Paper DS18

# Optimizing SDTM and ADaM: Remediation Techniques for Data Accuracy

Kishore Pothuri, Vita Global Sciences, Waltham, USA

## ABSTRACT

Remediation of SDTM (Study Data Tabulation Model) and ADaM (Analysis Data Model) datasets using legacy studies is a critical process in ensuring regulatory compliance and data consistency in clinical trials. Legacy studies often predate current CDISC (Clinical Data Interchange Standards Consortium) standards, necessitating the transformation and remediation of historical data to align with modern SDTM and ADaM formats. This process involves mapping legacy data structures to CDISC-compliant formats, addressing any discrepancies, and ensuring that the datasets are complete, accurate, and ready for submission. The remediation process enhances the reusability of legacy data, supports integrated analyses across multiple studies, and facilitates more efficient regulatory reviews. However, it presents challenges, including the need for deep domain expertise, extensive quality checks, and careful handling of data nuances.

## INTRODUCTION

In the context of Clinical Data Standards, remediation in SDTM (Study Data Tabulation Model) and ADaM (Analysis Data Model) the process of identifying, correcting, and improving data inconsistencies, structural errors, and compliance issues to ensure that the datasets adhere to latest CDISC (Clinical Data Interchange Standards Consortium) guidelines and regulatory requirements. This paper explores the methodologies and best practices for effectively remediating SDTM and ADaM datasets using legacy studies, emphasizing the importance of accuracy and compliance.

## IMPORTANCE OF REMEDIATION IN SDTM AND ADaM

Remediation in SDTM and ADaM is crucial because it ensures that clinical trial data is accurate, consistent, and compliant with regulatory standards like those set by the FDA and PMDA. Here's why it matters:

1. **Regulatory Compliance**

- Regulatory agencies require SDTM (Study Data Tabulation Model) and ADaM (Analysis Data Model) datasets to adhere to CDISC standards.
- Pinnacle 21 validation often reveals issues like missing values, incorrect variable formats, or domain mismatches. Remediation addresses these errors, ensuring submission readiness.

2**. Data Quality and Integrity**

- Clean, accurate datasets are essential for reliable statistical analysis.
- Remediation helps fix issues like data inconsistencies, mapping errors, and logical discrepancies, leading to more trustworthy study results.

3. **Analysis Accuracy**

- ADaM datasets drive statistical analyses and outputs like Tables, Listings, and Figures (TLFs).
- Errors in data can lead to incorrect conclusions, which might delay submissions or cause regulatory rejections.

4. **Submission Efficiency**

- Proactively remediate data during the study life cycle reduces the time needed for submission preparation.

- Clean datasets minimize back-and-forth communication with regulatory bodies, speeding up drug approval timelines.

5. **Cross-Study and Cross-Submission Consistency**

- Standardized, remediated datasets allow better comparison across studies, which is critical for integrated analyses (e.g., ISS/ISE).
- Consistent data structure aids in long-term data utilization, such as for meta-analyses or future submissions.

## REMEDIATION IN SDTM:

**PURPOSE**:

Remediation in **SDTM** involves correcting errors in data tabulation, ensuring proper mapping from raw clinical data to SDTM domains, and addressing structural inconsistencies.

**OBJECTIVE:**

The goal is to ensure that the data meets FDA and CDISC submission requirements.

**COMMON ISSUES IN SDTM:**

- **Mapping Errors**: Incorrect source-to-target mapping of raw data to SDTM domains.
- **Structural Issues:** Missing required variables or incorrect domain assignments.
- **Data Consistency Issues**: Discrepancies between related domains (e.g., AE domain not matching EX domain for adverse events and exposure).
- **Controlled Terminology Non-Compliance:** Use of non-standard terminology in variables requiring CDISC-approved terminology.
- **Date Issues:** Incorrect date formats, missing dates, or mismatched date relationships (e.g., AE start date after study discontinuation).
- **Duplicate Records:** Repeated or redundant records in SDTM datasets.
- **Key Variable Inconsistencies:** Mismatch in USUBJID across different domains.

