PharmaSUG 2019 - Paper AP157

Practices in CDISC End-to-End Streamlined Data Processing

Chengxin Li, Luye Pharma Group, Princeton, NJ, USA

ABSTRACT

From a programming perspective, principles and practices in end-to-end streamlined data processing are described under the CDISC umbrella. Besides compliance, there are several common practices across the clinical data processing lifecycle: traceability, controlled terminology, end-in-mind philosophy, structured design, and reusable solutions. The components of end-to-end streamlined data processing are also introduced: data collection, SDTM transformation, ADaM development, and TLF generation. The ISS/ISE programming model, MDR, and production harmonized with submission are depicted as well. To illustrate concepts, two examples are discussed, one on a specific data element (AE start date), and the other on efficacy analyses in pain therapeutic areas. The end-to-end streamlined data processing with CDISC is the optimized programming model to achieve high-quality and efficient deliverables.

INTRODUCTION

The Clinical Data Interchange Standards Consortium (CDISC)[1] has defined a series of data models from data collection to submission. Clinical Data Acquisition Standards Harmonization (CDASH) is a data collection standard that is harmonized with the Study Data Tabulation Model (SDTM). SDTM standardizes the collected data in tabulations and should fully reflect the collected data (e.g., mapping for any collected data and deriving a limited number of variables, but no imputation for missing data). Furthermore, the Analysis Data Model (ADaM) should only be derived from SDTM. The key endpoint analyses, inferential analyses, and complicated analyses should be designed with ADaM datasets.

The CDISC standards are widely implemented in industries complying with CDISC standards and conforming to agency requirements. With CDISC, it is a mission to achieve end-to-end streamlined data processing from data collection to analysis. From a programming perspective, the best practices are described in below sections with CDISC end-to-end streamlined data processing.

The Protocol Representative Model (PRM) will be very helpful and impacting in achieving end-to-end streamlined data processing. The discussion on PRM is beyond the scope of this paper.

COMMON PRACTICES

CDISC compliance becomes focused on CDISC implementations. The common tool used for compliance checks is the open source Pinnacle 21 Community Validator[2], which incorporates agency conformance rules. The interpretations to the output report from the tool are critical during the conduct of the study and submission. The unresolved issues, with explanations, should be stated in the reviewer's guides.

[1] http://www.cdisc.org

[2] https://www.pinnacle21.net; previously named as OpenCDISC Validator

Along with compliance with CDISC, there are other common principles across the data processing lifecycle: traceability, controlled terminology, end-in-mind philosophy, structured design, and reusable solution.

TRACEABILITY

To achieve data processing transparent and facilitate review, there are strict guidelines in place regarding traceability within CDISC data standards. With respect to traceability designed in CDISC, SDTM presents as the foundation. The CDASH design is harmonized with SDTM. The ADaM standard facilitates the traceability to SDTM. ADaM defines metadata traceability and data point traceability clarifying how the analysis datasets were created to assist review. It requires as many supportive SDTM variables from the SDTM domains as needed to facilitate the transparency and clarity of the derivations and analysis for statistical reviewers. CDISC with end-to-end streamlined implementations from data collection to SDTM, from SDTM to ADaM, and from ADaM to display assures traceability from the display back to ADaM, to SDTM, and further back to the collected raw data.

CONTROLLED TERMINOLOGY

The controlled terminology (CT) adoption is the critical step to achieve standardization and semantic interoperability in CDISC. The production terminology is published by the National Cancer Institute's Enterprise Vocabulary Services (NCIEVS) [3].

In CDASH, it's acceptable to use a subset of SDTM CT. It is a good practice to list the study specifically controlled terminologies in the corresponding fields of case report form (CRF). For non-CRF data with data transfer agreement (DTA), CT compliance can facilitate SDTM transformation. ADaM may carry over SDTM CT. And the CT can be directly used or further categorized in the displays with friendly readings (e.g., possibly using CT-defined NCI preferred terms in tables). In this way, the traceability is assured from the display possibly back to the collected raw data.

The CT is expressed in upper case text except: (a) when the CT is from an external reference, the CT must match the case text of the external reference (e.g., MedDRA); (b) units (e.g., mg/dL); (c) test names (e.g., LBTEST='Glucose').

The CT is designated as extensible or non-extensible. For an extensible code list, an existing process needs to be followed to add terms in the CT standard.

END-IN-MIND PHILOSOPHY

The CDASH User Guide gives recommendations on CRF design to common domains. However, multiplicity is widely presented in CRF design. Even though SDTM is a very rigid standard to tabulate the collected data, the variation also exists in implementation. Comparatively, the ADaM design is much more flexible.

The ADaM design is an art, not a science. However, the design should go with the "end-in-mind" philosophy. The "end-in-mind" philosophy refers to the submission orientation with regulatory agency's requirements in mind. The regulatory agency is the customer; therefore, the design needs to meet the agency requirements and facilitate agency reviews. With the "end-in-mind" philosophy, the best solution can be often determined among multiple options.

[3] https://www.cancer.gov/research/resources/terminology/cdisc

The agency requirements are transparent and available in the public domain. The Pinnacle 21 Community validator fully reflects the agency's compliance rules. The Study Data Technical Conformance Guide[4] provides detailed recommendations on CDISC practices as well.

STRUCTURED DESIGN

The structured design method divides a complex task or concept into several simpler modules (steps) and then inter-relating those modules (steps). The CDISC is designed with layers and the layers are interrelated by traceability. The layers of data collection, tabulation, analysis dataset, and display are developed with streamlined data flow. Within each layer, the implementation can be further developed with sequential manageable steps. The structured design method makes the CDISC end-to-end streamlined implementation less complexity, easy to understand, maintain, and use.

