

Diagnostics of technical errors in define.xml file

#SI07

PhUSE EU CSS 2018

Sergiy Sirichenko
November 6, 2018

Why define.xml is important?

- › CDISC standards are required for submitting study data to FDA and PMDA
- › Standards speeding up the review process and support automation
- › Define.xml file plays a critical role in automating processes as a major source of machine-readable study metadata
- › Some small errors in define.xml file may be critical for executing automated processes

FDA automation of validation

- › For validation programmers specify standards info manually
- › FDA DataFit extracts a version of SDTM standard from define.xml file

```
def:StandardName="SDTM-IG"  
def:StandardVersion="3.1.2">
```

- › If this version is incorrect, then a validation process will produce both false-positive and false-negative results
 - PMDA Rejection criteria and validation Errors reconciliation

Compensatory actions for missing info

- › If study metadata is invalid or missing then some compensatory actions are taken
 - the latest version of a standard or dictionary will be utilized instead

- › MedDRA example

```
<ExternalCodeList Dictionary="MedDRA" Version="19.0"/>
```

- Incorrectly implemented (e.g., like Comment instead of Dictionary)
 - MedDRA version also means missing machine-readable metadata
 - Missing decimal points in MedDRA version (e.g., “19” instead of “19.0”) is PMDA Rejection criteria
- › CT version info will be available only in Define-XML 2.1