**REMEDIATION STEPS FOR SDTM:**

- **Data Mapping Review:** Re-evaluate the mapping process from raw datasets to SDTM domains.
- **Validation Checks:** Use Pinnacle21 or internal validation tools to identify compliance issues.
- **Controlled Terminology Update:** Ensure all values align with CDISC-controlled terminology.
- **Cross-Domain Consistency Checks:** Ensure alignment between related datasets (e.g., AE and EX).
- **Structural Adjustments:** Add missing variables and correct data formats.
- **Sponsor and Regulatory Requirements Compliance:** Ensure all sponsor-specific and regulatory guidelines are met.
- **Revalidation and QC:** Conduct post-remediation validation using compliance tools.

## REMEDIATION IN ADaM:

**PURPOSE**:

Remediation in ADaM focuses on refining analysis datasets by ensuring traceability, accuracy in derived variables, and compliance with statistical analysis plans (SAP). This step is essential for regulatory submission and reliable statistical analysis.

**OBJECTIVE:**

Ensure datasets are structured correctly for statistical analysis, aligning with ADaM standards.

**COMMON ISSUES IN ADaM:**

- **Incorrect Dataset Structure:** ADaM datasets (ADSL, ADTTE, ADAE, etc.) not following ADaM rules (e.g., one record per subject in ADSL).
- **Traceability Issues:** Lack of clear traceability from SDTM to ADaM datasets.
- **Variable Naming and Metadata Errors**: Non-standard variable names, missing derivations, or incorrect metadata.
- **Incorrectly Derived Variables:** Errors in computation of analysis-ready variables (e.g., AVAL, AVALC).
- **Mismatch with SAP (Statistical Analysis Plan):** Derived datasets do not align with the pre-specified analysis plan.
- **Inconsistent Treatment Assignments:** Discrepancies between treatment assignments in ADSL and other analysis datasets.
- **Date Derivation Errors:** Incorrect calculations for derived dates such as analysis visit dates.

**REMEDIATION STEPS FOR ADAM:**

- **Align with ADaM Implementation Guide:** Ensure datasets follow ADaM structure and variable naming conventions.
- **Traceability Verification**: Ensure derivations are well-documented and traceable from SDTM datasets.
- **Validation with Pinnacle21:** Identify and correct issues flagged by ADaM compliance tools.
- **Check Analysis Plan Compliance:** Cross-check dataset derivations against SAP.
- **Recalculate Derived Variables:** Fix errors in variable derivations and transformations.
- **Ensure Dataset Interoperability:** Validate consistency between different ADaM datasets.
- **QC and Peer Review:** Conduct thorough QC checks before submission.

**REGULATORY AGENCIES:**







- **US FDA:** Food Drug Administration
- **EMA:** European Medicines Agency
- **Japan PMDA:** Pharmaceuticals and Medical Devices Agency
- **NMPA CHINA:** National Medical Products Administration
- **Health Canada**
- **TGA:** Therapeutic Goods Administration of Australia

