PharmaSUG 2015 - Paper 2015-37

Tips for efficient CDISC eCRT production

Lanting Li, PPD, Shanghai, China

Yu Zhu, PPD, Shanghai, China

Huan Zhu, PPD, Shanghai, China

ABSTRACT

CDISC Case Report Tabulation Data Definition Specification (Define-XML) is one of primary documents required by FDA for electronic submission, which describes the content and structure of the included data within a submission. eCRT production is helpful to increase the level of automation and improve efficiency of the Regulatory Review process.

In most cases, the original datasets and specifications may not be completely 'ready' for direct eCRT development. This will cause repetitive modification of the contents, and even have negative impact on the quality.

The objective of this paper is to present tips on improving the efficiency in eCRT production for two kinds of models (Study Data Tabulation Model (SDTM), Analysis Data Tabulation Model (ADaM)) and Integration Summary of Safety and Efficacy (ISS/ISE)). It includes raw data readiness check before creating define, tips in CRF annotation, data specification requirements to be used for eCRT purpose, tips in define creation, supplemental materials preparation and validation methods. In addition, this paper also broaches the standard workflow of development which can be helpful to accomplish eCRT production automation.

Key words: CDISC eCRT, specification, define.xml, OpenCDISC

INTRODUCTION

eCRT(Electronic Case Report Tabulation) is a package required by FDA for electronic submission, which acts a road map to guide regulatory reviewers to navigate study metadata and understand the clinical studies submitted for approval. CRT data definition specification commonly referred to as "define.xml" is one of the primary documents in this package.

A standard for providing CRT data definitions within submission has been prepared by Clinical Data Interchange Standards Consortium (CDISC) define.xml team referring to FDA electronic submission (eSub) guidance and the electronic Common Technical Document (eCTD) documents. eCRT package should be produced based on this standard, which not only increases the level of automation, but also improves the efficiency of the Regulatory Review Process[1].

Currently, define.xml can be generated using SAS code automatically. However, this task could be challenging due to several reasons. Firstly, the original datasets do not meet the CDISC standard before eCRT production, this will cause repetitive modification of the contents, even have negative impact on quality. For example, the size of original datasets may be too large for FDA to proceed; variables in each dataset may not be aligned in the order that FDA requests and so on. It is a good idea to perform raw data readiness check before eCRT production. Secondly, during the development process, annotated CRF is one of the primary documents in eCRT package. While CRF annotation may not be compliant with CIDSC standard, this could make the reviewers difficult to understand the relationship between the information collected and SDTM datasets. Thirdly, the description of datasets and variables may not be displayed in a manner that meets the FDA's requirement, like domain order, attributes or structure information for domains; origin or comment for each variable may not accurate or clear enough; Controlled Terminology (CT) definition is not matched with CDISC standard and so on. This inaccurate or nonstandard information could result in the deviation from the correct understanding of clinical trial. Fourthly, the essential supplemental materials may not be contained in the package. Fifthly, eCRT package cannot be double validated like other programming work, mistake is not easy to avoid completely. Lastly, the data definitions for (Study Data Tabulation Model (SDTM) model, Analysis Data Tabulation Model (ADaM) and Integrated Summary of Safety and Efficacy (ISS/ISE) could be different, although a routine process can be used to product, difference among them needs to be watched out. It would be helpful to improve the efficiency and quality for eCRT production if tips are summarized.

In this paper, we present tips on improving the efficiency in eCRT production for two kinds of models SDTM,

ADaM and collection of studies ISS/ISE. It contains raw data readiness check before creating define, tips in CRF annotation, data specification requirements to be used for eCRT purpose, tips in define creation, supplemental materials preparation and validation methods. In addition, this paper also discusses the standard work-flow of development which can be helpful to accomplish eCRT production automation.

INPUT DATA READINESS CHECK

In this section, we are going to concentrate on the SDTM, ADaM datasets readiness before define production. SDTM is the content standard of case report form data tabulations from clinical research studies. ADaM is the content standard of analysis datasets.