REUSABLE SOLUTION

The reusable solution can achieve programming in high efficiency and high quality. Any process can be automated, i.e., the solution reusable, if both input and output are standardized. This actually provides working direction in production by checking and standardizing both input and output to achieve a reusable solution.

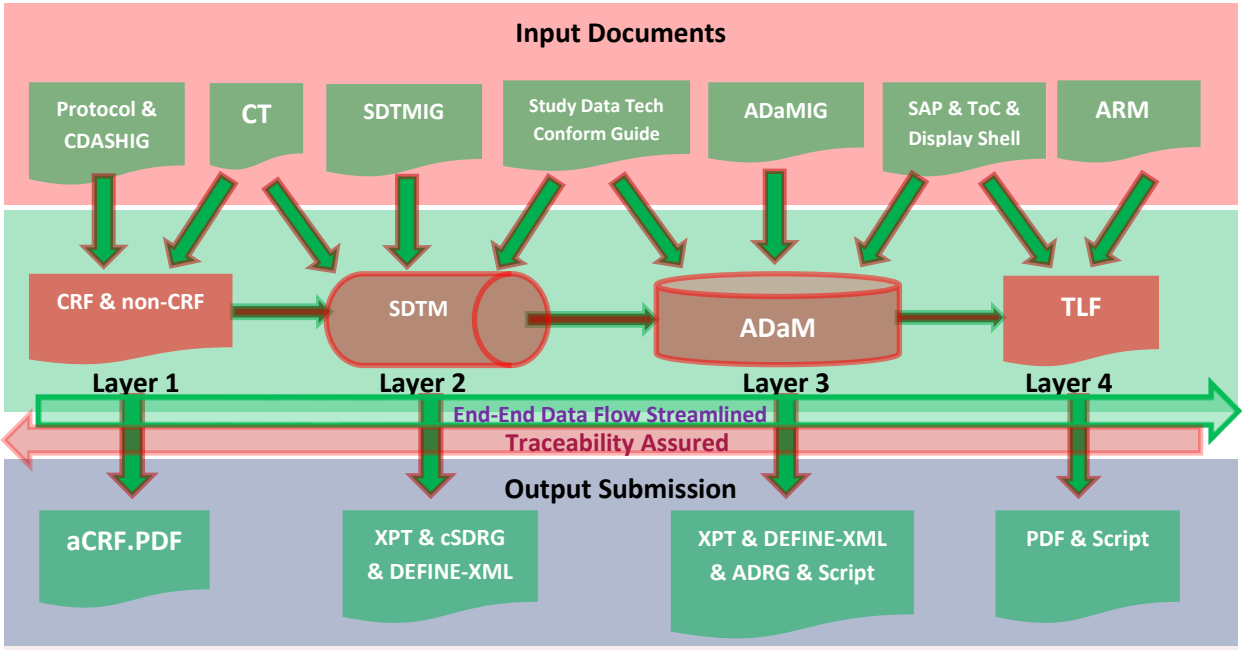
With traceability development, CDISC provides streamlined standard data models for data collection, tabulation, and analysis. By CDISC end-to-end streamlined data flow, it's feasible and practical to work out a reusable solution to automate the data processing. The CDISC compliance is very important but not all. In production, besides the solution analysis usable, it can further go to reusable. The CDISC standard data structures and metadata lead to standard programs and further result in reusable solution.

To achieve the reusable solution, technically, the **G**lobal/**P**roject/**S**tudy (GPS) navigation method can be applied to most developments. ADSL development is used as an example to illustrate this method. The multiplicity of information in ADSL requires multiple domains as the sources. However, coming into implementation, the common variables based on the common domains can be formalized as "**g**lobal" variables across all the studies, thus specified, derived, and validated only once but used across all the studies. The other ADSL variables such as indication or study specific baselines or covariates can be designated as "**p**roject" or "**s**tudy". In production, after global variables are generated with a global macro call, the study programming team focuses on study specific add-ons. Approximately 80% of ADSL variables can be targeted as "global". The GPS navigation method has also conformed with the universal '80/20' rule.

STREAMLINED DATA PROCESSING

With the above common practices in implementing CDISC, Figure 1 depicts the end-to-end streamlined data processing architecture. The process flow lies out with streamlined sequential layers, from data collection (layer 1), to data tabulation (layer 2), then to analysis dataset (layer 3), and further to displays (layer 4). Each layer is supported by corresponding CDISC standard(s) or study documents. Under this architecture, reverse traceability is assured from any display (e.g., a p-value in a table), back to the ADaM analysis dataset, to the tabulated SDTM dataset, and further back to the collected data. Below, several subsections detail the processes.

[4] http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM384744.pdf



CDASHIG=Clinical Data Acquisition Standards Harmonization Implementation Guide; **CT**=Controlled Terminology; **SDTMIG**=Study Data Tabulation Model Implementation Guide; **ADaMIG**=Analysis Data Model Implementation Guide; **SAP**=Statistical Analysis Plan; **ToC**=Table of Contents; **ARM**=ADaM Analysis Results Metadata; **aCRF**=Annotated Case Report Form; **TLF**=Table, Listing, and Figure; **XPT**=SAS Transport Version 5 Format File; c**SDRG**=clinical Study Data Reviewer's Guide; **ADRG**=Analysis Data Reviewer's Guide; **PDF**=Portable Document Format; **XML**=eXtensible Markup Language

Figure 1 End-to-end Streamlined Data Processing Architecture

DATA COLLECTION

The data processing is initiated with data collection with CRF. CRF design, supporting the adequate assessment of safety and efficacy, is driven by the clinical study protocol. The non-CRF data, specified with a predefined data structure and data transfer format, is another mechanism of data collection with electronic transfer of data (eTD) from a vendor.

