PharmaSUG China 2018 - Paper CD-43

Standardize Study Data for Electronic Submission

Qin Li, Regeneron Pharmaceuticals Inc.

ABSTRACT

In December 2014, FDA announced that "submissions under NDAs, ANDAs, BLAs, and INDs must be in electronic format specified in FDA guidance." The binding implementation date is March 2017. In the past years, even before the blinding requirement, Regeneron has been following the guidance as close as possible and successfully submitted number of electronic data packages for BLAs and sBLAs. We are happy to take this opportunity to share our experience with others at PharmaSUG China.

INTRODUCTION

In December 2014, FDA announced that *"submissions under NDAs, ANDAs, BLAs, and INDs must be in electronic format specified in FDA guidance."* The binding implementation date is March 2017. FDA issued following guidance and technical specifications document on electronic submission requirements to support implementation:

- Providing Regulatory Submissions in Electronic Format – Submission Under Section 745(a) of the Federal Food, Drug, and Cosmetic Act
- Guide for Industry: Providing Regulatory Submissions in Electronic Format – Standardized Study Data
- Study Data Technical Conformance Guide
- Data Standards Catalog
- Providing Regulatory and Submissions in Electronic Format – Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications

Standardized Study Data are required for INDs, NDAs, ANDAs, BLAs, and all subsequent submissions. Data must be in a format that the Agency can process, review and archive, as specified in the Data Standard Catalog. The Catalog provide guidance on the following standards:

- Exchange Format Standards
- Study Data Standard – Clinical Data Interchange Standards Consortium
- Controlled Terminology Standard
- Medical Dictionary for Regulatory Activities (MedDRA, WHO Drug, UNII etc.)

In the past Regeneron Pharmaceuticals collaborated with business partners on FDA and global submissions. The tools and processes of our business partners were most time in legacy standards. We started to build internal tools and processes for electronic submission according to current standards and guidelines in 2015. We have supported two major BLAs and a number of supplement BLAs by our tools and processes since then.

STUDY DATA TECHNICAL CONFORMANCE GUIDE

The Study Data Technical Conformance Guide (The Guide) supplements FDA Guide for Industry: Providing Regulatory Submissions in Electronic Format – Standardized Study Data. It provides specifications, recommendations, and general considerations on how to submit standardized study data

using FDA-supported data standards located in the Data Standards Catalog. It clarifies interpretations on CDISC standards and its implementation. The Guide is intended to complement and promote interactions between sponsors and FDA review divisions. Communication with review divisions is recommended as early as the pre-IND meeting.

A brief summary of sections covered in the Guide is as below.

PLANNING AND PROVIDING STANDARDIZED STUDY DATA section provides guidance on Study Data Standardization Plan (SDSP) and Reviewer's Guide (RG).

The Study Data Standardization Plan (SDSP) is a plan describing the submission of standardized study data to FDA. It will assist FDA in identifying potential data standardization issues early in the development program.

RGs describe any special considerations or directions or conformance issues that may facilitate an FDA reviewer's use of the submitted data and may help the reviewer understand the relationships between the study report and the data. Typical RGs for clinical trials are Study Data Reviewer's Guide (cSDRG) and Analysis Data Reviewer's Guide (ADRG).

EXCHANGE FORMAT - ELECTRONIC SUBMISSIONS section lists acceptable file formats: XML, PDF, and XPT (SAS SAS Transport Format XPORT Version 5). XLM format is used for define.xml; PDF format for acrf, RGs, and other supporting documents; XPT for all SAS datasets.

This section also provides details for datasets and variables, e.g.:

- Dataset size, names, labels, and descriptor length
- Variable length, names, labels, and descriptor length
- Special characters

STUDY DATA SUBMISSION FORMAT section indicates that Clinical Data Interchange Standards Consortium (CDISC) standards and corresponding implementation guide (IG) for Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM) should be followed.

In this section, the Guide provides clarifications on SDTMIG for imputation, numerically coded variables, subject identifier (SUBJID), unique subject identifier (USUBJID), adjudication data, and selected SDTM domains; and clarifications on ADaMIG for relationship of SDTM datasets and ADaM datasets, key efficacy and safety variables, timing variables, core variables, numeric date variables, dataset labels, data imputation, and submission of software programs.

THERAPEUTIC AREA STANDARDS THERAPEUTIC AREA (TA) section is introduced to extend the CDISC foundational standards to represent data that pertains to specific disease areas. CDISC publishes a TA User Guide (TAUG) for each therapeutic area which includes the extensions as disease-specific metadata, examples and recommendations for use. Currently there are thirteen TA extensions incorporated into FDA supported CDISC foundational standards.

