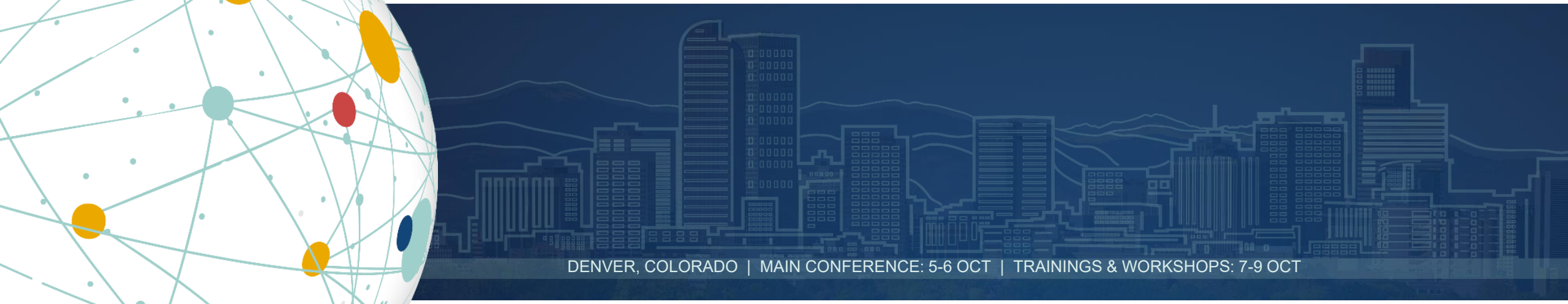


cdisc®

2026 US Interchange

THE FUTURE IS CONNECTED: STANDARDS AND AI POWERING DIGITAL TRANSFORMATION



Transforming Vendor Biomarker Data into CDISC Compliant SDTM Datasets

Devi Gohimukkula | Director, Global Clinical Data Standards, BeOne Medicines, USA

Saravanan Kandasamy | Senior Manager, Global Clinical Data Standards, BeOne Medicines, USA

Agenda

1. Biomarker Data Landscape
2. Challenges in Vendor Data Standardization
3. BeOne Standards-Driven Approach
4. Real-World DTS-to-SDTM Examples

Vendor biomarker data become difficult to standardize when standards are applied too late in the process

Speakers



Devi Gohimukkula, PhD is Director of Clinical Data Standards at BeOne Medicines, where he leads enterprise standards for protocol design, CRF development, vendor data integration, and SDTM transformation. An active member of the **CDISC SDS Team** and former **CDISC Pharmacogenomics Working Group** member, he is passionate about advancing standards-driven automation, data quality, and interoperability across clinical research

Disclaimer and Disclosures

The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC or BeOne Medicines, its affiliates, or employees.

Nothing in this presentation should be construed as an official statement of CDISC or BeOne Medicines.

Biomarkers: definition and key examples

A biomarker is a measurable biological characteristic used to understand disease, treatment response, safety, or patient selection.

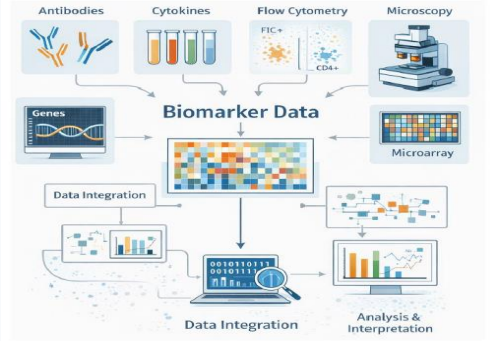
Core definition

Defined characteristic measured as an indicator of:

- normal biological processes
- pathogenic processes
- response to exposure or intervention

Not a direct patient-reported outcome or survival endpoint

Biomarker data modalities



Key biomarker examples

1 Molecular/genomic

- EGFR mutation
- ALK / ROS1 fusion
- BRAF V600E

2 Protein / tissue

- HER2 expression
- PD-L1 expression
- CLDN6 IHC status / H-score

3 Soluble / lab

- CA-125, PSA, AFP
- HbA1c
- Troponin

For vendor biomarker data transfer: define the biomarker name, specimen type, assay method, units/result format, allowable results, and CSR/analysis relevance before mapping to SDTM-ready structures.

