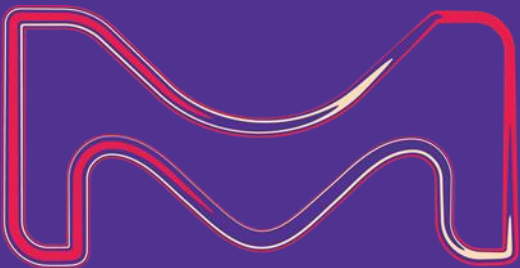


DS02: The Significance of Generic Data Transfer Specification for Non-CRF Data in Clinical Trials

PHUSE EU Connect 2024

Samaneh Sabiani and Kristina Zweier
Merck KGaA
12.11.2024



MERCK

Agenda

1. What is a Data Transfer Specification (DTS)?
2. Why do we need a Standardized Generic DTS Template for non-CRF data?
3. Implementing a Standardized Generic DTS Template
4. Different Data Structure Scenarios
5. DTS Solution and SDTM Validation for Scenario 1
6. DTS Solution and SDTM Validation for Scenario 2
7. DTS Solution and SDTM Validation for Scenario 3
8. Conclusion
9. Acknowledgments

What is a Data Transfer Specification (DTS)?

- A set of guidelines and technical specifications that outlines the process for transferring data between systems.
- Defines the structure, format, and mapping necessary for secure data transmission, providing a standardized approach for data exchange.
- Adhering to established DTS enables seamless integration and ensures data accuracy, integrity, and security.

Why do we need a Standardized Generic DTS Template for non-CRF data?

Background:

Non-CRF data, such as Safety, ePRO, Medical Imaging, Pharmacokinetics, Biomarker, Genomic, ECG and ICF are often less structured, posing standardization and integration challenges.

Problem Statement:

Historically, relying on manual processes has led to late identification of non-compliant data, necessitating time-consuming corrections during the SDTM programming phase. This reactive approach complicates the integration of external data and increases the workload for SDTM programming, often leading to submission delays.

Solution:

Ensure compliance from the beginning by implementing a Standardized Generic DTS Template.

Benefits:

- Standardization Across Data Types
- Adaptability for Diverse Assays
- Reduced Validation Efforts

Implementing a Standardized Generic DTS Template

The **DMMI*** Standardized Generic DTS Template is an Excel file which includes at least the following mandatory sheets:

- 1. Cover**
- 2. Contact Information**
- 3. Transfer Specifications**
- 4. Data Structure**
- 5. Value Level Metadata (VLM),** in case the vendor provides data in adherence with Merck's non-CRF collection data standards or **VLM-lookup**, in case the vendor cannot provide the data in adherence with Merck's non-CRF collection data standards
- 6. Visits**

* Data Monitoring, Management & Innovation

Different Data Structure Scenarios

1. Scenario 1: Data belongs to one SDTM domain and can be provided with full adherence to Merck's data collection standards.
2. Scenario 2: Data belongs to different SDTM domains and can be provided with full adherence to Merck's data collection standards.
3. Scenario 3: Data cannot be provided in adherence to Merck's data collection standards

DTS Solution and SDTM Validation for Scenario 1

Data Structure

SDTM-Like Data Strcture of the Genomics Findings domain

Variable Name	Variable Label	Type	Length	Codelist or format	SUPP --	Comments
STUDYID	Study Identifier	Char	20			XXXXXXXX
SUBJID	Subject Identifier	Char	20			e.g. 1001001
GFTESTCD	Short Name of Genomic Measurement	Char	8	(TESTCD)		See VLM tab!
GFTEST	Name of Genomic Measurement	Char	40	(TEST)		See VLM tab!
GFTSTDTL	Measurement, Test, or Examination Detail	Char	200	(GFTSDTL)		See VLM tab!
GFORRES	Result or Finding in Original Units	Char	200			See VLM tab!
GFORRESU	Original Units	Char	15	(UNIT)		See VLM tab!
GFSPEC	Specimen Material Type	Char	20	(GENSMP)		See VLM tab!
GFMETHOD	Method of Test or Examination	Char	30	(METHOD)		See VLM tab!

