From Raw Data to Regulatory Trust: Innovating Clinical Data Standards Without Compromising Integrity

Swaroop Kumar Koduri, Ephicacy Lifescience Analytics, Bengaluru, India

BENGALURU, INDIA PharmaSUG SDE APRIL 11, 2026

Abstract

As clinical trials grow more complex and timelines accelerate, programming teams must deliver faster results while maintaining regulatory trust. Health authorities require consistent standards, transparent derivations, and end-to-end traceability across the data lifecycle.

This poster presents a standards-driven framework that transforms raw clinical data into submission-ready outputs without compromising integrity. The approach integrates early SDTM and ADaM alignment, metadata consistency, and macro-light SAS programming to enhance clarity, reproducibility, and scalability. Embedded validation checkpoints ensure traceability from source data to final tables, listings, and figures.

A pharmacokinetic and safety case study demonstrates reduced rework, improved audit readiness, and fewer late-stage findings. Embedding integrity throughout the pipeline enables sustainable innovation and strengthens regulatory confidence.

Introduction

Clinical data integrity is a critical determinant of regulatory acceptance

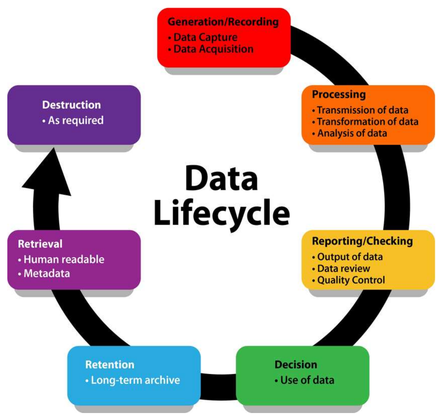
🔷 Key Considerations

- Increasing protocol complexity
- Integrated multi-study data
- Rising regulatory scrutiny
- Risk of late-stage validation findings

🔷 Regulatory Expectation

- Consistent SDTM implementation
- Traceable ADaM derivations
- Metadata transparency
- Define.xml alignment
- Reproducible TLF generation

END-TO-END DATA PIPELINE FRAMEWORK



Embedding integrity early in the data lifecycle minimizes rework and strengthens regulatory confidence.

Integrity Design Principles

🔷 Early Standards

SDTM Pre-Lock | Controlled Terminology | Domain Validation

🔷 Metadata-Driven

SAP→ADaM Traceability | Variable Origin | Define Sync

🔷 Macro-Light SAS

Explicit Logic | Minimal Dynamic Macros | QC Reproducible

🏆 **Integrity is preserved when transformation logic is controlled — not when issues are fixed downstream.**

VALIDATION IMPACT & REGULATORY CONFIDENCE

🔷 Built-In Validation Architecture

Validation is embedded at each transformation layer — not deferred to the end.

SDTM Layer

- Early Pinnacle21 checks
- Controlled terminology reconciliation
- Domain-level structural review

ADaM Layer

- SAP cross-verification
- Population flag harmonization
- Baseline & treatment-emergent logic validation

TLF Layer

- Output-to-ADaM reconciliation
- Independent double programming
- Denominator consistency verification

Regulatory Traceability Controls

- SDTM → ADaM linkage documentation
- Parameter-level trace matrices
- Define.xml synchronization
- Reviewer's Guide alignment

Every analysis result is traceable back to source SDTM variables.

🔷 High-Impact Data Integrity Case Studies

🧩 **1 SUBJID Overwrite During Merge**
Issue: SUBJID overwritten when merging by USUBJID
Control: Pre-merge rename (DM_SUBJID vs SUBJID)
Impact: Preserved lineage | Prevented subject-level corruption

🧩 **2 Many-to-Many Merge Duplication**
Issue: AE + EX merge creates duplicate records
Control: Merge using proper keys (USUBJID + date logic)
Impact: Accurate event counts | Prevented safety inflation

🧩 **3 Source Data Manipulation in ADaM**
Issue: SDTM values modified directly in ADaM
Control: Derive new variables (no overwrite of source)
Impact: Maintained SDTM traceability | Audit compliance

Each transformation layer incorporates validation checkpoints to ensure consistency, traceability, and compliance

