Validating Study Data

LPO Support Webinar

20 May 2019

NIH NATIONAL CANCER INSTITUTE

Upon Completion of This Webinar You Should Be Able To

- Briefly explain the use of validation tools to support FDA submissions
- Find and use free validation tools
- Describe different sources for validation rules on which the validation tools are based
- Apply good practices for reviewing, dispositioning and documenting validation report results
- Explain what validation tools can and can’t do
- Describe and apply practices LPOs can follow to support the creation of high quality SDTM submission data

Focus of this Webinar

- How validation is related to ***conformance*** to ***CDISC standards*** and ***compliance*** with ***FDA and PMDA requirements***
- ***What LPOs can do*** to facilitate successful SDTM (ADaM and SEND) preparation (even if the LPO is *not* doing the SDTM preparation…)

- Out of Scope:
  - Foundational principles of CDISC Standards
  - eCTD
  - Other standards (e.g., MedDRA, LOINC)

Why Validate?

- *FDA Rule: "Sponsors whose studies started after **December 17, 2016, must use** the data standards listed in the FDA Data Standards Catalog for NDAs, BLAs and ANDAs. For Commercial INDs, the requirement applies to studies started after **December 17, 2017**."*
- FDA (and PMDA) **will** validate your data to determine whether you are in compliance to the above Rule.
- For any issue they identify during their validation, FDA will expect to see an explanation from the Sponsor
  - Study Data Reviewer's Guide (SDRG)
  - Analysis Data Reviewer's Guide (ADRG)
  - Non-clinical Study Data Reviewer's Guide (nSDRG)
- For PMDA - certain validation errors will cause them to REJECT the application

High Level View of FDA Validation Process

Step 1: Technical Rejection Criteria validation

- Timing: Before the official submission
- Checks *very high level* conformance:
  - eCTD
  - CDISC standards

Step 2: Pinnacle 21 Enterprise validation

- Timing: After Successful Pass for Technical Rejection Criteria, *at official submission*
- Checks *detailed* conformance:
  - SDTM + Define.xml + CT
  - ADaM + Define.xml + CT
  - SEND + Define.xml + CT
  - FDA Business Rules

Step 1: Technical Rejection Criteria

- Process by which FDA can assess the overall quality of a submission *before* it goes through the actual submission process
- Criteria are published and periodically updated:
  - https://www.fda.gov/industry/study-data-standards-resources/study-data-submission-cder-and-cber
- Current version is ~10 pages
- Applies to eCTD Sections 4.2 (CDER) and 5.3 (CDER and CBER)
- Currently two rules that are directly related to study data:
  - Must include a valid DM dataset (DM.xpt) and ADSL, plus associated metadata (Define.xml)
  - Must include valid Study Start Date record in the Trial Summary domain (TS.xpt)

Study Start Date helps FDA determine which standards / versions apply to the study

Example: DM.xpt and ADSL Files Technical Rejection Criterion

Table 4: Validation 1736

Number:	1736
Group:	General
Description:	For Standard for Exchange of Nonclinical Data (SEND) data, a Demographic (DM) dataset and define.xml must be submitted in Module 4, sections 4.2.3.1, 4.2.3.2, 4.2.3.4 For Study Data Tabulation Model (SDTM) data, a DM dataset and define.xml must be submitted Module 5, sections 5.3.1.1, 5.3.1.2, 5.3.3.1, 5.3.3.2, 5.3.3.3, 5.3.3.4, 5.3.4, 5.3.5.1, 5.3.5.2 For Analysis Data Model (ADaM) data, an ADaM Subject level analysis dataset (ADSL) dataset and define.xml must be submitted in Module 5, sections 5.3.1.1, 5.3.1.2, 5.3.3.1, 5.3.3.2, 5.3.3.3, 5.3.3.4, 5.3.4, 5.3.5.1, 5.3.5.2
Severity Description:	High
US DTD Version	2.01 and 3.3
Effective Date:	TBD
Problem:	You have not submitted SEND DM and corresponding define.xml for each study in Module 4, section 4.2. You have not submitted SDTM DM and corresponding define.xml for each study in Module 5, section 5.3. You have not submitted ADSL, and corresponding define.xml for each study in Module 5, section 5.3.
Corrective Action:	Resubmit the submission with the SEND DM and corresponding define.xml for each study in Module 4, section 4.2 Resubmit including SDTM DM and corresponding define.xml for each study in Module 5, section 5.3 Resubmit including ADSL and corresponding define.xml for each study in Module 5, section 5.3
Guidance Source:	*Providing Regulatory Submissions in Electronic Format—Standardized Study Data; Study Data Technical Conformance Guide*

CDER only

CDER and CBER

Technical Rejection Criteria: What LPOs Can Do

- Understand the Technical Rejection Criteria
  - Read the document published by FDA (updated periodically):
  - https://www.fda.gov/industry/study-data-standards-resources/study-data-submission-cder-and-cber
- LPO collects and sends all participants' Demographics (DM) data for all studies to the SDTM programmers (with no imputed dates)
  - SDTM Programmers ensure the DM domain is valid - following all relevant rules for SDTM and FDA
    - *Includes one record for each study participant*
    - *Core designations (Include all required / expected variables)*
    - *Proper formats and rules(e.g., Dates should be in ISO 8601 and not imputed)*
    - *Using CDISC Submission Values for CT (e.g., M, F)*
    - *Correct file and variable naming conventions, etc.*

Example: SDTMIG TS.XPT Technical Rejection Criterion

Table 2: Validation 1734

Number:	1734
Group:	General
Description:	A dataset named ts.xpt with information on SSD must be present for each study in Module 4, sections 4.2.3.1, 4.2.3.2, 4.2.3.4, and in Module 5, sections 5.3.1.1, 5.3.1.2, 5.3.3.1, 5.3.3.2, 5.3.3.3, 5.3.3.4, 5.3.4, 5.3.5.1, 5.3.5.2
Severity Description:	High
US DTD Version	2.01 and 3.3
Effective Date:	TBD
Problem:	You have not submitted a dataset named ts.xpt with information on SSD for each study in Module 4, section 4.2, or in Module 5, section 5.3
Corrective Action:	Resubmit, including a dataset named ts.xpt with information on SSD for each study in Module 4, section 4.2, and Module 5, section 5.3
Guidance Source:	*Providing Regulatory Submissions in Electronic Format—Standardized Study Data; Study Data Technical Conformance Guide*

Technical Rejection Criteria: What LPOs Can Do

- Understand the Technical Rejection Criteria
  - Read the document published by FDA (updated periodically):
  - https://www.fda.gov/industry/study-data-standards-resources/study-data-submission-cder-and-cber
- LPO clearly communicates correct Study Start Date for each study to the SDTM Programmers

Study Start Date is defined in CDISC CT *(NCI C69208)* as: *The earliest date of informed consent among any subject (Date/Time of Informed Consent, RFICDTC) that enrolled in the study. For studies conducted without informed consent (ie. emergency use) use the date of treatment. Dates for subjects who were screen failures are not included.*

  - SDTM Programmers should ensure valid Trial Summary dataset created (TS.xpt) with *at least* one record to indicate the study start date
    - TSPARMCD= STSTDTC / TSPARM = Study Start Date
    - Date has to be in proper ISO 8601 format (YYYY-MM-DD)

For Studies starting after Dec 2016, **all** TS values should be in TS.xpt

Step 2: Study Data Validation

- Process by which FDA can assess conformance to all published rules for standardized submission data during the *actual submission process*
- Documentation of FDA requirements is ***published*** and ***periodically updated***:
  - https://www.fda.gov/industry/fda-resources-data-standards/study-data-standards-resources
  - FDA Data Standards Catalog (Dec 2018)
    - With specific rules in the relevant standards (e.g., SDTMIG)
  - FDA Study Data Technical Conformance Guide (March 2019)

Where do the Validation Rules Come From?

- Validation Rules are **based on**
  - Conformance rules in the CDISC Standards
    - Normative and Informative content in the standards documents (e.g., Implementation Guide)
    - Conformance rules described in the Model or Implementation Guide
  - CDISC Validation Rules
    - CDISC team publishes these in conjunction with a version of the standard
  - Additional FDA-specific published validation rules (based on Technical Conformance Guide and published Business Rules)
  - Additional PMDA-specific published validation rules (based on Technical Conformance Guide and other published information)

~ yearly

See relevant website for publication schedules

Data Standards Catalog

- ALL of the data standards expected for various types of submissions (*not just CDISC standards*)
  - EXAMPLE: Standards for Data Exchange:
    - CDISC Standards (SDTM, ADaM, SEND, CT, Define-xml)
    - ICSR, SPL, ASCII, XML
  - EXAMPLE: Standards for Terminology
    - CDISC/NCI Terminology
    - MedDRA, WHO
    - LOINC, UNII

FDA Data Standards Catalog v5.1 (08-2-2018) - Supported and

For full description of column headings, see Instr.& Column D

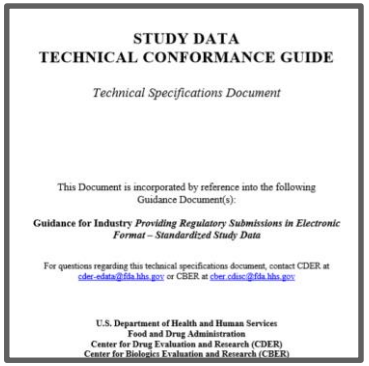
Supported mplementation Guide Version	FDA Center(s)	Date Support Begins (MM/DD/YYYY)	Date Support Ends (MM/DD/YYYY)	Date Requireme Begins (MM/DD/YYYY)
tuctured Product Labeling (SPL) ementation Guide with Validation cedures Version 1 Revision 01412101457	CBER	06/10/2015	n/a	06/10/2015
al Unique Device Identification tabase (GUDID) Release 1.2.3 alth Level 7 (HL7) tuctured Product Labeling (SPL) mplementation Specification Version 1.3	CDRH	1/13/2015		

FDA Data Standards Catalog

- Information in Data Standards Catalog includes
  - Name / source / version of standard
  - Purpose / usage of each standard for regulatory submissions
  - Dates
    - Support start/end dates
    - Requirement start/end dates
  - Which FDA Centers require the standard (CBER, CDER, CDRH, CVM…)
  - References (e.g., related Guidance) and other resources (examples of usage, URLs for source)

https://www.fda.gov/industry/fda-resources-data-standards/study-data-standards-resources

Study Data Technical Conformance Guide


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)

- Started out (2014) as a document to describe
  - Common mistakes and issues seen in SDTM data submissions (CDER)
  - How to avoid those common mistakes and issues
- Evolved (2014 to present) to a twice-yearly published Guide
  - Describes FDA's preferences that go *beyond* the published standards rules
  - Emphasizes several published rules (e.g., SDTMIG) that are **very** important to FDA

https://www.fda.gov/industry/fda-resources-data-standards/study-data-standards-resources

FDA Business Rules

- Rules for FDA study data standards and quality conformance
- Published: https://www.fda.gov/industry/fda-resources-data-standards/study-data-standards-resources

FDA Business Rule ID	FDA Business Rule
Clinical and Nonclinical	
FDAB005	Age or age range should be provided for all subjects, except for Screen Failures.
FDAB008	All exposure records should occur between First and Last Study Treatment dates.
FDAB009	All paired variables should have a one-to-one relationship. Examples include Short Name and Name of Test; Parameter Name and Parameter Code or Number; Variable Name and Variable Label, etc.
FDAB011	All Trial Design data should be submitted as specified in the Technical Conformance Guide (TCG).
FDAB012	Assessment results should include units whenever a unit of measure is available.
FDAB013	Baseline flags for Laboratory results, Vital Signs, ECG, Pharmacokinetic Concentrations, and Microbiology results should be submitted if the data was collected or can be derived.
FDAB015	Character values should not have leading spaces or only have a period character.
FDAB016	Collection Study Day should be populated when Date/Time of Collection is
FDAB017	Controlled terms should use the exact same case used by the terminology maintenance organizations (e.g., MedDRA, CDISC controlled terminology).
FDAB018	A variable's length across a study should be no longer than the maximum length of the actual data (except for SUPPQUAL).
FDAB019	SUPPQUAL variable length should be no longer than the maximum length of the actual data within the dataset.

FDA Business Rules

- FDA business rules are used as the basis for writing FDA validator rules
- May be a one:many relationship

FDA Business Rule ID	FDA Business Rule
Clinical and Nonclinical	
FDAB005	Age or age range should be provided for all subjects, except for Screen Failures.
FDAB008	All sure records should occur between First and Last Study Treatment dates.
FDAB009	A variables should have a one-to-one relationship. Examples include Short

SD1129	Neither AGE nor AGETXT variables are present	FDA	FDAB005	Age or age range must be provided for all subjects, except for Screen Failures.	At least one of Age (AGE) or Age Text (AGETXT) variables should be included into Demographics (DM) domain.	DM
SD1121	Neither AGE nor AGETXT values are populated	FDA	FDAB005	Age or age range must be provided for all subjects, except for Screen Failures.	Value for Age (AGE) or Age Range (AGETXT) variables should be populated for all subjects with only exception for Screen Failures (ARMCD=SCRNFAIL) and Not Assigned (ARMCD=NOTASSGN)	DM
SD2023	AGE is not provided	FDA	FDAB005	Age or age range must be provided for all subjects, except for Screen Failures.	Age (AGE) variable values should be provided, when Date/Time of Birth (BRTHDTC) variable values are populated.	DM

FDAB013	Baseline flags for Laboratory results, Vital Signs, ECG, Pharmacokinetic Concentrations, and Microbiology results should be submitted if the data was collected or can be derived.
FDAB015	Character values should not have leading spaces or only have a period character.
FDAB016	Collection Study Day should be populated when Date/Time of Collection is
FDAB017	Controlled terms should use the exact same case used by the terminology maintenance organizations (e.g., MedDRA, CDISC controlled terminology).
FDAB018	A variable's length across a study should be no longer than the maximum length of the actual data (except for SUPPQUAL).
FDAB019	SUPPQUAL variable length should be no longer than the maximum length of the actual data within the dataset.