## REGULATORY AGENCIES: SUBMISSION REQUIREMENTS

| Regulatory Agency | Region | Submission Standard | Submission Type | SDTM Requirements | ADaM Requirements | Submission Format | Key Guidelines |
|---|---|---|---|---|---|---|---|
| FDA (Food and Drug Administration) | USA | CDISC (SDTM, ADaM) | NDA, BLA, ANDA, IND | Required for all clinical trial data submissions & compliance validated using Pinnacle21 | Required for efficacy and statistical analysis & compliance validated using Pinnacle21 | eCTD with Define-XML, XPT files | FDA Technical Conformance Guide & 21CFR Part 11 complaint for electronic records |
| EMA (European Medicines Agency) | Europe | CDISC (SDTM, ADaM recommended | MAA | Datasets must comply with ICH E3 & E9 guidelines. SDTM Recommended but not mandatory | Datasets must comply with ICH E3 & E9 guidelines. ADaM Recommended but not mandatory | eCTD with Define-XML, XPT files | EMA Guidance on Clinical Data Submission |
| PMDA (Pharmaceuticals and Medical Devices Agency) | Japan | CDISC (SDTM, ADaM) | NDA, J-NDA | Required for electronic submissions | Required for statistical analysis | eCTD with Define-XML, XPT files | PMDA Technical Conformance Guide |
| NMPA (National Medical Products Administration) | China | CDISC (SDTM, ADaM recommended) | NDA, IND | Required for NDAs and some INDs | Encouraged but not strictly required | eCTD with Define-XML, XPT files | NMPA Data Submission Guidance |
| HC (Health Canada) | Canada | CDISC (SDTM, ADaM recommended) | NDS, SNDS, BLA | Highly Recommended but not mandatory | Highly Recommended but not mandatory | eCTD with Define-XML, XPT files | Health Canada Guidance on Data Standards |
| TGA (Therapeutic Goods Administration) | Australia | CDISC (SDTM, ADaM optional) | NDA, IND | Preferred but not mandatory | Optional | eCTD with Define-XML, XPT files | TGA Clinical Data Submission Guidelines |

## INDUSTRY GUIDANCE: SDTM & ADAM REMEDIATION

**Is there an industry provided practice or work instructions for SDTM & ADaM remediation?**

- While CDISC and FDA do not provide a direct guide for up-versioning SDTM and ADaM mappings, they provide foundational principles that can be applied. Following these principles ensures alignment with expectations
    - **SDTMIG/ADaMIG Change Log**: Carefully review the documentation for differences between versions.
    - **CDISC Community Forums**: Discussions often address challenges and solutions in version migration.
    - **Working Groups**: Join CDISC working groups or consult with vendors that specialize in data standards.
- Even though there is no gold standard method or practice for remediation of legacy data, there are several industry guidance documents, standards, and tools available for SDTM and ADaM remediation
- These resources are primarily provided by CDISC (Clinical Data Interchange Standards Consortium) and regulatory agencies such as the FDA, EMA, and PMDA.
- **Key resources:**
    - **CDISC Guidance Documents**: SDTM Implementation Guide & ADaM Implementation Guide

- **FDA Guidance**: FDA Technical Rejection Criteria for Study Data & Study Data Technical Conformance Guide

- **Validation and Remediation Tools**: Pinnacle 21 Enterprise & Pinnacle 21 Community, SAS Programming

## CASE STUDY: REMEDIATION OF LEGACY DATA IN SDTM AND ADaM

- Remediation of SDTM (Study Data Tabulation Model) and ADaM (Analysis Data Model) datasets using legacy studies is a critical process in ensuring regulatory compliance and data consistency in clinical trials.
- Legacy studies often predate current CDISC (Clinical Data Interchange Standards Consortium) standards, necessitating the transformation and remediation of historical data to align with modern SDTM and ADaM standards.
- This process involves mapping legacy data structures to CDISC-compliant formats, addressing any discrepancies, and ensuring that the datasets are complete, accurate, and ready for submission.
- The remediation process enhances the reusability of legacy data, supports integrated analyses across multiple studies, and facilitates more efficient regulatory reviews.

## APPROACH FOLLOWED FOR REMEDIATION OF LEGACY DATA

The remediation of legacy data from older versions of SDTM (Study Data Tabulation Model) and ADaM (Analysis Data Model) involves several key objectives and tasks to ensure the datasets comply with current standards, facilitate efficient regulatory review, and support the submission process.