When creating SDTM datasets and ADaM datasets, programmer must make sure that they are compliant with CDISC standards. The datasets must be kept as much stable as possible to avoid too much rework in define.xml. Below is a list of items which need programmers to pay attention to before they claim the datasets are 'ready':

1. File size of each dataset should be < 1GB

According to current guidance (Study Data Specifications, CDER Common Data Standards Issues Document, Draft Study Data Technical Conformance Guide) [2], each dataset should be provided in a single transport file. The maximum size of an individual dataset that FDA can process depends on many factors. Datasets greater than 1 gigabyte (gb) in size should be split into smaller datasets no larger than 1 gb in size. Datasets should be resized to the maximum length used for each character variable prior to splitting. Datasets divided to meet the maximum size restrictions should contain the same variable presentation so they can be easily combined. Split datasets should have matching column widths. This will ensure that the split datasets have matching variable lengths for merging data. Split data should be noted in the define.xml and the Study Data Reviewer Guide, clearly identifying the method used for the dataset splitting. Datasets that are split should be clearly named to aid the reviewer in reconstructing the original dataset (e.g., xxx1, xxx2, xxx3). For some examples of splitting Laboratory (LB), Questionnaire (QS) datasets, please refer to CDISC SDTMIG.

2. Sort variables should make every observation as unique key

When validating SDTM or ADaM datasets plus define.xml using openCDISC, the validation tool will check the key variables listed in define.xml vs. the actual datasets. Error message will be reported if duplicate observations are discovered. Programmers need to investigate the root cause of this error. If the sorting variables are correct and this issue is due to data issue, then we need to contact DM department to clean up the dataset. If the dataset is fine, then the sorting variables list need to be updated to make sure all observations are unique by the sorting variables.

3. SDTM dataset Variables order should comply with SDTM standard

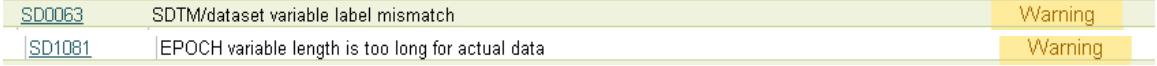
OpenCDISC validator (OCV) always checks the variable order in SDTM domains. Programmers should pay more attention when developing SDTM mapping spec especially when you need to add expected or permissible variables. These variables may not have example in the SDTM IG. Generally, the variables are listed by function in a specific domain, e.g. Identifier variables are always listed in the first batch, and timing variables are always listed in the last batch. The new added variables should be listed after the related required variables.

Display 1. Example for OCV result

4. Make sure the variable attributes are compliant

The variable attributes will be validated strictly by OCV including variable length, label and format. According to FDA request, the submitted variable length should be the largest variable value length, otherwise warning message will be reported. Programmers should 'shrink' all the variable length using the largest variable value length.

Sometimes the validation tool reports label mismatch. Programmers should double check the label in the SDTM IG. Before that the IG version should be confirmed. If the message still there, another option is to consult openCDISC.org [3], they shall provide a proper solution.



Display 2. Examples for OCV result

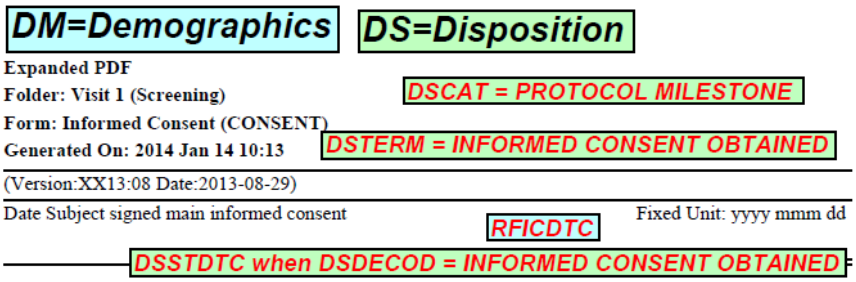
CRF ANNOTATION

A blank CRF annotation documents the location of the data in the corresponding tabulation datasets and the

variables [4]. Meanwhile, it provides a link between CRF questions and the database, which make the reviewers or programmers to understand the datasets easily [5]. The correct CRF annotation will greatly aid the reviewer to understand the relationship between the information collected and SDTM datasets. In define.xml, the variables originating from CRF can be efficiently traced back to where it was captured through annotation [6]. In this section, tips on correct CRF annotation are summarized as follows:

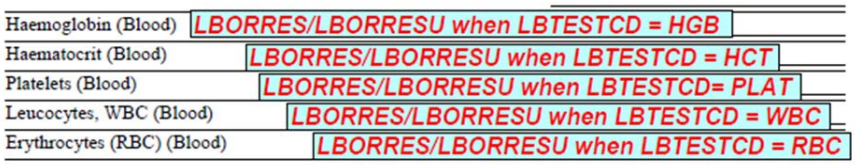
1. Annotation should be standard to meet FDA's requirement.