FDA Validator Rules

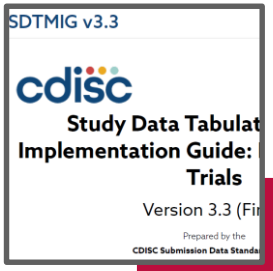
- One of the sources for validation rules used in validation software (e.g., Pinnacle 21)
- Published: https://www.fda.gov/industry/fda-resources-data-standards/study-data-standards-resources

version 1.3, finalized October 2018

FDA Validator Rule ID	FDA	Message	Publisher	Publisher ID	Business or Conformance Rule Validated	FDA Validator Rule	Domain
SD1097	No Treat						
SD1321	No						
SD1037							, CE, PP
SD0008	Value dictiona						
SD1114	Value for --BO	nd in MedDRA dictionary	FDA	FDAB003	Adverse Events must be coded using MedDRA dictionary.	Value for Body System or Organ Class (--BODSYS) variable must be populated using a System Organ Class of the MedDRA dictionary of a version specified in the define.xml (Case-insensitive).	AE, MH, CE
SD2007	Value for --PTCD not fou	in MedDRA dictionary	FDA	FDAB003	Adverse Events must be coded using MedDRA dictionary.	Value for Preferred Term Code (--PTCD) variable must be populated using a Preferred Term Code of the MedDRA dictionary of a version specified in the define.xml.	AE, MH, CE
SD2008	Value for --LLT not found in MedDRA dictionary		FDA	FDAB003	Adverse Events must be coded using MedDRA dictionary.	Value for Lowest Level Term (--LLT) variable must be populated using a Lowest Level Term of the MedDRA dictionary of a version specified in the define.xml (Case-insensitive).	AE, MH, CE
SD2010	Value for --HLT not found in MedDRA dictionary		FDA	FDAB003	Adverse Events must be coded using MedDRA dictionary.	Value for High Level Term (--HLT) variable must be populated using a High Level Term of the MedDRA dictionary of a version specified in the define.xml (Case-insensitive).	AE, MH, CE
SD2011	Value for --HLTCD not found in MedDRA dictionary		FDA	FDAB003	Adverse Events must be coded using MedDRA dictionary.	Value for High Level Term Code (--HLTCD) variable must be populated using a High Level Term Code of the MedDRA dictionary of a version specified in the define.xml.	AE, MH, CE

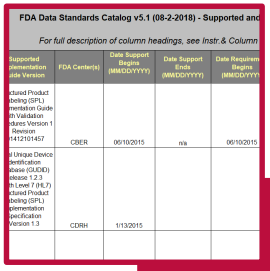
FDA Validator Rule ID is referenced in the published Pinnacle 21 validation rules https://www.pinnacle21.com/validation-rules/sdtm

Conformance to SDTM for FDA Submissions



Foundational Stds

- SDTMIG
- CDISC/NCI Terminology
- Define-XML

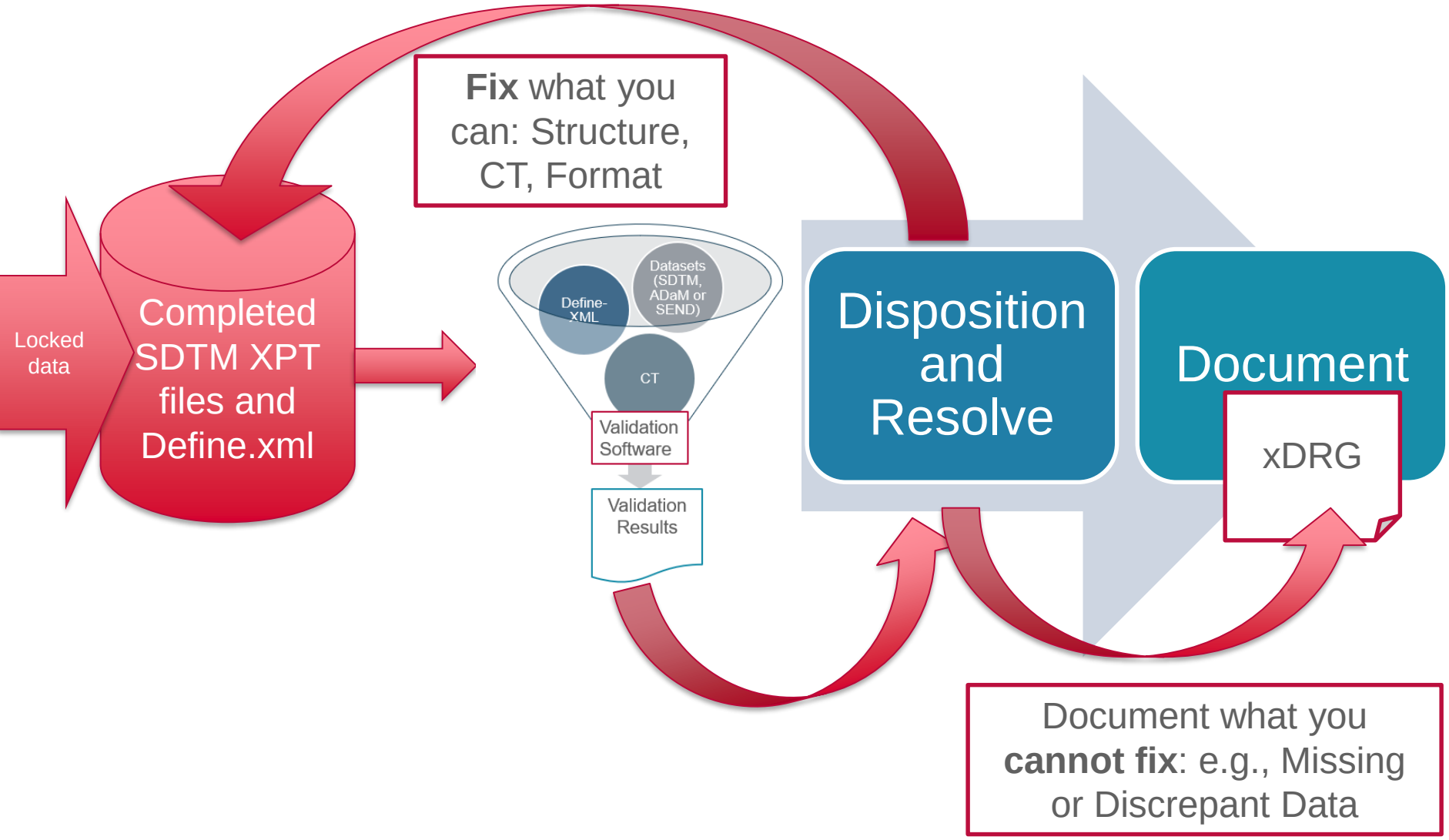


FDA Requirements

- FDA Data Standards Catalog
- FDA Technical Conformance Guide
- FDA Business and Validator Rules

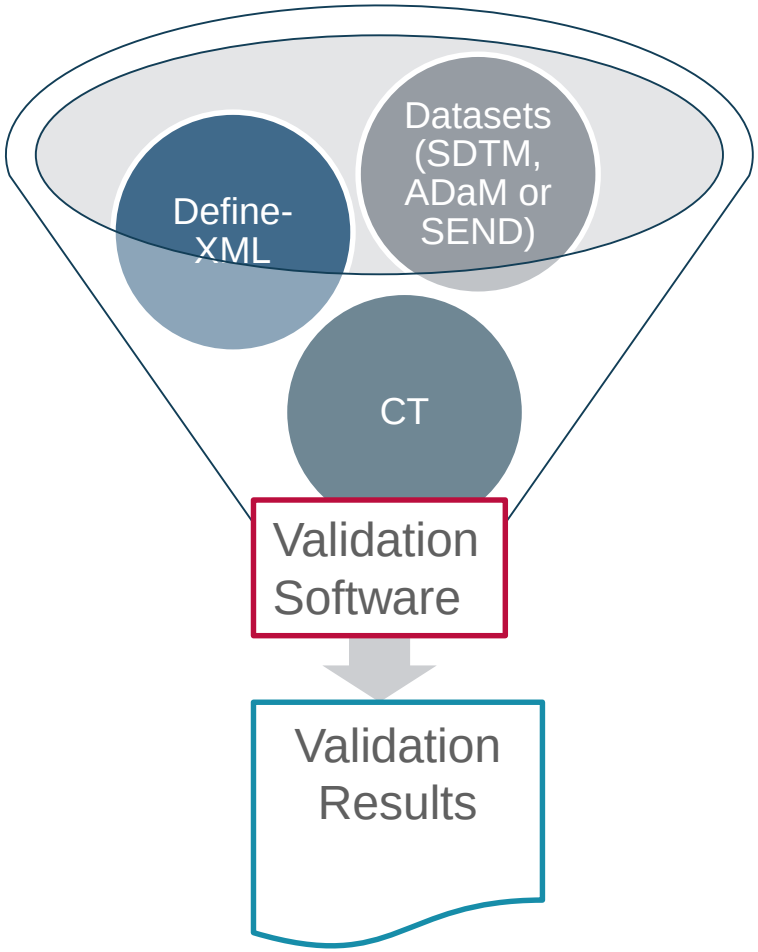
All of these rules are built into the standard validation software

The Basics of Validating Study Data for Submission



The Basics of Validating Study Data for a Submission

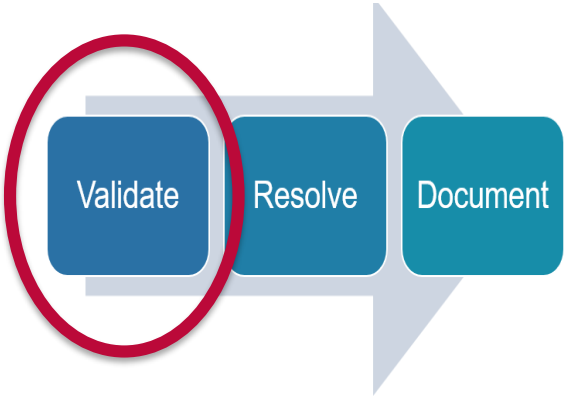
- Validation software is available for all **CDISC** data standards that are required by FDA and PMDA



The Basics of Validating Study Data for Submission

- Using software, check your data for
  - Valid structure (based on rules for the relevant standard, like SDTM)
    - Inclusion of required domains
    - Inclusion of required and expected variables
    - Population of required variables
    - Use of required terminology
  - Conformance to published rules from FDA (or PMDA)
    - Inclusion of PERM SDTM variables requested by FDA (EPOCH, --DY)
    - Inclusion of criteria for SAEs

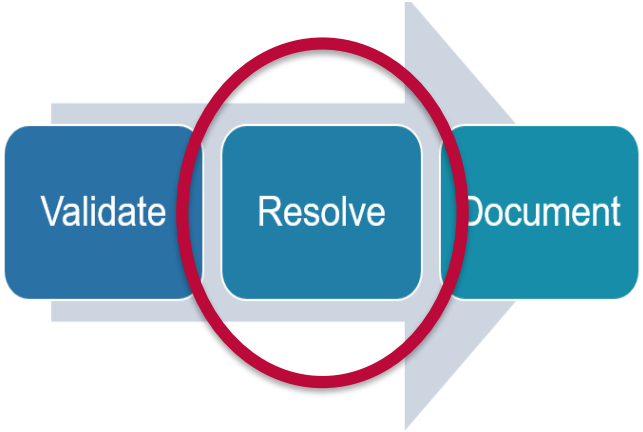
Validation Software has executable rules that are based on *the software vendor's interpretation* of the published rules. Interpretations may vary slightly from vendor to vendor. They *should* all be based on the same published rules.



The Basics of Validating Study Data for Submission

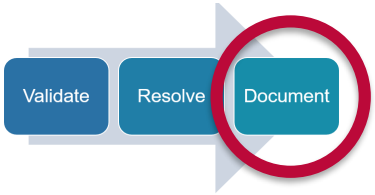
- Review the results / output from your validation
  - Resolve structure, format and terminology issues
  - Determine which results are either
    - Remaining discrepancies we cannot fix, or
    - Errors/Warnings that are false positives

All of these should be documented in the appropriate Data Reviewer's Guide

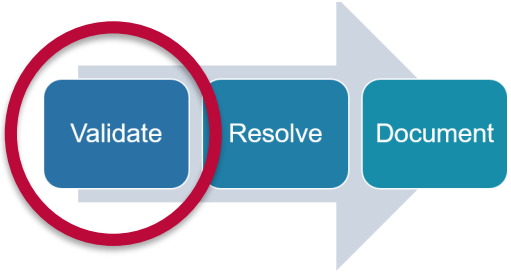


The Basics of Validating Study Data for Submission

- Even if your data are very high quality, validating your submission datasets will usually produce errors or warnings, some of which may be false positives
  - Example False Positives:
    - Lab results with no units (program would have to be test specific to avoid)
    - Comparing partial start dates to complete end date (software incorrectly interprets as end date before start date, i.e., poor data quality)
    - Other software “bugs” or programming deficiencies (e.g., software may be extra finicky about Define.xml structure)
- We are expected to proactively validate data and explain all issues in the relevant Data Reviewer’s Guide (SDRG, NSDRG and ADRG)
- FDA/PMDA will **identify** these issues by using validation software during the submission process and look for explanations in the xDRG



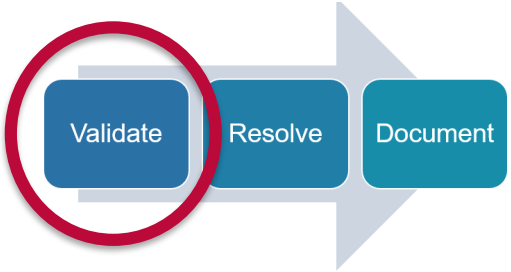
Example Validation Software



- https://www.pinnacle21.com/
- FDA* uses Pinnacle 21 *Enterprise* to validate your submission, but *they do not officially endorse it, and they may use other tools, too*

Feature	Community Version	Enterprise Version
Validate data and provide timestamped discrepancy report	X	X
Generate Define-XML	X	X
Save reports and manage validation processes (open issues)		X
Validate against *your* standards		X
Up to Date versions of standards included	*Eventually*	X
Manage validation for multiple studies in parallel		X
Windows and Mac	*X*	X

