

Achieving an error free OpenCDISC report – realistic or idealistic?



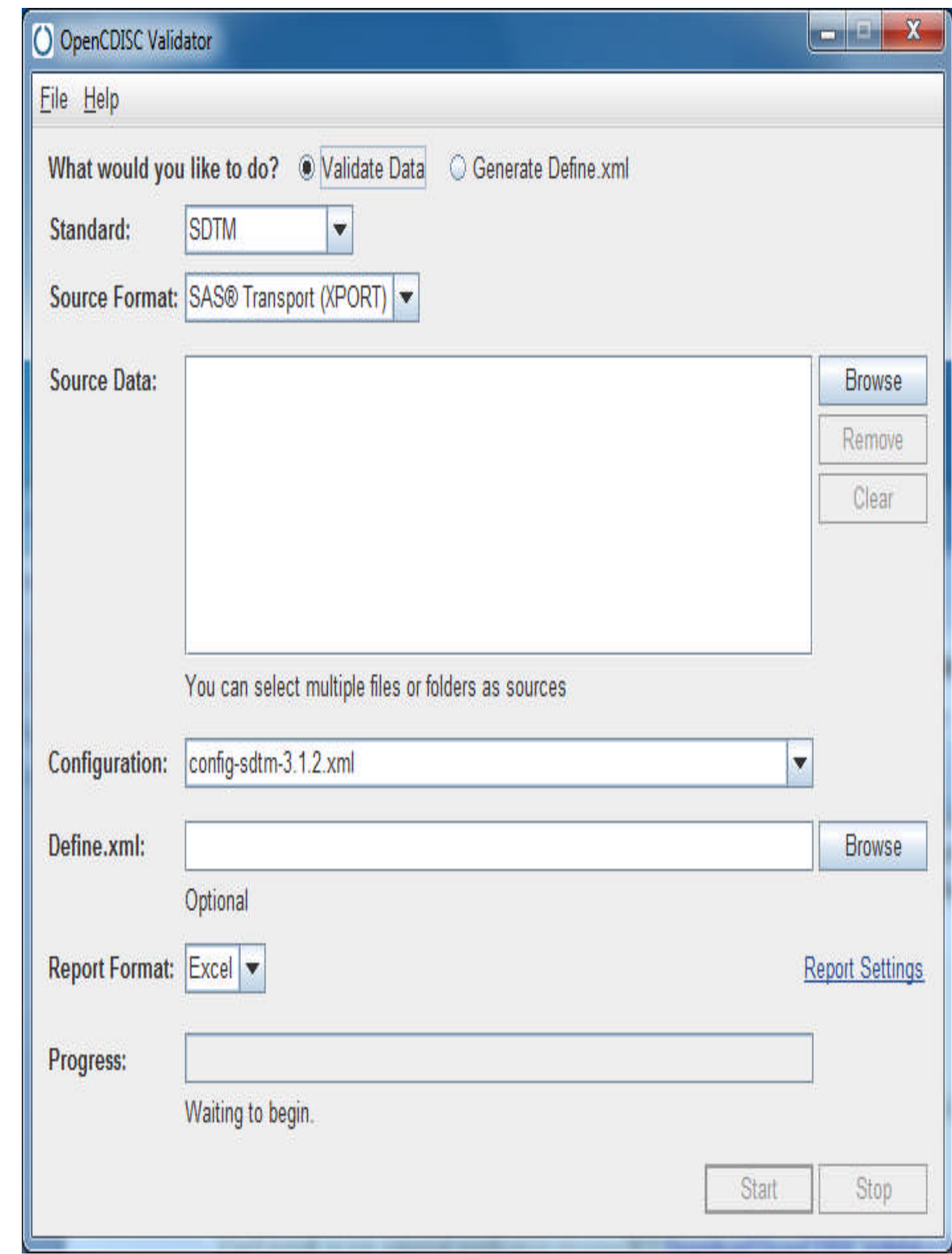
Ratnakar Gundlapalli: Ratnakar.Gundlapalli@iconplc.com
Jennie McGuirk: Jennie.McGuirk@iconplc.com

Clinical Research

Introduction

OpenCDISC is an open source community focused on building extensible frameworks and tools for the implementation and advancement of CDISC Standards. One of the useful tools offered is the OpenCDISC Validator. This poster will focus on the SDTM and define.xml portion of the validator. We will discuss the many advantages to using the OpenCDISC Validator tool, and also provide some real study examples of when an ‘error free’ OpenCDISC Validator report has not been achievable.

OpenCDISC Validator Tool



Advantages of OpenCDISC Validator :

- Supports SDTM, ADAM and SEND standards
- Supports generating/validating define.xml
- Validates the Controlled Terminology and also checks the terminology against define.xml
- Includes WebSDM and Janus checks
- Can be used by anyone who has basic understanding of CDISC standards, with minimal training
- Supports SDTM+ and custom datasets
- Generates validation report in Excel, CSV and HTML formats
- Tool is of commercial quality, freely available and easy to use
- Can be deployed on a server or hosted environment as an enterprise wide solution

Unavoidable error/warnings :

This poster includes real study examples describing instances where an error free OpenCDISC Validator report could not be produced due to nature of the data for a study and/or due to enhanced versions of the SDTM standards. In addition, examples of warnings that could be classified as lower risk are presented, with the recommendation that it is unnecessary to add explanations for the warnings to the reviewer’s guide.

