

OpenCDISC Live Forum

June 9th 2014

Majdoub Haloui Principal Analyst, Statistical Submissions Management Biogen idec



Disclaimer

Any views or opinions presented in this presentation are solely those of the author and do not necessarily represent those of Biogen Idec.

Topics

- Modes of Validation
- How to run from GUI
- How to automate execution from CLI or SAS
- How to analyze and interpret OpenCDISC reports

Modes of Validation



Running OpenCDISC Validator GUI

- Start in the main Validator directory where you unzipped the files
- Locate client.bat and click on it to start the program



Running OpenCDISC Validator GUI

OpenCDISC Valid	Jator		Standa
<u>F</u> ile <u>H</u> elp			-
What would you	u like to do?		-
Standard:	SDTM		Source
Source Format:	SAS® Transport (XPORT)		-
Source Data:	C:\Work\QC\Study1\SDTM\xpdata\AE.xpt	Browse	-
	C:\Work\QC\Study1\SDTM\xpdata\DM.xpt	Remove	Config
		Clear	_
			-
			-
	You can select multiple files or folders as sources		-
Configuration:	config-sdtm-3.1.2.xml		-
Define.xml:		Browse	Report
	Optional		-
Report Format:	Excel Report Settings CDISC CT: 2014-03-28 MedDRA: 15.1]	CDISC
Progress:]	-
	Waiting to begin.		-
	Start	Stop	-
			-

rd

- SDTM, ADaM,
- Define.xml, Send
- Custom (Tabular)

Format

- SAS Transport (XPORT)
- Delimited
- Dataset-XML (R 1.5)

uration

- config-adam-1.0.xml,
- config-define-1.0.xml,
- config-define-2.0.xml (R 1.5)
- config-sdtm-3.1.1.xml,
- config-sdtm-3.1.2.xml
- config-sdtm-3.1.3.xml
- config-sdtm-3.2.xml
- config-send-3.0.xml

Format

Excel and CSV

СТ

- 2013-04-12
- 2013-06-28
- 2013-10-04
- 2013-12-20
- 2014-03-28

MedDRA

MedDRA versions installed in the -MedDRA config folder

Running OpenCDISC Validator CLI

C:\opencdisc-validator_1	_5\lib>java -jar validator-	-cli-1.5.jar -help
OyenC	DISC Validator Command Line	e Client
The following parameter certain parameters may For additional help, or	es may be passed to the Vali be required. • to submit suggestions, ple	idator. Note that ease visit our community
at http://www.opencdisc	.org/	
General Parameters —task —type	Validate¦Generate (Validat SDTM¦ADaM¦SEND¦Define¦Cust	ce) com (SDTM)
Source Data Parameters -source -source:type -source:delimiter -source:qualifier	<path> SAS:Delimited (SAS) <delimiter> (,) <qualifier> (")</qualifier></delimiter></path>	
Configuration Parameter -config -config:define -config:codelists -config:cdisc -config:meddra Report Parameters -report -report:type -report:cutoff -report:overwrite	<pre> s</pre>	Tasks specific to generating Define.xml are highlighted in yellow. They cannot be combined with validation tasks.
Generation Parameters -output -output:overwrite C:\opencdisc-validator_1	<path> yesino _5\lib></path>	

Running OpenCDISC Validator SAS

x java -jar "C:\opencdisc-validator_1_5\lib\validator-cli-1.5.jar" -task="validate"

-type="sdtm"

-source="C:\Work\QC\study1\SDTM\xpdata*.xpt"

-source:type="sas"

-config="C:\opencdisc-validator_1_5\config\config-sdtm-3.1.2.xml"

-config:define="C:\Work\QC\study1\SDTM\xpdata\define.xml"

```
-config:cdisc="2014-03-28"
```

-config:meddra="15.1"

-report="C:\Work\submission\openCDISC_v1.5_data_&_Define.xls"
-report:type="excel"

-report:cutoff="1000"