The CDASH, which is harmonized with SDTM and incorporating agency regulations such as ICH guidelines or FDA guidance, provides the best practice recommendations on CRF design for common domains.

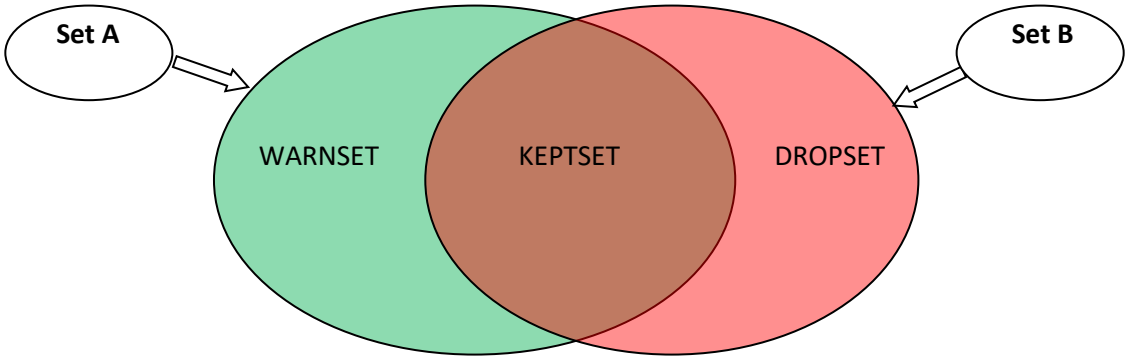
The aCRF in PDF format is a submission required document facilitating reviews with SDTM-compliant annotations. In production, besides servicing traceability from SDTM to data collection, the aCRF also provides instructions to the clinical database setup and SDTM transformations.

SDTM TRANSFORMATION

The collected data is stored in a data warehouse with a specific data management system. The database with well-designed structure and metadata can greatly facilitate SDTM transformations.

SDTM is not a data collection standard but standardizing the collected data with tabulations. Therefore, SDTM should fully reflect the collected data without any imputations. SDTM goes with CHAR type associated with CTs for text values or ISO8601 format for timing values. SDTM is a rigid standard requiring compliance with SDTM Implementation Guide and conformance to agency technical conformance guide as well.

Standardized CRF design and further formalized database setup can facilitate automation of SDTM transformations, optimally by sub-setting the database and further deriving a limited number of variables. To generate a specific SDTM domain, the sequential steps would be 1). setting/merging one or more input datasets; 2). deriving domain specific variables; 3). deriving domain common variables; 4). aligning metadata with standard (ordering variables, labeling, optimizing variable length, etc.). It is a challenge in production to fully and properly map the collected data to SDTM in a generic and reusable solution and incorporating the validation process. However, to develop such a solution, it will be helpful to keep the concepts clear of KEPTSET/WARNSET/DROPSET as illustrated below in Figure 2.



Set A = all the variables of a specific domain defined in SDTMIG; Set B = all the variables populated in the raw dataset for the same domain. KEPTSET (brown) = Set A and Set B; KEPTSET is the set of all shared variables by Set A and B and will be populated in the final SDTM domain, including SUPP (supplemental) domain or CO (comment) domain. DROPSET (pink) = Set B and not Set A; DROPSET is the set of all the variables which will be dropped from the raw dataset(s), including CDASH specific variables and administration variables. WARNSET (green) = Set A and not Set B; WARNSET is the set of all the variables defined in SDTMIG but not in raw dataset(s)

Figure 2 Relationships between Standard Set and Collected Set

For SDTM submission, XPT file format is required. The specification for SDTM transformation goes with define-xml. The corresponding SDTM review guide is named Study Data Reviewer's Guide, specifically cSDRG for clinical, and nSDRG for non-clinical. The DEFINE-XML and cSDRG (or nSDRG) are required to submit as well.

ADaM DEVELOPMENT

According to end-to-end streamlined data processing, ADaM should be derived from SDTM. ADaM datasets fully support analyses defined in SAP. Traceability and analysis-ready concepts are the two core features of ADaM development.

There are several dataset structures defined in the ADaM Standards: Subject-Level Analysis Dataset (ADSL), Basic Dataset Structure (BDS), and Occurrence Dataset Structure (OCCDS). Subject evaluation tables such as demographic, disposition, baseline, overall exposure, and overall compliance can go with ADSL. Safety analysis datasets such as using a BDS for Laboratory analyses or an OCCDS for Adverse Events tend to be more straightforward in sense of development. The efficacy dataset (ADEFFOUT) is often developed with the BDS structure and presents challenges to programmers.

However, with an appropriate design approach such as the structured design method, it is possible to relieve the challenges regarding efficacy dataset development. Below, Figure 3 introduces one two-layer ADaM design approach with the structured design method for efficacy endpoints analyses.

