Using the OpenCDISC validator for Ongoing Studies

Stefan Bordasch, 18th FEB 2014

for the 18th German CDISC User Group Meeting

PAREXEL.

Right where you need us ™

OpenCDISC Validator – what is that

- Free-ware tool – used for SDTM conformance check by FDA
  - FDA: Report should be part of final SDTM submission package

- Three different modes:
  - Check SDTM datasets only (in XPT format)
  - Check define.xml
  - Check both components plus additional crossover checks

- Option for different SDTM versions in place – 3.1.1, 3.1.2, 3.1.2 amendment and 3.1.3

- Output: OpenCDISC Validator Report (as Excel or CSV file)

What is the OpenCDISC validator (2)

- OpenCDISC Validator Report based on predefined checks
  - Different versions have a different set of these checks (or rules):
    - Version 1.2:   126 rules
    - Version 1.3:   243 rules
    - Version 1.4:   317 rules
    - Version 1.4.1: 543 rules (most new checks are for QS domain and the QS-CT)

- Two different group of checks:
  - CT checks for correct use of SDTM Controlled Terminologies
  - SD checks for syntax and (partially) semantic

- Each rule has an unique code
  - …and an explanation what it is about

What is the OpenCDISC validator (3)

- Examples

  - CT checks:

Rule ID	Message	Description	Category	Severity
CT0001	Value for --ACN not found in (ACN) CT codelist	Action Taken with Study Treatment (--ACN) variable values should be populated with terms found in 'Action Taken with Study Treatment' (C66767) CDISC controlled terminology codelist	Terminology	Warning
CT0002	Value for AESEV not found in (AESEV) CT codelist	Severity/Intensity (AESEV) variable values should be populated with terms found in 'Severity/Intensity Scale for Adverse Events' (C66769) CDISC controlled terminology codelist	Terminology	Warning
CT0003	Value for TSVAL not found in (AGESPAN) CT codelist	TS dataset Parameter Value (TSVAL) values should be populated with terms found in 'Age Span' (C66780) CDISC controlled terminology codelist, when Trial Summary Parameter Short Name (TSPARMCD) value is 'AGESPAN'	Terminology	Warning
CT0004	Value for AGEU not found in (AGEU) CT codelist	Age Units (AGEU) variable values should be populated with terms found in 'Age Unit' (C66781) CDISC controlled terminology codelist	Terminology	Error
CT0005	Value for TSVAL not found in (AGEU) CT codelist	TS dataset Parameter Value (TSVAL) values should be populated with terms found in 'Age Unit' (C66781) CDISC controlled terminology codelist, when Trial Summary Parameter Short Name (TSPARMCD) value is 'AGEU'	Terminology	Warning
CT0006	Value for COUNTRY not found in (COUNTRY) CT codelist	Country (COUNTRY) variable values should be populated with terms found in 'Country' (C66786) CDISC controlled terminology codelist	Terminology	Warning

  - SD checks:

Rule ID	Message	Description	Category	Severity
SD0001	No records in data source	'Domain table should have at least one record	Presence	Warning
SD0002	NULL value in variable marked as Required	Required variables (where Core attribute is 'Req') cannot be NULL for any records	Presence	Error
SD0003	Invalid ISO 8601 value	Dates and times of day must conform to the ISO 8601 international standard	Format	Error
SD0004	Inconsistent value for DOMAIN	Domain Abbreviation (DOMAIN) variable should be consistent with the name of the dataset	Consistency	Warning
SD0005	Duplicate --SEQ	The value of Sequence Number (--SEQ) variable must be unique for each record within a subject	Consistency	Error
SD0006	No baseline result in -- for subject	All subjects should have at least one baseline observation (--BLFL = 'Y') in EG, LB, QS, and VS domains, except for subjects who failed screening (ARMCD = 'SCRNFAIL') or were not fully assigned to an Arm (ARMCD = 'NOTASSGN')	Presence	Warning
SD0007	Inconsistent value for Standard Units	Standard Units (--STRESU) must be consistent for all records with same Short Name of Measurement, Test or Examination (--TESTCD) and Category (--CAT)	Consistency	Error
SD0008	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	Terminology	Error

OpenCDISC Validator - how to read the output

- Report contains 3 level of details:

  - Dataset Summary

Processed Sources							
Name	Label	Class	Source	Records	Errors	Warnings	Notices
AE	Adverse Events	EVENTS	AE.xpt	778	0	2414	1
CM	Concomitant Medications	INTERVENTIONS	CM.xpt	104481	0	175809	0
CO	Comments	SPECIAL PURPOSE	CO.xpt	3354	0	0	0
DA	Drug Accountability	FINDINGS	DA.xpt	9280	0	7515	0
		SPECIAL					

  - Issue Summary (per domain)

Issue Summary				
Source	Rule ID	Message	Severity	Found
AE				
	CT0027	Value for AEOUT not found in (OUT) CT codelist	Warning	776
	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Warning	778
	SD0009	No qualifiers set to 'Y', when AE is Serious	Warning	37
	SD0037	Value for AESEV not found in (AESEV) user-defined codelist	Warning	778
	SD0080	AE start date is after the latest Disposition date	Warning	26