-report:overwrite="yes";

OpenCDISC Validator Report (1)

- Validator Reports are based on predefined checks
- Different versions have a different set of checks/rules
- Four types of validation checks
 - CT Controlled Terminology
 - SD Study Data
 - DD Data Definition
 - OD Operation Data Model
- Each rule has an unique Rule ID, Message, Description, Category and Severity
- The Report contains four tabs or worksheets

3981						
3982						
3983						
4 F F	🕨 🛛 Data	set Summary	🔬 Issue Su	ummary 🚶	Details Rules	<u></u>
Ready	Filter Mo	de 🛅				

OpenCDISC Validator Report (2) Dataset Summary

 The Dataset summary tab provides information about each dataset in a quick, snapshot view

		Open	CDISC Vali	dator Re	port		
Configuration Define.xml: C	: C:\opencdisc-validator_1_5\config\c :\work\updated_pilot_submission_pa	config-sdtm-3.1.2.xr ackage\Updated Pil	nl ot Submission P	ackage\9001	72\m5\datas	ets\cdiscpilo	t01\tabulations\sdtm\define.xml
Generated: 2	014-04-28T17:37:48	· ·					
Engine Versi	on: 1.5						
MedDRA Ver	sion: 15.1						
CDISC Contr	olled Terminology Version: 2014-03-2	28					
			Processed S	ources			
Domain	Label	Class	Source	Records	Errors	Warnings	Notices
GLOBAL	Global Metadata				0	0	0
AE	Adverse Events	Events	ae.xpt	1191	14	233	5
CM	Concomitant Medications	Interventions	cm.xpt	7510	81	3	212
DM	Demographics	Special Purpose	dm.xpt	306	9	1	5
DS	Disposition	Events	ds.xpt	596	i 1	60	8
EX	Exposure	Interventions	ex.xpt	591	0	2	3
LB	Laboratory Tests Results	Findings	lb.xpt	59580	1790	18287	19361
MH	Medical History	Events	mh.xpt	1818	3	10	8
QS	Questionnaires	Findings	qs.xpt	121749	0	9	0
RELREC	Related Records	Relationship	relrec.xpt	234	237	234	0
SC	Subject Characteristics	Findings	sc.xpt	254	254	0	254
SE	Subject Elements	Special Purpose	se.xpt	752	3	2	1
SUPPAE	Supplemental Qualifiers for AE	Relationship	suppae.xpt	1191	4	1	0
SUPPDM	Supplemental Qualifiers for DM	Relationship	suppdm.xpt	1197	4	1	0
SUPPDS	Supplemental Qualifiers for DS	Relationship	suppds.xpt	3	4	1	0
SUPPLB	Supplemental Qualifiers for LB	Relationship	supplb.xpt	64403	64407	64404	0
SV	Subject Visits	Special Purpose	sv.xpt	3559	0	3	0
TA	Trial Arms	Trial Design	ta.xpt	8	5	0	0
TE	Trial Elements	Trial Design	te.xpt	7	5	0	0
TI	Trial Inclusion/ Exclusion Criteria	Trial Design	ti.xpt	31	1	1	1
TS	Trial Summary	Trial Design	ts.xpt	33	9	14	8
TV	Trial Visits	Trial Design	tv.xpt	21	3	2	1
VS	Vital Signs	Findings	vs.xpt	29643	3	1	0
Total				294677	66837	83269	19867
- Ottai				201011		00200	10001
			Unprocessed	Sources			
Domain	Label	Class	Reason		Errors	Warnings	Notices
Total					0	0	0
1 ottai					0	0	0
Grand Total				294677	66837	83269	19867

OpenCDISC Validator Report (3) Issue Summary

 The Issues Summary tab provides a break down of the type of rules that have issues and how many have been reported.