For the standard CRF annotation, the names of dataset and variable should be annotated for the corresponding item collected. "free-text" annotation is recommended since it is searchable using standard PDF viewers. The variable name is written into a text boxes with specific appearance, including the attributes like color, size and text style next to the spaces. If one item has been mapped into more than one datasets, all the locations and variable names should be annotated and as in different background colors, the example below (Display 3) illustrates the points mentioned above for a standard CRF annotation.



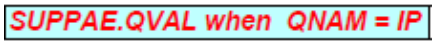
Display 3. Example for regular CRF annotation

Sometimes, Annotation needs to contain conditional statements to accurately convey a particular meaning, especially for the measure results for finding domains. For this situation, the items should be annotated as display 4 below.



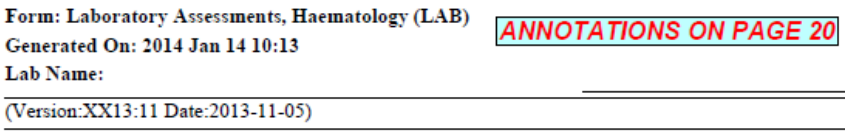
Display 4. Example for CRF annotation with conditional statements

In addition to the annotation mentioned above, for the items that cannot be mapped into SDTM standard variable, it should be mapped as supplemental qualifier variables, and annotated as follows:



Display 5. Example for CRF annotation for SUPP variable

For the repetitive variable annotation, the original location should be referred, and annotated as display as below:



Display 6. Example for CRF annotation for repetitive variable

2. Any variable should be ensured to be annotated, except for the operational or derived variables.

Any variable included in SDTM datasets for submission should be ensured to be annotated, except for the operational or derived variables. It is a good idea to check the inconsistencies after annotation, since it is possible that some variables do not need to map into datasets at last, these variables are permissible SDTM variables and are advised to drop due to no information collected eventually and are not required for analysis.

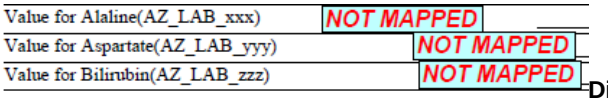
3. "NOT SUBMITTED" annotation.

For a variable which is not required referring to SDTM IG, if the corresponding information collected has been used to derive the other variable, or no information has been collected eventually, this variable should be dropped at last step. In this situation, the annotation should be annotated as “NOT SUBMITTED” on the CRF.

Display 7. Example for ”NOT SUBMITTED” annotation

4. “NOT MAPPTED” annotation.

Moreover, if data management department confirm that an item has not been collected from the beginning of clinical trials, it is recommended to annotate this item as “NOT MAPPTED”. At the same time, it should be documented and made explanation in the SDTM Reviewer’s Guide.



Display 8. Example for ”NOT MAPPED” annotation

DATA SPECIFICATION REQUIREMENTS

Since the specification is used as the basis of the generation of define.xml, it is important for the incorporation of the correct information into it. The specification defines the structure to describe the Case Report Tabulation Datasets and variables. The content includes four main parts, domain level metadata, variable level metadata, Controlled Terminology and Value Level Metadata. Domain level metadata lists all of the datasets included in the submission per study or collection of studies (e.g.,ISS/ISE). Whereas the Data Definition Tables provide detailed information about the variables contained within each of the domains. CT and VLM need to be provided per FDA’s requirement.

Domain level metadata

1. For TOC, attribute information should be compliant with CDISC SDTM IG.

For SDTM standard domain, the attributes displayed in SDTM define.xml should comply with CDISC SDTM IG [7]. For non-standard domain, the sponsor specifies the related information themselves and is recommended to describe the details in the SDTM Reviewer’s Guide.

2. Domain order in TOC.

For SDTM, domains listed in TOC should be ordered by class firstly, which are Trial Design Domains, Special-purpose Domains, Interventions Domains, Events Domains, Findings Domains and Relationship Domains. For domains within each class, they should be ordered alphabetically referring to the example below (Display 9).

Datasets for Study XXX

Dataset	Description	Class	Structure	Purpose	Keys	Location
TA	Trial Arms	Trial Design	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD	ta.xpt
TE	Trial Elements	Trial Design	One record per planned Element	Tabulation	STUDYID, ETCD	te.xpt
DM	Demographics	Special-Purpose	One record per subject	Tabulation	STUDYID, USUBJID	dm.xpt
CM	Concomitant Medications	Interventions	One record per recorded intervention occurrence or constant-dosing interval per subject	Tabulation	STUDYID, USUBJID, CMTRT, CMSTDTC, CMENDTC, CMSEQ	cm.xpt
AE	Adverse Events	Events	One record per adverse event per subject	Tabulation	USUBJID, AESPID, AETERM, AESTDTC	ae.xpt
EG	ECG Test Results	Findings	One record per ECG observation per time point per visit per subject	Tabulation	STUDYID, USUBJID, EGTESTCD, VISITNUM	eg.xpt
RELREC	Related Records	Relationship	One record per related record, group of records or dataset	Tabulation	RDOMAIN, USUBJID, IDVAR, IDVARVAL	relrec.xpt
SUPPAE	Supplemental Qualifiers AE	Relationship	One record per qualifier variable name per identifying variable value per identifying variable per subject	Tabulation	USUBJID, IDVAR, IDVARVAL, QNAM	suppae.xpt