VLM

VLM shows mapping of MSI and TMB according to SDTM structure and terminology

GFTESTCD	GFTEST	GFTSTDTL	GFORRES	GFORRESU	GFSPEC	GFMETHOD
MICRISTB	Microsatellite Instability	MICROSATELLITE INSTABILITY OVERALL STATUS	HIGH		DNA	POLYMERASE CHAIN REACTION
MICRISTB	Microsatellite Instability	MICROSATELLITE INSTABILITY OVERALL STATUS	LOW		DNA	POLYMERASE CHAIN REACTION
MICRISTB	Microsatellite Instability	MICROSATELLITE INSTABILITY OVERALL STATUS	STABLE		DNA	POLYMERASE CHAIN REACTION
TMB	Tumor Mutation Burden	NORMALIZED NUMBER OF SEQUENCE VARIANTS	numeric	/10^6 bp	ctDNA	NEXT GENERATION TARGETED SEQUENCING
TMB	Tumor Mutation Burden	VARIANT SEQUENCE BURDEN INTERPRETATION	HIGH		ctDNA	NEXT GENERATION TARGETED SEQUENCING
TMB	Tumor Mutation Burden	VARIANT SEQUENCE BURDEN INTERPRETATION	LOW		ctDNA	NEXT GENERATION TARGETED SEQUENCING

DTS Solution and SDTM Validation for Scenario 1

SDTM Validation

GFTESTCD	GFTEST	GFTSTDTL	GFORRES	GFORRESU	GFSPEC	GFMETHOD
MICRISTB	Microsatellite Instability	MICROSATELLITE INSTABILITY OVERALL STATUS	HIGH		DNA	POLYMERASE CHAIN REACTION
MICRISTB	Microsatellite Instability	MICROSATELLITE INSTABILITY OVERALL STATUS	STABLE		DNA	POLYMERASE CHAIN REACTION
TMB	Tumor Mutation Burden	NORMALIZED NUMBER OF SEQUENCE VARIANTS	<i>numeric</i>	/10^6 bp	ctDNA	NEXT GENERATION TARGETED SEQUENCING
TMB	Tumor Mutation Burden	VARIANT SEQUENCE BURDEN INTERPRETATION	HIGH		ctDNA	NEXT GENERATION TARGETED SEQUENCING
TMB	Tumor Mutation Burden	VARIANT SEQUENCE BURDEN INTERPRETATION	LOW		ctDNA	NEXT GENERATION TARGETED SEQUENCING



The VLM sheet of the DTS is run through the xpt converter tool in order to create an SDTM (like) domain to be validated through the Pinnacle 21 Community version

DTS Solution and SDTM Validation for Scenario 1

SDTM Validation



DTS file is read into xpt converter tool

VLM sheet is selected

Tool shows that all columns/variables are SDTM compliant

Xpt file is created based on VLM sheet

DTS Solution and SDTM Validation for Scenario 1

SDTM Validation

Variables	Values	Pinnacle 21 I	Message	Category	Severity	Standards Expert Comment
						Please use the SDTM submission value of "/MBP" as a unit. The listed unit in the DTS is only a synonym for the submission and not do be used.
GFORRESU	/10^6 bp	CT2002	GFORRESU value not found in 'Unit' extensible codelist	Terminology		
VARIABLE, DATASET	DOMAIN, GF	SD0056	SDTM Required variable DOMAIN not found	Metadata		okay
VARIABLE, DATASET	GFSEQ, GF	SD0056	SDTM Required variable GFSEQ not found	Metadata		okay
VARIABLE, DATASET	STUDYID, GF	SD0056	SDTM Required variable STUDYID not found	Metadata		okay
VARIABLE, DATASET	USUBJID, GF	SD0056	SDTM Required variable USUBJID not found	Metadata		okay
VARIABLE	GFDTCT	SD0057	SDTM Expected variable GFDTCT not found	Metadata		okay
VARIABLE	GFREFID	SD0057	SDTM Expected variable GFREFID not found	Metadata		okay
VARIABLE	GFSTRESC	SD0057	SDTM Expected variable GFSTRESC not found	Metadata		okay
VARIABLE	VISITNUM	SD0057	SDTM Expected variable VISITNUM not found	Metadata		okay

The xpt file is run through Pinnacle 21 (Community version) to check for SDTM compliance. Relevant findings are remarked and communicated.

