Paper PP25

Optimizing Vendor Data Mapping: A DTS-Driven Approach to SDTM Readiness

Chandan Patel Malyala, Cytel, Hyderabad, India
Prasad Attini, Cytel, Hyderabad, India

ABSTRACT

External vendor data (e.g., Labs, ECG, PK) are sourced from multiple providers with changing structures and conventions. Despite the availability of Data Transfer Specifications (DTS), external/vendor datasets frequently show inconsistencies in variable presence, attributes, visit labels, code-lists and duplicate records. Late detection of structural issues results in repeated data transfers, manual validation, increasing rework, and extended timelines. This poster summarizes the need for an early, and metadata-driven validation approach that supports external data compliance with approved DTS prior to downstream SDTM mapping and readiness.

INTRODUCTION

In clinical trials, statistical programmers often work with non-CRF or External/Vendor data from multiple external vendors, such as laboratories, ECG providers, and imaging services. These datasets often come in different formats and with unreliable levels of completeness, making it challenging to ensure consistency and standardization.

A proactive, metadata-driven approach harmonizes non-CRF/External data at receipt, establishes key derivations early, supporting SDTM mapping and readiness. This strategy streamlines processes and enhances data quality, especially when data continues to arrive throughout the trial.

OBJECTIVES

The objective of this work was to design and implement an automated, scalable validation framework for external clinical trial data that:

- Supports compliance with Data Transfer Agreements (DTA) and Data Transfer Specifications (DTS)
- Detects structural and content discrepancies early in the data lifecycle
- Integrates seamlessly with SAS-based clinical programming workflows
- Improves readiness for SDTM conversion and downstream activities

REQUIREMENTS AND CHALLENGES

To support downstream SDTM mapping and readiness, the following requirements and challenges were identified:

- Apply *consistent validation* across all external data domains
- Reduce dependency on *hard-coded* checks
- Use *approved DTS* as the authoritative metadata source
- *Early and transparent* issue detection
- Manage *various vendor formats* and changing DTS versions

TECHNICAL SOLUTION (DTS-DRIVEN MACRO)

A generic SAS macro was developed to validate external data dynamically against approved DTS metadata. The macro reads DTS tables (variable metadata, code list and Visits) at runtime and performs standardized checks on variable presence, attributes, visit mapping, VISITNUM consistency, code-list consistency, duplicate records, any non-printable characters and comparison of two external files (Previous/New) without specifying variables or domains in macro program. The framework is metadata-driven and scalable.

All checks are implemented using SAS Clinical programming standards, enabling repeatability and minimal customization across vendor datasets.

MACRO PROCESS

- Macro imports external data into SAS work library as per option provided in parameter based on type of data i.e. SAS, CSV or EXCEL.

- Based on DTS received, create 3 excel sheets manually detailing metadata for variables (name, label, length, and type), code-lists (test codes and test names) and visit mapping (VISITNUM and VISIT). Column names in the excel sheets are standard, to nullify effect of different formats, column names etc. in different vendor DTS documents.

