Paper 19

Creating a clinical trial data package for Regulatory submission

Pradip Sivaramakrishnan and Anbarasan Durairajan Navitas Life Sciences, India

ABSTRACT

Regulatory agencies increasingly require standardized clinical trial data submissions. Standardization enables consistent metadata definitions, improved traceability and streamlined downstream datasets. The CDISC SDTM provides a framework to ensure consistency, transparency, and interoperability across studies. The resulting SDTM package provides a clear, well-organized representation of collected study data for efficient review and analysis. In essence, SDTM e-submission is the bridge between clinical trial data and regulatory approval, ensuring that data is standardized, validated, and review-ready.

Introduction

👉 In essence, SDTM e-submission is the bridge between clinical trial data and regulatory approval, ensuring that data is standardized, validated, and review-ready. On comparing different guidance documents and from extensive study experience on regulatory submissions, we found several scenarios where programmers can utilize some guidance on how to bridge the gap to ensure compliance and regulatory submission efficiency. For example, there are differences between variable label and its length between CDISC IG and P21E, especially when working with lower versions. We also found parameters listed in FDA TCG that are currently not present in IG, among other findings.

Objective

The objective of SDTM e-Submission is to provide a standardized, structured format for clinical trial data that enables efficient regulatory review, data traceability, and global alignment across agencies like the FDA, PMDA, and EMA.

🎯 Key Objectives of SDTM e-Submission

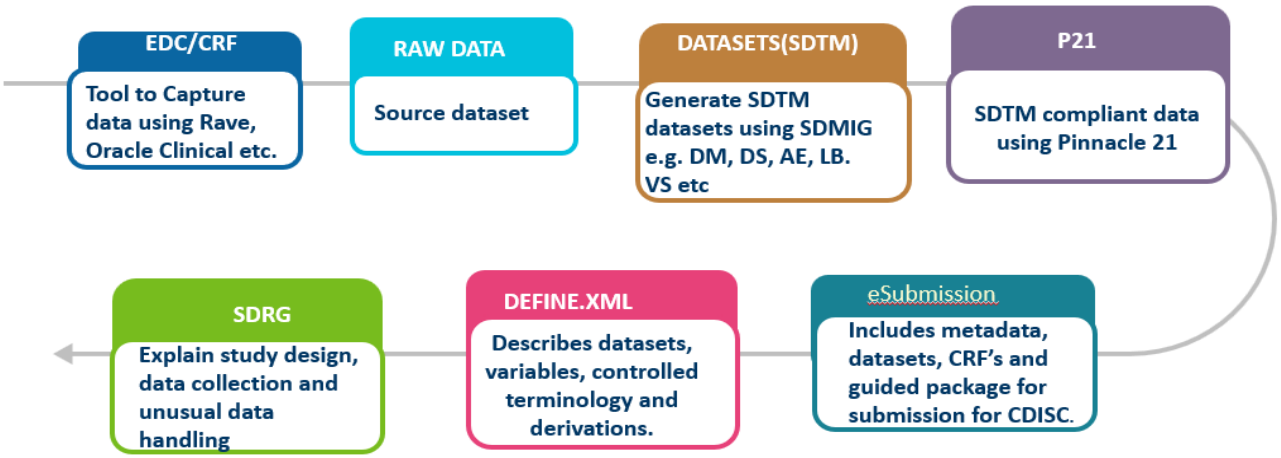
- **Regulatory Efficiency** SDTM (Study Data Tabulation Model) ensures that clinical trial data is organized in a consistent format, making it easier for regulators to review and assess submissions quickly.
- **Global Standardization** Since 2016, SDTM has been mandatory for most regulatory authorities (FDA, PMDA, NMPA). This harmonization reduces duplication of effort and aligns submissions worldwide.
- **Data Quality & Traceability** Structured datasets improve transparency, traceability, and reproducibility of results. Regulators can track data back to its source, which strengthens confidence in the findings.

- **Facilitates Analysis & Reporting** SDTM supports streamlined data collection, management, and reporting. It enables aggregation, warehousing, and reuse of data for statistical programming and safety/efficacy analysis.
- **Supports Risk Management & Portfolio Insights** Sponsors benefit from standardized data that can be mined for insights across trials, improving portfolio management and risk assessment.