TERMINOLOGY section recommends that sponsors use the terminologies supported and listed in the Data Standards Catalog. Common dictionaries should be used across all clinical studies and throughout the submission for each of the following: adverse events, concomitant medications, procedures, indications, study drug names, and medical history.

ELECTRONIC SUBMISSION FORMAT sections states that study datasets and their supportive files should be organized into a specific file directory. Details are displayed by the figure Folder Structure for Study Datasets and the table Study Dataset and File Folder Structure and Description.

DATA VALIDATION AND TRACEABILITY section explains that study data validation helps to ensure compliance, usefulness, and meaningfulness of the study data for review and analysis. Validation activities occur at different times during submission and review of study data, including submission receipt and at the beginning of the regulatory review.

Study data validation combine rules from Standards Development Organizations (e.g., CDISC), FDA eCTD Technical Rejection Criteria for Study Data, and FDA Business and Validator rules.

Traceability permits an understanding of the relationships between the analysis results (tables, listings and figures in the study report), analysis datasets, tabulation datasets, and source data. It is an important component of a regulatory review. This section also provides guidance on legacy data conversation for electronic submissions.

OUR PRACTICE

Our practice in study data standardization and preparation for electronic submission is centered around FDA guidelines. The activities are categorized in the following sections.

PLANNING

Our data standardization planning for a clinical study starts as early as CRF design. Our programmers participate in CRF design and review relevant data management documents, e.g. Data Management Plan (DMP), Data Transfer Specification (DTS), Architecture Loader Specification (ALS) of Electronic Data Capture system (EDC), etc. Our inputs on data standards, data relationships, regulatory requirements, and correct terminologies, etc. are valuable and important at study setup.

Between EDC go-live to First Patient First Visit (FPFV), SDTM specification will be drafted. When a solid draft of SAP is available, SDTM specification draft will be completed, and programming will start. ADaM specification draft starts after a solid draft of SAP, and complete with mock table shells. ADaM programming will start with a solid draft mock table shells and dry run planning. Both SDTM and ADaM will be validated on CDISC and regulatory compliance. Findings will be documented for future reference. Output programming follows ADaM programming.

All programming and validating should be completed before database lock in the development area on our server. After database lock, same programming and validating will be conducted in the production area on server.

CREATION

The simple flow chart below (Display 1, Process Flow Chart) describes our basic process.

Display 1, Process Flow Chart

Source data include data from EDC, external vendors (e.g. lab, ECG, ePRO, etc.), and other (e.g. adjudicated data). It also includes relevant documentations for these data. We find the following documents helpful in programming standardized data:

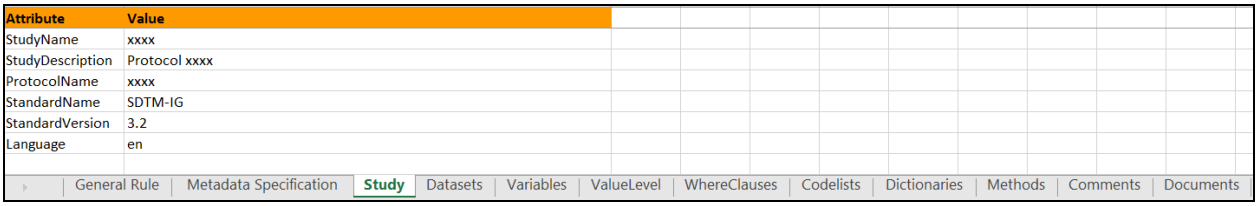
- Case Report Form (CRF) maps all data entered by investigational sites into EDC. The annotated version (annotated CRF) links each variable in EDC data transfer to certain field(s) on CRF page. A blank version of complete unique CRF will be used to create SDTM annotated CRF (acrf.pdf).
- Architecture Loader Specification (ALS), or Study Design Specification (SDS) for EDC provides a complete list of variables collected in EDC, their attributes, parameter values, and code list.
- Data Management Plan (DMP) includes all details in data processing on clinical data of a study. It covers EDC system, data processing and validation, vendor data, Serious Adverse Events (SAE)

reconciliation, dictionary coding, data transfers, and database lock process. It is a good reference to understand study data.