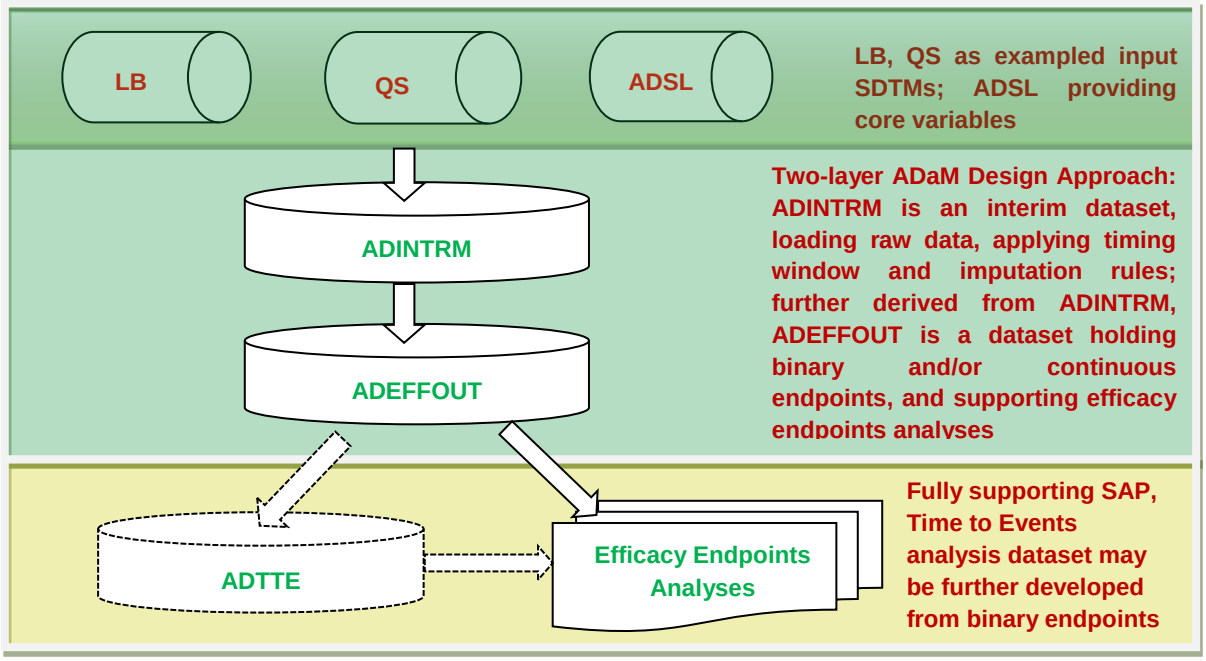


Figure 3 Architecture of Two-Layer Efficacy Datasets

The efficacy ADaM development is divided into two layers: interim dataset and endpoint dataset. Under the two-layer efficacy ADaM design architecture, sequential manageable steps are further developed within each individual layer, and each step may possibly be macrotized. With this approach, in the design stage, it's a breaking down process; in implementation, the dataset generation becomes assembling and polishing (metadata alignment) processes.

For ADaM submissions, XPT file format is required. The specification for ADaM goes with define-xml. The corresponding ADaM review guide is named Analysis Data Reviewer's Guide (ADRG). The DEFINE-XML and ADRG are required to submit. The programming codes in text format are recommended to be submitted as well.

TLF GENERATION

Any core tables and figures, (e.g., primary/secondary endpoint analyses, inferential analyses, and complicated analyses) should be developed with ADaM datasets. However, for a simple summary table, it can be developed with the programming model of SDTM plus ADSL.

ADaM has defined analysis results metadata (ARM), describing the main attributes of the analysis results and providing the link between analysis results and analysis datasets to facilitate reviews. The ARM functions as specifications for an analysis display. There are associations between the TLF programming flow and fields of ARM as depicted below in Figure 4. With this association, the generation for a table or a figure can be driven by the corresponding ARM, thus achieving reusable codes. For instance, one SAS® macro can produce multiple tables of the same type by changing macro parameters aligned with the fields of ARM.

There are two prerequisites to achieve high reuse of programming codes in generating displays: ADaM dataset designed with analysis-ready, and display template featured with structure-stable. As the analysis-ready is required in an ADaM development, there should not have to be any data processing between data

subsetting and the statistical procedure invoking. The display template provides instructions for formatting and outputting the statistical data. Therefore, normalizing the display template can facilitate display auto-generations.