  - Details

Name	Record	Count	Variables	Values	Rule ID	Message	Category	Severi
AE			VARIABLE, LENGTH	AEPRESP, 8	SD1054	Non-recommended variable length	Metadata	Warning
AE			VARIABLE, LENGTH	AESER, 8	SD1055	Non-recommended variable length	Metadata	Warning
AE			VARIABLE, LENGTH	AESDTH, 8	SD1055	Non-recommended variable length	Metadata	Warning
AE			VARIABLE, LENGTH	AECONTRT, 8	SD1055	Non-recommended variable length	Metadata	Warning
AE		166	AEOUT	NOT RECOVERED	CT0027	Value for AEOUT not found in (OUT) CT codelist	Terminology	Warning
AE		547	AEOUT	RECOVERED	CT0027	Value for AEOUT not found in (OUT) CT codelist	Terminology	Warning
AE		3	AEOUT	RECOVERED WITH SEQUELAE	CT0027	Value for AEOUT not found in (OUT) CT codelist	Terminology	Warning
AE		60	AEOUT	RECOVERING	CT0027	Value for AEOUT not found in (OUT) CT codelist	Terminology	Warning
AE		3	AEBODSYS	Blood and lymphatic system disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		45	AEBODSYS	Cardiac disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		5	AEBODSYS	Ear and labyrinth disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		2	AEBODSYS	Endocrine disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		10	AEBODSYS	Eye disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning
AE		70	AEBODSYS	Gastrointestinal disorders	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Terminology	Warning

- "Dataset Summary" is only for general overview

- "Issue Summary" is the main working area
  - Check all messages
  - Check all lines and find the "reason" for this issue:
    - Is it a data issue? If yes, then report it to the CDA for resolution
    - Is it a programming issue? If yes, then report it to the SDTM programmer
    - Is it based on a sponsor specification? If yes, then this is "acceptable"
  - FDA expects to get a comment to each line of the report

- "Details" gives you the information what might be wrong

- "Rules" defines the logic behind a check

What you have to do with that output

- Add an additional column in the report and add your comments about all issues:

Sou	Rule ID	Message	Severity	Found	
Issue Summary					
AE					
	CT0027	Value for AEOUT not found in (OUT) CT codelist	Warning	776	non-extensible code list has not expected entries
	CT0037	Value for AEBODSYS not found in (SOC) CT codelist	Warning	778	not sure why this exist
	SD0009	No qualifiers set to 'Y', when AE is Serious	Warning	37	data issue - AESEV=N even when severity flag is set
	SD0037	Value for AESEV not found in (AESEV) user-defined codelist	Warning	778	mismatch between dataset and define.xml ("Code Value" has to be in upper case too)
	SD0080	AE start date is after the latest Disposition date	Warning	26	no issue
	SD1021	Unexpected character value	Warning	2	data issue
	SD1031	Value for AEENRF is populated, when RFENDTC is NULL	Warning	13	RFENDTC=NULL means to have a screeing failure - is there an AE follow-up? Else this is a mapping issue (depending on the sponsor definition of end reference period date)
	SD1054	Non-recommended variable length	Warning	1	Y/N flag needs length $1 - here it is 8
	SD1055	Non-recommended variable length	Warning	3	Y/N flag needs length $1 - here it is 8
	SKIP_SD0008	The source data for C:\Studien\opencdisc-validator_V1_3\config/data/MedDRA/12.1/pt.asc is	Notice	1	???
CM					
	CT0018	Value for CMDOSFRQ not found in (FREQ) CT codelist	Warning	4755	extensible code list - no issue
	CT0019	Value for CMDOSFRM not found in (FRM) CT codelist	Warning	96258	extensible code list - no issue
	SD0021	Missing value for CMENRF, when both CMENDTC and CMENRTPT are NULL and CMOCCUR is not 'N'	Warning	36206	mapping issue - medication has been taken but there is no documentation when

The limits of the OpenCDISC checker

- OpenCDISC Checker is no „magic bullet"

  - Is not able to check the content for free text fields –
    e.g. AETERM => does not find typos or wrong entries as „AE1"

  - Needs review of the different issues, example:

SD0065	Invalid Subject Visit/Visit Number	All Unique Subject Identifier (USUBJID) + Visit Name (VISIT) + Visit Number (VISITNUM) combination values in data should be present in the Subject Visits (SV) domain

  could be a not defined entry – so an Error
  …or: that message flags an Unscheduled Visits that is not part of the
  planed visits structure in Trial Design – so all ok

- Another problem: SDTM is planned for clean, locked studies

  - Some messages appear due missing information, which could not be
    in place at the moment, e.g. end of study date per subject

Rule ID	Message	Severity	Found	Programmer comment
SD0050	Missing value for CMENTPT, when CMENRTPT is populated	Error	29	as long as the study is ongoing and some subjects don't have finished the study, these entries will exist

Thank you