PharmaSUG 2014 – DS14

CDISC Electronic Submission

Kevin Lee, Accenture Life Sciences, Berwyn, PA

ABSTRACT

FDA singed FDA Safety and Innovation Act (FDASIA) into law on July 9th, 2012 and announced Prescription Drug user Fee Act (PDUFA) V. And, FDA strongly recommended CDISC as electronic submission formats. The paper will introduce FDASIA, PDUFA V and Data Standard Strategy from FDA. Then, the paper will discuss what it means to programmers who prepare FDA submission. The paper will introduce where programmers can find FDA electronic submission guidelines and CDISC guidelines such as eCTD (electronic common technical document) specification, SDTM implementation guideline, ADaM implementation guideline and etc.

First, the paper will provide the brief introduction of regulatory electronic submission such as its methods, five modules in CTD especially m5, technical deficiencies in submission and etc. And, the paper will discuss what programmers need to prepare for the submission according to FDA and CDISC guidelines for CSR, Protocol, SAP, SDTM annotated eCRF, SDTM datasets, ADaM datasets, ADaM datasets SAS® programs and Define.xml.

And secondly, the paper will discuss how programmers can prepare the submitted materials - length, naming conventions and file formats of electronic files. For examples, SAS data sets should be submitted as SAS transport file formats and SAS programs should be submitted as text format. Finally, the paper will discuss the latest FDA concerns and issues about the electronic submission such as the size of SAS data sets, the length of character variables in SAS datasets, CDISC compliance checks and etc.

INTRODUCTION OF eCTD

The eCTD (electronic common technical document) is CDER/CBER's standard format for electronic regulatory submission. The eCTD can be sent two different methods: thru the submission on a CD-ROM or into the electronic submission gateway (ESG).

Five modules in eCTD according to comprehensive table of contents headings and hierarchy are

1. Module 1 – Administrative information
2. Module 2 – eCTD summary document
3. Module 3 – Quality
4. Module 4 – Non-clinical Study Reports
5. Module 5 – Clinical Study Reports.

Usually, the programmers will deal with module 5.

TECHNICAL DIFICIENCIES IN SUBMISSION

The followings could be the technical deficiencies in the submission

- Defect in the media
- An electronically non-readable from thru ESG
- A previously-submitted sequence number
- No index.xml
- A presence of a virus
- An incompatible file format

The technical deficiencies in the submission can lead to

- "Not Received" –Until the technical deficiencies are resolved, the submission is considered as not received. It will delay the review process.
- "Refuse to File" – If the technical deficiency is not resolved by 60 days after the submission, FDA will refuse to file

Refuse to file can also happen for the absence of electronic datasets, inadequate analysis and illegible, uninterpretable and inadequate submission.

CDISC COMPONENTS IN ELECTRONIC SUBMISSION

The followings are usual CDISC submission components.

- Protocol
- SAP
- eCRF
- SDTM

- ADaM
- SEND
- CSR
- Define.xml
- ADaM SAS programs
- Efficacy SAS programs (sometimes)

File formats

There are a couple of recommendations by FDA.

- PDF for reports and forms like CSR reports.
  1. FDA accepts PDF versions 1.4 to 1.7, and PDF files should be readable by Adobe Acrobat 8.0.
  2. The font size is recommended from 9 to 12 pints.
  3. There is no restriction of font type, but Arial, Courier New and Times New Roman are recommended.
  4. For naming files, one should use lower case characters and avoid a special characters such as hyphen, underscore, punctuation, spaces and non-alphanumeric variables (e.g., ? #$% < > +) .
  5. The print area of pages should be 11 by 8.5 inches and at least ¾ margins should be applied to all sides of pages. The titles and footnotes should be outside of margin
- ASCII Text file for SAS program files like c-dm.txt or c-dm.sas.txt
- XML for documents, data, and document information files like Define.xml
- XSL (Style sheets) and DTD (document type definition) for XML document

Name of Electronic file

The name of electronic files such as SAS programs and TFL names should follow the followings.

- The file name should be less than or equal to 64 characters including extension.
- It should contain only letters (lower case), numbers or hyphens.
- It should not contain blank.
- Its path cannot be longer than 230 characters.