The approach followed includes the following major objectives and steps:

**OBJECTIVES:**

- Conversion of Legacy Data from SDTM IG Version 3.1.3 to 3.4
- Use of the Latest SDTM Controlled Terminology
- Generation and Validation of SDTM define.xml v 2.1.0
- Conversion of Legacy Data from ADaM IG Version 1.1 to 1.3
- Use of the Latest ADaM Controlled Terminology
- Generation and Validation of ADaM define.xml v 2.1.0
- Generation of Other Regulatory Submission Components
    - **SDTM aCRF (Annotated Case Report Form):** This document maps the SDTM datasets to the case report forms used in the clinical trial, showing the relationship between data collected in the trial and the standardized SDTM format.
    - **cSDRG (clinical Study Data Reviewer's Guide):** The cSDRG provides detailed information about the structure and content of the SDTM datasets, enabling regulatory reviewers to understand the datasets quickly.
    - **ADRG (Analysis Data Reviewer's Guide):** Like the cSDRG, the ADRG provides a detailed guide to the ADaM datasets, including the statistical analysis plan and explanations of derived variables.
    - **Supporting Documentation:** Additional supporting documentation, including metadata, variable definitions, data transformations, and validation reports, is also prepared to assist regulatory reviewers and ensure compliance with submission requirements.

## STEPS IN LEGACY DATA REMEDIATION

**OBJECTIVE:** To assess the differences between legacy datasets and current SDTM/ADaM standards.

- **Gap Analysis**
  - The first step in the remediation process is to perform a gap analysis. This involves comparing the legacy datasets, which were often collected under outdated standards, with the current versions of SDTM and ADaM Implementation Guides (IG).
  - The gap analysis identifies:
    - Missing or incorrectly mapped variables.
    - Discrepancies in domain structures and data formats.
    - Variations in controlled terminology and coding.
    - Outdated or non-compliant data fields.

The result of this analysis is a clear understanding of the scope of changes required to bring the legacy data into compliance with current standards.

- **Mapping and Transformation**
  - **Objective**: To update legacy data to comply with SDTM and ADaM standards.
  - **Variable Names and Formats**: Legacy data often contains outdated variable names and formats. During this step, the datasets are updated to match the latest SDTM and ADaM variable naming conventions and data formats. This includes changes such as aligning variable names with the CDISC specifications and converting data types (e.g., changing date formats or numeric values).
  - **Controlled Terminology**: Legacy datasets may contain outdated or inconsistent coding for variables like adverse events, medications, and lab results. These are updated to use the latest SDTM controlled terminology and ADaM controlled terminology to ensure uniformity across datasets and regulatory compliance.
  - **Domain Structures**: The domain structures in legacy data must be transformed to align with the CDISC standards. This includes ensuring that data categories like demographics, adverse events, and laboratory results are organized in the correct domains as specified by the current SDTM and ADaM guidelines.
  - **Data Mapping**: Data from legacy systems are mapped into their appropriate SDTM and ADaM domains, ensuring that each piece of data corresponds to the right field in the standardized format.
- **Validation:**
  - **Objective:** To ensure that the transformed datasets are compliant with regulatory standards and CDISC guidelines.
  - **After the mapping** and transformation of legacy data, the next step is to validate the datasets for compliance with CDISC standards, regulatory guidelines, and industry best practices.
  - **Validation Tools:** Tools like Pinnacle 21 are used to automatically check for compliance issues, errors, and inconsistencies. These tools validate:
    - SDTM and ADaM datasets for standardization.
    - Controlled terminology consistency.
    - Completeness of metadata and variable definitions.
    - Adherence to regulatory requirements.
  - **Manual Review:** Along with automated validation, a manual review is often conducted to ensure that the datasets are accurately transformed, with no discrepancies or errors that could affect regulatory submissions.

- The validation process also ensures the integrity and quality of data, making it ready for regulatory submission.

- **Documentation Update**

  - **Objective**: To revise all necessary documentation to reflect the changes made during the remediation process.

