A natural keys: STUDYID, USUBJID, VISTNUM, PETESTCD
 - Sponsor-B natural keys: STUDYID, USUBJID, PEDTC, PETESTCD, PELOC, PEMETHOD

Tabulation Datasets for Study █████ (SDTM-IG 3.2)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
TA	Trial Arms	TRIAL DESIGN	One record per planned element per arm	Tabulation	STUDYID, ARMCD, TAETORD	ta.xpt	
TE	Trial Elements	TRIAL DESIGN	One record per planned element	Tabulation	STUDYID, ETC	te.xpt	
TI	Trial Inclusion/Exclusion Criteria	TRIAL DESIGN	One record per I/E criterion	Tabulation	STUDYID, IETESTCD, TIVERS	ti.xpt	

Define-XML-2-0-Specification

- Natural keys may not always match standard keys used
- Types of duplicate records due to fixed standard keys
 - True double entry: Two or more records with same keys (USUBJID, VISTNUM, PETESTCD) and only PEORRES is different
 - Insufficient standard Keys: The above example of PE falls under this category, and should be added in define.xml for easy identification by reviewers.
 - Incorrect standard Keys: Example: Sponsor Key variables USUBJID FACAT FASCAT FATESTCD VISITNUM, standard keys are USUBJID FATESTCD FAOBJ –should be fixed before submission
 - Missing values in keys: Example: Standard keys are STUDYID USUBJID DATESTCD and DADTC in DA domain. If the data has missing values of DADTC, multiple records will have same DADTC (i.e., missing) as per Pinnacle tool and are identified as duplicates.

Tabulation Datasets for Study [REDACTED] (SDTM-IG 3.2)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
TA	Trial Arms	TRIAL DESIGN	One record per planned element per arm	Tabulation	STUDYID, ARMCD, TAETORD	ta.xpt	
TE	Trial Elements	TRIAL DESIGN	One record per planned element	Tabulation	STUDYID, ETCD	te.xpt	
TI	Trial Inclusion/Exclusion Criteria	TRIAL DESIGN	One record per I/E criterion	Tabulation	STUDYID, IETESTCD, TIVERS	ti.xpt	

6/13/18

Cytel Inc.

18

Study Data - Technical Conformance guide (TCG)

- **Traceability Issues:** Traceability permits an understanding of the relationships between the analysis results and datasets
- Report → Define.xml → ADaM → SDTM → CRF.
- Traceability has been a proven issue for Legacy data conversions.
- **Legacy Data:** *Study data in a format that is not supported in FDA review process, and not ever listed in the FDA data standards catalog.*
- Sponsors should use processes for legacy data conversion that account for traceability.
- As per TCG, there are issues like “**Limited traceable path from SDTM to the ADaM datasets**” and/or “**Limited ability to replicate ADaM datasets using SDTM datasets**” when legacy study data and legacy analysis data are independently converted to SDTM and ADaM formats, respectively

Study Data - Technical Conformance guide (TCG)

- A tool to check if all raw data is mapped to SDTM was developed
- It read all existing SDTM logs to find input data library references
- An excel file output will be generated with columns for 'Raw FILE NAME' and 'SDTM DOMAIN'.
- Refer to paper for code

	A	B	C	D	E	F
1	FILENAME	NAME	TYPE	LABEL	FORMAT	DOMAIN
2	AB	ASSAYNO	char	Assay Number	\$20.	DM,IS,LB
3	AE	ACNOTSP	char	Other specify	\$600.	DM
4	AETARG	ACNOTSP	char	Other specify	\$600.	DM
5	ANTIBSA	ASSAYNO	char	Assay Number	\$28.	NOT SUBMITTED
6	ANTITRT	ASCT	char	Conditioning chemotherapy for ASCT	\$9.	DM
7	BSYMPT	DATAPAGEID	num	Internal id for data page	10.	DM,DOV,FA
8	BSYMPT2	DATAPAGEID	num	Internal id for data page	10.	DM,DOV,FA
9	CHEMLABS	BASEHIGH	num	Baseline Grade for Positive Shift		LB

How to Prepare High-quality Metadata



1. Knowledge on standards and latest requirements
2. Efficient preparation and Extensive Validation
3. Interpretation of issues
4. Evaluation of Impact of Issues
5. Fixing Critical Issues
6. Explaining non-fixable issues

Interpret Validation Results

- P21 Interpretation of severity (like ‘Error’, ‘Warning’, and ‘Notice’).
 - errors as issues reported with 100% confidence (e.g., AE start date is after the latest Disposition date)
 - warnings as potential issues requiring manual review (e.g., EXDOSU value not found in 'Unit' extensible codelist).
- Review agencies like FDA interprets
 - Error/Warning as issues with severe impact on review process
 - Reject as critical issues that prevent review and automation processes.
- Example: FDA Business Rules uses words like “must” and “should”, P21 reports these issues with a severity as ‘Warning’ but not as an ‘Error’ or ‘Reject’.
- Despite having differences in interpretation, sponsors are encouraged to fix as many issues that are fixable

FDA Business Rule ID	FDA Business Rule	FDA Validator Rule	Domains	SDTM 3.1.2	SDTM 3.1.3	SDTM 3.2
FDAB001	A treatment-emergent flag must be submitted.	A treatment-emergent flag should be included in SUPPAE according to SDTM IG v3.1.2 #8.4.3.	SUPPAE	X	X	X
FDAB022	EPOCH should be included for clinical subject-level observations (e.g., adverse events, laboratory, concomitant medications, exposure, and vital signs).	Variables requested by FDA in policy documents should be included in the dataset. E.g., EPOCH and ELEMENT.	ALL	X	X	X