Example Validation Software



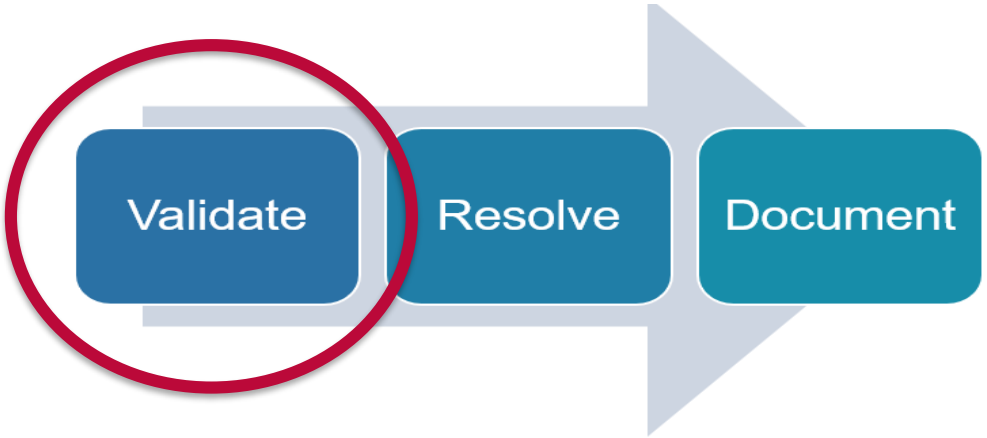
- info.pointcrosslifesciences.com/mysend
- Free download

Feature
Validates all published FDA, PMDA and CDISC conformance and validator rules plus additional rule for Nonclinical Define.xml.
Saving Dataset to Excel(xpt, sas7bdat)
Trial Summary (TS.xpt) generation for current and legacy studies
Multiple reporting functions and export function
nSDRG Template Generation
Data Visualization – Data Viewer to view SEND data in tabulations, graphical charts, and interactive graphics
Windows **only**

Example Validation Software

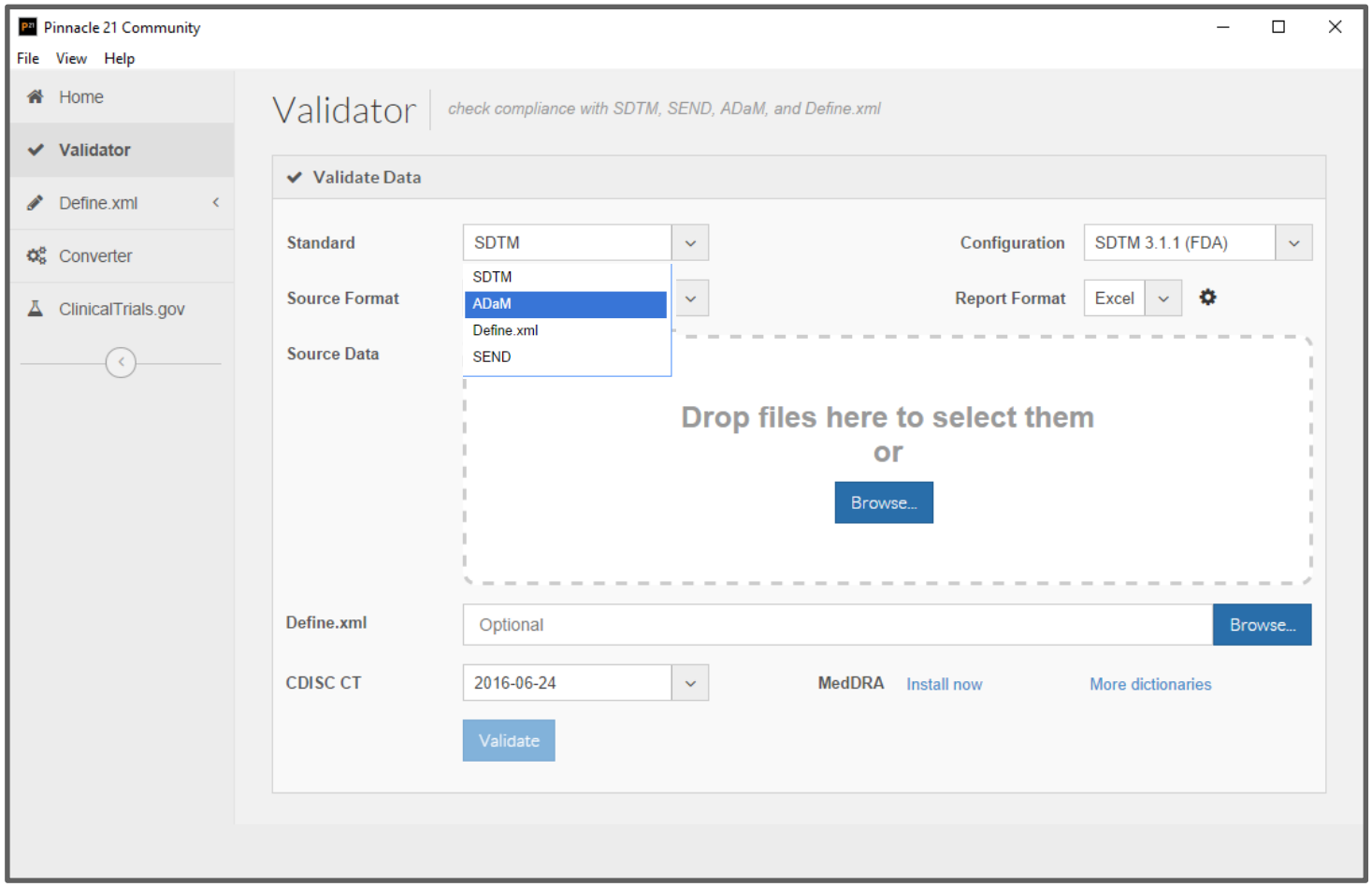
- https://www.entimo.com/products/entimice/sdtm-checker
- https://www.formedix.com/verifying-study-deliverables
- https://www.edetek.com/conform-tm/conform-components/

- Etc…

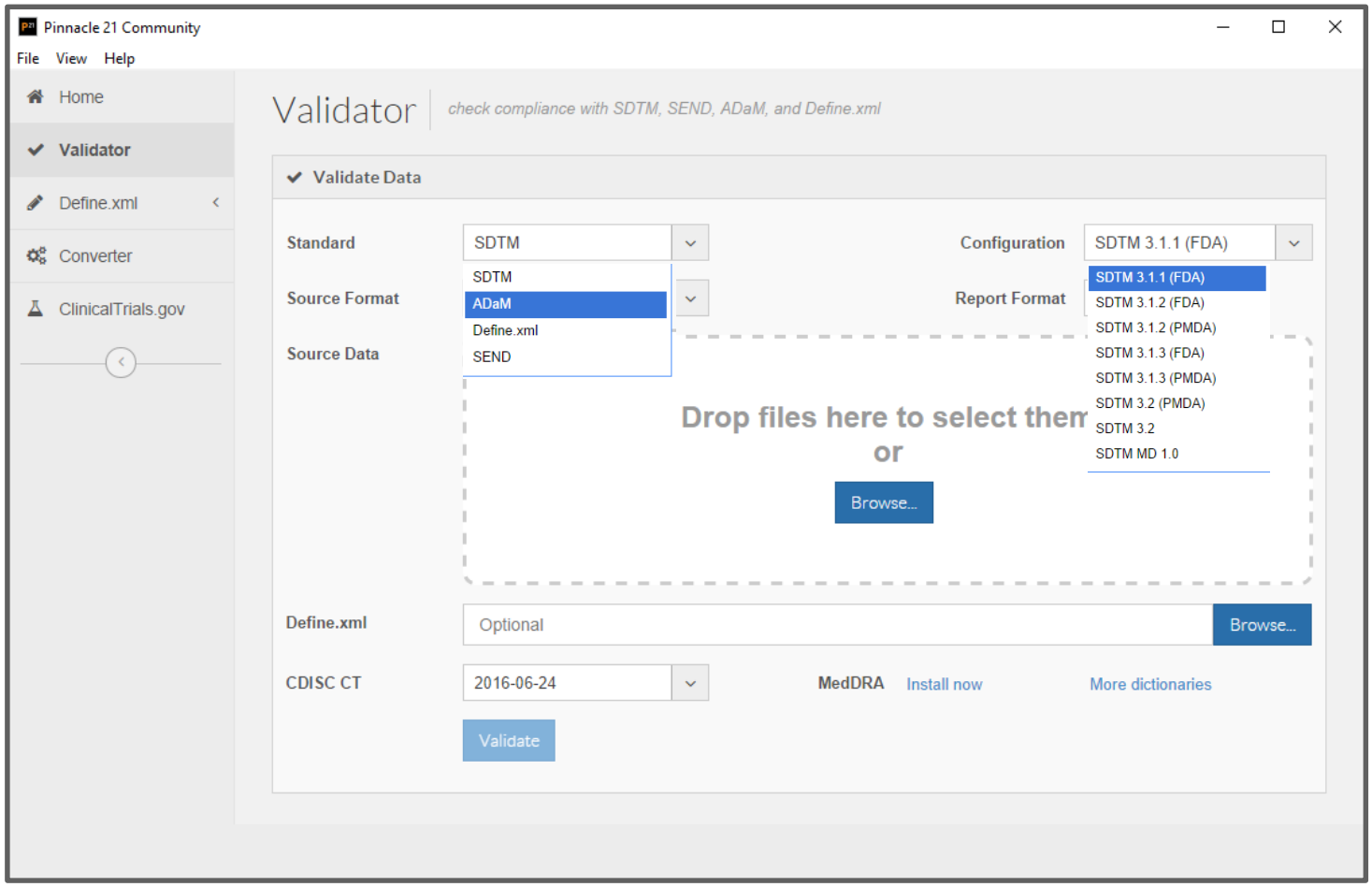


- There is no requirement for which validation software you use, but many organizations use Pinnacle 21 because it is used by FDA

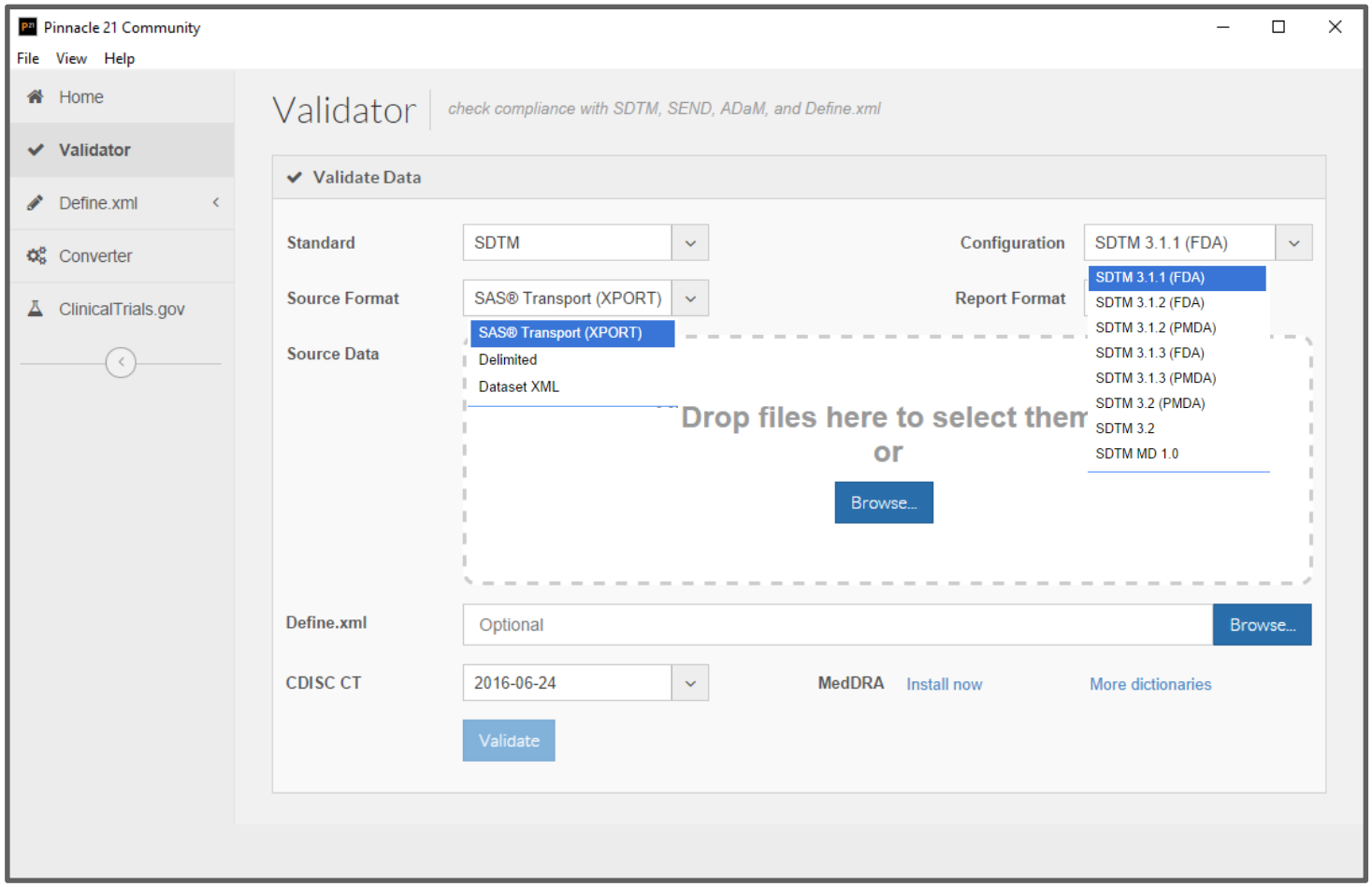
Validation Demo using Pinnacle 21 (Community Version)



