Comparison of FDA and OpenCDISC SEND Validator v1.4.1 Rules

PP05

Laura Kaufman, Angela Popescu, Frederic Mura, Michael Wasko,
Maro Schuster, Rich Buchanan, Amelia Bedeoan

PDS Preclinical Data Systems Inc.; Basel Switzerland and Mt Arlington, New Jersey USA

Summary

A rule-by-rule comparison was made between FDA and OpenCDISC v1.4.1 SEND validation rules to determine which were present in the FDA set but not in OpenCDISC, which in OpenCSDISC, and which in both. This analysis was a first step in a process to extend the OpenCDISC validator to also include FDA validation rules.

There was considerable overlap for the two rule sets, especially for controlled terminology, as would be expected. FDA rules are more specific regarding Trial Sets.

Methods

OpenCDISC v1.4.1 SEND validation rules were obtained from source code. FDA SEND validation rules (source code not available) were obtained from:
http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm

Each OpenCDISC rule was mapped to a corresponding FDA rule, and rules in both rule sets that could not be mapped were identified. In instances when such mapping was unclear, specific test cases were run through the OpenCDISC validator.

Results

There were approximately 300 rules in each rule set.

As shown in the Table and Figure to the right, most of the FDA rules could be mapped to OpenCDISC v1.4.1 SEND validation rules. Our analysis identified 13 FDA rules for which OpenCDISC v1.4.1 equivalents were not evident. It is notable that many of these 13 rules pertained to Trial Sets.

Many of the FDA and OpenCDISC shared rules were identical and pertained to CDISC controlled terminology, as would be expected. For some of the other rules, there was not always a one-to-one correlation between FDA and OpenCDISC rules, and relationships between portions of the 2 rule sets could be described as "one to many" for FDA as compared with OpenCDISC and vice versa.

OpenCDISC validator v1.4.1 appeared to include some rules that could not be mapped to the FDA validation set. These included a series of rules on a domain-by-domain basis for consistently populating --TEST and --TESTCD values according to their one-to-one relationships defined in CDISC CT. OpenCDISC also includes rules pertaining to dataset size (>1 GB) and split datasets.

Case Study

The FDA issued SEND validation rules in ~September 2013. At that time, we were preparing several SEND datasets (SENDIG v3.0/ Define 1.0) for a client. After submission of these datasets to FDA, which were free of errors according to the OpenCDISC validator v1.4.1, we were informed by FDA of the presence of errors according to their validator that prevented the SEND datasets from being loaded into the Agency's software.

Errors identified by FDA but not OpenCDISC v1.4.1:

- Missing TCNTRL in TXPARMCD
- SEENDTC was not in ISO 8601 format (this is a known bug in OpenCDISC 1.4.1 and will be fixed in the next release according to OpenCDISC validator forum)

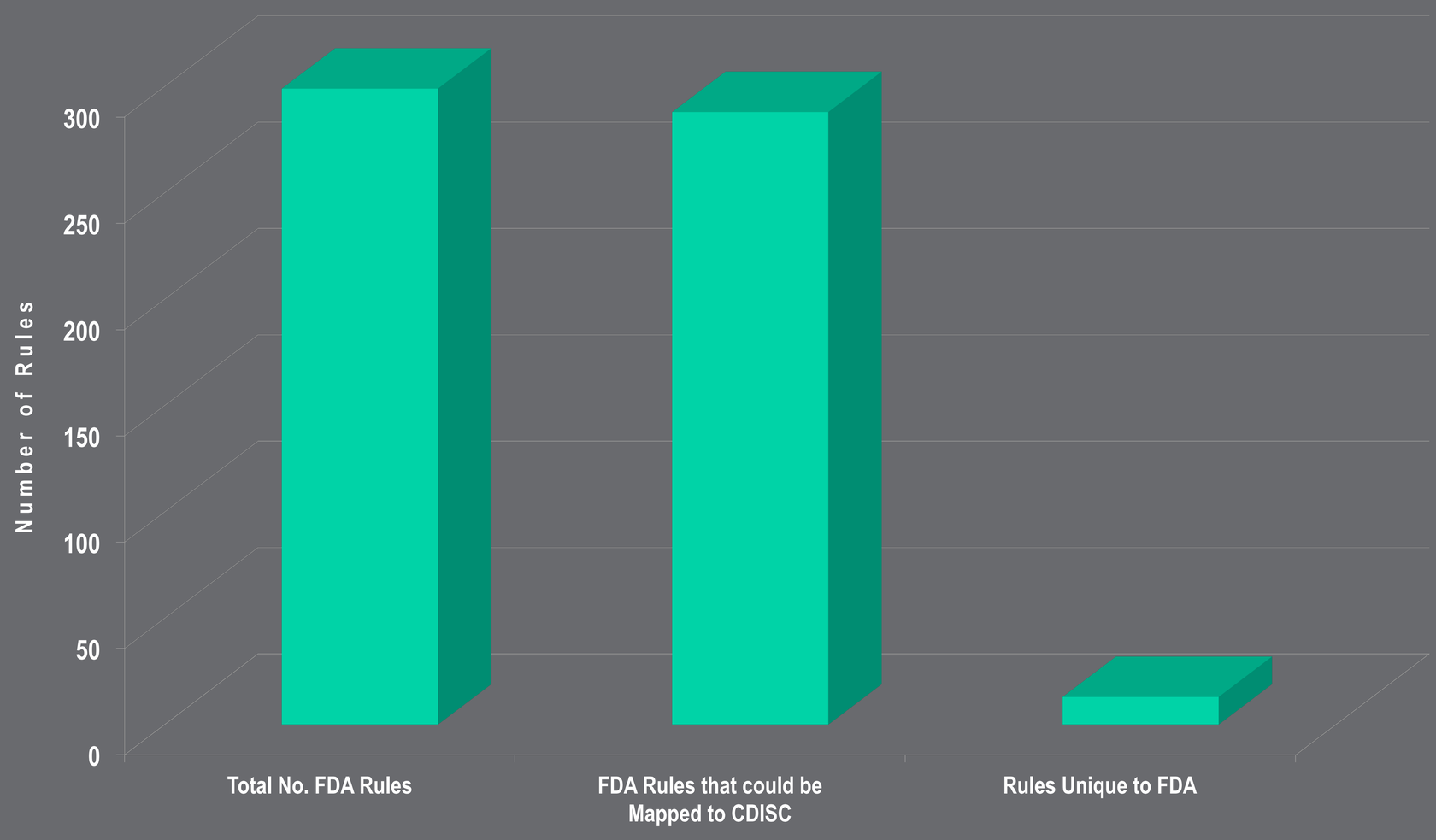
We corrected these errors and resubmitted the files, which were successfully loaded by the FDA's software without errors.