- Data Transfer Plan (DTS) could be a section in DMP or a separate document. DTS of vendor data is very important to programming, as it provides details in variables, their attributes, parameter values, and code list. It also specifies transfer timelines, reconciliation process, data format, and route (method) of transfer
- Electronic Patient Report Outcome (ePRO), or questionnaire data collected outside EDC, has been challenging in data standardization, especially those custom designed questionnaires. A well-defined DTS is must. Annotated screen shots of the electronic devices are very useful.

Source data could be in various formats. Most commonly we have data in SAS dataset (.sas7bdat), SAS transport files (.xpt), excel (.xls or .xlsx), or common delimited files (.csv).

Tabulation Data (SDTM) are created from source data on SDTM specifications. Our specifications (SDTM and ADaM) follow the templates provided by Pinnacle 21. It is a multi-sheet excel file built mainly upon CDISC SDTMIG and FDA Study Data Technical Conformance Guide. Guidance and requirements from other regulatory authorities, e.g. PMDA, are also considered. The specification includes all details for tabulation data mapping, i.e. datasets, variables, values, code list, dictionaries, and any additional information. The screen shot below (Display 2, REGN_SDTM_define_v3.2) demonstrates our generic SDTM specification template.

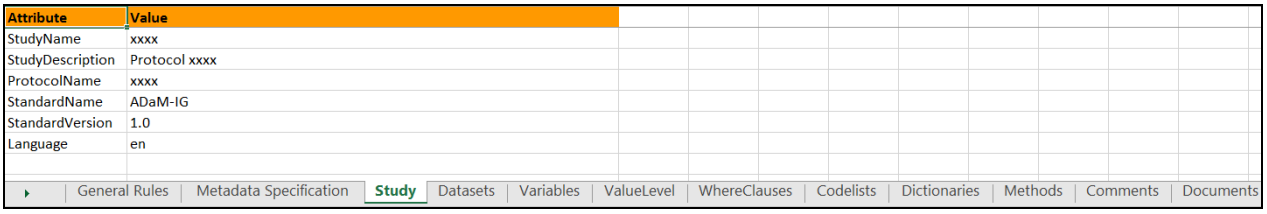


Attribute	Value
StudyName	xxxx
StudyDescription	Protocol xxxx
ProtocolName	xxxx
StandardName	SDTM-IG
StandardVersion	3.2
Language	en

General Rule | Metadata Specification | Study | Datasets | Variables | ValueLevel | WhereClauses | Codelists | Dictionaries | Methods | Comments | Documents

Display 2 REGN_SDTM_define_v3.2

Analysis Data (ADaM) are created from SDTM data on ADaM specifications. Similar as SDTM speficiation, ADaM specification is a multi-sheet excel file built mainly upon CDISC SDTMIG and FDA Study Data Technical Conformance Guide. Guidance and requirements from other regulatory authorities are also considered. The specification includes all details for analysis data mapping and derivations, i.e. datasets, variables, values, code list, dictionaries, and any additional information. The screen shot below (Display 3, REGN_ADaM_define_v1.0) demonstrates our generic ADaM specification template.



Attribute	Value
StudyName	xxxx
StudyDescription	Protocol xxxx
ProtocolName	xxxx
StandardName	ADaM-IG
StandardVersion	1.0
Language	en

General Rules | Metadata Specification | Study | Datasets | Variables | ValueLevel | WhereClauses | Codelists | Dictionaries | Methods | Comments | Documents

Display 3 REGN_ADaM_define_v3.2

Study level specifications for both SDTM and ADaM are built upon the generic template for each clinical trial. The specification excel files will be read into each SAS program for SDTM or ADaM creation to apply standard attributes for datasets and variables. Each SDTM or ADaM program is named same as the dataset to be created, e.g. dm.sas for creating of dm and suppdm domains, adae.sas for creating of adae.sas7bdat.