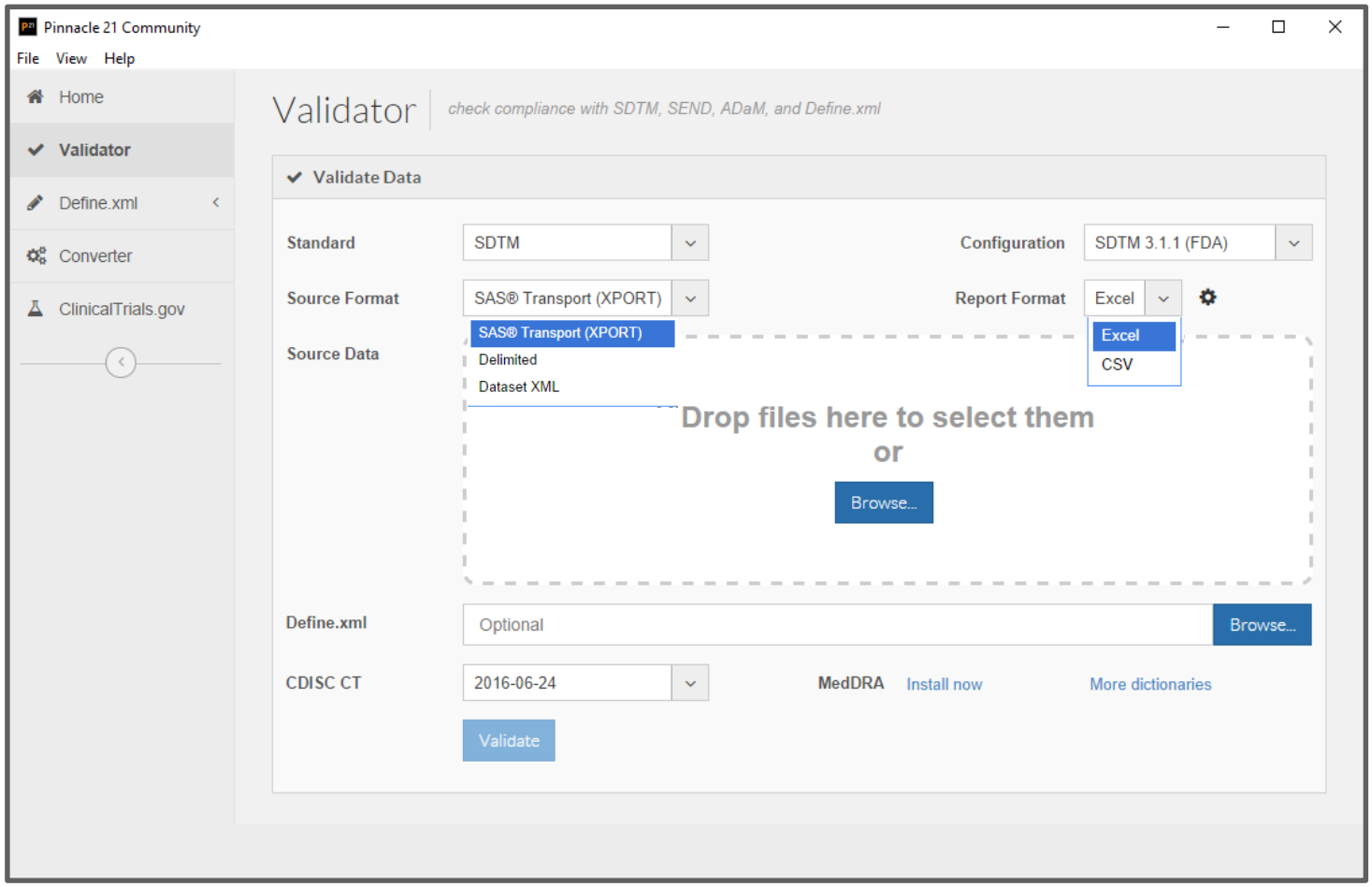
Validation Demo using Pinnacle 21 (Community Version)



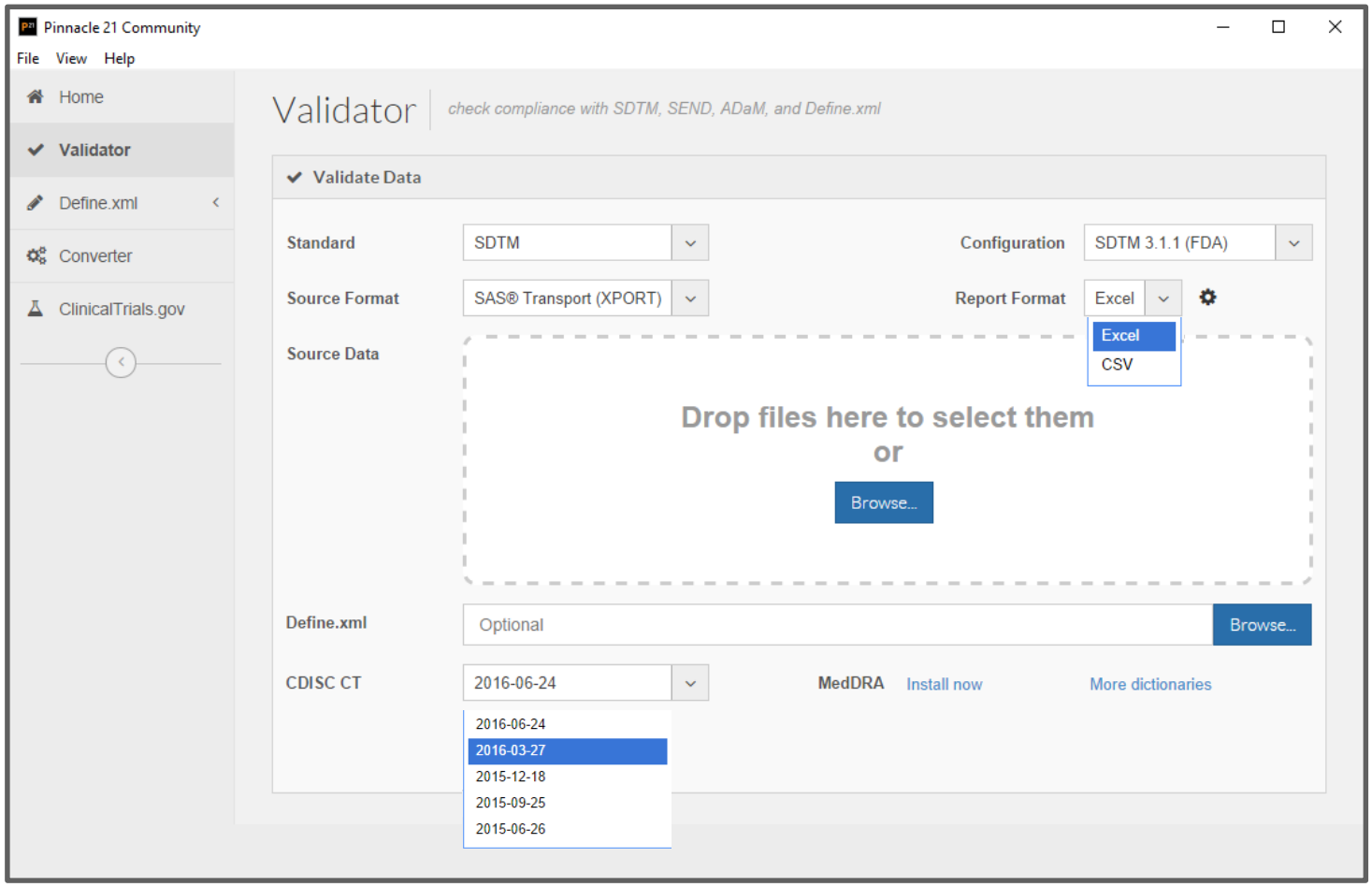
Validation Demo using Pinnacle 21 (Community Version)



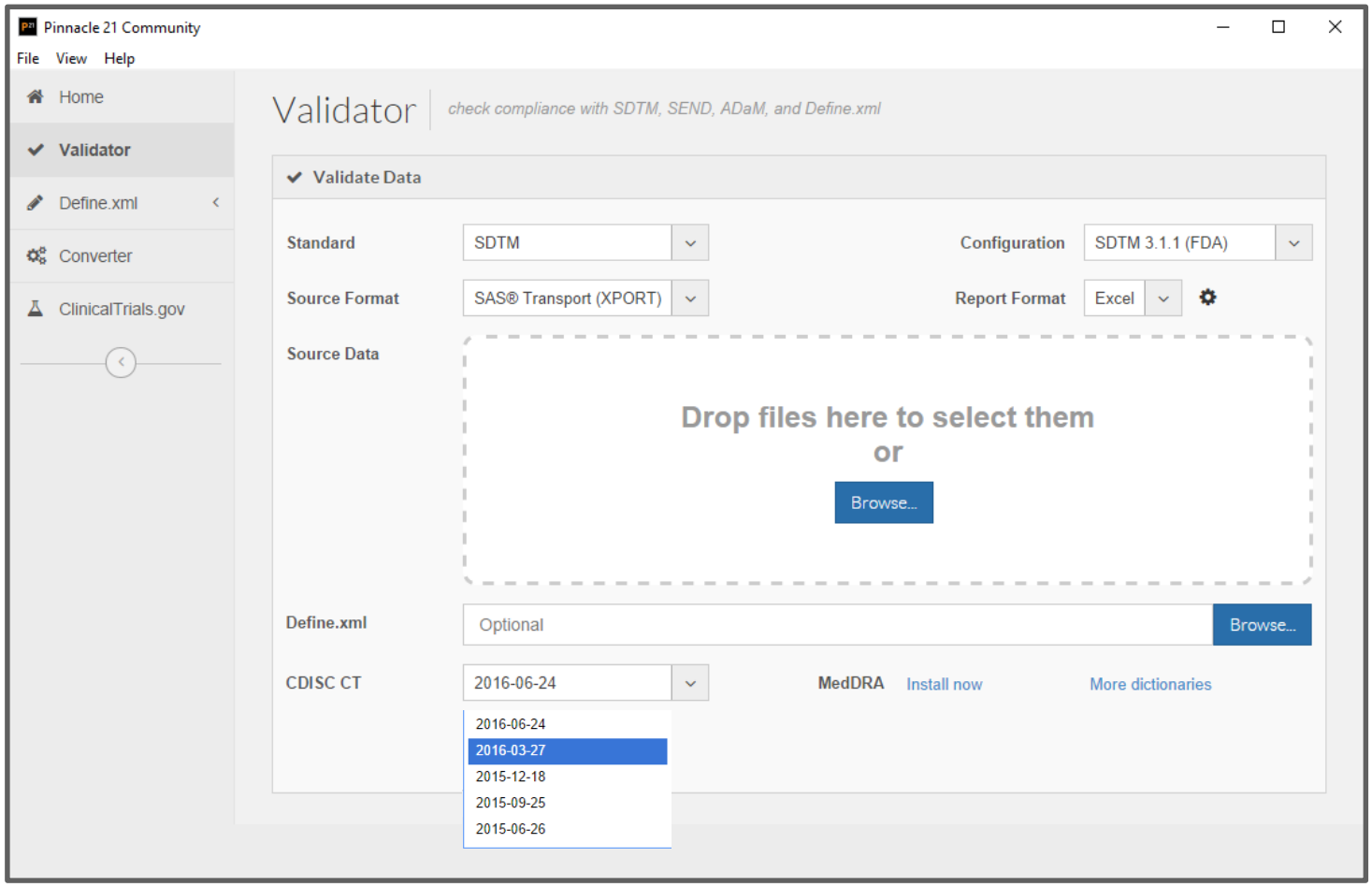
Validation Demo using Pinnacle 21 (Community Version)



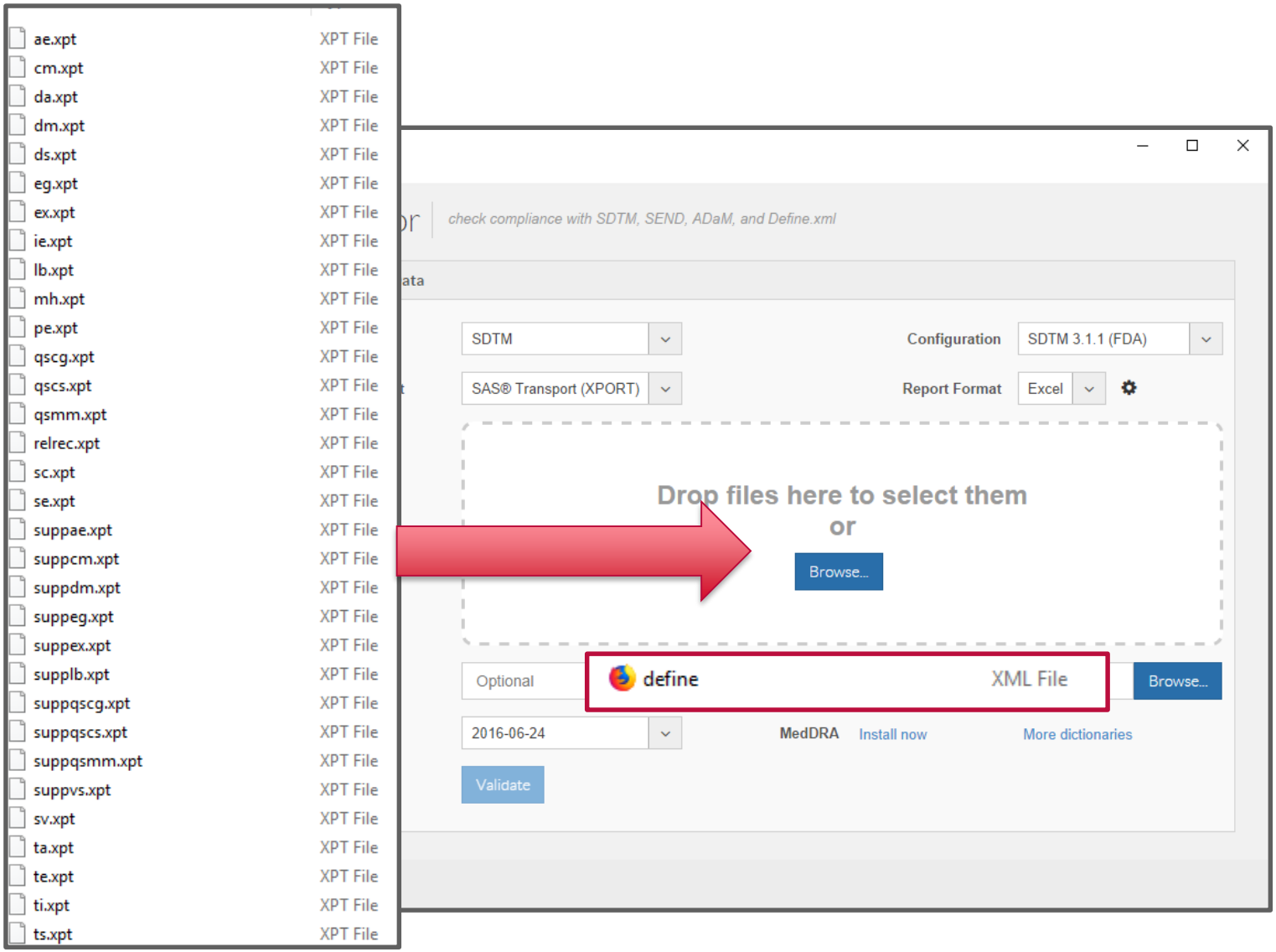
Validation Demo using Pinnacle 21 (Community Version)