From this experience, we concluded that there are important differences between the OpenCDISC v1.4.1 and FDA SEND validation rule sets that resulted in the initial inability of FDA to load SEND datasets, despite the absence of errors according to OpenCDISC. We therefore decided to undertake a gap analysis to determine which validation rules are present in the FDA set but not in OpenCDISC. This is a first step towards extending the OpenCDISC validator to also include FDA SEND validation rules.

Summary of Rules in FDA SEND Validation SET but not OpenCDISC v1.4.1

FDA Rule Number	Message	Domains	Severity
NONCLIN-0059	Missing value or multiple values with the text "control" in the TCNTRL.	TX	Warning
NONCLIN-0253	Missing value for ARMCD. The TXPARMCD listed as "Yes" in the "Should Include" column provide a relationship between Trial Sets and other information (the trial protocol or the Trial Arms dataset) and should be included in a well-formed TX domain submission.	TX	Warning
NONCLIN-0254	Missing value for TRTDOS. The TXPARMCD listed as "Yes" in the "Should Include" column provide a relationship between Trial Sets and other information (the trial protocol or the Trial Arms dataset) and should be included in a well-formed TX domain submission.	TX	Warning
NONCLIN-0255	Missing value for TRTDOSU. The TXPARMCD listed as "Yes" in the "Should Include" column provide a relationship between Trial Sets and other information (the trial protocol or the Trial Arms dataset) and should be included in a well-formed TX domain submission.	TX	Warning
NONCLIN-0256	Missing value for GRPLBL. The TXPARMCD listed as "Yes" in the "Should Include" column provide a relationship between Trial Sets and other information (the trial protocol or the Trial Arms dataset) and should be included in a well-formed TX domain submission.	TX	Warning
NONCLIN-0257	Missing value for SPGRPCD. The TXPARMCD listed as "Yes" in the "Should Include" column provide a relationship between Trial Sets and other information (the trial protocol or the Trial Arms dataset) and should be included in a well-formed TX domain submission.	TX	Warning
NONCLIN-0022	Invalid SETCD – (SETCD). Subjects should only be assigned to trial sets listed in the Trial Sets (TX) dataset.	DM	Error
NONCLIN-0024	Invalid POOLDEF – (POOLID). All pool identifiers should be present in POOLDEF Dataset.	EX, CL, FW, LB, PC, PP, RELREC, SUPP--	Error
NONCLIN-0046	Invalid --CNTRY value. Value not found in (COUNTRY) code list.	TS	Warning
NONCLIN-0053	Missing values for DSSTDY and --DY and --DTC & RFSDTC. Either of these three timing variables should be populated: [DSSTDY] or [--DY] or [--DTC and DM, RFSTDTC].	OM, MA, MI, TF, DM	Warning
NONCLIN-0057	Missing values for --RESMOND. The SUPPQUAL domain records should have QNAM variable populated with --RESMOND.	MA, MI, TF	Warning
NONCLIN-0060	Invalid value for LBSTRESN. Value must be >0.	LB	Warning
NONCLIN-0177	Invalid format for IDVAR. The variable value for IDVAR should not be more than 8 characters in length.	CO, RELREC, SUPPQUAL	Warning

Description of FDA SEND Validation Rules and Relationship to OpenCDISC 1.4.1



Conclusions and Recommendations

- We identified 13 rules in the FDA SEND validation rule set that do not appear to have equivalents in the OpenCDISC SEND validator, version 1.4.1.
- Many of these 13 rules pertain to Trial Sets.
- In our experience, SEND datasets with 0 errors according to OpenCDISC validator v1.4.1 may not meet FDA SEND validation rules with respect to Trial Sets.
- CDISC has announced a new release of their validator for the end of February, 2014. We plan to reach out to them and compare our analyses.
- It is important to appreciate that both OpenCDISC and FDA SEND validation rules will continue to evolve as the standards and guidances evolve and as the community gains additional SEND experience.