📊 Comparison Table: Benefits for Stakeholders

Stakeholder	Objective of SDTM e-Submission	Benefit
Regulators (FDA, PMDA, EMA)	Standardized format for review	Faster, more consistent approvals
Sponsors/Pharma Companies	Data traceability & reuse	Better portfolio insights, reduced risk
Clinical Programmers	Streamlined analysis & reporting	Efficient statistical programming
Patients/Public	Reliable regulatory decisions	Faster access to safe and effective therapies

1. Process Overview



📑 Purpose of e-Submission in SDTM

Regulatory Compliance SDTM is a required format for data submission to the FDA (U.S.) and PMDA (Japan). It ensures submissions meet the FDA's Data Standards Catalog requirements for NDAs, ANDAs, and BLAs CDISC.

Streamlined Review Process Standardized datasets make it easier for regulators to review, compare, and analyze data across studies. This accelerates the drug approval process by reducing ambiguity and manual reformatting CDISC.

Data Consistency & Quality SDTM enforces uniform structure and terminology, minimizing errors and inconsistencies. Facilitates data aggregation, mining, and reuse across multiple studies CDISC.

Transparency & Sharing Enables clear communication of study results between sponsors, regulators, and other stakeholders. Supports due diligence and data sharing for audits, cross-study analysis, and collaborative research CDISC.

Global Standardization SDTM is not limited to clinical trials—it extends to non-clinical (SEND), medical devices, and pharmacogenomics/genetics studies, making it a versatile global standard CDISC.

🔑 Key Components of SDTM e-Submission

1. SDTM Datasets

- Standardized tabulation data sets organized according to CDISC SDTM domains (e.g., DM, AE, LB, VS). Must follow CDISC standards and FDA's Data Standards Catalog.

2. Define.xml (Metadata File)

- Provides metadata describing datasets, variables, controlled terminology, and derivations. Essential for reviewers to understand the structure and meaning of the data.

3. Study Data Reviewer's Guide (SDRG)

- Explains study design, data collection, and any deviations from SDTM standards. Helps regulators interpret the datasets and understand context.

4. Annotated Case Report Forms (CRFs)

- Shows how data maps are collected to SDTM datasets. Ensures traceability from source data to submission datasets.

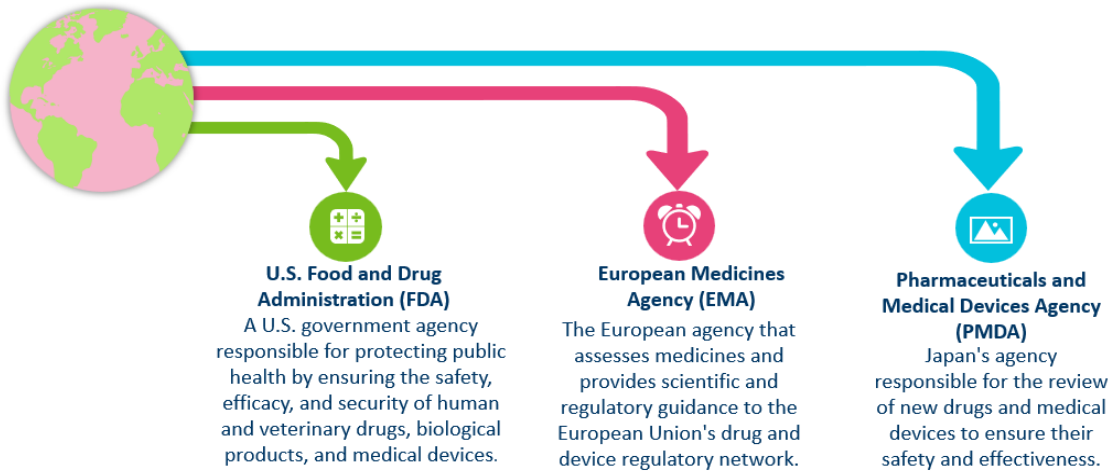
5. Compliance with Technical Conformance Guide

- FDA's *Study Data Technical Conformance Guide* specifies requirements for electronic submission format, file organization, and validation checks. Includes rules for eCTD (electronic Common Technical Document) submission.