  - **Define-XML:** The Define-XML file, which describes the structure and metadata of SDTM and ADaM datasets, is updated to reflect the changes made during the remediation process. This includes updating the definitions of variables, value labels, controlled terminology, and derivations

  - **Annotated CRFs:** The Annotated Case Report Forms (aCRF) are revised to reflect how data collected in the clinical trial maps to the SDTM domains. This is critical for regulatory reviewers to understand the link between the trial data and the SDTM datasets.

  - **Reviewer Guides:** Both the clinical Study Data Reviewer's Guide (cSDRG) for SDTM datasets and the Analysis Data Reviewer's Guide (ADRG) for ADaM datasets are updated to provide detailed explanations of the dataset's contents, structure, variable derivations, and analysis methodologies. These guides facilitate the regulatory review process by offering transparency into how the datasets were transformed and structured.

  - **Supporting Documentation:** Other supporting documents, including metadata, variable definitions, data mapping documentation, and validation reports, are updated to ensure that all changes and transformations are clearly documented and available for regulatory submission.

**P21Report**: Legacy SDTM data in SDTM IG version 3.1.2 vs SDTM IG version 3.4

- **Key Findings:** Few examples

  - Subject Elements (SE) domain missing

  - Adverse Events (AE) missing the coding variables such as AEBDSYCD, AEHLT etc.

  - Regulatory Expected Variable EPOCH is not found

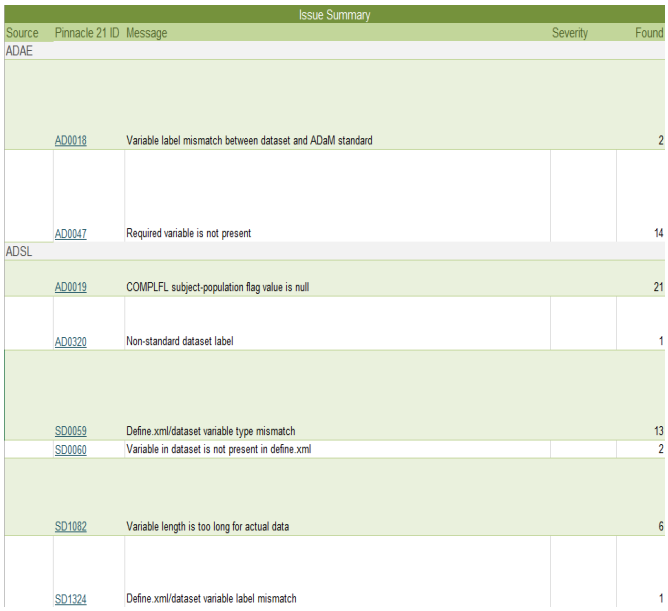
  - Demographics (DM) missing expected variables such as ACTARM/CD, ACTARMUD, ARMNRS, RFICDTC, RFPENDTC etc.

Issue Summary

| Source | Pinnacle 21 ID | Message | Severity | Found |
|---|---|---|---|---|
| GLOBAL | | | | |
| | SD1111 | Missing SE dataset | | 1 |
| AE | | | | |
| | SD0057 | SDTM Expected variable AEBDSYCD not found | | 1 |
| | SD0057 | SDTM Expected variable AEHLGT not found | | 1 |
| | SD0057 | SDTM Expected variable AEHLGTCD not found | | 1 |
| | SD0057 | SDTM Expected variable AEHLT not found | | 1 |
| | SD0057 | SDTM Expected variable AEHLTCD not found | | 1 |
| | SD0057 | SDTM Expected variable AELLT not found | | 1 |
| | SD0057 | SDTM Expected variable AELLTCD not found | | 1 |
| | SD0057 | SDTM Expected variable AEPTCD not found | | 1 |
| | SD0057 | SDTM Expected variable AESOC not found | | 1 |
| | SD0057 | SDTM Expected variable AESOCCD not found | | 1 |
| | SD1077 | Regulatory Expected variable EPOCH not found | | 1 |
| DM | | | | |
| | SD0057 | SDTM Expected variable ACTARM not found | | 1 |
| | SD0057 | SDTM Expected variable ACTARMCD not found | | 1 |
| | SD0057 | SDTM Expected variable ACTARMUD not found | | 1 |
| | SD0057 | SDTM Expected variable ARMNRS not found | | 1 |
| | SD0057 | SDTM Expected variable DTHDTC not found | | 1 |
| | SD0057 | SDTM Expected variable DTHFL not found | | 1 |
| | SD0057 | SDTM Expected variable RFICDTC not found | | 1 |
| | SD0057 | SDTM Expected variable RFPENDTC not found | | 1 |
| | SD0057 | SDTM Expected variable RFXENDTC not found | | 1 |
| | SD0057 | SDTM Expected variable RFXSTDTC not found | | 1 |
| | SD1082 | Variable length is too long for actual data | | 10 |
| | SD1083 | Missing DMDY variable when DMDTC variable is present | | 1 |