SD1077	FDAC021	FDA Expected variable EPOCH not found	Warning	1
SD1097	FDAC022	No Treatment Emergent info for Adverse Event	Warning	6794

How to Prepare High-quality Metadata



1. Knowledge on standards and latest requirements
2. Efficient preparation and Extensive Validation
3. Interpretation of issues
4. Evaluation of Impact of Issues
5. Fixing Critical Issues
6. Explaining non-fixable issues

Evaluate Impact of Issues on Data

- Meetings with review divisions (e.g., CDER and CBER) play a crucial role on approvals
- Meetings can be organized at different stages of drug development process.
 - Pre-IND meetings
 - End-of-phase2 meetings
 - Pre-submission meetings
 - Post-submission meetings
 - Adhoc meetings to resolve uncertainty about how some issues are going to affect their approval
 - Type C meeting to discuss substantive data standardization issues for NDAs and BLAs.
Example: a sponsor's desire to use a standard (e.g., therapeutic area standard in SDTM format) that is not currently supported by FDA
- “Sponsors and applicants may submit technical questions related to data standards at any time to the technical support team identified by each Center” - Providing Regulatory Submissions In Electronic Format — Standardized Study Data

How to Prepare High-quality Metadata



1. Knowledge on standards and latest requirements
2. Efficient preparation and Extensive Validation
3. Interpretation of issues
4. Evaluation of Impact of Issues
5. Fixing Critical Issues
6. Explaining non-fixable issues

Fix Critical Issues

- Critical issues if not fixed leads to rejection
- Study data standards catalog must be followed for studies that start after December 17, 2016.
- Technical rejection criteria is additional to the existing validation criteria
- There are two rules so far.
 - Rule #1734 – Trial Summary (TS) dataset must be present for each study in Module 4 and 5
 - Rule # 1736 – Demographic (DM) dataset, Subject level analysis dataset (ADSL) and define.xml must be submitted in Module 5 for clinical data
- The above 2 rules are covered by P21 tool set as below.

Pinnacle 21 ID	Message	Description	Category	Severity
SD1020	Missing DM dataset	Demographics (DM) dataset must be included in every submission.	Presence	Error
AD0001	Missing ADSL dataset	ADaM Subject level (ADSL) dataset should be included in every submission.	Presence	Error
SD1115	Missing TS dataset	Trial Summary (TS) dataset must be included in every submission.	Presence	Error
DD0101	Missing Define.xml	FDA eCTD submissions must include a define.xml file for each study in Module 4 (nonclinical) and Module 5 (clinical)	Presence	Error

How to Prepare High-quality Metadata



1. Knowledge on standards and latest requirements
2. Efficient preparation and Extensive Validation
3. Interpretation of issues
4. Evaluation of Impact of Issues
5. Fixing Critical Issues
6. Explaining non-fixable issues

Explain Non-Fixable Issues



- Explain non-fixable issues in reviewers guide (RG) (optional, but recommended)
- RG helps understanding of complex issues associated with data review and validation.

Issue types

- CRF design or data collection issues - should be explained Section 3.3 in clinical SDRG or section 5.2 in ADRG.
- Mapping issues could be mentioned in section 4.1 of SDRG.
 - Example1: Compound adverse events (that are combination of simple adverse events) are mapped to a **Custom Events domain** (e.g., XC), and it is used as input to ADAE along with AE, then P21 report on ADAE may throw errors due to the domain variable having values other than AE.
 - Example 2: **Study day 0** is not ideal as per SDTM model, but some sponsors may prefer to have study day zero for treatment onset day.

Sponsors are encouraged to include and explain issues generated on **Define xml alone, and data + define.xml issues** to be added to reviewer's guides.

Conclusion



- It is important to have a balance between use of standard guidance from FDA, and automatic tools like Pinnacle 21 for creation of high quality metadata.
- All sponsors should educate their Standards implementation teams about the available resources and how they could be used to have successful approvals for submissions for regulatory review.

References

- [1] Data standards catalog <https://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm>
- [2] Technical rejection criteria <https://www.fda.gov/downloads/drugs/developmentapprovalprocess/formssubmissionrequirements/electronic submissions/ucm523539.pdf> .
- [3] Providing Regulatory submissions <https://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm>
- [4] Define.xml schema validation https://www.cdisc.org/system/files/all/standard_category/application/pdf/definereport_v1_0.pdf
- [5] SDRG release package V1.2 http://www.phusewiki.org/wiki/index.php?title=Study_Data_Reviewer%27s_Guide
- [6] FDA Business rules/Validator rules <https://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm>
- [7] SDTM-Metadata submission guidelines <https://www.cdisc.org/standards/foundational/study-data-tabulation-model-implementation-guide-sdtmig/metadata-submission>
- [8] SDTM IG 3.2 <https://www.cdisc.org/standards/foundational/sdtmig>
- [9] Define-XML-2.0-Specification <http://www.phusewiki.org/wiki/images/8/89/Define-XML-2-0-Specification.pdf>
- [10] eCTD Technical conformance guide <https://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm>
- [11] Study data – Technical conformance guide <https://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm>

Questions