6. Validation & Quality Checks

- Datasets must pass validation tools like Pinnacle 21 or FDA's validation rules. Ensures consistency, completeness, and compliance with regulatory standards

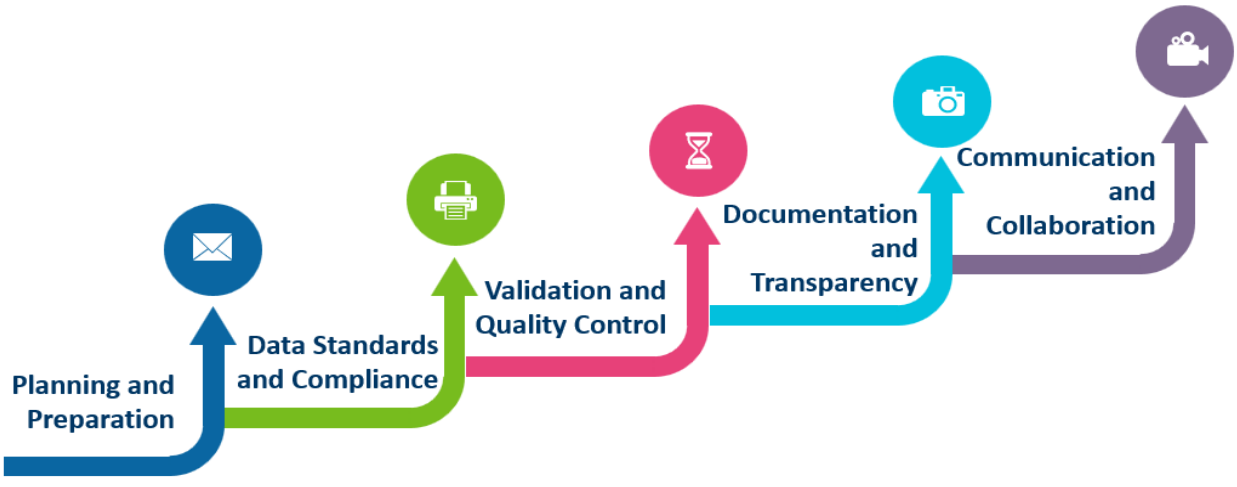
2. Regulatory Agencies



📌**Regulatory Expectations FDA: Requires SDTM datasets for all clinical trials in NDAs, BLAs, and ANDAs.**

- PMDA (Japan): Also mandates CDISC-compliant submissions.
- EMA (Europe): Encourages CDISC standards but requirements vary.
- Validation: Agencies use tools like Pinnacle 21 to check compliance with CDISC and FDA rules CDISC ClinChoice.

3. Lessons Learnt



- **Planning & Preparation**

Start early: Define submission requirements and timelines at the beginning of the study. Late conversions to SDTM often cause errors and delays.

Understand regulatory expectations: FDA and PMDA require SDTM compliance; check the latest Data Standards Catalog and guidance documents CDISC.

Scope management: Clearly identify which datasets, domains, and metadata are required for submission.

- **Data Standards & Compliance**

Follow CDISC SDTM rules strictly: Misinterpretation of variables, domains, or controlled terminology leads to rejection.

Consistency across studies: Harmonize variable naming conventions and metadata to avoid discrepancies.

Use controlled terminology: Ensure alignment with CDISC code lists and FDA requirements.

- **Validation & Quality Control**

Run validation tools: Use Pinnacle 21 or similar tools to detect compliance issues before submission.

Check reviewer guides: Ensure the Define.xml and Study Data Reviewer's Guide (SDRG) are complete and accurate.

Iterative QC: Validate datasets multiple times during development, not just at the end.

- **Documentation & Transparency**

Define.xml clarity: Provide clear metadata, derivations, and controlled terminology references.

Reviewer's Guide: Explain study design, data handling decisions, and any deviations from standards.

Traceability: Maintain links between raw data, SDTM datasets, and analysis datasets (ADaM).

- **Communication & Collaboration**

Cross-functional teamwork: Involve statisticians, programmers, data managers, and regulatory experts early.

Agency feedback: Address FDA/PMDA comments quickly and thoroughly.