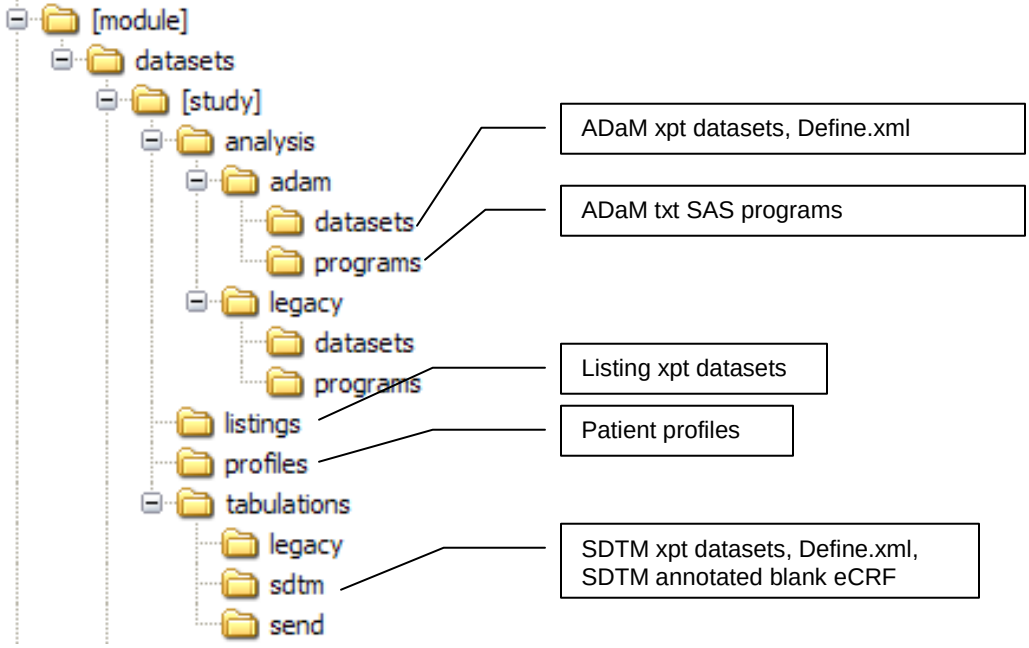
FORMATS OF ELECTRONIC FILES IN ELECTRONIC SUBMISSION

The followings components are usual CDISC submission.

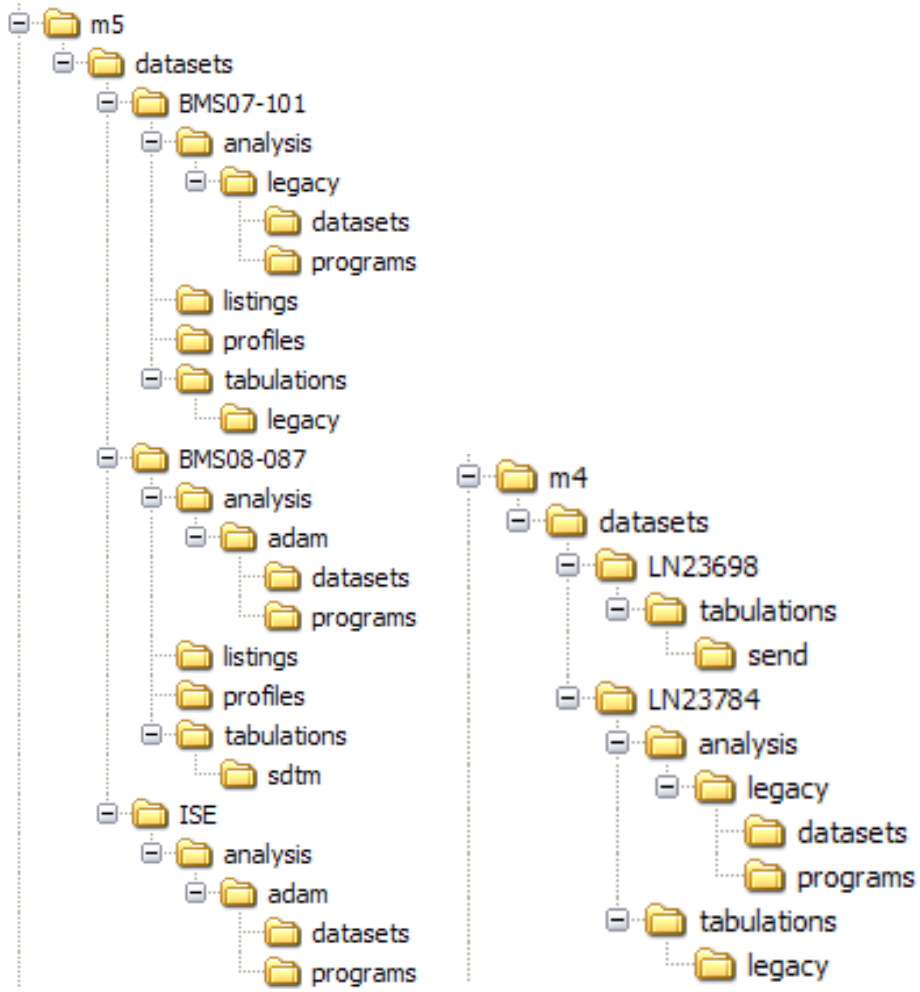
- Protocol – pdf (e.g., study001-protocol.pdf)
- SAP – pdf (e.g., sutdy001-sap.pdf)
- eCRF – pdf (e.g., sutdy001-blankecrf.pdf)
- SDTM – xpt (e.g., dm.xpt, ae.xpt and ds.xpt)
- ADaM – xpt (e.g., adsl.xpt, adae.xpt and adtteos.xpt)
- SEND – xpt (e.g., dm.xpt, se.xpt, and bw.xpt)
- CSR – pdf (e.g., sutdy001-csr.pdf)
- Define file – xml or pdf (e.g., define.xml/define.pdf)
- ADaM SAS programs – txt (e.g., c-adsl-sas.txt)
- Efficacy SAS programs – txt (e.g., t-14-01-001-ds-sas.txt)