The DTS is updated and reviewed until it is SDTM compliant, respectively, needed values that are not part of the current SDTM CT version are documented as sponsor-specific and submitted to CDISC for the inclusion in future SDTM CT versions.

DTS Solution and SDTM Validation for Scenario 2

Data Structure

Data Structure according to SDTM requirements belonging to 4 different SDTM domains

Variable Name	Variable Label	Type	Length	Codelist or Format	SUPP--	Comments
Domain	Domain Abbreviation	Char	2			See VLM tab!
STUDYID	Study Identifier	Char	20			XXXXXXXX
SUBJID	Subject Identifier	Char	20			e.g. 1002010
TESTCD	Short Name of Measurement, Test or Examination	Char	8	(TESTCD)		See VLM tab!
TEST	Name of Measurement, Test or Examination	Char	40	(TEST)		See VLM tab!
BDAGNT	Binding Agent	Char	200	(ISBDAGT)		See VLM tab!
MRKSTR	Marker String	Char	200			See VLM tab!
TSTPNL	Test Panel	Char	200			See VLM tab!
CAT	Category	Char	200			See VLM tab!
ORRES	Result or Finding in Original Units	Char	200			See VLM tab!
ORRESU	Original Units	Char	15	(UNIT)		See VLM tab!
SPEC	Specimen Material Type	Char	20	(SPECTYPE)		See VLM tab!
METHOD	Method of Test or Examination	Char	30	(METHOD)		See VLM tab!
VISIT	Visit Name	Char	30			See Visit tab!

VLM

VLM shows content and mapping according to SDTM structure and terminology of parameters with their respective domain abbreviation in the first column

Domain	TESTCD	TEST	BDAGNT	MRKS TR	TSTPNL	CAT	ORRES	ORRESU	SPEC	METHOD
MB	MTB	Mycobacterium tuberculosis				TUBERCULOSIS SCREENING	Positive		BLOOD	MICROSCOPY
MB	MTB	Mycobacterium tuberculosis				TUBERCULOSIS SCREENING	Negative		BLOOD	MICROSCOPY
MB	MTB	Mycobacterium tuberculosis				TUBERCULOSIS SCREENING	Indeterminate		BLOOD	MICROSCOPY
LB	WBC	Leukocytes				HEMATOLOGY		10^9/L	BLOOD	AUTOMATED COUNT
IS	MBAB	Microbial-induced Antibody	HEPATITIS B VIRUS CORE ANTIGEN			SEROLOGY	Reactive		SERUM	ELECTROCHEMILUMINESCENCE
IS	MBAB	Microbial-induced Antibody	HEPATITIS B VIRUS CORE ANTIGEN			SEROLOGY	Non-reactive		SERUM	ELECTROCHEMILUMINESCENCE
CP	TLYCE	T-Lymphocytes		CD3+ ABS	A167 QTBNK Assay	IMMUNOPHENOTYPING		10^6/L	BLOOD	FLOW CYTOMETRY



DTS Solution and SDTM Validation for Scenario 2

SDTM Validation

Domain	TESTCD	TEST	BDAGNT	MRKS TR	TSTPNL	CAT	ORRES	ORRESU	SPEC	METHOD
MB	MTB	Mycobacterium tuberculosis				TUBERCULOSIS SCREENING	Positive		BLOOD	MICROSCOPY
MB	MTB	Mycobacterium tuberculosis				TUBERCULOSIS SCREENING	Negative		BLOOD	MICROSCOPY
MB	MTB	Mycobacterium tuberculosis				TUBERCULOSIS SCREENING	Indeterminate		BLOOD	MICROSCOPY
LB	WBC	Leukocytes				HEMATOLOGY		10^9/L	BLOOD	AUTOMATED COUNT
IS	MBAB	Microbial-induced Antibody	HEPATITIS B VIRUS CORE ANTIGEN			SEROLOGY	Reactive		SERUM	ELECTROCHEMILUMINESCENCE
IS	MBAB	Microbial-induced Antibody	HEPATITIS B VIRUS CORE ANTIGEN			SEROLOGY	Non-reactive		SERUM	ELECTROCHEMILUMINESCENCE
CP	TLYCE	T-Lymphocytes		CD3+ ABS	A167 QTBNK Assay	IMMUNOPHENOTYPING		10^6/L	BLOOD	FLOW CYTOMETRY



The VLM sheet of the DTS contains several SDTM domain destinations for the data and needs to be split up into individual sheets with a domain splitter tool before they can be run through the xpt converter tool in order to create an SDTM (like) domain to be validated through the Pinnacle 21 Community version.