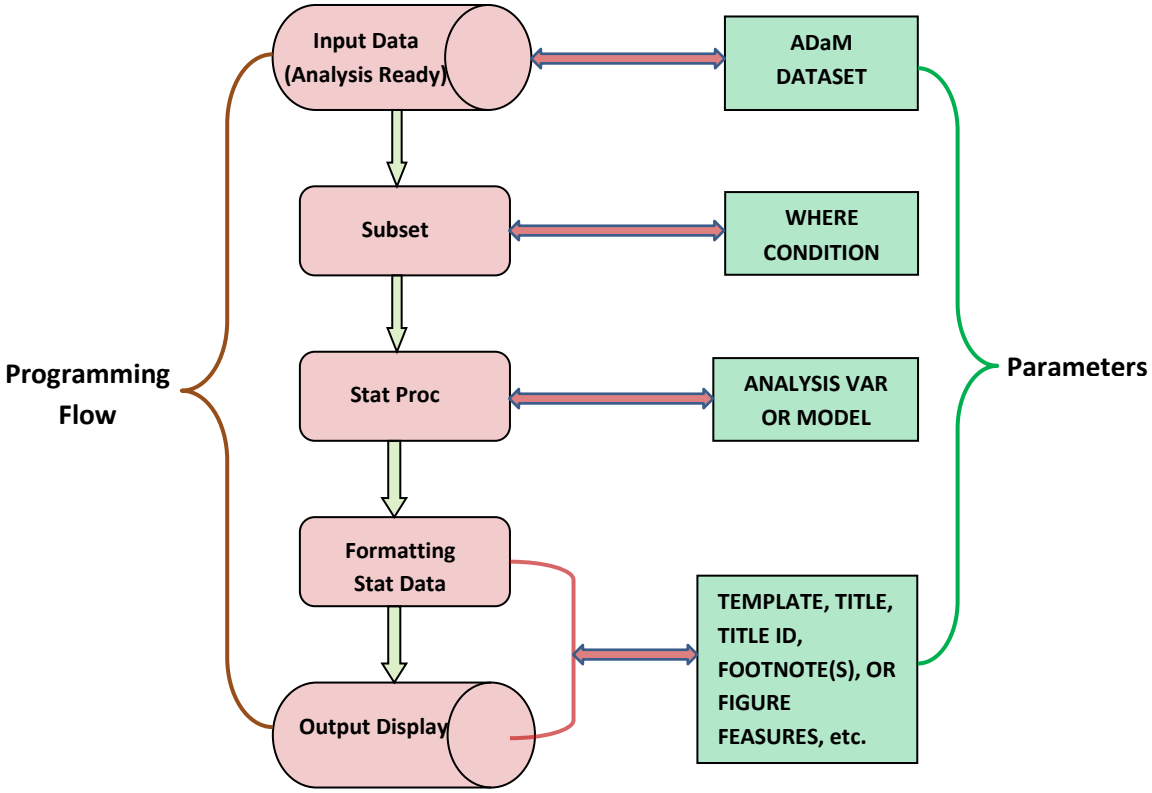


Figure 4 Programming Process Associated with ARM

There are multiple options to manage ARM for TLF auto-generations such as embedded in a reporting system, consolidated in the Table of Contents (ToC), or explicitly with macro parameters. Based on the parameters aligned with the ARM, the corresponding DEFINE-XML of ARM can be auto-generated with high efficiency. To facilitate agency reviews, along with the programming codes, the ARM is recommended to submit with DEFINE-XML, especially for key endpoints analyses.

INTEGRATED ANALYSES

In an integrated summary of safety (ISS) or an integrated summary of efficacy (ISE), fully supporting integrated SAP, there can be two basic models to generate integrated ADaMs. One option is that, based on individual study SDTM domains (from CDISC studies or converted legacy studies), project SDTM database is built up; and further based on the project SDTM, the ADaM ISS or ISE datasets are derived correspondingly. Another option is, based on individual study ADaM datasets, ADaM ISS or ISE datasets are further derived. Figure 5 depicts the two basic models to produce the ADaM datasets for pooled analyses.

Choosing which model to use in production can be determined on a case-by-case basis. For instance, if the ISE is only comprised of two sister-pivotal studies, the ISE ADaM datasets may be produced by setting the two individual ADaM datasets together as the metadata of two datasets are normally very close. As

another example, if ISS AE analyses are across different clinical phases with different MedDRA dictionary versions (e.g., across legacy phase II studies and CDISC phase III studies), the ADAE may be developed on project SDTM AE domain. Regardless, the pooled SDTM model is universal and preferred especially to pooled AE analyses, with which simplifies the development in most cases.

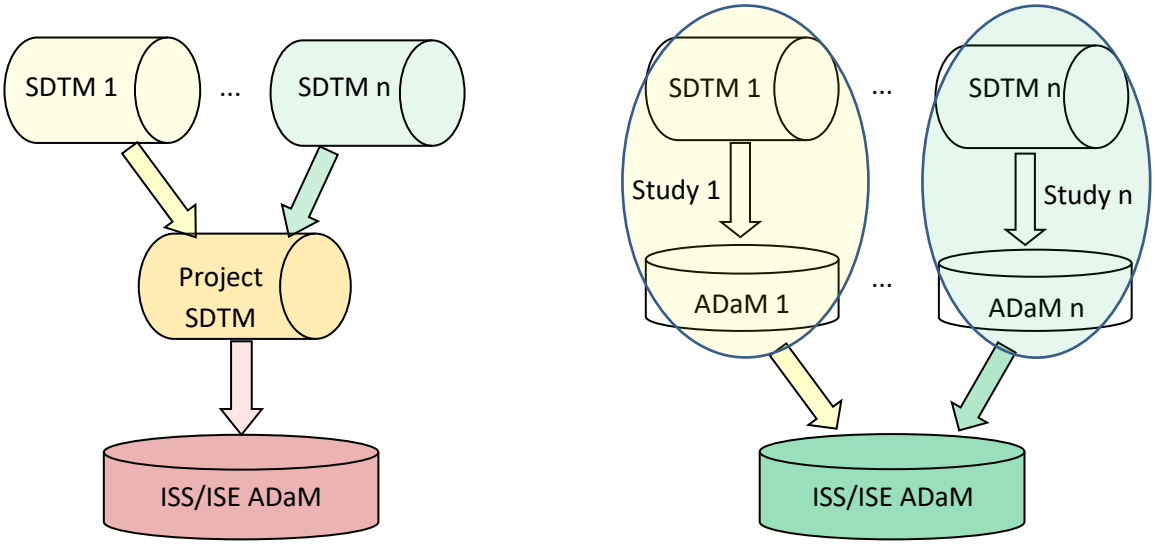


Figure 5 Programming Models for ISS/ISE Analysis

PRODUCTION HARMONIZED WITH SUBMISSION

Documentation is always the core component in data processing and submission. It seems very common that submission documentations become post-processing to conform to the agency or regulatory requirements. However, the post-processing isolates the submission from production and may result in heavy workloads like, extra validation efforts on any deviations between submission and production. With the “end-in-mind” philosophy and end-to-end streamlined data processing, it’s possible that the production are harmonized with submissions by incorporating the agency requirements into production process, thus achieving submission ready documents or documents that require minor changes after the programming deliveries. In CDISC, the FDA Study Data Technical Conformance Guide gives detailed recommendations on the submission requirements and CDISC implementation practices.