ELECTRONIC FILE STRUCTURE in eCTD

eCTD Module File Structure



eCTD m4 and m5 Folder Structure



FDA CONCERNS AND ISSUES ON ELECTRONIC SUBMISSION

SAS Dataset

There are a couple of recommendations by FDA.

1. FDA accepts SAS xport (Version 5) transport file. Since we need to prepare SAS transport file, there are a couple of things to consider
    a. The length of variables should be less than or equal to 8.
    b. The length of variable labels should be less than or equal to 40.
    c. The length of dataset label should be less than or equal to 40.

2. The size of dataset is recommended to less than 1 GB. Especially, SDTM LB dataset can be too huge, so FDA recommended splitting LB into such as LB1, LB2 and LB3 with the same variable length.
3. The lengths of character variables are recommended to be reduced to the maximum length used. For example, we should reduce the length of ADSL.SUBJID to 20 if the maximum length used is 20, rather than keeping it to 200 long.
4. No special characters on datasets and variables.

Study Data Standardization Plan

The sponsor needs to discuss about Study Data Standardization Plan with FDA at pre-IND stage. For INDs, the Study Data Standardization Plan should be located in general investigational plan.

1. List of the planned studies
2. Type of studies
3. Study Design
4. Planned data standards, formats and terminologies and their revisions
5. List of and justification for studies that may not conform to the standards

Study Data Reviewer's Guide (SDTM and ADaM)

The Study Data Reviewer's Guide needs to be placed in each study of eCTD in Module 5. It should provide study data standards and their conformance validation of each study.

1. Study Protocol title, number, and version
2. Study Design
3. Standards, formats, and terminologies and their versions
4. Description of study datasets
5. Data Standards conformance validation rules, versions and issues.

SDTM

Below are FDA recommendations on Study Data Tabulation Model.

- SDTM Implementation Guide (SDTMIG) is main document
- USUBJID
    - Consistent thru entire application (studies and datasets)
    - No spaces
- SUBJID : consistent in study report
- DM
    - One single record for each subject
    - For screen failures, ARM is blank
    - If a subject is randomized, but not treated, ARM is "Planned Treatment" and ACTARM is blank.
- DS : EPOCH for more than one disposition event
- SE should be included
- AE
    - Include all AE records
    - If AESER ='Y', serious AE criteria (i.e., death and hospitalization) should be included.
- Custom domain needs to be informed to the reviewers prior to submission
- LB : submit both split data (LB1, LB2,, in../SPLIT) and original data(LB)
- TD should be included in SDTM submission

ADaM

- ADaM Implementation Guide (ADaMIG) is the main document
- General consideration
    - Traceability : ADaM should be from SDTM
    - ADaM programming codes should be submitted in txt or pdf (e.g., adsl_sas.txt)
    - Key Efficacy and Safety ADaM datasets should be submitted.
- Timing Variables

    - If multiple records per subject, timing variables (e.g., AVISIT, AVISITN, ADT and ATM) and its relative day (e.g., ADY) should be included.
    - Other timing variables beside AVISIT could be used mainly for plotting data. For example, Visit (week 8) and a corresponding numeric variables (8)
- Core Variables should be included in all ADaM datasets - study/protocol, center/site, region, country, treatment assignment, sex, age, race, analysis population flag and other important baseline.
- Numeric date formats
- For same domain, ADaM dataset label should be different from SDTM domain label.
- ADSL is required for each study.
- Imputed data should be documented in ADRG or define.xml

SEND

SEND follows SEND Implementation Guide and SDTM. SUPPQUAL and custom domain could be used as well.