Display 9. Example for domain order in TOC in SDTM define

3. Variables listed in Keys should be matched with structure description, and obtain unique observation.

It is worth noting that the variables listed in keys should be compliant with the structure description. Moreover, the key variables should be able to obtain unique observation in the dataset, and the validation tool mentioned in the first part can help to check this structure information.

4. Comparison among SDTM, AdaM and ISS/ISE

Comparing with SDTM, the domain level metadata in ADaM and ISS/ISE include all the same information with ones in SDTM. However, the display order is flexible. It is recommended that the domains are ordered alphabetically except ADSL display firstly referring to CDISC ADaM IG [8](Display 10).

Dataset	Description	Class	Structure	Purpose	Keys	Location
ADSL	Subject-Level Analysis Dataset	Analysis	One record per subject	Analysis	STUDYID, USUBJID	adsl.xpt
ADAE	Adverse Events Analysis Dataset	Analysis	One record per adverse event per subject	Analysis	STUDYID, USUBJID, ASTDT, AEDECOD	adae.xpt

Display 10. Example for domain order in TOC in ADaM define

Variable level metadata

1. Attribute information should be referred to SDTM and ADaM IG.

For the variable metadata, several metadata attributes to describe SDTM data are described, including the variable name, label, data type, controlled terminology, origin, role and the comments. For the SDTM standard variables, the information like name, label, type and role should be used based on the CIDSC SDTM IG.

2. Origin and Comment information need to be watched out.

It is notable that the origin and comment information are not easy to illustrate clearly and always need to repetitive modification. Since the wrong origin will make the reviewer confused about the variable definition, also, the comment information may present the derivation rules as SAS code which is convenient for programmer to create datasets. It is a good idea to make clear how to incorporate the information correctly. There are five kinds of origin including CRF (only in SDTM model), eDT, Derived, Assigned, and Protocol. The page number should be correctly direct to the variables mapped if the origin of this variable is “CRF”; for the origin of the variable belongs to “eDT” or “CRF”, the comment information should leave blank; the further clarification can be added if need for the origin of “Assigned” and “Protocol”; for the origin is “Derived”, the comment about data derivation rules or computational method used in the study must be described, such as data imputation rules or conventions. In addition, the comment should avoid putting SAS code inside. The example in Display 11 illustrates the tips summarized above.

3. Comparison among SDTM, AdaM and ISS/ISE

It is worth noting that the comments for derived variables are different among SDTM model, ADaM model and ISS/ISE model. In ADaM model, the comments should be a concatenation of domain name and variable if computational rules use the SDTM variable directly (Display 12). In ISS/ISE situation, it just needs to clarify the origin of the variable without detailed derivation rules input if there is no new rules used in the integrated analysis (Display 13).

Demographics Dataset (DM)						dm.xpt
Variable	Label	Type	Controlled Terminology	Origin	Role	Comment
STUDYID	Study Identifier	text		Assigned	Identifier	D12345678
USUBJID	Unique Subject Identifier	text		Derived	Identifier	Concatenation of study and subjid
SUBJID	Subject Identifier for the Study	text		CRF Page 223	Topic	
COUNTRY	Country	text	COUNTRY	Protocol	Record Qualifier	CHN

Display 11. Example for SDTM define

Subject-Level Analysis Dataset Dataset (ADSL)					adsl.xpt
Variable	Label	Type	Controlled Terminology	Source	Comment
STUDYID	Study Identifier	text		DM.STUDYID	
SITEIDN	Study Site Identifier (N)	integer		Derived	Numeric version of SITEID
TRT01A	Actual Treatment for Period 1	text		Derived	= DM.ACTARM