VALIDATION

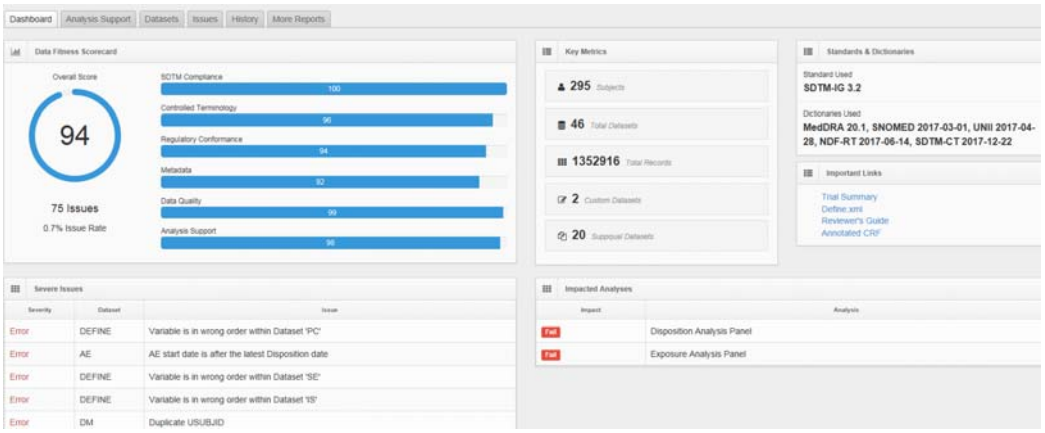
Generally, we work on every clinical study as if it is going for regulatory submission. We use Pinnacle 21 Enterprise to validate standardized study data (SDTM and ADaM) and supporting documentation. Validation starts when draft datasets are created. The following steps iterate throughout development of SDTM and ADaM:

- Load SDTM or ADaM specifications to Pinnacle 21 system to generate define.xml file
- Covert SAS datasets (sas7bdat) to SAS transport files (xpt)
- Load xpt files and define.xml to Pinnacle 21 system and generate final validation report
- Export validation report, review and document explanations as appropriate

Pinnacle 21 Enterprise provides an overall report on compliance status for all study validated. For each individual study, it provides a dash board with the following panel:

- Data Fitness Scorecard, which reports fitness scores for SDTM Compliance, Controlled Terminology, Regulatory Conformance, Metadata, Data Quality, and Analysis Support
- Severe Issues, categorized as Rejects, Errors, Warnings, and Notices
- Key Metrics, e.g. number of subjects, datasets and suppquals, records, etc.
- Standards & Dictionaries used
- Impacted Analyses

In separate tabs, the report goes further into details on Analysis Support, Datasets, and Issue. The screen shot below (Display 4, Standardized Study Data Validation Report by Pinnacle 21 Enterprise) shows how a study level report looks.

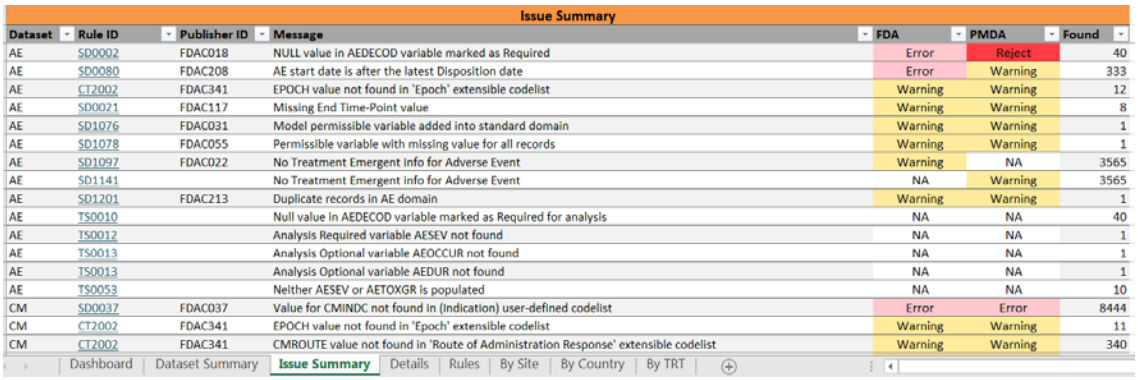


Display 3 REGN_ADaM_define_v3.2

When developing SDTM and ADaM programming during study ongoing, we will export the issues identified by Pinnalce 21 tool as an excel file. The excel file will help to:

- Identify and report data issues
- Update programming, e.g. code list, conformance issues, etc.

The screen shot below (Display 5, Pinnacle 21 Enterprise Validation Report) shows an example.