Knowledge sharing: Document lessons learned internally to improve future submissions

Tools for e-validation: LORENZ

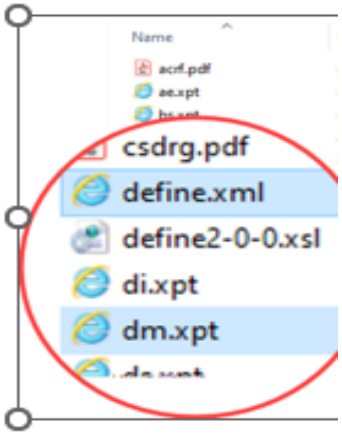
LORENZ e-Validator is a software application that verifies and validates electronic Common Technical Document (eCTD) applications, as well as non-eCTD applications such as NeeS, according to set parameters, dynamic verification rules, and international standards. The validation process produces a report on the analyzed items which can be viewed on screen or exported as a file. Users can then decide to accept or reject the application based on the report.

Using LORENZ e-Validator helps avoid **submission rejections** due to technical non-compliance. For pharmaceutical companies, this means faster approvals and smoother regulatory interactions across multiple regions.

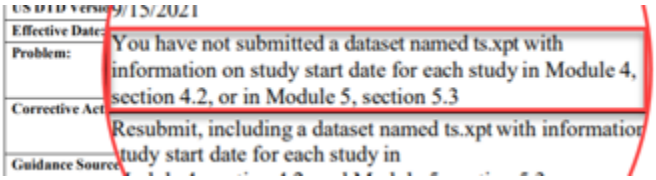
SDTM checks on Technical Rejection criteria:

Example1: A DM dataset and define.xml must be submitted

Number:	1736
Group:	General
Description:	DM dataset and define.xml must be submitted in module 4, sections 4.2.3.1, 4.2.3.2, 4.2.3.4. DM dataset, ADSL dataset, define.xml must be submitted in module 5, sections 5.3.1.1, 5.3.1.2, 5.3.3.1, 5.3.3.2, 5.3.3.3, 5.3.3.4, 5.3.4, 5.3.5.1, 5.3.5.2
Severity Description:	High
US DTD Version	2.01 and 3.3
Effective Date:	TBD
Problem:	You have not submitted Demographic dataset (DM) and define.xml for each study in Module 4, section 4.2. You have not submitted Demographic dataset (DM), Subject level analysis dataset (ADSL), and define.xml for each study in Module 5, section 5.3
Corrective Action:	Resubmit the submission with the Demographic dataset (DM) and define.xml for each study in Module 4, section 4.2. Resubmit including Demographic data, Subject level analysis dataset, and define.xml for each study in Module 5, section 5.3

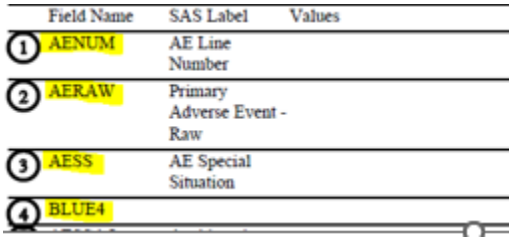


Example 2: SDTM.TS exists with TS.SSTDTC populated with the earliest Informed Consent date in the correct format (yyyy-mm-dd)



Example 3:

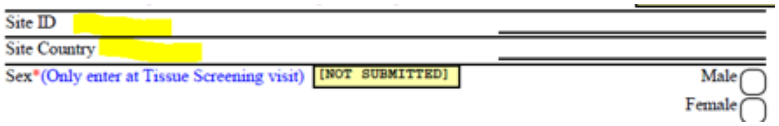
Ensure the blank CRF used to do the annotations correspond to your study data - it should be unique forms only, with no printed field or variable OIDs and no



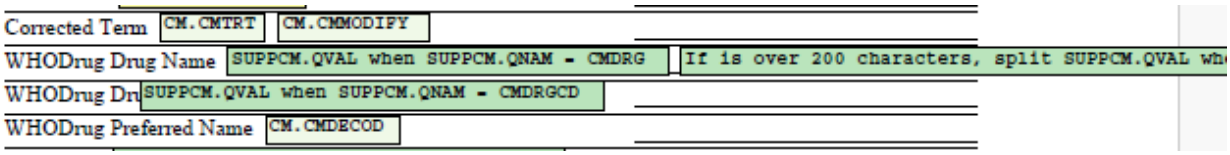
	Field Name	SAS Label	Values
1	AENUM	AE Line Number	
2	AERAW	Primary Adverse Event - Raw	
3	AESS	AE Special Situation	
4	BLUE4		