Display 12. Example for ADaM define

Subject Level Analysis Data, Integrated Dataset (ADSL)					adsl.xpt
Variable	Label	Type	Controlled Terminology	Source	Comment
STUDYID	Study Identifier	text		ADSL.STUDYID	From individual study ADSL dataset (study XX1and study XX2)
QTCFL	QTC Analysis Set Population Flag	text	NY	ADSL.QTCFL	From individual study ADSL dataset (only study XX2)
BLBSA	Baseline BSA (m2)	float		Derived	Equals to 0.007184 * BLWGHT^(0.425) * BLHGHT^(0.725)

Display 13. Example for ISS/ISE define

Controlled Terminology (Code Lists)

1. Code List in the define.xml should match CDISC standard code lists.

CDISC Controlled Terminology is a set of standardized possible values for a variable, which is an important part in the domain definition section when building define.xml. For the internal CT, if a Term is defined by CDISC, the name of Code List in the define.xml must match standard name in the CDISC CT list. If it is not included in the standard list and the name of CT is permissible, the values should be used as the CT. At the same time, the notable extensions to CDISC CT should be described in the Reviewer's Guide.

2. The coded values need to represent in a meaningful order in an analysis or display context.

In addition, for the coded values need to represent in a meaningful order in an analysis or display context, the rank order should be provided. For example, since the CT for TOXGRV4 includes five kinds of values, the order needs to add to display it meaningfully (Display 14).

TOXGRV4, Reference Name (TOXGRV4)	
Code Value	Code Text
1	Mild AE
2	Moderate AE
3	Severe AE
4	Life-threatening or disabling AE
5	Death related to AE

Display 14. Example for Internal CT

3. Some specific CTs should be split into each domain for displaying transparently.

Per implementation guide, some specific CT is required to provide in define.xml as the same name. For example, CT name is "LOC" which means Location information used for measurement or exam is required to include in several domains. It will be better if it includes into different domains as different CT names.

4. The accurate dictionary name and version should be provided for external CTs.

For example, ISO8601 is used for date variables, MedDRA is used for AE and MH variables, and WHODD is also used for CM variables. The dictionary name and version need to be provided for these external CTs (Display 15).

Controlled Terminology (External Dictionaries)	
CMDICT_F, Reference Name (CMDICT_F)	
External Dictionary	Dictionary Version
AZ Drug Dictionary	14.2
ISO8601, Reference Name (ISO8601)	
External Dictionary	Dictionary Version
ISO 8601	Current Standard
MedDRA, Reference Name (MedDRA)	
External Dictionary	Dictionary Version
MedDRA	17.1

Display 15. Example for External CT

5. Comparison among SDTM, ADaM and ISS/ISE

CDISC CT population for SDTM model should refer to CDISC SDTM IG. Comparing with SDTM, ADaM should contain all the CTs populated in the SDTM. In addition, for the new derived variables in ADaM, CT population should refer to CIDSC ADaM IG. The same CT population method should be performed for ISS/ISE.

Value Level Metadata (Value List)

For the Value Level Metadata (VLM), it often describes AVAL or AVALC in BDS data structures based on the values of PARAMCD.

1. VLM information should be complete enough to provide additional metadata support in order to be useful for review and analysis

In order to provide additional metadata support in VLM, information specially the origin of the value should be specified.

Value Level Metadata (ValueList.PE.PETESTCD)							
Source Variable	Value	Label	Type	Controlled Terminology	Origin	Role	Comment
PETESTCD	ABDOMEN	Abdomen	text		CRF Page 18, 34		
PETESTCD	CV	Cardiovascular	text		CRF Page 17, 33		
PETESTCD	GENERAL	General appearance	text		CRF Page 14, 30		
PETESTCD	HEAD	Head and neck	text		CRF Page 15, 31		
PETESTCD	LYMPH	Lymph nodes	text		CRF Page 15, 31		
PETESTCD	MUSC	Musculoskeletal / Extremities	text		CRF Page 16, 32		
PETESTCD	NEURO	Neurological	text		CRF Page 18, 34		
PETESTCD	RESP	Respiratory	text		CRF Page 17, 33		
PETESTCD	SKIN	Skin	text		CRF Page 14, 30		
PETESTCD	THYROID	Thyroid	text		CRF Page 16, 32		

Display 16. Example for VLM

2. Consistence across variable level metadata, VLM and CRF annotation should be check.

Besides values of PARAMCD or TESTCD in VLM, the value of QNAM in SUPP domain also should include in VLM. In this situation, there is some overlap between variable level metadata and VLM, the variables included in SUPP domain should comply with the value of QNAM in VLM. Meanwhile, a cross check should be performed among annotated CRF, SUPP domains in variable level metadata, and value of QNAM in VLM.