Issue Summary

Dataset	Rule ID	Publisher ID	Message	FDA	PMDA	Found
AE	SD0002	FDAC018	NULL value in AEDECOD variable marked as Required	Error	Reject	40
AE	SD0080	FDAC208	AE start date is after the latest Disposition date	Error	Warning	333
AE	CT2002	FDAC341	EPOCH value not found in 'Epoch' extensible codelist	Warning	Warning	12
AE	SD0021	FDAC117	Missing End Time-Point value	Warning	Warning	8
AE	SD1076	FDAC031	Model permissible variable added into standard domain	Warning	Warning	1
AE	SD1078	FDAC055	Permissible variable with missing value for all records	Warning	Warning	1
AE	SD1097	FDAC022	No Treatment Emergent info for Adverse Event	Warning	NA	3565
AE	SD1141		No Treatment Emergent info for Adverse Event	NA	Warning	3565
AE	SD1201	FDAC213	Duplicate records in AE domain	Warning	Warning	1
AE	TS0010		Null value in AEDECOD variable marked as Required for analysis	NA	NA	40
AE	TS0012		Analysis Required variable AESEV not found	NA	NA	1
AE	TS0013		Analysis Optional variable AEOCCUR not found	NA	NA	1
AE	TS0013		Analysis Optional variable AEDUR not found	NA	NA	1
AE	TS0053		Neither AESEV or AETOXGR is populated	NA	NA	10
CM	SD0037	FDAC037	Value for CMINDC not found in (Indication) user-defined codelist	Error	Error	8444
CM	CT2002	FDAC341	EPOCH value not found in 'Epoch' extensible codelist	Warning	Warning	11
CM	CT2002	FDAC341	CMROUTE value not found in 'Route of Administration Response' extensible codelist	Warning	Warning	340

Dashboard | Dataset Summary | Issue Summary | Details | Rules | By Site | By Country | By TRT

Display 5 Pinnacle 21 Enterprise Validation Report

Before database lock, all issues categorized as Reject must be fixed in data or programming. Errors and Warnings will be fixed in data or programming as much as possible. Explanation for issues remained will be documented in an additional column in the excel file. The documentation will be used to:

- Complete Data Conformance Summary section in SDRG or ADRG
- Prepare responses for future inspection

An example of the Data Conformance Summary in SDRG for electronic submission is displayed here (Display 6, Data Conformance Summary in SDRG). Wordings in Explanation column were extracted from Pinnacle 21 Enterprise Validation Report for this study.

4. Data Conformance Summary

4.1 Conformance Inputs

Was Pinnacle21 used to evaluate conformance?	**Yes**
If yes, specify the versions of Pinnacle21 and the Pinnacle21 validation rules:	**Pinnacle 21 Enterprise version 3.2.4**
Were sponsor-defined validation rules used to evaluate conformance?	**No**
If yes, describe any significant sponsor-defined validation rules:	**N/A**
Were the SDTM datasets evaluated in relation to define.xml?	**Yes**
Was define.xml evaluated?	**Yes**

4.2 Issues Summary

Check ID	Diagnostic Message	Severity	Dataset	Count (Issue Rate)	Explanation
CT2002	EPOCH value not found in 'Epoch' extensible codelist	Warning	AE	16 (0.45%)	Values have been added into the extensible codelist. This is a precautious warning by Pinnacle21
CT2002	CMDOSFRQ value not found in 'Frequency' extensible codelist	Warning	CM	1162 (10.88%)	Values have been added into the extensible codelist. This is a precautious warning by Pinnacle21
CT2002	CMDOSU value not found in 'Unit' extensible codelist	Warning	CM	1434 (13.42%)	Values have been added into the extensible codelist. This is a precautious warning by Pinnacle21