DTS Solution and SDTM Validation for Scenario 2

SDTM Validation

Q2_Domain_Splitter

Import

File Name

DTS_Scenario2.xlsx

Select

Exit

Excel Sheets

modify prefix

Contact Information

Transfer Specifications

Data Structure

VLM

Visits

Controlled Terminology

LB
MB
IS
CP

DTS file is uploaded into the domain splitter tool

VLM sheet is selected

Domain splitter tool shows the individual SDTM domains

By clicking „Select“, a new window pops up with the individual columns/variables of the VLM sheet

DTS Solution and SDTM Validation for Scenario 2

SDTM Validation

Q2_Domain_Splitter

Import

File Name

DTS_Scenario2.xlsx

Select

Exit

Excel Sheets

Contact Information

Transfer Specifications

Data Structure

VLM

Visits

Controlled Terminology

Toplevel Window

Split

By clicking on the „Split“ button, an excel output file is generated with the individual SDTM domains from the VLM sheet documented in their own sheets

Reset

Select

Exit

Domain
TESTCD
TEST
TSTDTL
TSTOPO
BDAGNT
MRKSTR
TSTPNL
CAT
SCAT
ORRES
ORRESU
ORNRLO
ORNRHI
STRESC
STRESU
STNRC
STAT
REASND

DTS Solution and SDTM Validation for Scenario 2

SDTM Validation

1	MBDomain	MBTESTCD	MBTEST	MBBDAGNT	MBMRKSTR	MBTSTPNL	MBCAT	MBORRES	MBORRESU	MBSPEC	MBMETHOD
2	MB	MTB	Mycobacterium tuberculosis				TUBERCULOSIS SCREENING	Positive		BLOOD	MICROSCOPY
3	MB	MTB	Mycobacterium tuberculosis				TUBERCULOSIS SCREENING	Negative		BLOOD	MICROSCOPY
4	MB	MTB	Mycobacterium tuberculosis				TUBERCULOSIS SCREENING	Indeterminate		BLOOD	MICROSCOPY
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											

- The individual SDTM domains are shown in their own sheets.
- The variable names now include the two letter prefix of the individual SDTM domains.
- The sheets can now be prepared for the conversion to xpt files:
- Update variable names (remove two letter prefix from the domain column)
 - Remove variables that are not used in the individual SDTM domain
- Afterwards the same steps apply as in scenario 1.

DTS Solution and SDTM Validation for Scenario 3

Vendor Data Structure

Dataset does not follow any existing standards for structure and terminology:

- Data are provided horizontally
- Lab Tests might follow by their associated units
- No controlled Terminology

Study_ID	Patient_ID	MSI_High	TMB_score (10^6 bp)	TMB_category	Specimen	Test Method
XXXXXXXX	1001001	Detected	<i>numeric</i>	Low	DNA	PCR