SUPPLEMENTAL MATERIALS PREPARATION

Currently most of sponsors provide supplemental data definition information as an appendix in the Data Definition Document, since this additional or supplemental information provide regulatory reviewers further explanations or descriptions of the variables contained within the datasets.

1. Reviewer's Guide

Reviewer's Guide is the most important part in these supplemental materials, which need to be placed in each study of eCTD in Module 5.

For SDTM, the Study Data Reviewer's Guide needs to provide protocol description, such as protocol title, number and version, protocol design, description for trial design datasets. Subject Data Description also needs to be included, like notification of annotated CRF, the detailed description for SDTM domains. At last, Data Standard conformance validation rules, versions and issues summary are involved in the guide.

For ADaM, in addition to the information mentioned above, Analysis Data Reviewer's Guide needs to focus on some information description, for example, comparison of SDTM and ADaM content, core variables, treatment variables, subject issues that require special analysis rules, use of visit windowing, unscheduled visits and record selection, and imputation/derivation methods. Also, the issues about analysis data creation and processing needs to summarize in guide, like split datasets, data dependencies, intermediate datasets or variable conventions. At last, the programs need to submit should be listed.

2. Other optional supplemental materials

In addition, if some clinical algorithms as expressed in the Statistical Analysis Plan(SAP) which are complex and may need additional explanations, or are bested described in a flow chart or in some other graphical depiction, an external PDF file that contains this additional information are helpful.

Moreover, some other supplemental materials such as conversion factor which are used to normalize LBSTRESN from LBORRES, or International SOC order number based on AEBODSYS used for reporting is also added into the package as a supplemental material.

DEFINE.XML CREATION

Once the specification is finished, define.xml and define.pdf are ready to be generated. It is recommended to perform cross check in the define.xml after generation. Firstly, the TOC list should be consistent with datasets presented in the define; secondly, the name, label and type of the variable should be identical to the information in tabulation datasets; thirdly, the format of variables for similar types of data should be consistent within and across study, such as the date variables; lastly, CT and VLM should be consistent within and across the whole study.

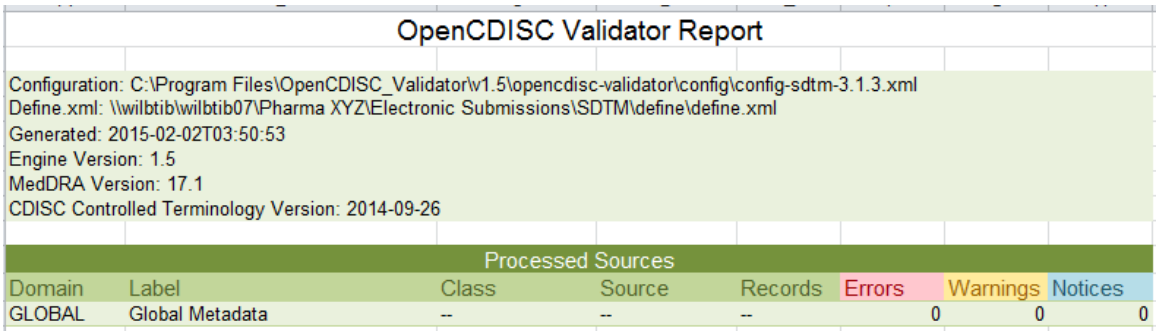
VALIDATION METHOD

OpenCDISC's Validator checks define.xml against the schema and CRT-DDS. It is rated Easy in Set-up and Ease of Use categories as it is very straightforward to download, install, and use. This tool has a good user interface, allowing the user to click and select the options and file to test. Additionally, OpenCDISC provides a full list of checks which it performs for define.xml. The thoroughness of checks gave this tool an Excellent rating for validation checks. It certainly helped us identify several inconsistencies during our initial testing phases which could then be addressed systematically once we put our define.xml process into production. The reports can be cumbersome to interpret at times, leading to a Moderate rating for the reports category. Several of our define.xml files initially failed in OpenCDISC after passing in DefineValidator, so it seems that OpenCDISC has additional checks beyond that of DefineValidator. Below are some tips when conducting OCV check:

Version confirmation

Since the validation tool is updating frequently as well as standards and dictionaries. Confirming the correct version is the first thing to do before validation. Some trials may last for several years, below tool and standard version need to be considered:

- OpenCDISC validator version (latest: v2.0.1)
- Config SDTM version (latest: v3.2)
- MedDRA version (latest: v8.0)
- CDISC Controlled Terminology Version [9] (latest: 2015-06-26)