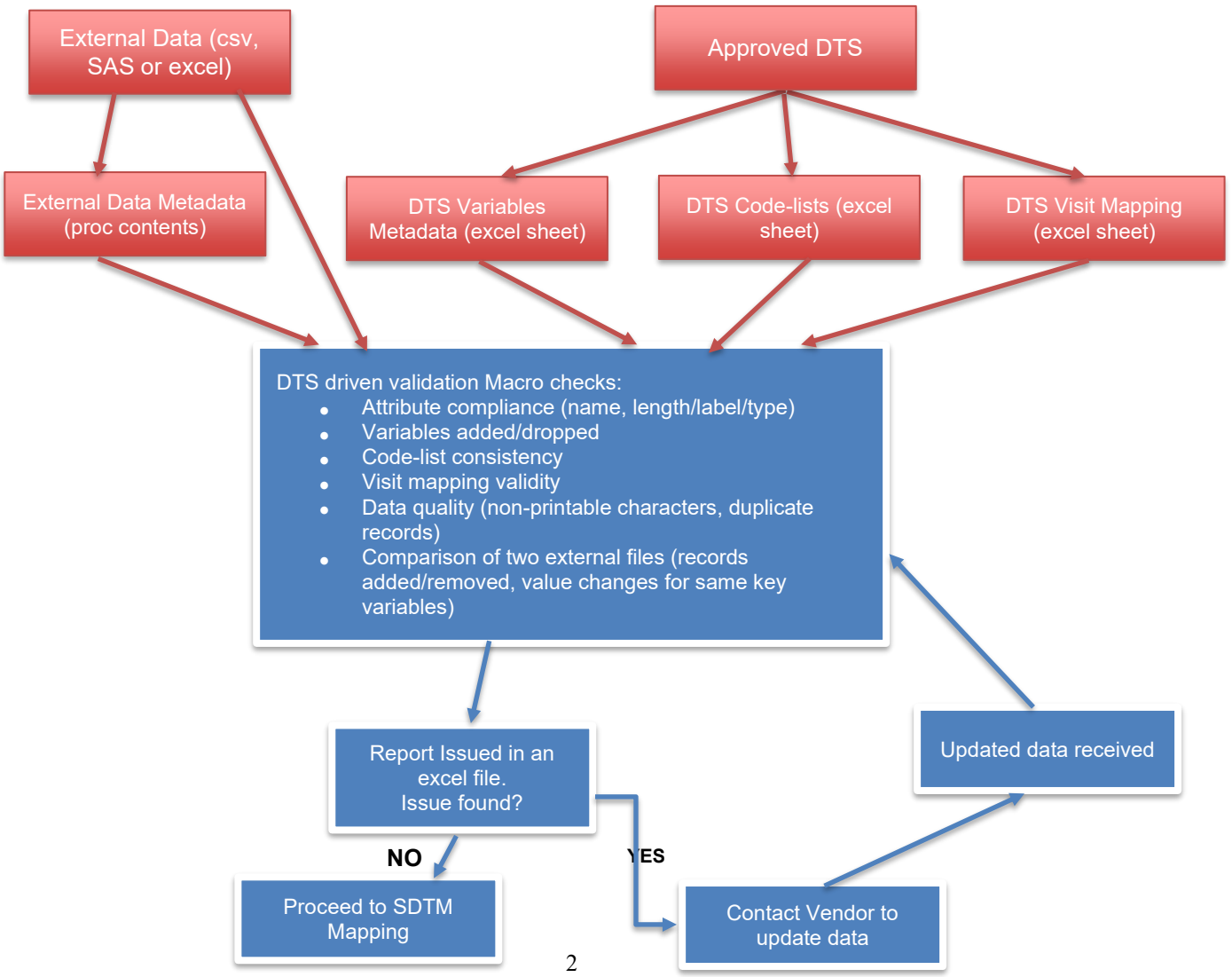
- Macro uses the excel sheets and the external data to perform following checks: The macro runs one external file at a time and generates a consolidated report by external data type.

    - Attribute compliance (name, length/label/type) using metadata of both DTS/External data
    - Variables added/dropped using metadata of both DTS/External data
    - Code-list consistency using DTS Code-lists sheet
    - Visit mapping validity using DTS Visit Mapping sheet
    - Data quality (non-printable characters, duplicate records) using External data
    - Comparison of two external files (records added/removed, value changes for same key variables)
- Macro creates a consolidated report detailing issues found for all the above checks. If no issues are found, the report is blank and external data can safely be used for SDTM mapping.

- If issues are found, vendor can be contacted to update the external data to ensure compliance with DTS.

- Once updated data is received from vendor, macro can be rerun to check for issues.

- Once the report is clear, external data can safely be used for SDTM mapping.



RESULTS

The macro performs following checks automatically: The approach helped early identification of structural issues prior to SDTM mapping.

1. Check for missing or unexpected variables in the external dataset and verify variable attributes (type, length, label) against the Data Transfer Specification.

Domain	Issue Category	Variabe Name	Label - DTS	Type - DTS	Length - DTS	Label - External	Type - External	Length - External	Issue
QS	Variable existence & attribute checks	DOMAIN	Domain Abbreviation	Char	2	Domain Abbreviation_	Char	2	LABEL MISMATCH
QS	Variable existence & attribute checks	QSCAT	Category of Question	Char	25	Category of Question	Char	16	LENGTH MISMATCH
QS	Variable existence & attribute checks	SITE	Site Identifier	Char	6				VARIABLE MISSING IN EXTERNAL
QS	Variable existence & attribute checks	QSDESCR				Description	Char	100	NEW VARIABLE IN EXTERNAL
QS	Variable existence & attribute checks	QSDESCR				Description	Char	100	TYPE MISMATCH

2. Validate VISIT → VISITNUM mapping according to defined rules and confirm VISITNUM logical consistency across records.

Domain	Issue Category	Visit - DTS	Visit - External	Visitnum - DTS	Visitnum - External	Issue
QS	Visit mapping consistency		Visit 4		5	VISIT NOT IN DTS MAP
QS	Visit mapping consistency	Visit 3		4		VISIT IN DTS MAP BUT NOT IN EXTERNAL
QS	Visit mapping consistency	Unscheduled	Unscheduled	8	6	VISITNUM INCONSISTENT WITH DTS