🧩 **4 Population Flag Drift (SAFFL/ITTFL)**
Issue: Flags re-derived differently across datasets
Control: Derive once in ADSL and propagate
Impact: Consistent denominators | Stable TLF outputs

🧩 **5 Treatment-Emergent Flag Misclassification**
Issue: TRTEMFL inconsistent due to date handling
Control: Centralized exposure merge + unified logic
Impact: Correct safety classification | Reduced queries

🧩 **6 Partial Date Imputation Variability**
Issue: Different imputation rules across programs
Control: Single standardized imputation logic
Impact: Consistent baseline & TRTEMFL derivation

🧩 **7 Duplicate Key Records Not Detected**
Issue: Duplicate records in SDTM domains
Control: Mandatory DUPKEY validation before ADaM
Impact: Prevented record multiplication | Data stability

🧩 **8 Baseline Flag Inconsistency**
Issue: Different baseline definitions across parameters
Control: Central baseline derivation rule
Impact: Accurate change-from-baseline results

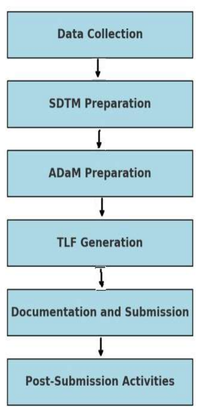
🧩 **9 Visit Window Misalignment**
Issue: Visit assignment differs across programs
Control: Metadata-driven visit mapping
Impact: Consistent visit-based analysis

🧩 **10 Define.xml Misalignment**
Issue: Dataset variables not matching metadata
Control: Pre-submission Define reconciliation check
Impact: Improved reviewer clarity | Fewer findings

🔷 Consequences of Weak Validation & Traceability

- 🚨 Increased reviewer queries
- 🚨 Define.xml inconsistencies
- 🚨 Traceability gaps in submission
- 🚨 Audit/inspection findings
- 🚨 Submission delays or rejection
- 🚨 Loss of stakeholder confidence
- 🚨 Reduced scalability across studies
- 🚨 Significant rework and timeline delays
- 🚨 Increased cost of validation cycles

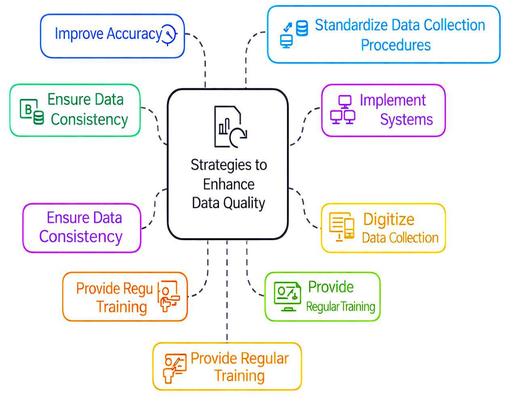
Clinical Trial Data Workflow: From SDTM to ADaM to TLFs



CONCLUSIONS

✔ Integrity must be **embedded early**, not validated late
✔ SDTM standardization enables **structured and compliant data flow**
✔ ADaM ensures **traceability and reviewer transparency**
✔ Macro-light SAS programming improves **clarity and audit readiness**
✔ Embedded validation checkpoints reduce **rework and late-stage risks**
✔ Traceability from SDTM → ADaM → TLF is critical for **regulatory trust**

🚀 **Regulatory confidence is not achieved by correction — it is achieved by design.**



References

- CDISC. Study Data Tabulation Model (SDTM) Implementation Guide, Version 3.4.
- CDISC. Analysis Data Model (ADaM) Implementation Guide, Version 1.3.
- CDISC. *Define-XML Specification*, Version 2.1.
- European Medicines Agency (EMA). *Guideline on Data Integrity and Traceability in Clinical Trials*.

Acknowledgments

The author gratefully acknowledges Ephicacy Lifescience Analytics for providing the opportunity and support to develop this work.

Contact Information

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

Swaroop Kumar Koduri
Ephicacy Lifescience Analytics, Bengaluru, India
E-mail: swaroop.koduri@ephicacy.com
www.ephicacy.com