OpenCDISC Validator Report

Configuration: C:\Program Files\OpenCDISC_Validator\v1.5\opencdisc-validator\config\config-sdtm-3.1.3.xml
Define.xml: \\wilbtib\wilbtib07\Pharma XYZ\Electronic Submissions\SDTM\define\define.xml
Generated: 2015-02-02T03:50:53
Engine Version: 1.5
MedDRA Version: 17.1
CDISC Controlled Terminology Version: 2014-09-26

Processed Sources

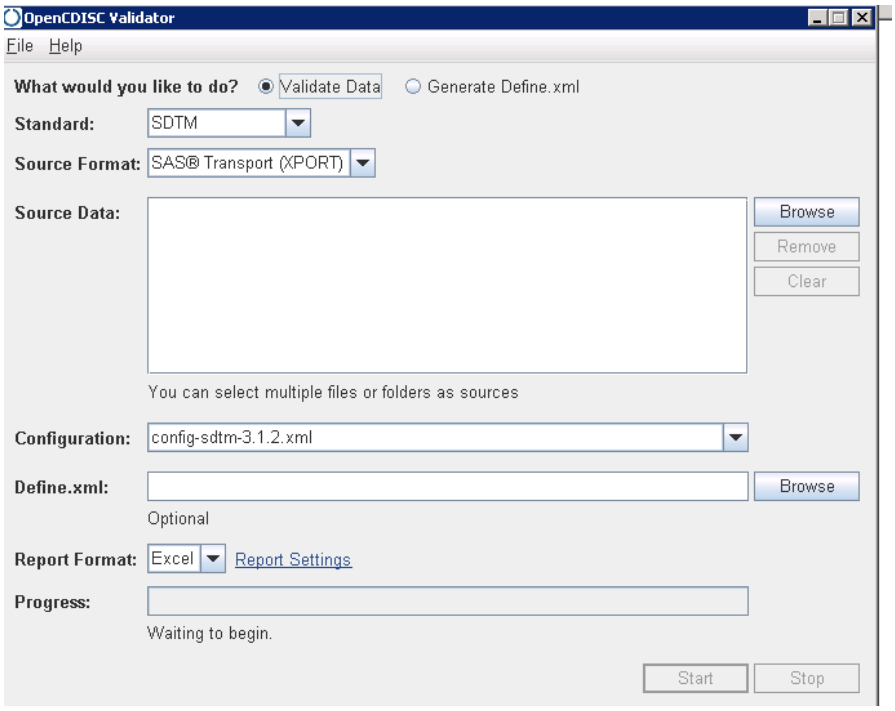
Domain	Label	Class	Source	Records	Errors	Warnings	Notices
GLOBAL	Global Metadata	--	--	--	0	0	0

Display 17. OpenCDISC Validation Report header

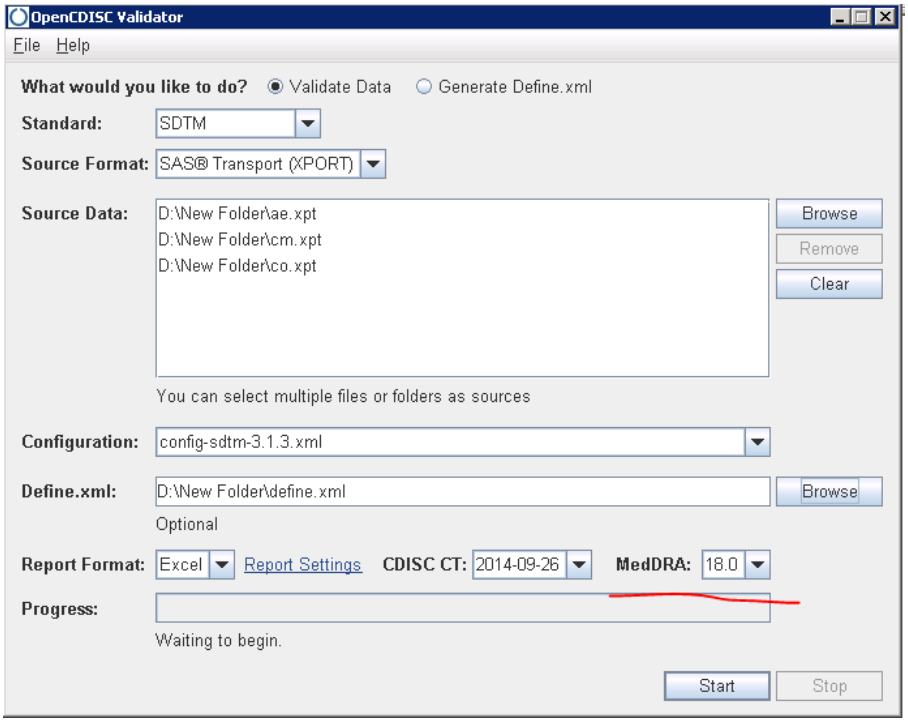
Install dictionary

The openCDISC Validator does not embed external dictionaries (e.g. MedDRA) as default. Users shall install manually. openCDISC official website provided a detailed instruction.