EDC vs External Vendor Data Collection

Aspect	CRF (EDC) Data Collection	Vendor Data Collection
Collection design	Sponsor / CDISC-defined	Vendor database / assay design
Data structure	Predefined, standardized, reusable	Different by vendor; limited reuse across vendors
Data collected	Only data required for analysis	Core analysis results plus exploratory, QC, tracking, reconciliation, sample handling
Terminology	Controlled terminology wherever possible	Vendor-specific terms or free text; may vary by vendor
Result representation	Structured dropdowns / coded values	Variable formats, e.g., Positive, +, +ve, +++
Missing data	Clear distinction between not done / not available	Absence of results may be represented inconsistently
SDTM mapping	Predefined or planned upfront	Often determined after DTS is finalized

Design upstream to reduce downstream transformation burden

Why Vendor Data Are Difficult to Standardize?

1 Lack of standardization

Different variable names, result formats, units, and terminology across vendors

2 Proprietary codes

Vendor-specific values may not align directly to CDISC controlled terminology

3 Assay-specific structure

Flow cytometry, NGS, IHC, ctDNA, FISH, PCR, ELISA, and cytokine panels may require different SDTM handling.

4 Incomplete metadata

Missing specimen type, method, unit, sample ID, or collection date limits traceability.

5 Specimen reconciliation

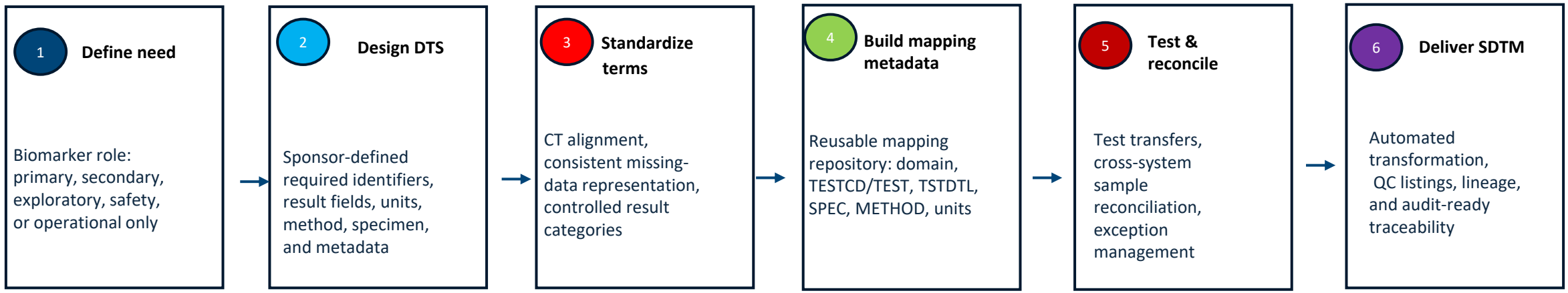
Sample collected in EDC may not match vendor result; vendor result may lack matching CRF sample record.

6 Late mapping decisions

If mapping starts after DTS finalization, transformation becomes reactive and study-specific.

More interpretation + more mapping decisions + more reconciliations = greater risk to analysis readiness and audit traceability

How to Move Vendor Data Toward CDISC Compliance



Governance layer

- Standard SME
- Biomarker Scientist
- Data management
- Statistical programming
- Vendor/lab
- QC/regulatory readiness

A practical upstream governance model connects science, data transfer design, metadata, SDTM mapping, and QC.

Where should standards intervene?

1	What should be defined before DTS finalization?	Define required identifiers, expected result fields, specimen/method/units, and missing-data conventions.
2	Which data are analysis-critical vs operational?	Separate CSR/endpoint-relevant data from QC, tracking, reconciliation, and sample handling fields.
3	Where can we standardize once and reuse?	Create reusable mapping file or database for recurring assays, vendor terms, TESTCD/TEST, TSTDTL, units
4	How do we preserve traceability	Maintain mapping metadata and reconciliation evidence from source through SDTM and down stream analysis

Early standards engagement turns vendor variability into governed SDTM transformation — without losing scientific flexibility

Standard data transfer specifications: What each vendor row must carry

The DTS organizes each record around identifiers, collection timing, assay metadata, results, and quality/status information

Subject & visit definition

STUDYID, SITEID, SUBJID, R_VISIT, R_TPT

Purpose: connect vendor result to study, site, subject, and protocol visit/time point

Sample Ids & timing

REFID, DTC, SMPRCDT, TESTDT

Purpose: preserve sample collection, receipt, and test timing for traceability and stability checks.

Assay/test metadata

R_CAT, R_SCAT, R_TESTCD, R_TEST, R_TSTDTL, R_SPEC, R_METHOD

Purpose: define panel/category, exact test, specimen, and assay method.

Result & unit

R_ORESS, R_ORRESU, ORNRLO, R_ORNHI

Purpose: capture original result, units, reference range limits

Status/exceptions/comments

STAT, REASND, SPCCND, NAM, TSTINF

Purpose: capture not-done status, reason, specimen condition, vendor/lab, and extra comments.