Example 4:

All fields (or pages) should be annotated (if a variable is not submitted annotate with [NOT SUBMITTED])



Example 5: No annotation is truncated, blocking CRF text or off the page



Example 6: Variables referenced in the define.xml should be prefixed by the dataset name

EXSTDY	Study Day of Start of Treatment	integer	Timing	8		Derived Start date (--STDTC) - reference start date (DM.RFSTDTC). Add 1 if start date (--STDTC) is greater than or equal to reference start date (DM.RFSTDTC)

Example 7: Ensure conformance issue explanations documented in the RG's are meaningful and interpretable

Insufficient explanation

Check ID	Diagnostic Message	Severity	Dataset	Count (Issue Rate)	Explanation
SD0063	SDTM/dataset variable label mismatch	Warning	PE	1 (6.67%)	Label is fine.

Good – meaningful and informative explanation

Check ID	Diagnostic Message	Severity	Dataset	Count (Issue Rate)	Explanation
SD0063	SDTM/dataset variable label mismatch	Warning	PE	1 (6.67%)	Per the SDTMIG, the variable label in the dataset should match the label in the SDTMIG. In SDTMIG v3.2, there are several labels that were erroneously defined to be >40 characters. The variable label for PESTRESC, 'Character Result/Finding in Standard Format', exceeds 40 characters. To avoid truncation, PESTRESC was assigned the label 'Character Result/Finding in Std Format' in the dataset so that it is <= 40 characters due to SAS restrictions.

Example 8:

Ensure alignment and consistency of Meddra / WHO dictionary versions and SDTM IG version across the equivalent information in the cSDRG, SDTM and define files, Pinnacle21 and Trial Summary (TS) domain.

CSDRG

Study Data Standards and Dictionary Inventory

Standard or Dictionary	Versions Used
SDTM	•SDTM v1.4 •SDTM-IG v3.2
Controlled Terminology	CDISC SDTM Controlled Terminology, 2018-12-21
Data Definitions	Define-XML v2.0
Medications Dictionary	LOINC 2.69, SNOMED 2020-09-01, UNII 2020-12-14, NDF-RT 2021-01-04
Medical Events Dictionary	MedDRA v23.1

TS

VIEWTABLE: Work.Ts (Trial Summary)

	TSPARMCD	TSPARM	TSVAL	TSVAL1	TSVALNF	TSVALCD	TSVCDREF	TSVCDVER
1	ACTSUB	Actual Number of Subjects	114					
2	ADAPT	Adaptive Design	N			C49487	CDISC	2018-11-08
3	ADDON	Added on to Existing Treatments	Y			C49488	CDISC	2018-11-08
4	AEDICT	Adverse Event Dictionary	MedDRA 23.1					
5	AGEMAX	Planned Maximum Age of Subjects			PINF		ISO 21090	
6	AGEMIN	Planned Minimum Age of Subjects	P18Y				ISO 8601	
7	AGESPAN	Age Group	>=18					
8	AGEU	Age Unit	YEARS			C29848	CDISC	2018-11-08
9	CMDICT	WHODrug Dictionary Version	WHODRUG GLOBAL B3 SEPTEMBER 1, 2020					

Key Takeaways

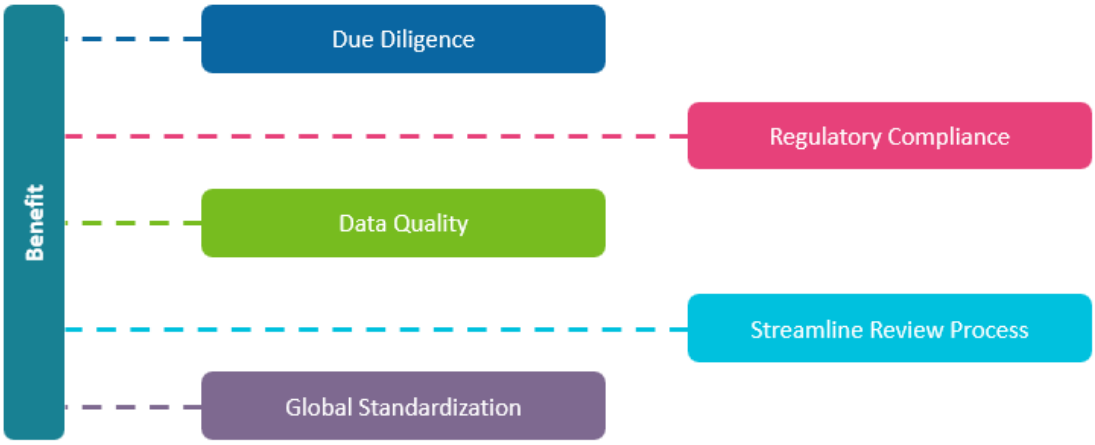
Data quality is important for efficient submission
Provide valid explanations for issues in SDRG.