		OpenCDISC Validator Report		
Configurat Define.xm Generated Engine Ve	ion: C:\opencdisc l: C:\work\updated l: 2014-04-28T17:: ersion: 1.5	-validator_1_5\config\config\sdtm-3.1.2.xml d_pilot_submission_package\Updated Pilot Submission Package\900172\m5\datasets\cdiscpilot01\tabulatio 37:48	ns\sdtm\define.xn	nl
MedDRA	Version: 15.1			
CDISC Co	ntrolled Terminolo	gy Version: 2014-03-28		
		Issue Summany		
Source	Rule ID	Message	Severity	Found
AE	rtaio ib	moorage	oorong	. ound
	SD0058	Variable appears in dataset, but is not in SDTM standard	Error	10
	SD1082	AEACN variable length is too long for actual data	Error	1
	SD1082	AEDECOD variable length is too long for actual data	Error	1
	SD1082	AEOUT variable length is too long for actual data	Error	1
	SD1082	AETERM variable length is too long for actual data	Error	1
	SD1077	FDA Expected variable not found	Warning	1
	SD1083	Missing AEDY variable, when AEDTC variable is present	Warning	1
	SD1090	Incorrect value for AESTDY variable	Warning	1
	SD1201	Duplicate records	Warning	230
	SD1076	Model permissible variable added into standard domain	Notice	1
		C:\opencdisc-validator_1_5\config/data/MedDRA/8.0/pt.asc is missing or lacks necessary variables and		
	SKIP_SD0008	cannot be used for this cross-dataset validation	Notice	1
	SKIP_SD0008	C:\opencdisc-validator_1_5\config/data/MedDRA/8.0/pt.asc is missing or lacks necessary variables and		
	С	cannot be used for this cross-dataset validation	Notice	1
		C:topencdisc-validator_1_5/contg/data/MedDRAV8.0/soc.asc is missing or lacks necessary variables and	Notico	4
	SKIP_SD1114	Compandiscivalidator 1 Sponfindata/MedDRA/8 0/soc asc is missing or lacks necessary variables and	Notice	
	C	cannot be used for this cross-dataset validation	Notice	1
СМ				
	SD0035	Missing value for CMDOSU, when CMDOSE, CMDOSTXT or CMDOSTOT is provided	Error	80
	SD1082	CMROUTE variable length is too long for actual data	Error	1
	SD1077	FDA Expected variable not found	Warning	1
	SD1081	CMDOSU variable length is too long for actual data	Warning	1
	SD1083	Missing CMDY variable, when CMDTC variable is present	Warning	1
	CT0018	Value for CMDOSFRQ not found in (FREQ) CT codelist	Notice	192
	CT0049	Value for CMDOSU not found in (UNIT) CT codelist	Notice	16
	SD1076	Model permissible variable added into standard domain	Notice	4