DTS Solution and SDTM Validation for Scenario 3

Data Structure

Vendor's Data Specification on the left side

SDTM-Like Data Structure of the Genomics Findings domain on the right side

Vendor's Data Specification			Merck's Non-CRF Data Collection Standards						
Field Name	Type	Field description	Variable Name	Variable Label	Type	Length	Codelist or format	SUPP --	Comments
Study_ID	varchar	Study identifier from customer	STUDYID	Study Identifier	Char	20			XXXXXXXX
Patient_ID	varchar	Patient identifier from customer	SUBJID	Subject Identifier	Char	20			e.g. 1001001
Visit	varchar	Visit name from customer	VISIT	Visit Name	Char	20			See Visit tab!
MSI_High	varchar	MSI status indicates whether the sample is MSI-High "Detected" or "Not Detected"	GFTEST	Name of Genomic Measurement	Char	40	(TEST)		See VLM-Lookup tab!
TMB_score (10^6 bp)	double precision	TMP score	GFTEST	Name of Genomic Measurement	Char	40	(TEST)		See VLM-Lookup tab!
TMB_category	varchar	TMP score categories are "High" and "Low"	GFTEST	Name of Genomic Measurement	Char	40	(TEST)		See VLM-Lookup tab!
Specimen	varchar	Specimen Material	GFSPEC	Specimen Material Type	Char	20	(SPECTYPE)		See VLM-Lookup tab!
Test Method	varchar	Test Method	GFMETHOD	Method of Test or Examination	Char	30	(METHOD)		See VLM-Lookup tab!

DTS Solution and SDTM Validation for Scenario 3

VLM-Lookup

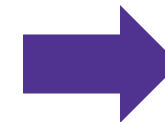
VLM-Lookup with Vendor Lab Test Name on the left side and their content and mapping according to SDTM structure and terminology on the right side

Vendor Lab Test Name	GFTESTCD	GFTEST	GFTSTDTL	GFORRES	GFORRESU	GFSPEC	GMETHOD
MSI_High	MICRISTB	Microsatellite Instability	MICROSATELLITE INSTABILITY OVERALL STATUS	HIGH		DNA	POLYMERASE CHAIN REACTION
MSI_High	MICRISTB	Microsatellite Instability	MICROSATELLITE INSTABILITY OVERALL STATUS	STABLE		DNA	POLYMERASE CHAIN REACTION
TMB_score (10^6 bp)	TMB	Tumor Mutation Burden	NORMALIZED NUMBER OF SEQUENCE VARIANTS	<i>numeric</i>	/10^6 bp	ctDNA	NEXT GENERATION TARGETED SEQUENCING
TMB_category	TMB	Tumor Mutation Burden	VARIANT SEQUENCE BURDEN INTERPRETATION	HIGH		ctDNA	NEXT GENERATION TARGETED SEQUENCING
TMB_category	TMB	Tumor Mutation Burden	VARIANT SEQUENCE BURDEN INTERPRETATION	LOW		ctDNA	NEXT GENERATION TARGETED SEQUENCING

DTS Solution and SDTM Validation for Scenario 3

SDTM Validation

Vendor Lab Test name	GFTESTCD	GFTEST	GFTSTDTL	GFORRES	GFORRESU	GFSPEC	GFMETHOD
MSI_High	MICRISTB	Microsatellite Instability	MICROSATELLITE INSTABILITY OVERALL STATUS	HIGH		DNA	POLYMERASE CHAIN REACTION
MSI_High	MICRISTB	Microsatellite Instability	MICROSATELLITE INSTABILITY OVERALL STATUS	STABLE		DNA	POLYMERASE CHAIN REACTION
TMB_score (10^6 bp)	TMB	Tumor Mutation Burden	NORMALIZED NUMBER OF SEQUENCE VARIANTS	<i>numeric</i>	/10^6 bp	ctDNA	NEXT GENERATION TARGETED SEQUENCING
TMB_category	TMB	Tumor Mutation Burden	VARIANT SEQUENCE BURDEN INTERPRETATION	HIGH		ctDNA	NEXT GENERATION TARGETED SEQUENCING
TMB_category	TMB	Tumor Mutation Burden	VARIANT SEQUENCE BURDEN INTERPRETATION	LOW		ctDNA	NEXT GENERATION TARGETED SEQUENCING



In scenario 3, only the SDTM part of the VLM lookup table is extracted for the SDTM alignment review. Afterwards, the same steps apply as in scenarios 1 or 2.

Conclusion

A Standardized Generic DTS Template for requesting non-CRF data in SDTM format from laboratories plays a crucial role in standardizing, streamlining, and improving data integration quality in clinical studies. By promoting consistency and compliance, this approach contributes to more efficient and effective clinical research outcomes.

Acknowledgements

We would like to thank our colleague **Francis de la Poza Arredondo**, who is responsible for the creation of the xpt converter and the domain splitter tool.



Questions