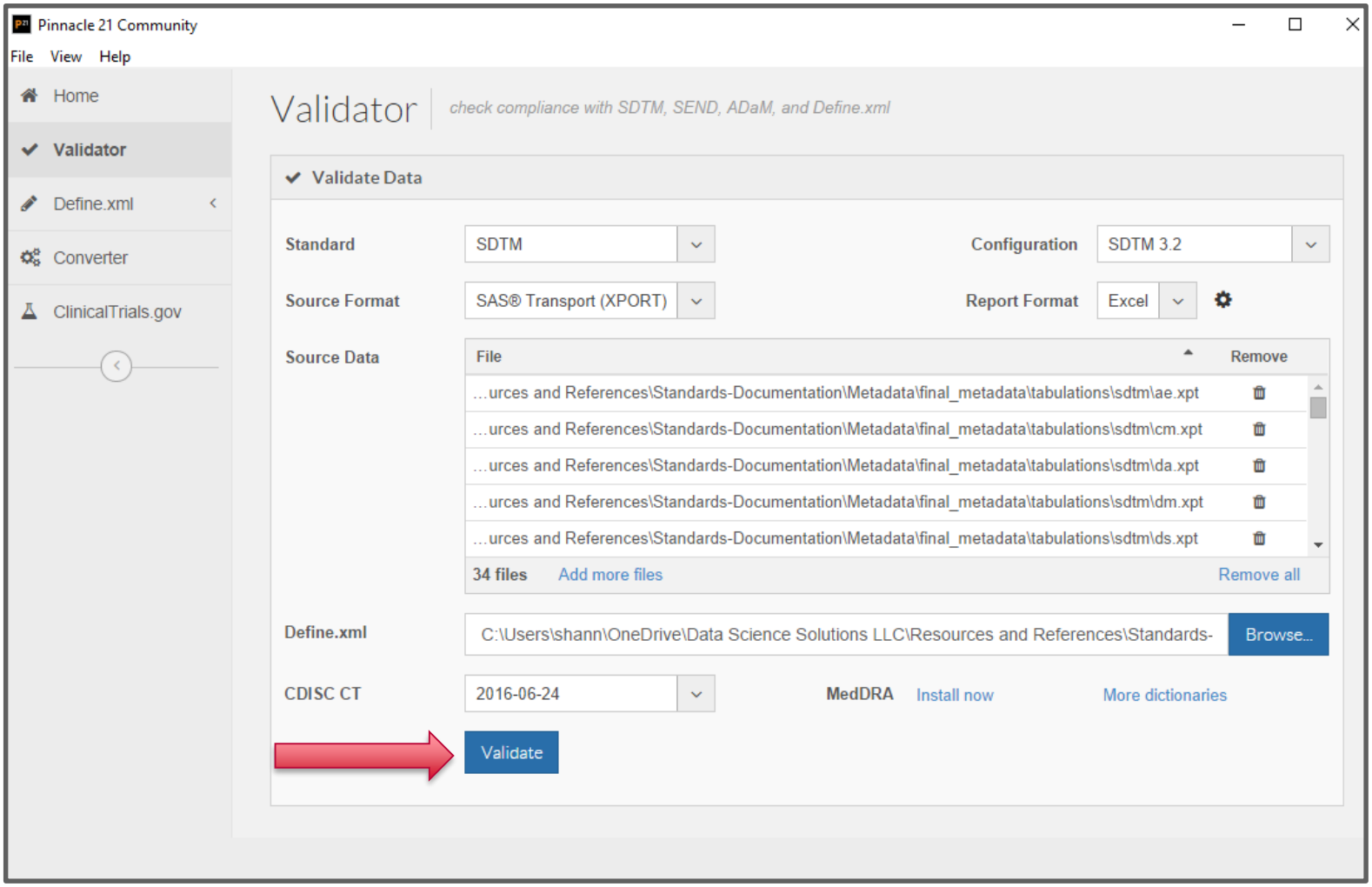
Validation Demo using Pinnacle 21 (Community Version)



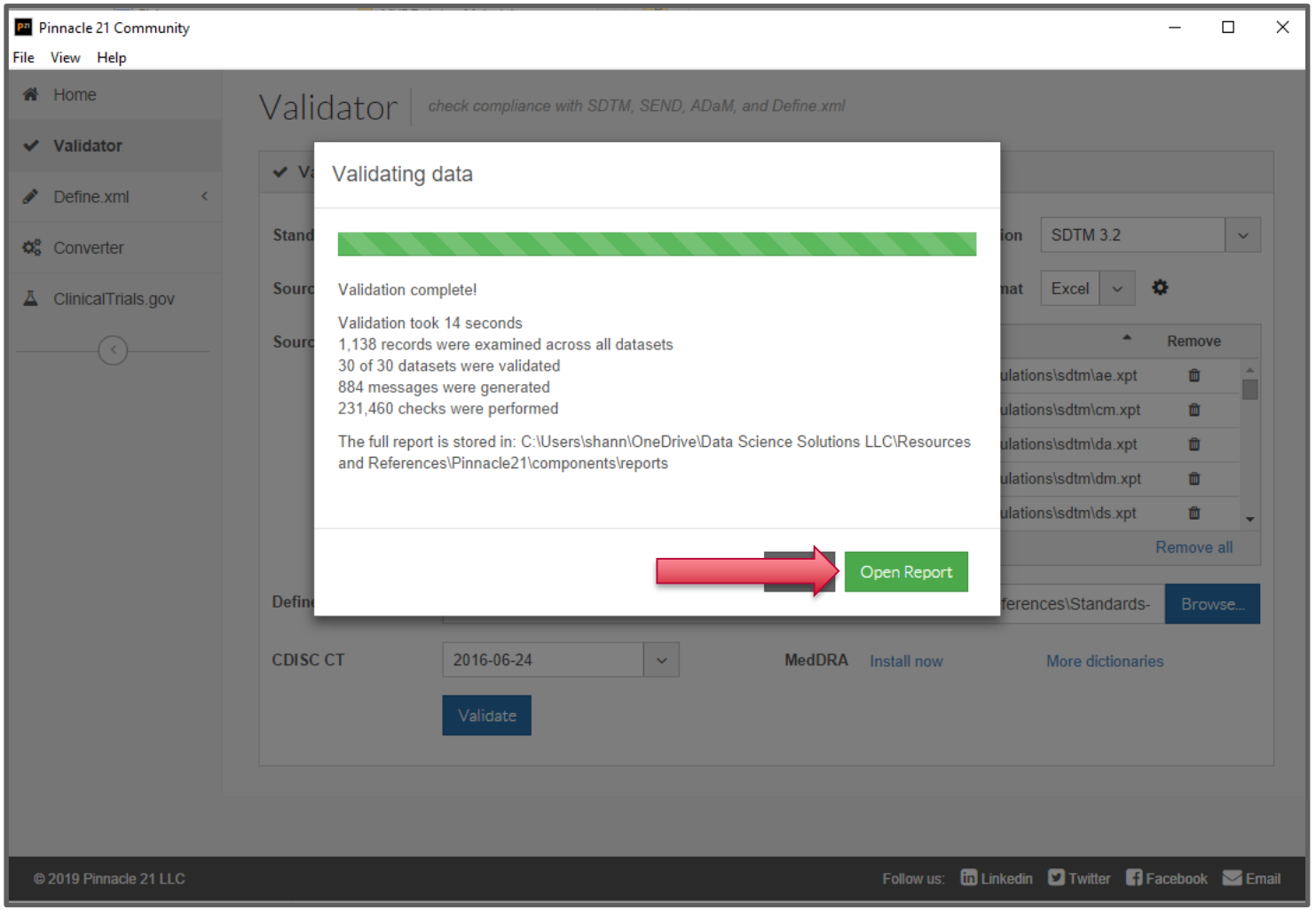
Validation Demo using Pinnacle 21 (Community Version)



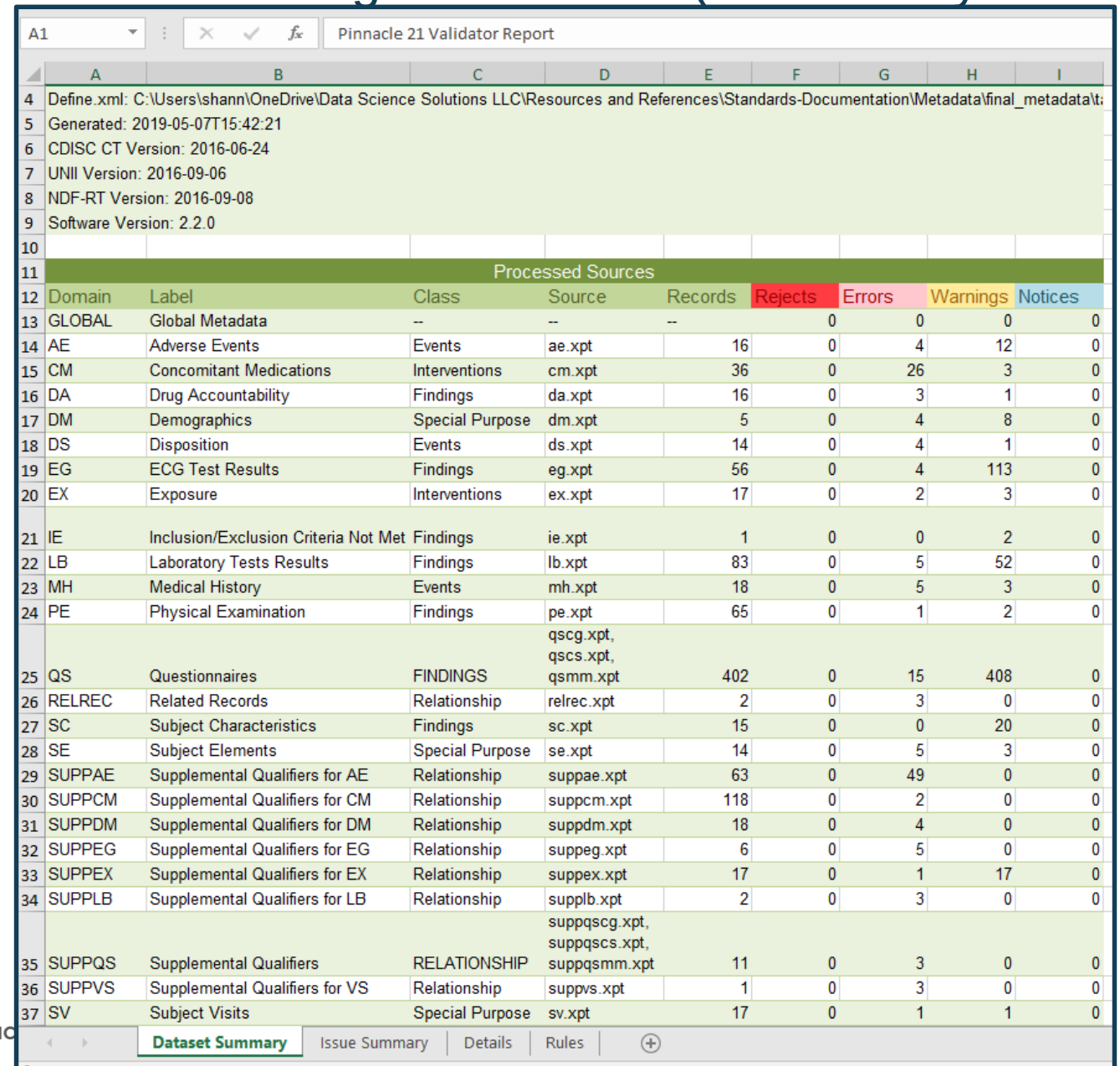
Validation Demo using Pinnacle 21 (Community Version)



Validation Demo using Pinnacle 21 (Community Version)



Validation Demo using Pinnacle 21 (Community Version)



A1 | Pinnacle 21 Validator Report

	A	B	C	D	E	F	G	H	I
4	Define.xml: C:\Users\shann\OneDrive\Data Science Solutions LLC\Resources and References\Standards-Documentation\Metadata\final_metadata\t								
5	Generated: 2019-05-07T15:42:21								
6	CDISC CT Version: 2016-06-24								
7	UNII Version: 2016-09-06								
8	NDF-RT Version: 2016-09-08								
9	Software Version: 2.2.0								
10									
11	Processed Sources								
12	Domain	Label	Class	Source	Records	Rejects	Errors	Warnings	Notices
13	GLOBAL	Global Metadata	--	--	--	0	0	0	0
14	AE	Adverse Events	Events	ae.xpt	16	0	4	12	0
15	CM	Concomitant Medications	Interventions	cm.xpt	36	0	26	3	0
16	DA	Drug Accountability	Findings	da.xpt	16	0	3	1	0
17	DM	Demographics	Special Purpose	dm.xpt	5	0	4	8	0
18	DS	Disposition	Events	ds.xpt	14	0	4	1	0
19	EG	ECG Test Results	Findings	eg.xpt	56	0	4	113	0
20	EX	Exposure	Interventions	ex.xpt	17	0	2	3	0
21	IE	Inclusion/Exclusion Criteria Not Met	Findings	ie.xpt	1	0	0	2	0
22	LB	Laboratory Tests Results	Findings	lb.xpt	83	0	5	52	0
23	MH	Medical History	Events	mh.xpt	18	0	5	3	0
24	PE	Physical Examination	Findings	pe.xpt	65	0	1	2	0
25	QS	Questionnaires	FINDINGS	qscg.xpt, qscs.xpt, qsmm.xpt	402	0	15	408	0
26	RELREC	Related Records	Relationship	relrec.xpt	2	0	3	0	0
27	SC	Subject Characteristics	Findings	sc.xpt	15	0	0	20	0
28	SE	Subject Elements	Special Purpose	se.xpt	14	0	5	3	0
29	SUPPAE	Supplemental Qualifiers for AE	Relationship	suppae.xpt	63	0	49	0	0
30	SUPPCM	Supplemental Qualifiers for CM	Relationship	suppcm.xpt	118	0	2	0	0
31	SUPPDM	Supplemental Qualifiers for DM	Relationship	suppdm.xpt	18	0	4	0	0
32	SUPPEG	Supplemental Qualifiers for EG	Relationship	suppeg.xpt	6	0	5	0	0
33	SUPPEX	Supplemental Qualifiers for EX	Relationship	suppex.xpt	17	0	1	17	0
34	SUPPLB	Supplemental Qualifiers for LB	Relationship	supplb.xpt	2	0	3	0	0
35	SUPPQS	Supplemental Qualifiers	RELATIONSHIP	suppqscg.xpt, suppqscs.xpt, suppqsmm.xpt	11	0	3	0	0
36	SUPPVS	Supplemental Qualifiers for VS	Relationship	suppvs.xpt	1	0	3	0	0
37	SV	Subject Visits	Special Purpose	sv.xpt	17	0	1	1	0