OpenCDISC Validator Report (4) Details

 The Details tab provides detailed information about errors and warnings

Domain	Record	Count	Variables	Values	Rule ID	Message	Category	Severity
GLOBAL			DOMAIN	SE	SD1111	Missing SE dataset	Presence	Warning
GLOBAL			DOMAIN	TA	SD1112	Missing TA dataset	Presence	Warning
GLOBAL			DOMAIN	TE	SD1113	Missing TE dataset	Presence	Warning
AE			VARIABLE, DATASET	EPOCH, AE	SD1077	FDA Expected variable not found	Metadata	Warning
AE		15	5 AEOUT	NOT RESOLVED	CT0027	Value for AEOUT not found in (OUT) CT codelist	Terminology	Error
AE			Variable, Excess	AESEV, 12	SD1081	AESEV variable length is too long for actual data	Metadata	Warning
AE			Variable, Excess	AEREFID, 8	SD1081	AEREFID variable length is too long for actual data	Metadata	Warning
AE			Variable, Excess	AEENDTC, 10	SD1081	AEENDTC variable length is too long for actual data	Metadata	Warning
AE			Variable, Excess	AEACN, 9	SD1081	AEACN variable length is too long for actual data	Metadata	Warning
AE			Variable, Excess	AESPID, 18	SD1081	AESPID variable length is too long for actual data	Metadata	Warning
AE			Variable, Excess	AESTDTC, 10	SD1081	AESTDTC variable length is too long for actual data	Metadata	Warning
AE			Variable, Excess	DOMAIN, 6	SD1081	DOMAIN variable length is too long for actual data	Metadata	Warning
AE			Variable, Excess	AEENTPT, 31	SD1082	AEENTPT variable length is too long for actual data	Metadata	Error
AE			Variable, Excess	AETERM, 46	SD1082	AETERM variable length is too long for actual data	Metadata	Error
AE			Variable, Excess	AEOUT, 38	SD1082	AEOUT variable length is too long for actual data	Metadata	Error
AE			Variable, Excess	AEBODSYS, 148	SD1082	AEBODSYS variable length is too long for actual data	Metadata	Error
AE			Variable, Excess	AEDECOD, 161	SD1082	AEDECOD variable length is too long for actual data	Metadata	Error
AE			Variable, Excess	AEENRTPT, 33	SD1082	AEENRTPT variable length is too long for actual data	Metadata	Error
AE	8		AESER	Y	SD0009	No qualifiers set to 'Y', when AE is Serious	Consistency	Warning
AE	18		AESER	Y	SD0009	No qualifiers set to 'Y', when AE is Serious	Consistency	Warning
AE	30		AESER	Y	SD0009	No qualifiers set to 'Y', when AE is Serious	Consistency	Warning
AE	41		AESER	Y	SD0009	No qualifiers set to 'Y', when AE is Serious	Consistency	Warning
AE	68		AESER	Y	SD0009	No qualifiers set to 'Y', when AE is Serious	Consistency	Warning
AE	69		AESER	Y	SD0009	No qualifiers set to 'Y', when AE is Serious	Consistency	Warning
				ACCIDENTAL FALL				
				RESULTING IN				
				SUPERFICIAL LAYERS				
				OF HEAD INJURY				
AE	84		AETERM		SD1021	Unexpected character value in AETERM variable	Format	Warning
				ACCIDENTAL FALL				
				RESULTING IN				
				SUPERFICIAL LAYERS				
				OF HEAD INJURY				

OpenCDISC Validator Report (5) Rules

- Contains a list of Rule IDs and their descriptions
- The links to each rule ID in the Issues Summary and Details worksheets take you to this worksheet

Rule ID	Message	Description	Category	Severity
		Action Taken with Study Treatment (ACN) variable values must be populated with terms found		
		in 'Action Taken with Study Treatment' (C66767) CDISC controlled terminology codelist. New		
CT0001	Value forACN not found in (ACN) CT codelist	values cannot be added into the CDISC CT non-extensible codelist.	Terminology	Error
		Severity/Intensity (AESEV) variable values must be populated with terms found in		
	Value for AESEV not found in (AESEV) CT	'Severity/Intensity Scale for Adverse Events' (C66769) CDISC controlled terminology codelist.		
CT0002	codelist	New values cannot be added into the CDISC CT non-extensible codelist.	Terminology	Error
		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		
	Value for TSVAL not found in (AGESPAN) CT	(TSPARMCD) value is 'AGESPAN'. New terms can be added as long as they are not duplicates,		
CT0003	codelist	synonyms or subsets of existing standard terms.	Terminology	Information
		Age Units (AGEU) variable values must be populated with terms found in 'Age Unit' (C66781)		
		CDISC controlled terminology codelist. New values cannot be added into the CDISC CT non-		
CT0004	Value for AGEU not found in (AGEU) CT codelist	extensible codelist.	Terminology	Error
		TS dataset Parameter Value (TSVAL) values must be populated with terms found in 'Age Unit'		
		(C66781) CDISC controlled terminology codelist, when Trial Summary Parameter Short Name		
		(TSPARMCD) value is 'AGEU'. New values cannot be added into the CDISC CT non-extensible		_
CT0005	Value for TSVAL not found in (AGEU) CT codelist	codelist.	Terminology	Error
		Country (COUNTRY) variable values must be populated with terms found in 'Country' (C66786)		
	Value for COUNTRY not found in (COUNTRY) CT	CDISC controlled terminology codelist. New values cannot be added into the CDISC CT non-		_
CT0006	codelist	extensible codelist.	Terminology	Error