**P21Report**: Legacy ADam data in ADaM IG version 1.1 vs ADaM IG version 1.3

- **Key Findings**: Few examples
  - Analysis Dataset of Adverse Events (ADAE) missing the coding variables such as AEBDSYCD, AEHLT etc.
  - Analysis Data Subject Level (ADSL) has the COMPLFL value as null
  - Analysis Data Subject Level (ADSL) has variables which are not present in define.xml



Issue Summary

| Source | Pinnacle 21 ID | Message | Severity | Found |
|---|---|---|---|---|
| ADAE | | | | |
| | AD0018 | Variable label mismatch between dataset and ADaM standard | | 2 |
| | AD0047 | Required variable is not present | | 14 |
| ADSL | | | | |
| | AD0019 | COMPLFL subject-population flag value is null | | 21 |
| | AD0320 | Non-standard dataset label | | 1 |
| | SD0059 | Define.xml/dataset variable type mismatch | | 13 |
| | SD0060 | Variable in dataset is not present in define.xml | | 2 |
| | SD1082 | Variable length is too long for actual data | | 6 |
| | SD1324 | Define.xml/dataset variable label mismatch | | 1 |

**REMEDIATION PLAN**

- Based on the GAP analyses performed, a combined remediation plan document for SDTM & ADaM was developed to document all the findings, planning, strategies, objectives, scope and resolutions etc. for remediation.
- This can be a living document but needs to be updated simultaneously along with remediation updates.
- After many iterations, the final version was agreed by all the stake holders and finally signed off for documentation.
- **Contents of a Remediation Plan can include the following but not limited to:**
  - Introduction
  - Objectives
  - Scope
  - Current State Assessment: SDTM & ADaM
  - Stakeholder Input
  - Comparison with Best Practices
  - Documentation of Findings
  - Remediation Strategy: aCRF, SDTM & ADaM
  - Action Plan
  - Resources Required
  - Timelines
  - Communication Plan
  - Potential Constraints/Risks
  - Quality Control

- Documentation
- Approval
- Appendix

## CHALLENGES

- Structural Differences Between Legacy Data and SDTM/ADaM Standards
- Incomplete or Missing Data Elements
- Controlled Terminology Updates
- Traceability Issues Between Raw, SDTM, and ADaM Datasets
- Handling Unstructured or Free-Text Data
- Regulatory Compliance and Validation Challenges
- Study Design Inconsistencies
- Define-XML and Metadata Challenges
- Lack of Documentation and Institutional Knowledge

## CONCLUSION

- Remediating legacy data into CDISC-compliant SDTM and ADaM formats is challenging but crucial for regulatory submissions and integrated analyses.
- The key to success is a structured approach, leveraging automation, expert validation, and adherence to the latest CDISC guidance.

## CONTACT INFORMATION

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

Author Name: Kishore Pothuri
Company: Vita Global Sciences
Address: 281 Winter St, First Floor
Waltham, MA 02451
Work Phone: 781-833-0259
Email: kpothuri@softworldinc.com