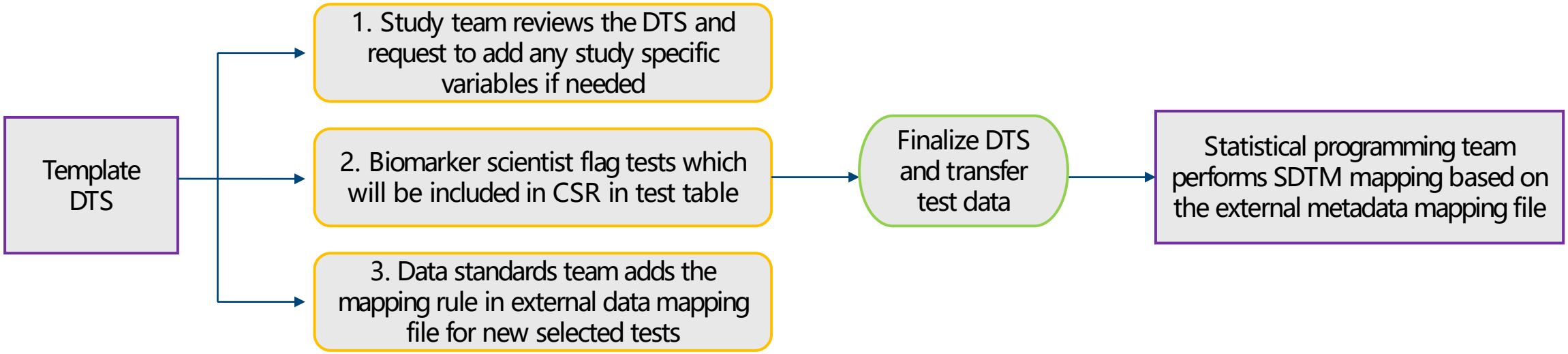
Genomic-only attributes

SYM, GENLOC, CHROM, SEQID, INHERT

Purpose: capture gene/variant/location/reference attributes when the biomarker data are genomic.

Required variables should remain in the study DTS; optional fields are included only when needed for the study or assay.

BeOne Operating Model: Standardize Once, Reuse Across Studies



External data mapping file:

- Current state : Maintained in excel
- Future state : Centralized metadata repository for governance, traceability, and automation

With the reusable mapping file in place, we are now exploring AI-assisted automation to apply these approved mapping rules and support more efficient vendor-data-to-SDTM transformation

Populating CSR Analysis Flag in DTS Expected Test Table

Used during DTS finalization to distinguish tests that definitely or possibly support CSR analysis

CORE RULE Flag Y when the test will definitely or possibly be used in CSR analysis. Flag N when it is not needed for analysis or logic is not deterministic. Data transfer from vendor to BeOne still occurs regardless of Y/N.

Decision logic for expected tests table

1 **Sample requisition details / sample management info / reconciliation / tracking** — Flag N

Not an actual lab test; mark N. Prefix examples: ADT / SMT, SM.

2 **Tests used for primary/secondary objectives of the study** — Flag Y / N

Mark Y only if explicitly listed in protocol / SoA for CSR-relevant analysis; otherwise, N (Some tests may not be used in the study, but may be provided by vendor as part of the larger test panel).

3 **Protocol-defined non-exploratory tests** — Flag Y

If identified in protocol non-exploratory analysis, mark Y. whether performed centrally or locally.

4 **Confirmatory tests used to validate subject's eligibility before randomization** — Flag Y / N

May likely be used in CSR to determine analysis population; obtain Stats / scientist / medical input.

5 **Specimen QC data used to validate results used in CSR** — Flag Y / N

Y only with deterministic programming logic and comments; N if uncertain or manual judgement needed.

Expected tests table: practical population pattern

R_CAT	R_TEST	R_ORRES	CSR Flag
PATIENT INFORMATION	Site/organ of biopsy:	ENLARGED LYMPH NODES,BONE MARROW	N
PATIENT INFORMATION	Fresh tissue is shipped in	10% formalin,70% ethanol	N
12p11.1-q11 CLL FISH	Reviewed by	Dr. E. Abdellatif,Dr. R. Ahmad,Dr.	N
SPECIMEN MANAGEMENT	PK Primary	RECEIVED,SITE DID NOT COLLECT,NOT	N
SPECIMEN MANAGEMENT	PK Backup	RECEIVED,SITE DID NOT COLLECT,NOT	N
SPECIMEN MANAGEMENT	Cytokine/Soluble protein 1	RECEIVED,SITE DID NOT COLLECT,NOT	N
BLOCK QC	Block Qc performed-CL	Free Text	N
BLOCK QC	Block technical QC-CL	Free Text	N
PD-L1 (SP263)	% Pos TC / Total Tumor Cells	0-100; <1	N
PD-L1 (SP263)	% Pos IC / Tumor Area	0-100; <1	N
PD-L1 (SP263)	Tumor area positivity (TAP) Raw Score	0-100; <1; Not evaluable	Y
PD-L1 (SP263)	PD-L1 TAP Status at ⩾1%	Negative, Positive, Not Evaluable	Y
PD-L1 (SP263)	Pathologist	free-text	N
PD-L1 (SP263)	Score Date	DD-MMM-YYYY	N
PD-L1 (SP263)	Comment	Free Text	N