Define

- Define.xml describes SDTM, SEND and ADaM data sets.
- Clear and accessible to Code list, origin and derivation for each variables
- Data definition specification for submitted datasets
- Version 1 : Define.xml and Define.pdf
- Version 2 and later: only Define.xml

aCRF

Annotated CRF should include treatment assignment forms and map all variables that are in submitted datasets.

Terminology

- Supported in Standard Catalog
- Sponsor-defined terminology in Define.xml or Data Reviewer Guide
- Verbatim in eCRF and coded term in submitted datasets
- Discourage the custom term, but if present, it should be documented in SDRG.
- Most current versions for each study, but for pooling coded data, a single version is recommended. If not, the justification should be documented in Standardization Plan.
- Terms and Versions for each study should be documented in SDRG.

CDISC Terminology

- CT : from NCI
- AE : MeDRA
- Medication : FDA UNII (Unique Ingredient Identifier)
- Pharmacologic Class : Veterans Administration's NDF-RT (National Drug File – Reference Terminology)
- Indication : SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms)

Data Fitness

FDA recommends three Data Fitness focus.

1. Conformance to supported data standards : FDA will post common conformance errors
2. Study Data Validation Rules
    - Ensure that submitted data are both compliant and useful.
    - There are two Data Validation Rules
        - Conformance validation rule : submitted data conform to standards.
        - Quality Checks : submitted data support meaningful analysis.
3. Study Data Traceability
    - Relationship between analysis results, analysis datasets, SDTM datasets and source data.
    - Maybe legacy data in submission
    - Common Issues involving traceability
        - Legacy data conversion to SDTM only
            - No traceable path from legacy analysis data to SDTM.
            - No ability to confirm analysis variable imputation or derived variables.
            - Unable to replicate tables, listings and figures (TLFs) and legacy analysis datasets using SDTM datasets.
            - No ability to confirm derivation of intermediate datasets or custom domains.
            - No ability to determine location of collected CRF variables in the converted SDTM data.
            - Difficulty in understanding the source or derivation methods for imputed or derived variables in integrated/pooled data, supplemental qualifiers, and related records.

- Independent legacy data conversion to SDTM and ADaM
  - No traceable path from legacy to SDTM to ADaM and to Study Report.
  - No explanation or source for analysis imputed or derived variables.
  - No traceable path to ISS and ISE / pooled data.
  - TLFs do not match datasets (analysis datasets or SDTM datasets (when used)).
  - No traceable path to intermediate datasets or custom domains.
  - No explanation or source for imputed or derived variables in datasets.
- Legacy data conversion to SDTM and ADaM in sequence
  - May not be able to replicate the results in the TLFs and CSR using the ADaM or the SDTM datasets.

– FDA recommendation
- Submit a Legacy Data Conversion Plan and Report
- aCRF for legacy data
- Documentation in SDRG
- Submission of Legacy data

CONCLUSION

The use of the word of "SHOULD" in the regulatory documents did not mean required, but suggested or recommended. However, some of future FDA documents such as FDA Guidance for Industry for Electronic Formats will have a binding effect on submission. So, some of standard (e.g., eCTD, SDTM 3.1.2, SDTM 3.1.3 and ADaM 2.1) will be "REQUIRED", rather than recommended. The FDA Data Standard Catalog (DSC) will provide the support date, support end date, requirement date of each standard. Once standards were required in FDA DSC, sponsors can plan their submission and submit appropriate standards to FDA.

REFERENCES

Study Data Technical Conformance Guide (Draft)
Guidance for Industry : Providing Regulatory Submissions in Electronic Format – Submissions Under Section 745A (a) of the Federal Food Drug, and Cosmetic Act (Draft)
Guidance for Industry : Providing Regulatory Submissions in Electronic Format – Standardized Study Data (Draft)
The ADaM Implementation Guide, Version V 1.0 (ADaMIG v1.0)
The Analysis Data Model, Version 2.1 (ADaM 2.1)
The ADaM Basic Data Structure for Time-to-Event Analyses
Study Date Specification (v2.0) in
FDA PDF Specification v3.1
Providing Regulatory Submission in Electronic Format – Receipt Date
Providing Regulatory Submissions in Electronic Format – Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications Revision 2

CONTACT INFORMATION

Your comments and questions are valued and welcomed. Please contact the author at

Kevin Lee
Accenture Life Sciences
Berwyn, PA
(610) 407 - 1767
Email:Kevin.s.lee@accenture.com

TRADEMARKS