For blinded trials, ensure unblinding is accurately done.
An overall P21 validation score of minimum 90% or more should be achieved on the final run.

Check all tabulation data provided have dataset & variable labels and that they are consistent if referenced within the Define and cSDRG files.

Check for references of intermediate datasets or external files (*.xls, *.doc) within the RGs and associated define files to ensure such files are documented according to our guidance.

Benefit



🧭 What is Due Diligence in Clinical Trials?

- **Definition**: A structured process of assessing risks, verifying feasibility, and ensuring compliance in clinical research.
- **Scope**: Covers scientific design, vendor selection, regulatory requirements, financial planning, and ethical safeguards.

- **Purpose**: To minimize risks, protect participants, and maximize the chances of regulatory approval and commercialization.

🧭 What is Regulatory Compliance in Clinical Trials?

- **Definition**: Adhering to all applicable regulations (FDA, EMA, ICH-GCP, local laws) and ethical principles (Declaration of Helsinki).
- **Purpose**: Safeguard patient rights, ensure scientific validity, and maintain transparency in trial conduct

📊 Why Data Quality Matters in Clinical Trials

- **Patient Safety:** Accurate data prevents misinterpretation of drug safety and efficacy.
- **Regulatory Approval:** Agencies like the FDA and EMA require robust, validated data for drug approval.
- **Scientific Integrity:** Reliable data ensures reproducibility and credibility of trial outcomes.
- **Operational Efficiency:** Clean data reduces delays, rework, and costs.
- **Market Confidence:** Investors and healthcare providers trust results backed by strong data management.

🚀 Why Streamlining Matters

- **Faster Patient Access**: Delays in review mean patients wait longer for potentially life-saving treatments.
- **Cost Reduction:** Streamlined processes cut administrative overhead and trial costs.
- **Global Competitiveness**: Efficient reviews help sponsors bring therapies to market faster.
- **Improved Collaboration:** Harmonized processes across institutions and regulators reduce duplication.

🌍 What is Global Standardization in Clinical Trials?

- **Definition**: The adoption of harmonized international guidelines (such as ICH-GCP) and frameworks to ensure consistency in trial design, conduct, and reporting worldwide.
- **Purpose**: To reduce duplication, improve transparency, and make trial results valid across borders.

Results

Successful e-Submission improves regulatory review efficiency, reduces queries, and enhances data traceability. Common issues include metadata mismatches, controlled terminology errors, and incomplete documentation.

Conclusion

There is occasionally a delay between an update in the FDA requirement and the corresponding IG/P21E update, even though these documents are generally consistent. Discrepancies may also result from other factors. The regulatory submission process can be streamlined by offering some understanding of some of these and strategies to reconcile the disparities.

REFERENCES

[1] Clinical Study Data Reviewer's Guide (cSDRG) Package: Clinical Study Data Reviewer's Guide (cSDRG) Package - WORKING GROUPS - PHUSE Advance Hub

[2] FDA Data Standards Catalog: Data Standards Catalog | FDA

[3] PMDA Data Standards Catalog: New Drug Review with Electronic Data | Pharmaceuticals and Medical Devices Agency (pmda.go.jp)

[4] FDA Study Data Technical Conformance Guide,2021. https://www.fda.gov/industry/fda-resources-data-standards/study-data-standards-resources.

CONTACT INFORMATION

- **Author:** Anbarasan Durairajan
- **Company:** Navitas Life Sciences
- **Email:** anbarasan.durairajan@navitaslifesciences.com

- **Co-Author:** Pradip Sivaramakrishnan
- **Company:** Navitas Life Sciences
- **Email:** pradip.s@navitaslifesciences.com

Website: www.navitaslifesciences.com