Display 6, Data Conformance Summary in SDRG

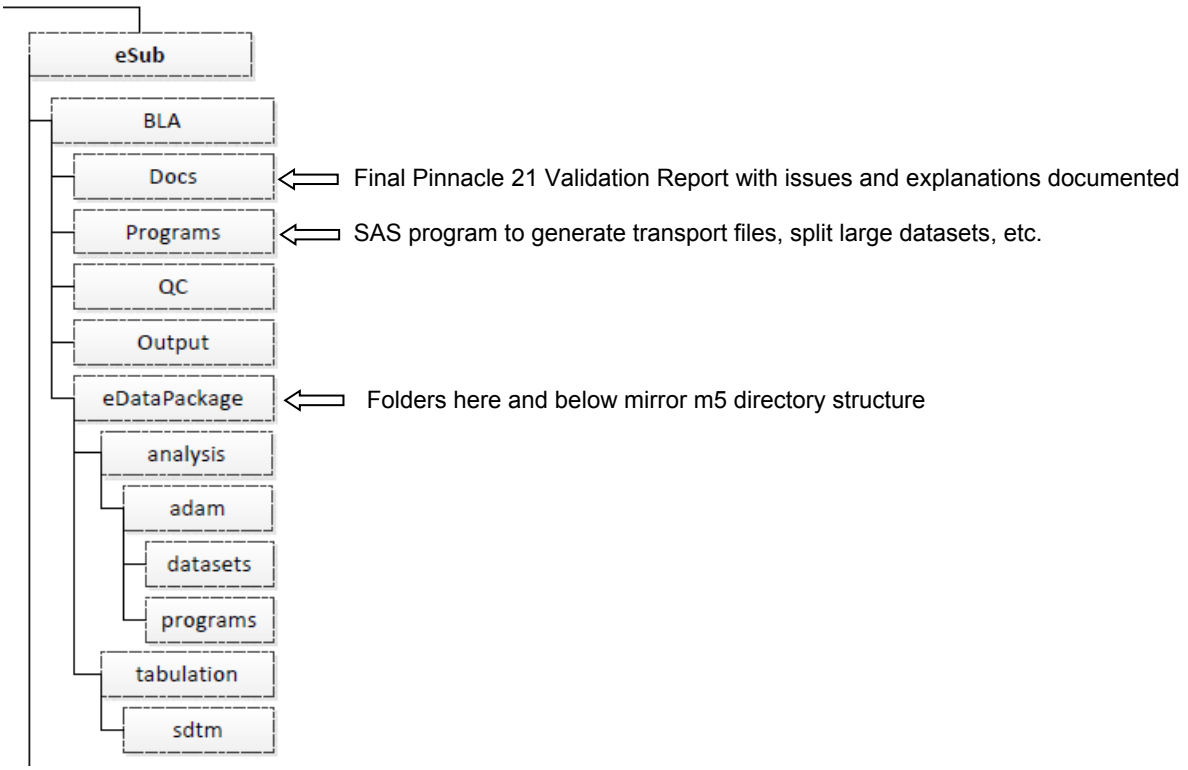
SUBMISSION

When a submission s planned, we will complete the Study Data Standardization Plan (Checklist), which will be included in pre-submission briefing book for FDA. We will start to prepare integrated analysis, and data packages for the electronic submission. Since the heavy liftings have been done for each individual

study with SDTM and ADaM development, preparation of electronic submission data package is mainly about finalizing documentations and assembling.

- Finalize documentations: acrf.pdf, SDRG, ADRG and other supportive documents (e.g. computation note)
- Split datasets larger than 5 GB
- Load all completed documents and SAS transport files to Pinnacle 21 system to generate final define.xml file with proper hyperlinks
- Review, and repeat steps above till results are satisfactory

The final data package for electronic submission is organized into folders mirror the directory structure proposed in Study Data Technical Conformance Guide (Display 7, Directory Structure for Electronic Submission)



Display 7, Directory Structure for Electronic Submission

Integrated analysis data standardized is like individual study. Assembled data packages will be transferred to regulatory operation for final posting.

CONCLUSION

Study data standardization for electronic submission is highly regulated. The Agency may Refuse To File (RTF) for NDAs and BLAs, or Refuse To Receive (RTR) for ANDAs an electronic submission that does not have study data in conformance to the required standards specified in the Catalog. There is tremendous amount of work for data management and programming. Early planning and developing with the end picture in mind is very important. Regular cross functional communications on topics such as Study Data Standardization Plan, data traceability, standard terminology, and possible integrated

analysis, etc., will help study team focus. Proper tool for data compliance validation and relevant documentation generation is critical.

Our current process and tool are efficient. But we never stop explore enhancement. For example, automate SDTM mapping, extend TA level standards, create master specifications to better facilitate integrated analysis.

REFERENCES

1. Guidance for Industry Providing Regulatory Submissions in Electronic Format — Certain Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications. (2014, July). Retrieved from http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm
2. Providing Regulatory Submissions in Electronic Format — Submissions Under Section 745A(a) of the Federal Food, Drug, and Cosmetic Act. (2014, December). Retrieved from http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm
3. Providing Regulatory Submissions In Electronic Format — Standardized Study Data. (2014, December). Retrieved from http://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm
4. STUDY DATA TECHNICAL CONFORMANCE GUIDE. (2015, March). Retrieved from http://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm
5. Data Standards Catalog v4.0. (Dec 17, 2014). Retrieved from http://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm
6. FDA Specific SDTM Validation Rules. (2014, November 13). Retrieved from http://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm

CONTACT INFORMATION

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

Qin Li
Regeneron Pharmaceuticals Inc.
Qin.Li@regeneron.com