Dataset Summary | Issue Summary | Details | Rules

Validation Demo using Pinnacle 21 (Community Version)

Pinnacle 21 Validator Report

Configuration: C:\Users\shann\OneDrive\Data Science Solutions LLC\Resources and References\Pinnacle21\components\config\SDTM 3.2.xml
Define.xml: C:\Users\shann\OneDrive\Data Science Solutions LLC\Resources and References\Standards-Documentation\Metadata\final_metadata\tabulations\sdtm\define.xml
Generated: 2019-05-07T15:42:21
CDISC CT Version: 2016-06-24
UNII Version: 2016-09-06
NDF-RT Version: 2016-09-08
Software Version: 2.2.0

Issue Summary

Source	Pinnacle 21 ID	Publisher ID	Message	Severity	Found
AE					
	SD0009	FDAC206	No qualifiers set to 'Y', when AE is Serious	Error	1
	SD1082	FDAC036	Variable length is too long for actual data	Error	2
	SD1089	FDAC130	AESTDY variable value is imputed	Error	1
	SD0057	FDAC020	SDTM Expected variable AEBDSYCD not found	Warning	1
	SD0057	FDAC020	SDTM Expected variable AEHLGT not found	Warning	1
	SD0057	FDAC020	SDTM Expected variable AEHLGTCD not found	Warning	1
	SD0057	FDAC020	SDTM Expected variable AEHLT not found	Warning	1
	SD0057	FDAC020	SDTM Expected variable AEHLTCD not found	Warning	1
	SD0057	FDAC020	SDTM Expected variable AELLT not found	Warning	1
	SD0057	FDAC020	SDTM Expected variable AELLTCD not found	Warning	1
	SD0057	FDAC020	SDTM Expected variable AEPTCD not found	Warning	1
	SD0057	FDAC020	SDTM Expected variable AESOC not found	Warning	1
	SD0057	FDAC020	SDTM Expected variable AESOCCD not found	Warning	1
	SD1077	FDAC021	FDA Expected variable EPOCH not found	Warning	1
	SD1097	FDAC022	No Treatment Emergent info for Adverse Event	Warning	1
CM					
	SD1082	FDAC036	Variable length is too long for actual data	Error	3
	SD1089	FDAC130	CMSTDY variable value is imputed	Error	22
	SD1093	FDAC135	CMENDY variable value is imputed	Error	1
	SD1031	FDAC138	Value for CMENRF is populated, when RFENDTC is NULL	Warning	2
	SD1077	FDAC021	FDA Expected variable EPOCH not found	Warning	1
DA					
	SD1082	FDAC036	Variable length is too long for actual data	Error	3
	SD1077	FDAC021	FDA Expected variable EPOCH not found	Warning	1
DM					

Dataset Summary | Issue Summary | Details | Rules

Validation Demo using Pinnacle 21 (Community Version)

	A	B	C	D	E	F	G	H	I	J
1	Domain	Record	Count	Variables	Values	Pinnacle 21 ID	Publisher ID	Message	Category	Severity
2	AE			VARIABLE	AEBDSYCD	SD0057	FDAC020	SDTM Expected variable AEBDSYCD not found	Metadata	Warning
3	AE			VARIABLE	AEHLGT	SD0057	FDAC020	SDTM Expected variable AEHLGT not found	Metadata	Warning
4	AE			VARIABLE	AEHLGTCD	SD0057	FDAC020	SDTM Expected variable AEHLGTCD not found	Metadata	Warning
5	AE			VARIABLE	AEHLT	SD0057	FDAC020	SDTM Expected variable AEHLT not found	Metadata	Warning
6	AE			VARIABLE	AEHLTCD	SD0057	FDAC020	SDTM Expected variable AEHLTCD not found	Metadata	Warning
7	AE			VARIABLE	AELLT	SD0057	FDAC020	SDTM Expected variable AELLT not found	Metadata	Warning
8	AE			VARIABLE	AELLTCD	SD0057	FDAC020	SDTM Expected variable AELLTCD not found	Metadata	Warning
9	AE			VARIABLE	AEPTCD	SD0057	FDAC020	SDTM Expected variable AEPTCD not found	Metadata	Warning
10	AE			VARIABLE	AESOC	SD0057	FDAC020	SDTM Expected variable AESOC not found	Metadata	Warning
11	AE			VARIABLE	AESOCCD	SD0057	FDAC020	SDTM Expected variable AESOCCD not found	Metadata	Warning
12	AE			VARIABLE, DATASET	EPOCH, AE	SD1077	FDAC021	FDA Expected variable EPOCH not found	Metadata	Warning
13	AE	8		AESER	Y	SD0009	FDAC206	No qualifiers set to 'Y', when AE is Serious	Consistency	Error
14	AE			Variable, Excess	AESPID, 3	SD1082	FDAC036	Variable length is too long for actual data	Metadata	Error
15	AE			Variable, Excess	AEACN, 14	SD1082	FDAC036	Variable length is too long for actual data	Metadata	Error
16	AE	1		SUB:RFSTDTC, AESTDTC, AESTDY	2003-04-29, 2003-05, 3	SD1089	FDAC130	AESTDY variable value is imputed	Presence	Error
17	AE	8		AESEQ, USUBJID	5, CDISC01.100014	SD1097	FDAC022	No Treatment Emergent info for Adverse Event	Presence	Warning
18	CM			VARIABLE, DATASET	EPOCH, CM	SD1077	FDAC021	FDA Expected variable EPOCH not found	Metadata	Warning
19	CM	35		CMENRF, USUBJID	AFTER, CDISC01.200005	SD1031	FDAC138	Value for CMENRF is populated, when RFENDTC is NULL	Consistency	Warning
20	CM	36		CMENRF, USUBJID	AFTER, CDISC01.200005	SD1031	FDAC138	Value for CMENRF is populated, when RFENDTC is NULL	Consistency	Warning
21	CM			Variable, Excess	CMDOSFRQ, 1	SD1082	FDAC036	Variable length is too long for actual data	Metadata	Error
22	CM			Variable, Excess	CMENRF, 1	SD1082	FDAC036	Variable length is too long for actual data	Metadata	Error
23	CM			Variable, Excess	CMDECOD, 4	SD1082	FDAC036	Variable length is too long for actual data	Metadata	Error
24	CM	1		SUB:RFSTDTC, CMSTDTC, CMSTDY	2003-04-29, 1986, -5963	SD1089	FDAC130	CMSTDY variable value is imputed	Presence	Error
25	CM	2		SUB:RFSTDTC, CMSTDTC, CMSTDY	2003-04-29, 1987, -5598	SD1089	FDAC130	CMSTDY variable value is imputed	Presence	Error
26	CM	3		SUB:RFSTDTC, CMSTDTC, CMSTDY	2003-04-29, 1995, -2676	SD1089	FDAC130	CMSTDY variable value is imputed	Presence	Error
27	CM	4		SUB:RFSTDTC, CMSTDTC, CMSTDY	2003-04-29, 1995, -2676	SD1089	FDAC130	CMSTDY variable value is imputed	Presence	Error
28	CM	5		SUB:RFSTDTC, CMSTDTC, CMSTDY	2003-04-29, 1995, -2676	SD1089	FDAC130	CMSTDY variable value is imputed	Presence	Error
29	CM	6		SUB:RFSTDTC, CMSTDTC, CMSTDY	2003-04-29, 1995, -2676	SD1089	FDAC130	CMSTDY variable value is imputed	Presence	Error
30	CM	7		SUB:RFSTDTC, CMSTDTC, CMSTDY	2003-04-29, 1995, -2676	SD1089	FDAC130	CMSTDY variable value is imputed	Presence	Error

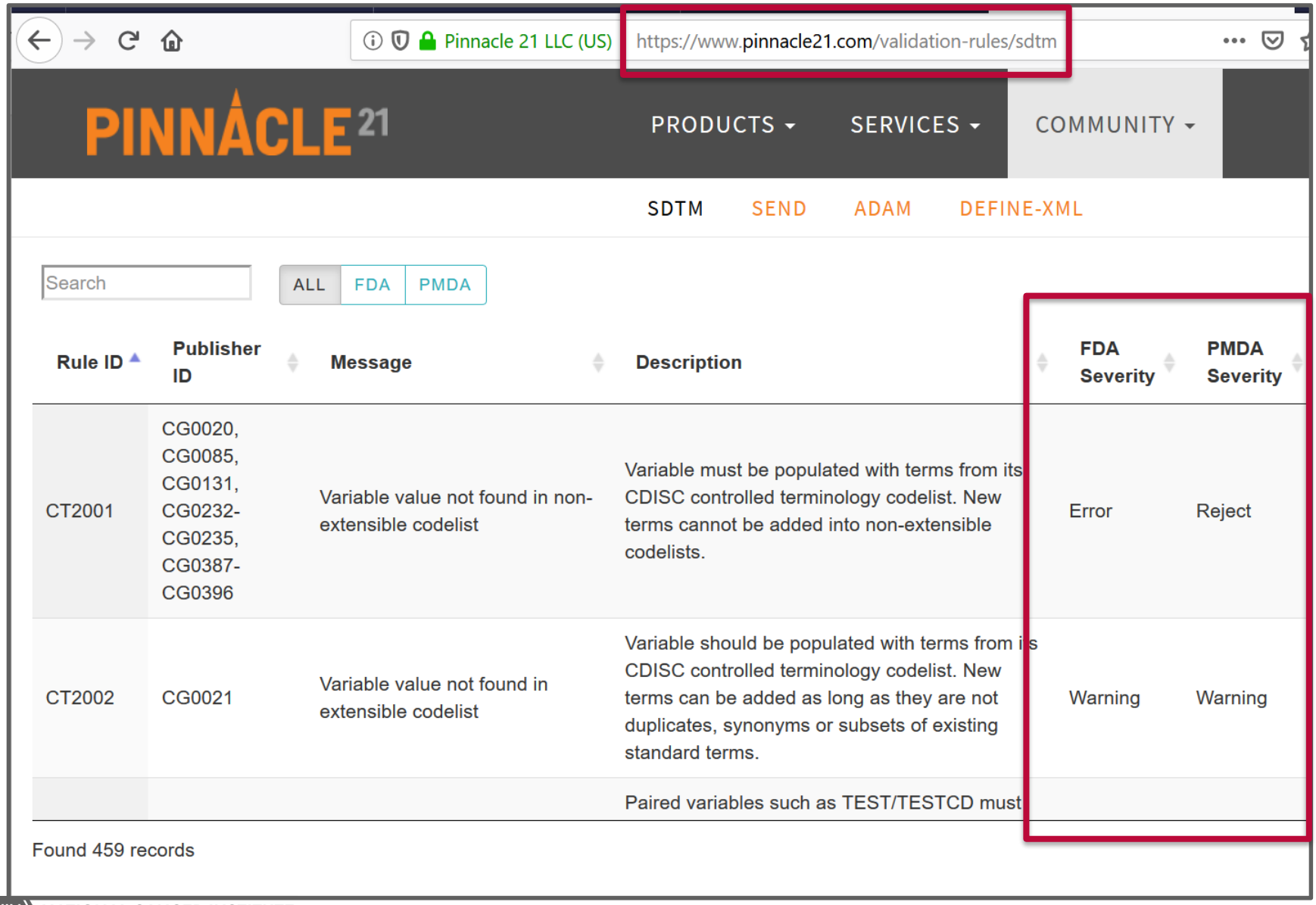
Dataset Summary | Issue Summary | **Details** | Rules

Validation Demo using Pinnacle 21 (Community Version)

Pinnacle 21 ID	Publisher ID	Message	Description	Category	Severity
CT2001	FDAC340	Variable value not found in non-extensible codelist	Variable must be populated with terms from its CDISC controlled terminology codelist. New terms cannot be added into non-extensible codelists.	Terminology	Error
CT2002	FDAC341	Variable value not found in extensible codelist	Variable should be populated with terms from its CDISC controlled terminology codelist. New terms can be added as long as they are not duplicates, synonyms or subsets of existing standard terms.	Terminology	Warning
CT2003	FDAC342	Coded and Decoded values do not have the same Code in CDISC CT	Paired variables such as TEST/TESTCD must be populated using terms with the same Codelist Code value in CDISC control terminology. There is one-to-one relationship between paired variable values defined in CDISC control terminology by Codelist Code value.	Terminology	Error
CT2004	FDAC343	Variable value not found in non-extensible codelist when value-level condition occurs	Variable must be populated with terms from its CDISC controlled terminology codelist, when its value level condition is met. New terms cannot be added into non-extensible codelists.	Terminology	Error
CT2005	FDAC344	Variable value not found in extensible codelist when value-level condition occurs	Variable should be populated with terms from its CDISC controlled terminology codelist, when its value level condition is met. New terms can be added as long as they are not duplicates, synonyms or subsets of existing standard terms.	Terminology	Warning
CT2006	FDAC345	Coded and Decoded values do not have the same Code in CDISC CT when value-level condition occurs	Paired variables such as TEST/TESTCD must be populated using terms with the same Codelist Code value in CDISC control terminology. There is one-to-one relationship between paired variable values defined in CDISC control terminology by Codelist Code value within the same value level condition.	Terminology	Error
SD0001	FDAC014	No records in data source	Domain table should have at least one record.	Presence	Error
SD0002	FDAC018	NULL value in variable marked as Required	Required variables (where Core attribute is 'Req') cannot be NULL for any records.	Presence	Error
SD0003	FDAC038	Invalid ISO 8601 value for variable	Value of Dates/Time variables (*DTC) must conform to the ISO 8601 international standard.	Format	Error
SD0004	FDAC056	Inconsistent value for DOMAIN	Domain Abbreviation (DOMAIN) variable should be consistent with the name of the dataset.	Consistency	Error
SD0005	FDAC044	Duplicate value for --SEQ variable	The value of Sequence Number (--SEQ) variable must be unique for each record within a domain and within each Unique Subject Identifier (USUBJID) , Pool Identifier (POOLID) or Sponsor Device Identifier (SPDEVID) variables value when they are present in the domain.	Consistency	Error
SD0006	FDAC113	No baseline result in Domain for subject	All subjects should have at least one baseline observation (--BLFL = 'Y') in EG, LB, MB, MS, PC and VS domains, except for subjects who failed screening (ARMCD = 'SCRNFAIL') or were not fully assigned to an Arm (ARMCD = 'NOTASSGN') or were not treated (ACTARMCD = 'NOTTRT').	Presence	Warning
SD0007	FDAC084	Inconsistent value for Standard Units	Standard Units (--STRESU) must be consistent for all records with same Short Name of Measurement, Test or Examination (--TESTCD), Category (--CAT), Specimen Type (--SPEC) and Method of Test or Examination (--METHOD).	Consistency	Error
SD0008	FDAC346	Value for --DECOD not found in MedDRA dictionary	Value for the Dictionary-Derived Term (--DECOD) variable must be populated using a Preferred Term of the MedDRA dictionary of a version specified in the define.xml (Case-insensitive).	Terminology	Error
SD0008C	FDAC347	Value for --DECOD is in incorrect case	Case for the Dictionary-Derived Term (--DECOD) variable must be sentence case using a Preferred Term of the MedDRA dictionary of a version specified in the define.xml (Case-sensitive).	Terminology	Error
SD0009	FDAC206	No qualifiers set to 'Y', when AE is Serious	When Serious Event (AESER) variable value is 'Y', then at least one of seriousness criteria variables is expected to have value 'Y' (Involves Cancer (AESCAN), Congenital Anomaly or Birth Defect (AESCONG), Persist or Signif Disability/Incapacity (AESDISAB), Results in Death (AESDTH), Requires or Prolongs Hospitalization (AESHOSP), Is Life Threatening (AESLIFE), or Other Medicaly Important Serious Event (AESMIE)).	Consistency	Error