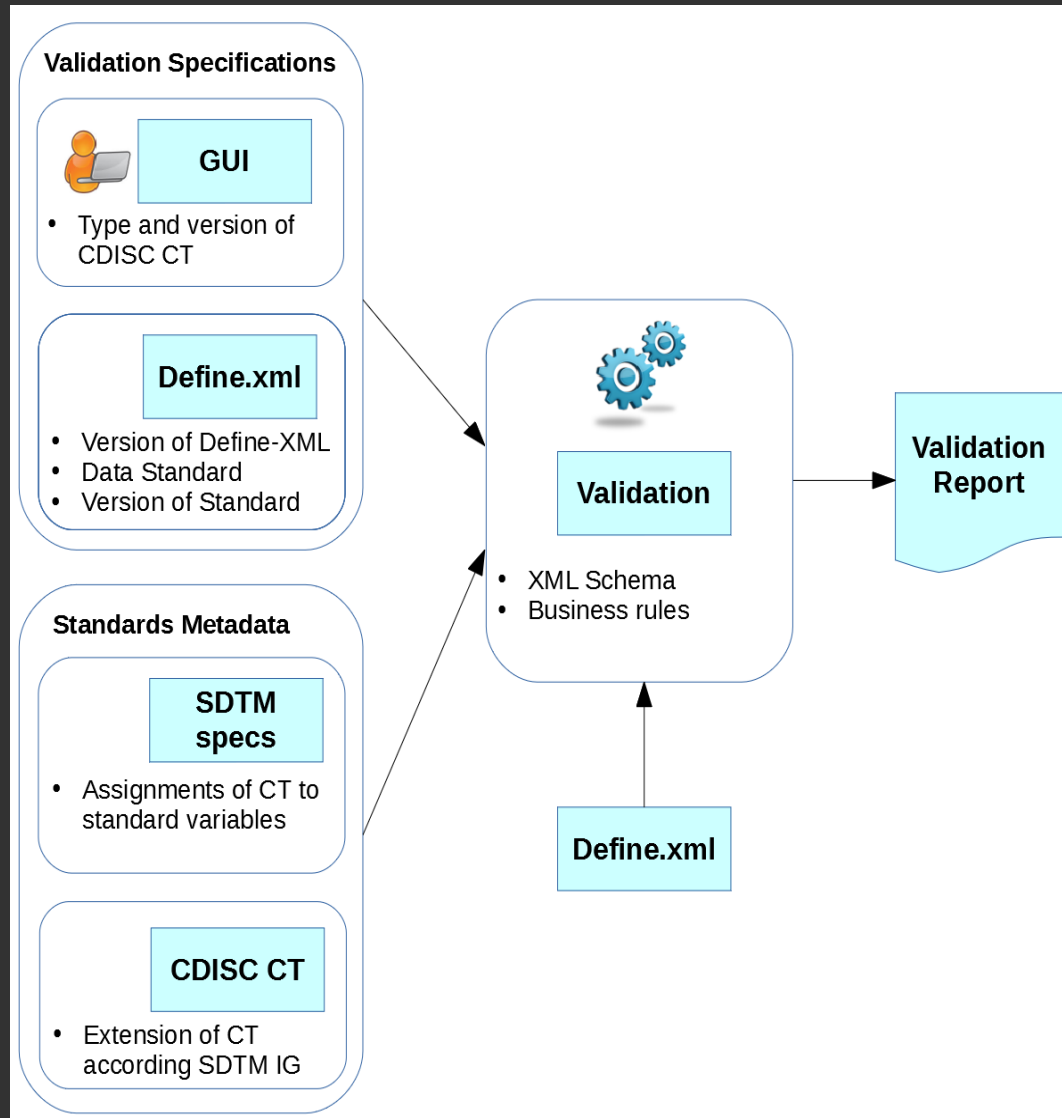
Types of validation rules

- CDSIC XML schema for Define-XML standard is machine-readable specifications
- Pinnacle 21 Validator
 - 27 XML schema related checks
 - 96 additional content-related checks
 - submission-specific business rules
 - ✓ *“MedDRA version should include decimal point”*
 - conforming correct implementation of CDISC CT
 - ✓ *“Term/NCI Code mismatch in Codelist”*
 - Define-XML rules which are not covered by XML schema

3 steps of validation

1. Ensure that there are no major structural problems in define.xml file and its content is readable
 2. Content related checks
 - If the first step failed, then content related checks cannot be executed or may produce unexpected results
 3. Consistency of study data and define.xml info
- › 2 separate validations
1. Define.xml
 2. SDTM (SEND or ADaM) datasets and define.xml

Configurations for define.xml validation



Standards and versions

```
<MetaDataVersion  
OID="MDV.CDISC01.SDTMIG.3.1.2.SDTM.1.2"  
  Name="Study CDISC01, Data Definitions"  
  Description="Study CDISC01, Data Definitions"  
  def:DefineVersion="2.0.0"  
  def:StandardName="SDTM-IG"  
  def:StandardVersion="3.1.3">
```

Standard extensions

- › SDTM (SEND, ADaM) specs have assignment of CDISC CT to standard variables
- › CT files include additional specifications from SDTM IG
 - For example, Baseline Flag variables may be populated only with a single value 'Y' from standard 'Yes/No' Codelist in CDISC CT which includes 5 terms instead of 1 and therefore cannot be directly utilized for validation of CT for Baseline Flag variables.

Incorrect configuration specs

- If there is something wrong with validation configuration specifications, then validation process may produce unexpected or confusing results
- It requires special diagnostics based on good knowledge of validation logic and computational algorithms for validation checks
 - For example, if version of utilized Define-XML standard is missing or provided as invalid value (`def:DefineVersion="1"` instead of correct `def:DefineVersion="1.0.0"`), then version 2.0 will be utilized for validation as the most recent version of Define-XML standard.

Validation errors due to incorrect version of CDISC CT

- Example 1. Error DD0028: *Term/NCI Code mismatch in Codelist 'Laboratory Test Code'*
 - A reported term 'CALB' and its NCI Code 'C125942' are actually correct standard term and corresponding NCI Code in CDISC SEND CT
- The actual problem is that this term was introduced in CDISC SEND CT version 2016-03-25, while according to sponsors Reviewer's Guide, CT version 2015-12-18 was utilized in the study

Validation logic

- P21 validator was configured to use SEND CT 2015-12-18 and utilized this standard metadata as a lookup table to find a match for '*CALB / C125942*' which does not exist yet in this version of SEND CT
- Solutions:
 1. Correct version of CT in Reviewer's Guide
 2. Consider '*CALB*' term as non-standard according to claimed CT version; correct define.xml file by removing NCI Code for this term and flag it as a formal extension of standard CT