Examples

Rule ID	Message	Description	Category	Type	Example/Explanation
SD0072	Referenced Domain not found	When comments are related to a domain, then the value of Related Domain Abbreviation (RDOMAIN) variable must reference the Domain Code of the parent domain.	Consistency	Error	Occurs for CO when using split domains e.g. QSMM, QSCG. The referenced domain (RDOMAIN = QS) does not exist
SD0074	Referenced Domain not found	Identifies Supplemental Qualifiers domain reference to a key variable that isn't defined in the target domain	Consistency	Error	Occurs for SUPP-- when using split domains e.g. QSMM, QSCG. The referenced domain (RDOMAIN = QS) does not exist
SD0035	Missing units on value	Identifies records that violate the condition [Dose Units is not null], limited to records where [Dose is not null]	Consistency	Error	This error often appears for the Conmeds domain (CM). Cleaning criteria on Conmed questions tend not to be as strict as other domains such as the Study Treatment Exposure (EX). For Conmeds this could be a warning only
SD0013	Begin day must be less than or equal to end day	Begin day must be less than or equal to end day	Limit	Error	This error often appears for the Conmeds domain (CM), where a conmed has a partial start date (CMSTDTC) and a complete end date (CMENDTC).
SD0055	SDTM/ dataset variable type mismatch	Variable Data Types in the dataset should match the variable data types described in SDTM.	Metadata	Warning	This is a valid warning, however the check should be extended to account for permitted extensions of domains. For example, within the Comments domain (CO) or Trial Summary domain (TS), when length \$200 restrictions apply, we often need to extend these domains to include variables such as COVAL1 and TSVAL1
SD0080	AE start date must be less than or equal to latest disposition date	Start Date/Time of Adverse Event (AESTDTC) must be less than or equal to the Start Date/Time of the latest Disposition Event (DSSTDTC)	Consistency	Warning	A large number of protocol will include a follow-up period for adverse events, post the last disposition event in the trial schedule
SD0087	RFSTDTC cannot be null for randomized subject	Subject Reference Start Date/Time (RFSTDTC) is required for all randomized subjects, those where Planned Arm Code (ARMCD) is not equal to 'SCRNFAIL' or 'NOTASSGN'.	Presence	Warning	This warning can appear if a subject is randomized in error, or if a subject has been randomized but never receives any study treatment
SD0006	No baseline result in dataset	All subjects should have at least one baseline observation (--BLFL = 'Y') in each Findings domain, except for subjects who failed screening (ARMCD = 'SCRNFAIL') or were not fully assigned to an Arm (ARMCD = 'NOTASSGN').	Presence	Warning	This warning can appear if a subject is randomized in error, or if a subject has been randomized but never receives any study treatment
SD0088	RFENDTC cannot be null for randomized subject	Subject Reference End Date/Time (RFENDTC) is required for all randomized subjects, those where Planned Arm Code (ARMCD) is not equal to 'SCRNFAIL' or 'NOTASSGN'.	Presence	Warning	This warning can appear if a subject is randomized in error, or if a subject has been randomized but never receives any study treatment.
SD0026	Missing units on value	Identifies records that violate the condition [Original Units is not null], limited to records where [Result or Finding in Original Units is not null]	Consistency	Warning	This is a valid warning, however we often have Lab assessments where units are either not available, or not applicable. Typical examples, include pH.

Examples (Continued)

Rule ID	Message	Description	Category	Type	Example/Explanation
SD0058	Variable appears in dataset but is not in SDTM standard	Only variables listed in SDTM should appear in a dataset. New sponsor defined variables must not be added, and existing variables must not be renamed or modified.	Metadata	Warning	This is a valid check, however we may want to include additional variables from the class definition. An example is the Medical History domain (MH). The Events Class includes the --REL. We may want to include the MHREL in the MH domain. The Implementation Guide does not restrict this 'the following Qualifiers would generally not be used in MH, --REL', so this warning should not appear
SD0009	AE is Serious but no qualifiers set to 'Y'	Identifies records where Serious Event='Y' but none of Involves Cancer, Congenital Anomaly or Birth Defect, Persist or Signif Disability/Incapacity, Results in Death, Requires or Prolongs Hospitalization, Is Life Threatening, Other Medically Important Serious Event, or Occurred with Overdose equals 'Y'	Consistency	Warning	This warning is useful, but often we find these reasons are not captured on the CRF. Permissible variables should not cause warnings to execute. Otherwise these should be considered expected

Enhancements to OpenCDISC Validator

- Downgrade extensible code list messages to information only
- Add a check if the order of variables is not per the SDTM IG
- Add a check if Permissible variables are null
- Add a check for QNAM values that are reserved parent variable names (e.g. CMMODIFY used for SUPPCM)
- Add a check for when RELTYPE is not null and IDVARVAL is populated
- Add a check to identify datasets with 0 observations
- Identify gap between elements in SE
- Publish list of valid ISO 8601
- Support future versions and extensions of SDTM (e.g. SDTM IG 3.1.2 amendment)
- Timely support of latest controlled terminology

Conclusion :

Although achieving an error free OpenCDISC Validator report seems to be unavoidable, the tools flexibility allows it to be modified accordingly. A unified understanding of how to validate SDTM, as well as the other CDISC content models, is needed to achieve a higher degree of verifiable consistency in SDTM-compliant data and submissions and reduce unnecessary information within the reviewers guide.