Dataset Summary | Issue Summary | Details | **Rules**

Pinnacle 21 Publishes Severity by Agency



Rule ID	Publisher ID	Message	Description	FDA Severity	PMDA Severity
CT2001	CG0020, CG0085, CG0131, CG0232-CG0235, CG0387-CG0396	Variable value not found in non-extensible codelist	Variable must be populated with terms from its CDISC controlled terminology codelist. New terms cannot be added into non-extensible codelists.	Error	Reject
CT2002	CG0021	Variable value not found in extensible codelist	Variable should be populated with terms from its CDISC controlled terminology codelist. New terms can be added as long as they are not duplicates, synonyms or subsets of existing standard terms.	Warning	Warning
			Paired variables such as TEST/TESTCD must		

Found 459 records

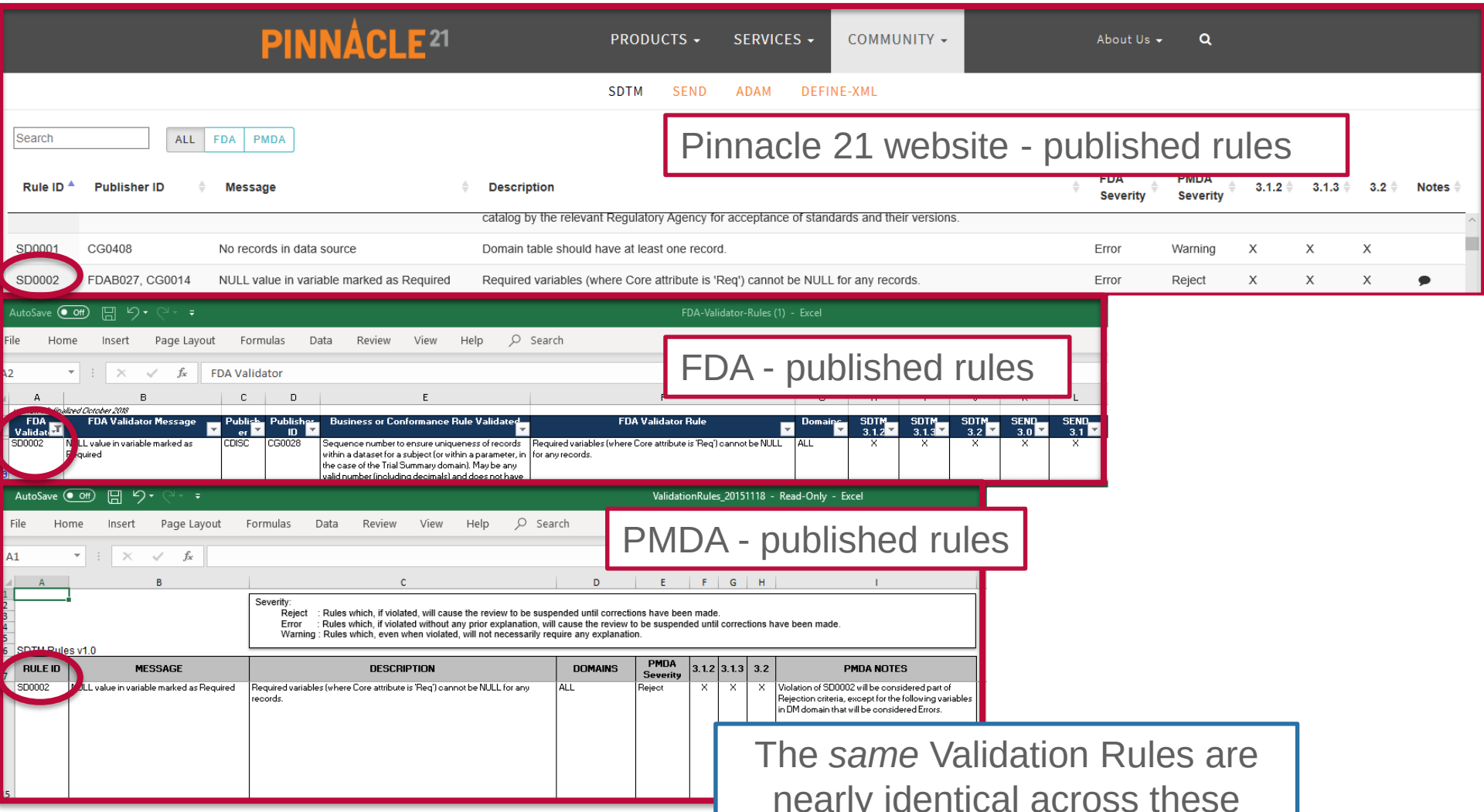
Severity of Validation Errors and Warnings - by Agency

Pinnacle 21 Severity	FDA	PMDA
Notice	FDA only. 2 in current version. Non-critical, just not "clean"- (Fix SD1078).	Both rules give a Warning for PMDA.
Error	-Fix issues -Document what cannot be fixed in xDRG	Rules which, if violated without any prior explanation (xDRG), will cause the review to be suspended until corrections have been made
Reject	*Not used - but if the data are deemed to be unusable for the review FDA can RTF/RTR*	Rules which, if violated, will cause the review to be suspended until corrections have been made

SD0063A	CG0303	SDTM/dataset variable label mismatch	Variable Label in the dataset should match the variable label described in SDTM IG. When creating a new domain Variable Labels could be adjusted as appropriate to properly convey the meaning in the context of the data being submitted.	Notice	Warning	X	X	X
SD1078	CG0015	Permissible variable with missing value for all records	Permissible variable should not be present in domain, when the variable has missing value for all records in the dataset.	Notice	Warning	X	X	X

Highlighted approach is the most conservative and will meet the requirements of both Agencies.

Example: Rule SD002 NULL Required Variables



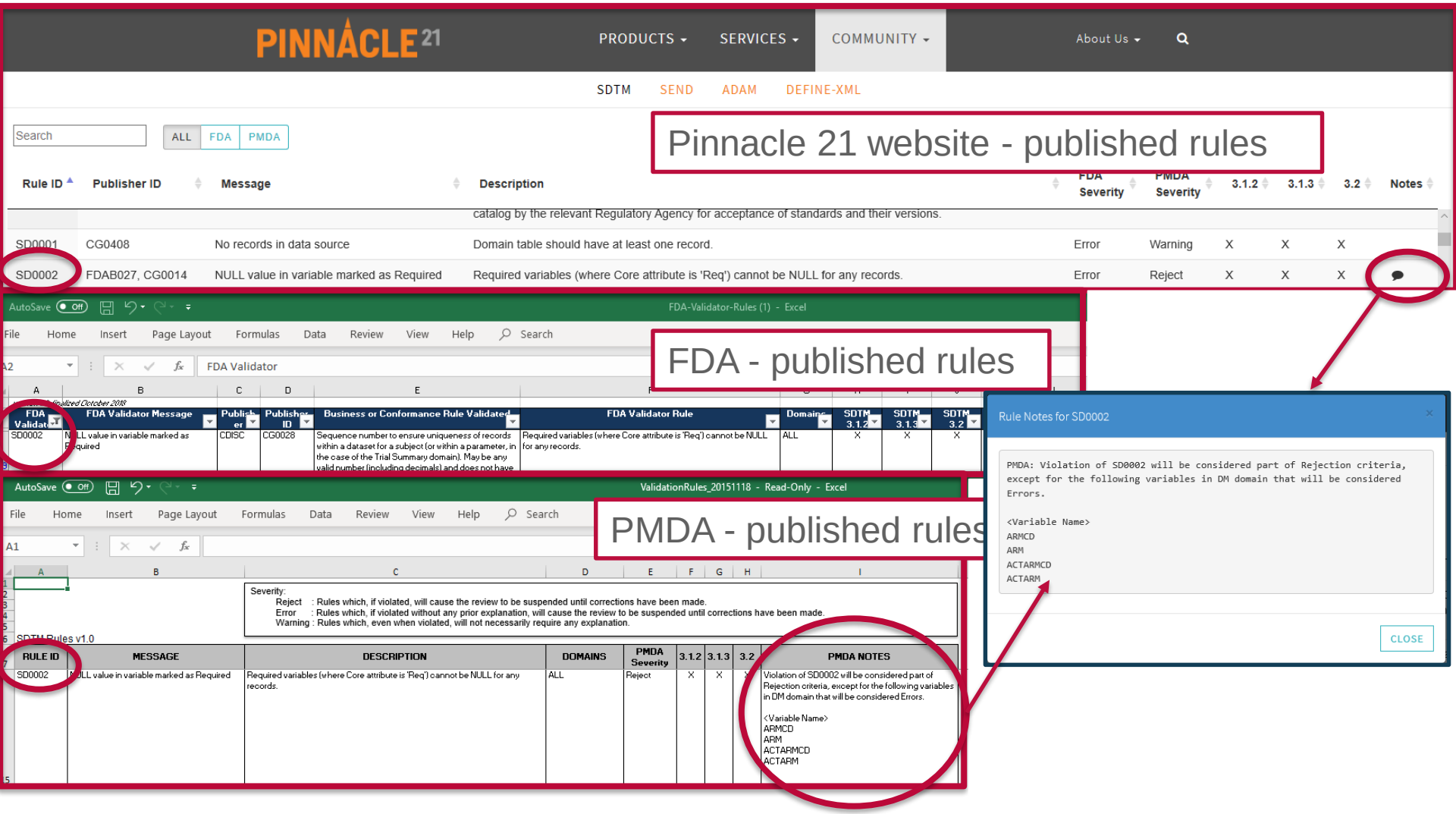
The *same* Validation Rules are nearly identical across these published sources, and use same reference numbers.

Example: Rule SD002 NULL Required Variables

Pinnacle 21 Severity	FDA	PMDA
Notice	FDA only. 2 in current version. Non-critical, just not "clean"- (Fix SD1078).	Both rules give a Warning for PMDA.
Warning	-Fix issues -Document what cannot be fixed in xDRG	Rules which, even when violated, will not necessarily require any explanation
Error	-Fix issues -Document what cannot be fixed in xDRG	Rules which, if violated without any prior explanation (xDRG), will cause the review to be suspended until corrections have been made
Reject	*Not used - but if the data are deemed to be unusable for the review FDA can RTF/RTR*	Rules which, if violated, will cause the review to be suspended until corrections have been made

PMDA Reject requirement is the most conservative approach. E.g., SDTMIG Required variables cannot be NULL, so this should be fixed in the data.

Other Differences Between FDA and PMDA Rules



FDA does not have this exception, so you should just fix all for FDA.

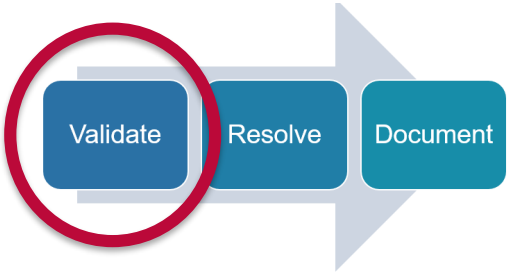
Other Differences Between FDA and PMDA Rules

Some rules have a Severity assignment from only one of the Agencies and not the other
(but you can still take a conservative approach to Resolve, or consult with the Agency to determine their preference)

Ru				FDA Severity	PMDA Severity
SD1				Warning	Warning
SD1071	FDAB024	Dataset is greater than 5 GB in size	Datasets greater than 5 gigabytes (GB) in size should be split into smaller datasets no larger than 5 GB.	Warning	

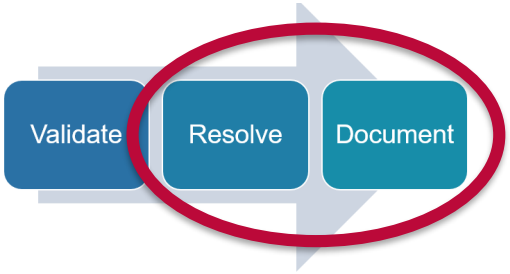
Rule ID	Publisher ID	Message	Description	FDA Severity	PMDA Severity
SD1137	CG0348	DRVFL='Y'	value is 'Y'.	Warning	Warning
SD1140		PMDA Expected variable not found	Variables requested by PMDA in policy documents should be included in the dataset. E.g., EPOCH.		Warning
SD1141		No Treatment Emergent info for Adverse Event	A treatment-emergent flag should be included in SUPPAE according to SDTMIG.		Warning
SD1142		Dataset is greater than 5 GB in size	Applicant should consult the PMDA beforehand if the dataset exceeds the file size specified by PMDA.		Warning
SD1143	CG0043	No Details info for AESMIE Adverse Event in SUPPAE domain	When a description of Other Medically Important Serious Adverse Events category is collected on a CRF, sponsors should place the description in the SUPPAE dataset using the standard supplemental qualifier name code AESOSP.	Warning	
SD1144	CG0079	MHSTDTC date is after RFSTDTC	The medical history dataset includes the subject's prior history at the start of the trial. Start Date/Time of Medica History Event (MHSTDTC) should be before Subject Reference Start Date/Time (RFSTDTC).	Error	
SD1145	CG0078	MHENDTC date is after RFSTDTC	The medical history dataset includes the subject's prior history at the start of the trial. End Date/Time of Medical History Event (MHENDTC) should be before Subject Reference Start Date/Time (RFSTDTC).	Error	
SD1147	CG0088	Missing --PRESP variable, when --OCCUR variable is present	Pre-specified (--PRESP) variable should be included into dataset, when Occurrence (--OCCUR) variable is present. (--PRESP/Events) Used to indicate whether the event describe by --TERM was pre-specified on a CRF Value is Y for pre-specified events, null for spontaneously reported events. (--PRESP/Interventions) Used when a specific intervention is pre-specified on a CRF. Values should be 'Y' or null. (--OCCUR) Used to record whether a pre-specified event occurred when information about the occurrence of a specific event is solicited.	Error	