In submission outputs illustrated in Figure 1, there are several programming related documents correlated to production: aCRF, SDRG and SDTM define-xml, ADRG and ADaM define-xml including ARM define-xml.

The aCRF can facilitate reviews and instruct SDTM transformations. The best practice is to create the aCRF during the study initiation other than with post-annotations. SDRG and the define-xml template of SDTM specification (e.g., in excel sheet) are directly used as production-live documents. The compliances are pre-checked, and the issues are resolved with best efforts during the conduct of study. Any unresolved issues are documented and updated in SDRG correspondingly. With this process, the SDTM implementation and aCRF are consistent. Meanwhile, SDRG used as a production document is ready for submission after database lock and programming deliveries.

Similarly, ADRG and define-xml template of ADaM specification are directly used as production-live documents for ADaM development as well. After ADaM is finalized, the ADaM define-xml is directly converted from the specification and is ready for submission. The efficacy ADaM design architecture, as

illustrated in Figure 3, may be put into ADRG to facilitate review. The ARM can facilitate reviews and drive display (table or figure) generations. After the display is finalized, the ARM of the selected core tables is embedded into ADaM define-xml and ready for submission.

Optionally, programming folder setup can be aligned with Electronic Common Technical Documentation (eCTD) module 5 dataset structure.

Production aligned with submission in programming achieves submission ready after programming deliveries, with which is one core component of CDISC end-to-end streamlined implementation.

METADATA REPOSITORY (MDR)

Metadata management, including CTs, is another critical step to achieve CDISC compliance, semantic interoperability, and high efficiency.

Traditionally, the sponsor maintains metadata by using Excel as the tool. This results in high costs to develop, implement, and maintain. Therefore, the MDR management system has been developing in the industry.

With the MDR management system, production standards can be easily kept abreast of the adopted CDISC version. One example would be directly downloading CDISC standards from SHARE or latest CTs from NCIEVS. In production, the MDR management system may also facilitate CDISC implementations in electronic CRF (eCRF) or aCRF creation, study database setup, SDTM compliance, and ADaM TA/global consistency, etc. It's expected MDR system will be widely deployed and adopted in the industry in the future.

USE CASES

EXAMPLE 1 – SREAMLINED TIMING VARIABLE IN ADVERSE EVENT ANALYSES

The example gives one CDISC-compliant implementation to illustrate the streamlined data processing and correspondingly assured traceability and reusability.

CDISC data models define "core" designations for variables as summarized in Figure 6. Figure 6 shows that CDASH "core" is designed for field(s) on a CRF, SDTM "core" is designed for submitted values, and ADaM "core" regards variable(s) in dataset.

There are logical connections between designations and streamlined data processing from CDASH to SDTM and then to ADaM. For instance, AESTDAT (Start Date of Adverse Event) is highly recommended (HR) in CDASH indicating this field must be presented in the AE CRF; however, still possibly not entered. When mapping to SDTM AE, the corresponding derived AESTDTC (consolidating date and time) is expected (Exp) showing that AESTDTC should appear in a SDTM AE dataset and may contain a NULL value for any AE start date that is not entered. In ADaM ADAE, the corresponding ASTDT (Analysis Start Date, converted from AESTDTC) is conditionally required (Cond), required if the start date is pertinent to the analysis. This pertinent may come from the interests of deriving ASTDY (Analysis Start Relative Day), ADURN (AE Duration (N)), APHASE (Phase),TRTEMFL (Treatment Emergent Analysis Flag), PREFL (Pre-treatment Flag), or FUPFL (Follow-up Flag). After data cleaning with best efforts, imputation rules defined in SAP may be applied to ADAE.ASTDT for the still missing start date. The corresponding imputations are

differentiated with ASTDTF (Analysis Start Date Imputation Flag) conforming to controlled terminology DATEFL with 'D' (day missing), or 'M' (month missing), or 'Y' (year missing or completely missing). After applying the data selection criteria associated with analysis results metadata, the specific summary of adverse event frequency table can be produced. For instance, the summary table of treatment emergent adverse events can be generated to count the number of distinct subjects by TRTA, AESOC, AEDECOD with WHERE=(TRTEMFL='Y' and APHASE='TREATMENT').

Standard	Core	Comments
CDASH	Highly Recommended (HR)	Regulatory requirement, field(s) must be on the CRF.
	Recommended/Conditional (**R/C**)	Conditionally required on CRF, e.g., date of birth.
	Optional (**O**)	On CRF as needed.
SDTM	Required (**Req**)	Values must be present, null value not allowed.
	Expected (**Exp**)	Must be mapped, null allowed
	Permissible (**Perm**)	Should be mapped and submitted if collected or derivable or agency expected, null value allowed
ADaM	Required (**Req**)	Must be included in the dataset, null allowed
	Conditionally required (**Cond**)	Must be included in the dataset in certain circumstances
	Permissible (**Perm**)	May be included in the dataset, but not required

Figure 6 Core Designations for CDASH, SDTM, and ADaM

For the summary table of treatment-emergent adverse events (TEAE), a statistic can be traced back to ADAE, checking which subjects with the events, new events or worsen events, any missing adverse event start dates, and what type of imputations on the missing dates. It can be further traced back to SDTM.AE and aCRF.AE page on how the raw data was collected and mapped. To support traceability and facilitate review, SDTM.AE variables related to derivation/imputation (e.g., AESTDTC) are recommended to be listed in ADAE. Therefore, to the reviewer, it will be sufficient to only check the ADAE for traceability with such implementation.