3. Identify duplicate records within the external data.

Domain	Issue Category	QSTEST	VISIT	QSDTC	Number of Duplicates	Issue
QS	Data Quality	Total Score	Visit 2	30JUN23:10:32:10	2	DUPLICATE RECORDS
QS	Data Quality	Total Score	Visit 4	02JUL23:12:10:25	3	DUPLICATE RECORDS

4. Ensure no non-printable characters are present.

Domain	Issue Category	Variable name	Value	Issue
QS	Data Quality	QSTEST	Mean Score for S® Sleep	NON-PRINTABLE CHARACTER
QS	Data Quality	QSDESCR	Where10 indicates High and 0® indicates Low	NON-PRINTABLE CHARACTER

5. Ensure code-list consistency.

Domain	Issue Category	Test name - DTS	Test name - External	Test code - DTS	Test code - External	Issue
QS	Test name/code consistency		QS Test name 8			TEST NOT IN DTS MAP
QS	Test name/code consistency	QS Test name 5				TEST NOT IN EXTERNAL
QS	Test name/code consistency				QSTEST8	TESTCD NOT IN DTS MAP
QS	Test name/code consistency			QSTEST5		TESTCD NOT IN EXTERNAL

6. Compare current vs. previous external files to detect structural or content changes.

In previous version

Domain	Issue Category	QSSEQ	QSTESTCD	QSORRES	VISITNUM	VISIT	QSDTC
QS	Version Comparisi	4	TSCORE	75	1	Baseline	10OCT24:10:30:25
QS	Version Comparisi	23	TSCORE	47	2	Visit 1	11OCT24:12:20:15
QS	Version Comparisi	42	TSCORE	88	6	Unscheduled	12OCT24:10:15:10

In new/current version

Domain	Issue Category	QSSEQ	QSTESTCD	QSORRES	VISITNUM	VISIT	QSDTC
QS	Version Comparisi	4	TSCORE	75	1	Baseline	10OCT24:10:30:25
QS	Version Comparisi	23	TSCORE	88	2	Visit 1	12OCT24:10:15:10
QS	Version Comparisi	42	TSCORE	85	3	Completion	14OCT24:09:20:09
QS	Version Comparisi	61	TSCORE	47	6	Unscheduled	11OCT24:12:20:15

Difference in New vs Previous version file comparison

Domain	Issue Category	QSSEQ	QSTESTCD	QSORRES	VISITNUM	VISIT	QSDTC	Issue
QS	Version Comparisi	23	TSCORE	47	2	Visit 1	11OCT24:12:20:15	DELETED/UPDATED RECORD COMPARED TO PREVIOUS TRANSFER
QS	Version Comparisi	23	TSCORE	88	2	Visit 1	12OCT24:10:15:10	NEW RECORD ADDED IN NEW TRANSFER
QS	Version Comparisi	42	TSCORE	85	3	Completion	14OCT24:09:20:09	NEW RECORD ADDED IN NEW TRANSFER
QS	Version Comparisi	61	TSCORE	47	6	Unscheduled	11OCT24:12:20:15	NEW RECORD ADDED IN NEW TRANSFER
QS	Version Comparisi	42	TSCORE	88	6	Unscheduled	12OCT24:10:15:10	DELETED/UPDATED RECORD COMPARED TO PREVIOUS TRANSFER

LESSONS LEARNED (BENEFITS AND LIMITATIONS)

Benefits:

- *Improved Data Quality:* Early detection of inconsistencies ensures clean datasets.
- *Operational Efficiency: Significant reduction in manual effort and faster turnaround.*
- *Adaptability: Can be applied across multiple studies and adapted to study specific DTS requirements.*
- *Future-Ready: Supports SDTM readiness for downstream activities.*
- *Compliance:* Adherence to DTS specification from the start.

Limitations:

- Effectiveness depends on DTS quality.
- Metadata compliance does not replace clinical validation against CRF data.
- Initial transfer compares External data (metadata) only
- Subsequent transfers compare External data metadata and previous/new files at value level.

CONCLUSION

A proactive, metadata-driven validation approach using SAS significantly improves the quality, efficiency, and compliance of non-CRF external data (vendor data) handling in clinical trials and supports SDTM readiness, while still requiring clinical validation against CRF data.

REFERENCES

1. Brian Armstrong and Renée Kerwin, 2017, PharmaSUG, Clinical and Vendor Database Harmony; Can't we all just get along? https://www.lexjansen.com/pharmasug/2017/DA/PharmaSUG-2017-DA05.pdf

ACKNOWLEDGMENTS

We would like to thank Saurabh Upalkar and Poorna Chander Kasarla for their help.

CONTACT INFORMATION

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

Chandan Patel Malyala
Cytel Statistical Software & Services Pvt Ltd
Chandanpatel.Malyala@cytel.com

Prasad Attini
Cytel Statistical Software & Services Pvt Ltd
Prasad.Attini@cytel.com

Website: https://cytel.com/