1. After successful installation of OpenCDISC Validator, navigate to the */components/config/data/MedDRA* folder.
2. The next step is to locate your company's licensed copies of MedDRA.
3. Copy the contents of the *MedAscii* folder to [version] folder under MedDRA.
4. The configuration is now complete, you are ready to validate data against MedDRA



Display 18. Before installing MedDRA dictionary



Display 19. After installing MedDRA dictionary

Understand the report code

OpenCDISC provides a simple naming convention for its validation rules. The first two characters identify the type of validation.

- CT Controlled Terminology
- SD Study Data
- DD Data Definition
- OD Operation Data Model

Detailed rules can be found in openCDISC website (e.g. SDTM 3.1.3 [10]):

Normally a CDISC compliant dataset should be free of Error and Warning (Any unsolved Error and Warning should be noted in Reviewer's Guide)

OpenCDISC breaks down validation checks into the following categories:

- **Consistency**: There are dependencies between certain variables.
- **Controlled Terminology**: Some variables require only specific values based on pre-defined code lists.
- **Format Compliance**: The values of certain variables must follow a specific format, such as the ISO 8601 standard used for dates, time, and periods of time.
- **Referential Integrity**: There are dependencies across domains.
- **Limits**: The values of certain variables must be within specified limits, such as the variable AGE which must be greater than zero.

DISCUSSION

Although we can create define files using SAS code, even a good use interface to help us generate the define files from the specification directly. During the creation process, parts of work still needs to perform manually, this will cause the mistake or inconsistence may happen, and will need the resource to review the work after generation. For example, CT, origin and comments cannot be generated automatically. Although the information is not the same for different study, is it possible to set up some way to standardize the information?

CONCLUSION

The task of eCRT production could be challenging to meet FDA's requirement, but it could save much time and effort if we obtain the key points. In this paper, we summarized the tips on eCRT package preparation. For example, we need to check the raw data readiness before eCRT production; some notification for CRF annotation should be watched out; how to incorporate the correct information into specification to avoid the repetitive modification; tips on creating define.xml and define.pdf; and validation method to make sure the final eCRT package for submission can meet FDA's requirement. These tips would be helpful in increasing the efficiency and quality during eCRT production.

REFERENCES

[1] CDISC Define-XML Team. "CDISC Define-XML Specifications." Version 2.0, http://www.cdisc.org/define-xml
[2] STUDY DATA TECHNICAL CONFORMANCE GUIDE. U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). 2015
[3] OpenCDISC open source community, http://www.opencdisc.org/
[4]CDISC Metadata Submission Guidelines v 1.0. (draft for comments – 25 Jul 2007) Appendix to SDTM IG v. 3.1.2 http://www.cdisc.org/models/sdtm/v1.1/index.html
[5]Study Data Specifications v 2.0 Jul 2012 http://www.fda.gov/cder/regulatory/ersr/ectd.htm
[6]Ryan Wilkins, Joel Campbell; A Regular Language: The Annotated Case Report Form, Proceedings of the Pharmaceutical SAS® Users Group Conference, PharmaSUG 2011
[7] CDISC Submission Data Standards Team. "Study Data Tabulation Model Implementation Guide." Version 3.2, http://www.cdisc.org/sdtm
[8] CDISC Analysis Data Model Team. "Analysis Data Model (ADaM) Implementation Guide." Version 1.0, http://www.cdisc.org/adam
[9] CDISC Controlled Terminology http://www.cdisc.org/terminology
[10] CDISC SDTM 3.1.3 Validation Rules http://www.opencdisc.org/projects/validator/cdisc-sdtm-3.1.3-validation-rules

ACKNOWLEDGMENTS

Appreciation goes to Zibao Zhang for his valuable review and comments.

RECOMMENDED READING

- Base SAS® Procedures Guide
- ☐ SAS® For Dummies®

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Lanting Li
Enterprise: PPD
Address: 23F, 5 Corporate Avenue, 150 Hubin Road,
City, State ZIP: Shanghai 200021, China
Work Phone:
Fax:
E-mail: Lanting.li@ppdi.com
Web:
Twitter:

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.