The above streamlined AE analysis process is further depicted in below Figure 7.

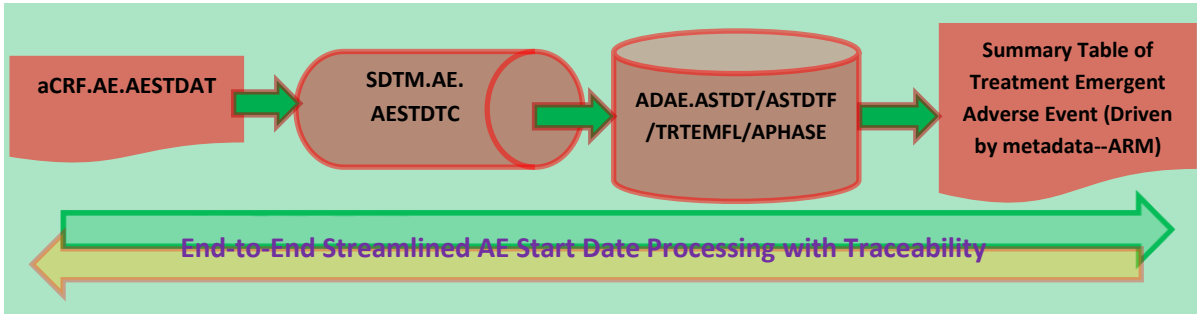


Figure 7 End-to-End Streamlined AE Start Date Processing and with TEAE Analysis

Adverse event analyses are often straightforward and normalized. The CRF for AE collection, SDTM mapping for AE, and ADAE including imputation rules can also be standardized, which lead to standard

programs for SDTM AE transformation and ADAE generation. If further stabilizing the AE table templates, driven by ADaM.ARM, the programs for AE summary tables can be standardized as well. Therefore, the reuse of codes or auto-generation for adverse event analyses is achieved across all studies. A similar approach can be applied to other safety analyses.

With the end-to-end streamlined CDISC-compliant implementation, the data processing is kept transparent. This approach provides programming with a major benefit: the reuse of codes for high efficiency and high-quality deliverables.

EXAMPLE 2 – STREAMLINED EFFICACY ANALYSIS IN PAIN THERAPEUTIC AREA

Therapeutic area (TA) CDISC is developing very fast. So far nearly 30 TA standards are available for use. From a CDISC application perspective, for a therapeutic area, the presented challenge is often on the mapping solution to efficacy SDTM with controlled terminologies. The example introduces one CDISC implementation in acute pain TA for efficacy analyses with the end-to-end streamlined data processing approach.

For pain efficacy assessments, there may be three concepts collected in the CRF: pain intensity (PI), pain relief (PR), and general clinical global impressions (GCGI). The collected data are mapped to SDTM findings questionnaire (QS) domain with the SDTMIG version 3.2, QRS (Questionnaires, Ratings and Scales) [5], and Pain Therapeutic Area User Guide v1.1[6] . The compliances are checked with Pinnacle 21 validator. The aCRF is skipped here; Instead, the mapping solution to QSCAT/QSTEST/QSTESTCD with project consistency in pain TA is provided below in Figure 8.

1	QSCAT	QSTESTCD	QSTEST	QSORRES	QSSTRESN	Note
2	PI	PI0104	PI0104-Pain Intensity	None	0	4-point pain severity
3	PI	PI0104	PI0104-Pain Intensity	Mild	1	4-point pain severity
4	PI	PI0104	PI0104-Pain Intensity	Moderate	2	4-point pain severity
5	PI	PI0104	PI0104-Pain Intensity	Severe	3	4-point pain severity
18	PR	PR010105	PR010105-Pain Relief	None	0	5-point pain relief
19	PR	PR010105	PR010105-Pain Relief	A Little	1	5-point pain relief
20	PR	PR010105	PR010105-Pain Relief	Some	2	5-point pain relief
21	PR	PR010105	PR010105-Pain Relief	A Lot	3	5-point pain relief
22	PR	PR010105	PR010105-Pain Relief	Complete	4	5-point pain relief
23	PR	PR0109	PR0109-Time to Perceptible Pain Relief	PTnHnMnS	nn	QSORRES in ISO8601 format; seconds converted for QSSTRESN
24	PR	PR0111	PR0111-Time to Meaningful Pain Relief	PTnHnMnS	nn	QSORRES in ISO8601 format; seconds converted for QSSTRESN
25	GCGI	GCGI0105	GCGI0105-Global Rating of Pain Mediation	Very Poor	0	6-point global evaluation; GCGI-General Clinical Global Impressions
26	GCGI	GCGI0105	GCGI0105-Global Rating of Pain Mediation	Poor	1	6-point global evaluation; GCGI-General Clinical Global Impressions
27	GCGI	GCGI0105	GCGI0105-Global Rating of Pain Mediation	Fair	2	6-point global evaluation; GCGI-General Clinical Global Impressions
28	GCGI	GCGI0105	GCGI0105-Global Rating of Pain Mediation	Good	3	6-point global evaluation; GCGI-General Clinical Global Impressions
29	GCGI	GCGI0105	GCGI0105-Global Rating of Pain Mediation	Very Good	4	6-point global evaluation; GCGI-General Clinical Global Impressions
30	GCGI	GCGI0105	GCGI0105-Global Rating of Pain Mediation	Excellent	5	6-point global evaluation; GCGI-General Clinical Global Impressions

Figure 8 Mapping Solutions for Acute Pain Study

[5] https://www.cdisc.org/foundational/qrs

[6] https://www.cdisc.org/standards/therapeutic-areas/pain

With the above mapping solutions, Figure 9 provides a simulated QS dataset with selected variables for the collected efficacy assessment data.