SD

CT

Α	В	C	D	E
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 for variable	Value of Dates/Time variables (*DTC) 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 value forSEQ variable	The value of Sequence Number (SEQ) variable must be unique for each record within a domain and within of Unique Subject Identifier (USUBJID) or Pool Identifier (POOLID) variables value when they are present in the domain.	Consistency	Error
SD0006	No baseline result in 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
		Standard Units (STRESU) must be consistent for all records with same Short Name of		

How to Interpret the Errors/Warnings (1)

Issue Summary Tab

	Issue Summary										
Source	Rule ID	Message	Severity	Found							
AE											
	CT0027	Value for AEOUT not found in (OUT) CT codelist	Error	274							
	SD1082	AEBODSYS variable length is too long for actual data	Error	1							
	SD1082	AEDECOD variable length is too long for actual data	Error	1							
	SD1082	AEENRTPT variable length is too long for actual data	Error	1							
	SD1082	AEENTPT variable length is too long for actual data	Error	1							
	SD1082	AEOUT variable length is too long for actual data	Error	1							

Details Tab

Domai 🝷	Recor	Count	Variables	-	Values	Rule ID 🗾	Message	*	Category 🔽	Severit 🗸
AE		15	AEOUT	NC	OT RESOLVED	CT0027	Value for AEOUT not found in (OUT) CT codelist		Terminology	Error
AE		259	AEOUT	ST	OPPED	CT0027	Value for AEOUT not found in (OUT) CT codelist		Terminology	Error

Rules Tab

		Outcome (OUT) variables values must be populated with terms found in 'Outcome of Event'			
		(C66768) CDISC controlled terminology codelist. New values cannot be added into the CDISC			
CT0027	Value forOUT not found in (OUT) CT codelist	CT non-extensible codelist.	Terminology	Error	

Terminology

C66768	OUT	Outcome of Event	A condition or event that is attributed to the adverse event and is the result or conclusion of the adverse event. (NCI)	No	
--------	-----	------------------	--	----	--

The allowed values for AEOUT are:

- FATAL, NOT RECOVERED/NOT RESOLVED, RECOVERED/RESOLVED, RECOVERED/ RESOLVED WITH SEQUELAE, RECOVERING/RESOLVING and UNKNOWN

How to Interpret the Errors/Warnings (2)

Issue Summary Tab

Source	Rule ID	Message	Severity	Found
AE				
	CT0027	Value for AEOUT not found in (OUT) CT codelist	Error	274
	SD1082	AEBODSYS variable length is too long for actual data	Error	1
	SD1082	AEDECOD variable length is too long for actual data	Error	1
	SD1082	AEENRTPT variable length is too long for actual data	Error	1
	SD1082	AEENTPT variable length is too long for actual data	Error	1
	SD1082	AEOUT variable length is too long for actual data	Error	1
	SD1082	AETERM variable length is too long for actual data	Error	1
	SD0009	No qualifiers set to 'Y', when AE is Serious	Warning	10

Details Tab

Domai 💌	Recor	Count	Var	riables	Values	🖌 🛛 Rule ID 🔽	Message 🚽	Category 💌	Severit 💌
AE			Variable, E	xcess	AEBODSYS, 148	SD1082	AEBODSYS variable length is too long for actual data	Metadata	Error
AE			Variable, E	xcess	AETERM, 46	SD1082	AETERM variable length is too long for actual data	Metadata	Error
					8HA02PED-587-603,				
AE	170		USUBJID, A	AESTDTC	2012-12-17	SD0080	AE start date is after the latest Disposition date	Consistency	Warning