Standard terms may be different across CT versions

- › NCI Code = C17988

Term	SDTM CT 2011-07	SDTM CT 2018-06
<i>U</i>	6	3
<i>UNK</i>	1	1
<i>UNKNOWN</i>	5	14

- › Term '*U*' was changed to new standard term '*UNKNOWN*' in SDTM CT 2014-03-28 for 3 non-extensible codelists (ENRTPT), (STENRF) and (STRTPPT)

Use of different CT codelist

- Error DD0028: *Term/NCI Code mismatch in Codelist 'PPSTUNIT'*
- Due to use of standard terms from a different CDISC CT codelist (PKUDMG), while referencing to CDISC CT codelist (PKUNIT)
 - '*nmol/L/(mg/kg)*' (C119458) is a standard term in CDISC CT (PKUDMG) "PK Units of Measure - Dose mg", NCI Code = C128685
 - In define.xml this term is used in codelist with reference to CDISC CT codelist (PKUNIT) "PK Units of Measure", NCI Code = C85494, which does not include this reported term

PK Units codelists in SEND CT

Codelist Code	Codelist Name	NCI Code	Number of Terms
PKUDMG	PK Units of Measure – Dose mg	C128685	144
PKUDUG	PK Units of Measure – Dose ug	C128686	127
PKUNIT	PK Units of Measure	C85494	292
PKUWG	PK Units of Measure – Weight g	C128684	56
PKUWKG	PK Units of Measure – Weight kg	C128683	54

- ‘*nmol/L/(mg/kg)*’ (C119458) was standard term in CDISC CT codelist (PKUNIT) “PK Units of Measure”, NCI Code = C85494 in older versions until 2016-09-30
- In this version of SEND CT, ‘*nmol/L/(mg/kg)*’ term was removed from (PKUNIT) codelist and added into newly introduced codelist (PKUDMG) and (PKUDUG)
- It was added back to codelist (PKUNIT) in the most recent versions (2017-12-22 +)

Standard is time-sensitive

- › '*nmol/L/(mg/kg)*' is formally considered as a standard term associated with PPSTRESU variable and representing by CDISC CT (PKUNIT) codelist
 - with exception of time period from 2016-09-30 to 2017-12-22
- › For this study sponsor utilized CT 2016-12-16 where a term '*nmol/L/(mg/kg)*' was treated as non-standard for CT (PKUNIT) codelist
- › The use of standard terms from different codelists does not allow formally treating these terms as standard within original target CT codelist

Other examples related to invalid CT version

- DD0032: *Missing NCI Code for Term in Codelist 'Unit' with reported terms 'L/L', 'ng/mL', 'ng/mL/mg'.*
 - These terms existed in CDISC SDTM CT 2014-03-28, but they were removed in more recent versions utilized for study
- DD0033: *Unknown NCI Code value for Codelist 'Specimen Material Type'*
- DD0034: *Unknown NCI Code value for Term in Codelist 'Epoch'*

Additional invalid standard terms

- › DD0024: *Invalid Term in Codelist 'No Yes Response'*
 - reporting 'N', 'U' or 'NA' terms which are standard terms in CDISC CT
- › Due to invalid utilization of non-relevant terms for Flag variables like DTHFL, --BLFL, --DRVFL or --PRESF
 - where these 'N', 'U' and 'NA' terms are not applicable according to SDTM/SEND IG documentation
- › P21 validator uses extensions for CT according SDTM specs

Validator uses CT codelists per standard specs

- › Variable codelist in define.xml file may have a reference to CDISC CT specific codelist
- › Pinnacle 21 uses independent internal codelist assignment as a part of SDTM metadata
 - To avoid potential mistakes done by programmers in assignment of CDISC CT codelist to standard variables in define.xml file
 - For example, if a codelist for LBTEST variable in define.xml has a reference to CDISC CT codelist (UNIT), the Validator will still use CDISC CT codelist (LBTEST) as specified by SDTM standard