	QSORRES	QSSTRESC	QSSTRESN	QSSTRESU	EPOCH	QSBLFL	QSDTC	QSTPT	QSTPTREF
1	Severe	3	3		RUN-IN	Y	2000-02-14T10:50	BASELINE	
2	None	0	0		TREATMENT		2000-02-14T10:56	5 MIN	FIRST DOSE
3	Complete	4	4		TREATMENT		2000-02-14T10:56	5 MIN	FIRST DOSE
4	PT1M4S	64	64	SECONDS	TREATMENT		2000-02-14T10:56	5 MIN	FIRST DOSE

	STUDYID	DOMAIN	USUBJID	QSSEQ	QSTESTCD	QSTEST	QSCAT	QSSCAT
1	STUDYX	QS	STUDYX-P0001	1	PI0104	PI0104-Pain Intensity	PI	4-Point Categorical Pain Intensity Rating Scale
2	STUDYX	QS	STUDYX-P0001	2	PI0104	PI0104-Pain Intensity	PI	4-Point Categorical Pain Intensity Rating Scale
3	STUDYX	QS	STUDYX-P0001	3	PR010105	PR010105-Pain Relief	PR	5-Point Categorical Pain Relief Rating Scale
4	STUDYX	QS	STUDYX-P0001	4	PR0109	PR0109-Time to Perceptible Pain Relief	PR	

Figure 9 Example of QS in Pain Therapeutic Area

Figure 9 shows the collected data with one subject mapped to the QS domain using standard controlled terminology for QSTESTCD, QSTEST, and QSCAT. QSSCAT uses sponsor defined extensible terms consistently in the pain TA. The CRF collected Time to Perceptible Pain Relief is presented in QSORRES with ISO8601 format (line 4 with PT1M4S) and further transformed into numeric value in seconds (line 4 with 64 SECONDS). The planned assessment time point is assumed 5 minutes with a fixed reference point of the first dose.

With two-layer efficacy ADaM design and implementation method, efficacy ADaM datasets can be developed based on SDTM.QS and ADaM.ADSL. Further, based on efficacy analysis dataset, TLF can be generated conforming to analysis results metadata. One implementation on ARM, which is consolidated in ToC to achieve table auto-generations, is presented in figure 10.

ADaM	PARAM	PARAMCD	AnalVar	WHERE	ProgStat	Shell	ProgName
ADaM.ADTEST	Gastri PH4	GASPH4	PHCHG	FASFL='Y' and PARAMCD='GASPH4' and AVISITN=12 and DTYPE in ('','LOCF') and anl01FL='Y'	proc glm; class sex; model PHCHG = AVAL AGE SEX BMI/solution clparm; run; quit;	Shell XYZ	test2

StudyID	Type	UDisplayID	DisplayID	DisplayName	Population	Footnote
STUDYX	T	ET143120	Table 14.3.1.2	Association between Improvement in Daily Pain Events and Change in Gastric pH at Week 12	FAS	P-value is obtained from the linear regression model with XYZ controlling for age, sex and BMI.

Figure 10 TLF Generations Driven by ARM of ToC

The unique display ID (UDisplayID), not like DisplayID (e.g., table number) or DisplayName (i.e., title), is required to be unique following conventions once the analysis entry is created. The conventions meet both uniqueness and easy identification (e.g., 'E' for efficacy, 'S' for safety, 'T' for table, 'F' for figure) and plus extensible numbers. The unique display ID will be referenced as the key in the programming to retrieve the analysis results metadata. In the ToC, the footnote column can be designed to dynamically hold multiple footnotes separated with a special character. In the green header section of figure 10, the columns are aligned with other core fields of ADaM analysis results metadata. Those analysis results metadata, converted to SAS dataset, will be used to facilitate automated programming. For a specific study, with changing data selection criteria (WHERE statement), different inferential analysis tables with the same inferential model (e.g., Analysis of Covariance (ANCOVA) model) can be generated. When implementation across studies, tables can be generated with other ANCOVA models. The following implements one way to make the model executed with the SAS macro:

```
ods output ParameterEstimates=param; /*SAS structured ODS OUTPUT needs to be fully utilized*/
%macro model;
  &ProgStat.;
%mend;
%model;
```

Based on specified ARMs in Figure 10, the DEFINE-XML for ARMs of core tables can be further automatically produced for submission.

DISCUSSION

CDISC implementation presents challenges but also provides opportunities to improve programming processes for high quality and efficient deliverables. From the statistical programming perspectives, the paper summaries the best practices in CDISC implementations accompanying optimal programming approach.

The programming standard of procedure (SOP) and guidelines are developed from programming and provide guidance to production programming. The programming SOP and guidelines shall be reexamined and updated, as needed, to reflect the best practices.

The production harmonized with submission is one of the best practices. For a high reuse of codes, SAS macros are commonly developed. This requires balance with code readability and facilitating usages. It helps with less macro layers, more notes in the program, and clear documentations. Corresponding independent QC programs may be used for submission instead of complex company macros.

ACKNOWLEDGEMENT

The author would like to thank Enecia Rivera for the invaluable inputs and paper reviews.

CONTACT INFORMATION

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

Name: Chengxin Li
Enterprise: Luye Pharma Group
Address: 502 Carnegie Center, Suite 100
City, State ZIP: Princeton, NJ 08540
E-mail: chengxin.li@luye.com