What Validation Tools CAN check



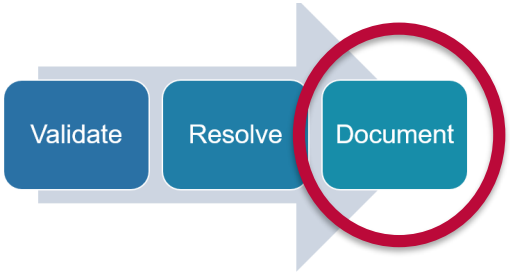
- Validation Tools can only check against the published rules (standards, FDA, PMDA)
- These rules are based on
  - Structured data rules, such as
    - Naming conventions for domains and variables
    - Variable order within each domain
    - Use of Controlled Terminology
    - Comparison between Define.xml and data in SAS xpt files
  - Agency (FDA, PMDA)-specific business rules
  - Software developer's interpretation of informative text in the published standards

What to DO With Validation Results



- Resolve:
  - Figure out which results are actually issues - investigate!
  - Fix: SDTM, CT and Define-XML structural problems, e.g.,
    - Correct order of variables
    - Use of correct CT (CDISC Submission Value)
    - Inclusion of required domains
    - Domain code, Variable Name and Variable Label
    - Consistent standardization for unique TESTCD
  - Document: anything that is inherent in the data and cannot be changed
    - **NOTE: You should NEVER modify data JUST to resolve a validation issue**
    - Document actual issues in the Study Data Reviewer's Guide (SDRG)

Study Data Reviewer's Guide (SDRG)

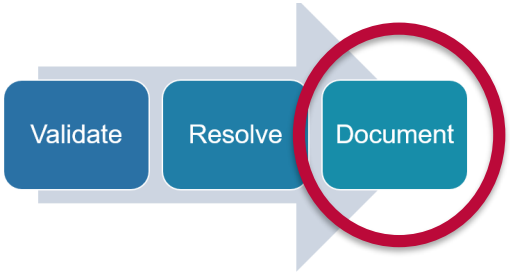


- Template available from PhUSE Wiki (Developed with FDA participation)
- https://www.phusewiki.org/wiki/index.php?title=Study_Data_Reviewer%27s_Guide

4.2 Issues Summary

Dataset	Diagnostic Message	Severity	Count	Explanation

Study Data Reviewer's Guide (SDRG)



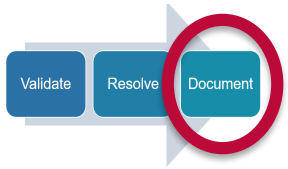
- Clearly explain the reason for each issue. Examples:

Dataset	Diagnostic Message	Severity	Count	Explanation
CM	Start Date/Time of Observation (--STDTC) or Start Relative to Reference Period (--STRF) should not be NULL, when End Date/Time of Observation (--ENDTC) or End Relative to Reference Period (--ENRF) is not NULL	Warning	37	The start date for historical corticosteroids was not reported for 37 subjects.

Dataset	Diagnostic Message	Severity	Count	Explanation
LB	Missing Units on Value	Error	22	**Not an error:** Lab results for pH and Specific Gravity have no units

Check ID	Diagnostic Message	Severity	Dataset	Count (Issue Rate)	Explanation
SD1118	Neither DSSTDTC, DSDTC nor DSSTDY are populated	Warning	DS	28(33.33%)	The date was not collected for these subjects.

Summary: Documenting Validation Issues in SDRG



- Thoroughly investigate each issue and make sure you understand the reason before you write an explanation for FDA
  - Make sure you fix structural SDTM issues and then re-validate
- Summarize the issues when you document them in the xDRG - don't write one for each record-level error or warning
- Clearly explain the reason for each remaining issue with sufficient detail that the Reviewer will understand
- Be transparent: insufficient or missing explanations will delay review and call your submission quality into question

What ***Else*** Needs to be done to confirm conformance?

- The structure of the SDTM datasets, and to a limited extent content (i.e., controlled terms) *can be* validated automatically

- HOWEVER

- There are other aspects of conformance that cannot be checked automatically

What ELSE needs to be done to confirm conformance

- Validation Tools **CANNOT** determine:
  - Domains were used for the right purpose
    - No way to detect that you put Medical History into AE or CM
  - Variables were used for the right purpose
    - No way to detect that the values in MHTERM are actually medical conditions
  - All of the data that should be in SDTM is actually in SDTM
  - Sufficient ADaM datasets to support the TLFs in your CSR
  - Uniqueness and consistency of USUBJID across studies
  - Correct standardization of unique --TESTCD/--TEST (because uniqueness may include other variables e.g., --METHOD, --SPEC)

These are areas that will be GREATLY supported by a good CDASH implementation

What ELSE needs to be done to confirm conformance

- Validation Tools **CANNOT** determine:
  - **Study Specific** Define-XML metadata is complete/correct
    - Origin
    - Meaningful, appropriate Comments
    - Study - specific codelists (all values available for data collection)
      - Variable-specific subsets
      - Codelist extensions
      - Sponsor-defined/custom CT
  - Human-readable form of Define.xml (appropriate style sheet)
    - Can you open it?
    - Do navigation links work as expected?

Automated Validation vs. Human Review

- We have to both manually review the data *and the* automated validation output
  - Did we use the standards correctly, and
  - Have all structural and content errors and warnings been addressed in the SDRG?
    - Data issues
    - False Positives
    - Study design
  - Are there any **quality** issues that surfaced during the validation (stop date before start date) that can (and should) still be fixed?
    - ➢ **THIS CAN ONLY HAPPEN if you start validation processes BEFORE DATABASE LOCK, while data changes can still be made**

Human Review: **Data**

- Are **all** of the data from your study represented in SDTM, √ or have some been left out?
- Do the Reference Dates in DM have a consistent definition (in Define.xml) and do they accurately reflect their defined purpose? √
- Have all deaths been reported using DTHFL in DM? √
- Have data been synthesized or imputed to avoid validation errors?
- Have data cleaning aids or other operational data been added to the SDTM dataset?

Human Review: **Data**

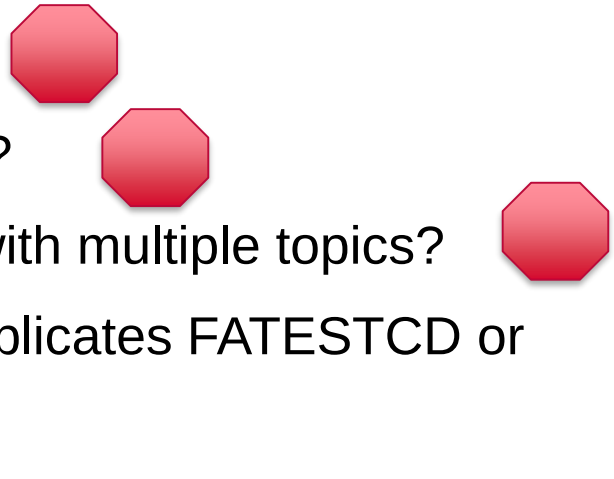
- Have codes been added to the dataset or used instead of controlled terminology? (*all values should be human interpretable - not coded)*
- Are the dates all in proper ISO 8601 format (without imputing missing parts, or zero-filling uncollected parts)? √
- Do all Supplemental Qualifiers make sense in the context and timing of the parent record? √
- Are there any extensible codelists to which values have been improperly added because they duplicate standard values?

Human Review: **Data**

- If you are using MODIFY, has it been populated correctly ? √
- Is USUBJID formatted consistently across your submission ? √
- Does USUBJID accurately reflect unique human subjects using a consistent value ? √

Human Review: **Domains**

- Are any data placed in the wrong domains?
- Are there any unnecessary custom domains?
- Are there any domains that are overloaded with multiple topics?
- Has FA been used improperly? (If FAOBJ duplicates FATESTCD or FATEST, then FA is not needed)



Human Review: **Variables**

- Have any of the variables been used for the wrong purpose?
- Has --STRESN always been populated when --STRESC contains a numeric value ? √ Is --STRESN blank when --STRESC is a character value ? √
- Have Permissible variables been added to any domain where they were not needed?
- Have Supplemental Qualifiers been used when Findings About would have been the appropriate data structure (e.g., when the SUPP-- records don't share the timing of the parent record), or vice versa?

Human Review: **Relationships**

- Does RELREC reflect the collected relationships or have they been represented somewhere else in the data?
- Have non-unique QNAMs been used for the same data in SUPP--?
- Are the dataset to dataset relationships (e.g., PC, PP) in RELREC accurately representing the relationships ? √

Human Review: **Trial Design**

- Do these trial design datasets reflect the study accurately ? √
  - Trial Summary
  - Trial Arms / Trial Elements
  - Trial Visits
  - Trial Inclusion/Exclusion

Human Review: **Define.xml**

- Do the codelists include all of the values that were available for entry during the study ? √
- Is the ORIGIN correct for *your* data ? √
- Have appropriate comments been added for all derived variables, and any other variables that need additional variable-level information provided to the reviewer ? √

Responsibility for Validation of Standard Data

- During and post- SDTM (ADaM, SEND) programming, whoever is preparing the submission documentation should ensure validation is done and documented in the appropriated xDRG
- May be an iterative process that ***could*** **involve LPOs** who collected data
  - To help investigate specific issues
  - To provide information that may not be explicit in the data

How LPOs Can Support Efficient Submission Preparation

- LPOs can PROACTIVELY support greater efficiency in SDTM prep by:
  - Ensuring *all collected* data are sent to submission programming function
    - CRF/eCRF
    - Wearables and other machine-generated data
      - Protocol should describe what will be collected since it is not usually practical to include ALL
    - Questionnaires, Ratings, Scales (PRO/ePRO, COA/eCOA)
      - Protocol should describe what will be *collected* – May or May Not be all responses (e.g., may be Total Score)
    - Safety core lab and local lab data
    - PK, MB, MI

Etc.

What LPOs can do to Support High Quality SDTM

- Use standardized "concepts" to collect data
  - *Alignment with SDTM concepts is crucial to successful, efficient SDTM programming*
  - *This is the primary value of the CDASH implementation*
- Use standardized variable names to export collected data for SDTM programming
  - No "mapping" required – only a few standardized transformations (e.g., --DAT/--TIM to --DTC)
  - Predictability allows SDTM programmers to write standard programs
  - *This is the primary value of using pattern-based variables names*

This is supported by NCI's CDASH implementation.

How LPOs Can Support Efficient Submission Preparation

- Send the SDTM programmers a clearly annotated CRF with
  - Intended mapping to SDTM (detailed mapping specification, or build into pattern-based variable names)
  - Complete codelists (i.e., all values *available* for data collection in this study)
    - Without this, the SDTM programmer has to rely on what they can see in the collected data, and it may not include all possible values
    - What versions of *published* terminologies were used in the study
    - Your extensions to extensible codelists and your custom codelists
  - Identify values that were Derived (e.g., BMI) in Rave
  - Identify values that were Assigned (e.g., internal AE and CM coding)
  - Identify values displayed on the eCRF that are from the protocol (e.g., SITTING position for blood pressure)

What LPOs can do to Support High Quality SDTM

- Proactively communicate to the SDTM programmers anything else that is not explicit in the data
  - **Origin**: How each datapoint was collected (eCRF? Loaded from core lab file? Entered by an in-house medical coder? Derived after data entry?)
    - Describe at the TEST level for Findings Class data
    - Describe for each QNAM/QVAL
  - **Derivations**: What algorithm was used to derive values (if you are deriving anything, like BMI)

What LPOs can do to Support High Quality SDTM

- Proactively communicate to the SDTM programmers anything else that is not explicit in the data
  - **Participants who were in more than one study for any given IP**
    - Which studies were they in and what was their identifier in each study
  - **RELREC**: Whether there are any collected relationships (e.g., AE and CM)
  - **--EVAL**: If any values were provided by someone *other than the PI*
    - *What was the role of the person who provided the value?*
    - *Was there an identifier used for the person in the study? If so, what was it (e.g., EVAL1, EVAL2, Reader 1, Reader 2)?*

What LPOs can do to Support High Quality SDTM

- Understand the FDA's rules for conformance and **proactively collect the necessary data**
  - E.g., FDA Technical Conformance Guide has this requirement in Section 4.1.1.3 for the Adverse Event (AE) domain:
  - *"The entry of a 'Y' for the serious adverse event variable, AESER, should **have the assessment indicated**, (e.g., as a death, hospitalization, or disability/permanent damage). Frequently, sponsors omit the assessment information, even when it has been collected on the CRF. The criteria that led to the determination should be provided. This information is critical during FDA review to support the characterization of serious AEs."*
  - In other words, FDA expects to see the criterion/criteria that made the event serious *in the AE domain data. The most efficient way to get that is to proactively collect it in the eCRF (and then perform appropriate CDM processes to verify it before database lock).*

Summary and Final Thoughts

- Whether or not you are directly responsible for preparing SDTM datasets, proactive planning and implementation by the LPOs will go a long way toward supporting and facilitating high quality submission data, which in turn can speed and enhance the FDA Review process
- Proactively create a mapping to SDTM for every study so you know where the data are going before you collect it, and so you can communicate that to the SDTM programmers early in the process
- **Communicate** necessary information to SDTM programmers
- Remain knowledgeable
  - Requirements from FDA and PMDA
  - Foundational standards - requirements, updates

Understand the importance of the NCI CDASH implementation to support high quality submission data.

Q&A

NCICDISCSupport@nih.gov