Invalid assignment of CT codelist

- DD0028: *Term/NCI Code mismatch in Codelist 'UNIT'* reporting 'IU, C48579'
 - 'U' is a standard term in CDISC SDTM CT codelist (UNIT)P
 - Provided NCI Code for this term is also correct
 - There is no issues around inconsistent version of CDISC CT
- An actual problem was due to invalid use of CDISC CT codelist (UNIT) for VSORRESU/VSSTRESU variables instead of CDISC CT codelist (VSRESU)

Validation logic

- Validator uses CDISC CT codelist assignments to standard variables as specified by CDISC standards
 - It ignores assignments of CDISC CT to standard variables in define.xml file as potentially unreliable information
- In this case, the tool uses CDISC CT codelist (VSRESU) as a lookup table for validation of 'IU, C48579' combination
 - It exists in CDISC CT codelist (UNIT), but does not exist in expected correct codelist (VSRESU)
- Similar DD0033: *Unknown NCI Code value for Codelist 'Unit'*

Invalid Source of NCI Code

- Issue DD0034: *Unknown NCI Code value for Term in Codelist*
 - reported for term '*COLOR*' with NCI Code=C37927 in '*Laboratory Test Code*' codelist of LB domain
 - Correct NCI Code for term '*COLOR*' in LB domain is C64546
 - The same standard terms may have different NCI Code across codelists

Codelist Code	Codelist Name	CT Term	NCI Code
LBTESTCD	Laboratory Test Code	<i>COLOR</i>	C64546
OETESTCD	Ophthalmic Exam Test Code	<i>COLOR</i>	C37927
PHSPRPCD	Physical Properties Test Code	<i>COLOR</i>	C37927
URNSTSCD	Urinary System Test Code	<i>COLOR</i>	C64546

Incorrect standard

- › DD0045: *Missing Domain value* reported for ADaM define.xml
 - However, Domain attribute is applicable only for tabulation data and is not used for analysis data

```
def:StandardName="SDTM-IG"  
def:StandardVersion="3.2">
```

- › This info is used as configuration for validation
 - If define.xml refers to SDTM standard, then all SDTM related business rules will be applied including a requirement for non-missing Domain values

Unsupported version of standard

- DD0055: *Invalid Class value* reported for ADaM define.xml
 - However, Class values look good

```
def:StandardName="ADaM-IG"  
def:StandardVersion="1.1">
```

- P21 Community 2.2.0 does not support new versions of standards
 - ADaM IG 1.1
 - SEND IG 3.1
 - If tool cannot recognize expected versions of standards, then validation process may produce unexpected results

Bug in P21 Community 2.2.0

- The Validator recognizes ODM element only if it starts with a new line of XML code like this modified code:

```
<?xml version="1.0" encoding="ISO-8859-1"?><?xml-  
stylesheet type="text/xsl" href="define2-0-  
0.xsl"?><!-- Produced from SAS data using the SAS  
Clinical Standards Toolkit 1.6 -->  
<ODM xmlns:xlink="http://www.w3.org/1999/xlink" ... >
```

- Observed in files created by SAS® Clinical Tool Kit
- Often this issue is false-positively reported due to leading comments <!-- .../> in front of ODM element

```
<!-- comments --><ODM ... >
```

Links to CRF pages do not work

- › Correctly implemented links to variable-specific CRFs pages in define.xml file do not work in browser
 - Annotated CRFs document always opens on page 1 instead of a page specified by a link
- › Ensure that your browsers (Internet Explorer or Chrome) are configured to open linked acrf.pdf file within the browser rather than in other viewers like Adobe Acrobat Reader
 - Try to use different browsers and computers as simple first-step diagnostics

Summary & Questions

- › Define.xml file is a major source of machine-readable study metadata used by automated processes at FDA and PMDA
- › You should ensure that define.xml files include all expected content and do not have any technical errors
- › Understanding of define.xml validation logic is needed for diagnostics of validation findings

Sergiy Sirichenko

Pinnacle 21

sergiy@pinnacle21.com

Max Kanevsky

Pinnacle 21

max@pinnacle21.com