AE Dataset

🛐 SAS Univer	al Viewer - [a	ae.xpt]									
											_ & ×
File Tools	Window	Help									
Address											-
Library Prope	ties										
Name			Obs	Vars	#	Variable	Туре	Length	Format	Informat	Label ^
AE			274	21	1	STUDYID	Character	8			Study Identifier
					2	DOMAIN	Character	8			Domain Abbrevi
					3	USUBJID	Character	20			Unique Subject
					4	AESEQ	Numeric	8			Sequence Num
					5	AEREFID	Character	20			Reference ID
					6	AESPID	Character	20			Sponsor-Define
					7	AETERM	Character	200			Reported Term f
					8	AEDECOD	Character	200			Dictionary-Deriv

How to Interpret the Errors/Warnings (3)

Issue Summary Tab

CE				
	<u>SD0037</u>	Value for CECAT not found in (CECAT) user-defined codelist	Error	9
	SD1082	CECAT variable length is too long for actual data	Error	1
	<u>SD1082</u>	CESCAT variable length is too long for actual data	Error	1
	SD1082	CETERM variable length is too long for actual data	Error	1
	<u>SD1077</u>	FDA Expected variable not found	Warning	1
			1.0.0	

Details Tab

A	В	С	D	E	F	G	H	
Domai 🕶	Recor	Count 🗸	Variables 🔽	Values 🔽	Rule ID 🔽	Message 🔽	Category 🔽	Severit 🔽
CE			VARIABLE, DATASET	EPOCH, CE	SD1077	FDA Expected variable not found	Metadata	Warning
CE			VARIABLE	CESTDY	SD1087	Missing CESTDY variable, when CESTDTC variable is present	Presence	Warning
CE		5	CECAT	INTERNAL	SD0037	Value for CECAT not found in (CECAT) user-defined codelist	Terminology	Error
CE		4	CECAT	JOINT	SD0037	Value for CECAT not found in (CECAT) user-defined codelist	Terminology	Error
CE			Variable, Excess	DOMAIN, 6	SD1081	DOMAIN variable length is too long for actual data	Metadata	Warning
CE			Variable, Excess	CECAT, 89	SD1082	CECAT variable length is too long for actual data	Metadata	Error

Dataset Values

	CECAT
I	NTERNAL
	JOINT
ľ	NUSCLE
Ş	SKIN/MUCOSA

Define Values

	Code Value
INTERNAL_TEST	
JOINT_TEST	
MUSCLE	
SKIN/MUCOSA	

How to Interpret the Errors/Warnings (4)

Issue Summary Tab

Source	Rule ID	Message	Severity	Found
DEFINE				
	DD0024	Invalid Codelist for 'AECONTRT' variable	Error	1
	DD0024	Invalid Codelist for 'AESER' variable	Error	1
	DD0024	Invalid Codelist for 'AESEV' variable	Error	1
	DD0024	Invalid Codelist for 'XRBLFL' variable	Error	1
	OD0048	Referenced 'CodeList' not found	Error	2

Details Tab

xpath 🚽	Variables 🔽	Values 🔽	Rule ID 🚽	Message 🔽	Category 🗸	Severit 🗸
//ItemDef[@OID='DM.SEX']/CodeListRef[@CodeListOID='SEX']	@CodeListOID	SEX	OD0048	Referenced 'CodeList' not found	Cross-reference	Error
//ItemDef[@OID='IE.IETESTCD']/CodeListRef[@CodeListOID='IET ESTCD']	@CodeListOID	IETESTCD	OD0048	Referenced 'CodeList' not found	Cross-reference	Error
	_					

Define

SEX	Sex	text	SEX	CRF Page <u>8</u>	<u>Record</u> <u>Qualifier</u>	

Thank you!

Majdoub Haloui Biogen Idec <u>Majdoub.haloui@biogenidec.com</u>