Who confirms the flag?

Study medical / biomarker scientist / biostatistician: confirms CSR relevance for each test.

Study SP/data standards teams reviews the flag marked for analysis by cross-checking against protocol, SoA and CRF coverage in DTS.

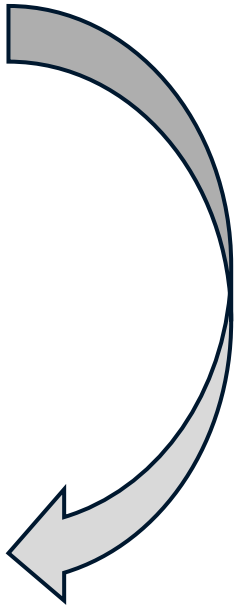
Standards team reviews the flag and adds mapping for the flagged rows in the mapping table.

cdisc®

Transforming vendor database to sponsor DTS structure

Vendor database structure

AssayName	AssayMeasure	SpecimenSource	Locus	SampleAmount	Mrd	Mrd1e4	Mrd1e5	Mrd1e6
B-cell 2.0 Tracking	MRD	PBMC	IGH	2153469.38	Positive	Positive	Positive	Positive



Sponsor DTS structure

	Panel Name / Category	Subcategory for Test	Vendor Test code	Vendor Test Name	Test Detail	Vendor - Specimen Type	Vendor - Method	Possible Results	Vendor Conventional units	CSR Analysis Flag? (Y/N)
	R_CAT	R_SCAT	R_TESTCD	R_TEST	R_TSTDTL	R_SPEC	R_METHOD	R_ORRES	R_ORRESU	IS_MAPPED
	MRD by NGS	MRD (TRACKING)	SampleAmount	SampleAmount		PERIPHERAL BLOOD MONONUCLEAR CELL	NEXT GENERATION SEQUENCING	2153469.38	Cells	Y
	MRD by NGS	MRD (TRACKING)	MRD	Minimal Residual Disease	VARIANT SEQUENCE BURDEN INTERPRETATION	PERIPHERAL BLOOD MONONUCLEAR CELL	NEXT GENERATION SEQUENCING	Positive Negative Indeterminate		Y
	MRD by NGS	MRD (TRACKING)	MRD	Minimal Residual Disease	Sensitivity of 10−4	PERIPHERAL BLOOD MONONUCLEAR CELL	NEXT GENERATION SEQUENCING	Positive Negative Indeterminate		Y
	MRD by NGS	MRD (TRACKING)	MRD	Minimal Residual Disease	Sensitivity of 10−5	PERIPHERAL BLOOD MONONUCLEAR CELL	NEXT GENERATION SEQUENCING	Positive Negative Indeterminate		Y
	MRD by NGS	MRD (TRACKING)	MRD	Minimal Residual Disease	Sensitivity of 10−6	PERIPHERAL BLOOD MONONUCLEAR CELL	NEXT GENERATION SEQUENCING	Positive Negative Indeterminate		Y

Vendor data structure variability, not the data itself, is the primary barrier to efficient SDTM conversion. Early alignment on standards and transfer requirements enables consistent, automated, and reusable transformations

Examples: Applying a common DTS Framework Across Biomarker Technologies

The next three examples demonstrate how vendor-specific biomarker data from diverse sources—including Histopathology (IHC), Genomics, and Flow Cytometry, can be transformed into CDISC SDTM-compliant datasets using a standardized DTS and a centralized global mapping repository. Despite scientific differences across biomarker technologies, a common standards framework enables consistent, traceable, and reusable SDTM transformations.

Different biomarker technologies. One DTS. One mapping repository. One standardized SDTM approach.

Example 1: Histopathology (IHC) Vendor Data to into Reusable CDISC SDTM standards

Vendor-provided source metadata

Panel Name / Category	Vendor Test code	Vendor Test Name	Vendor - Specimen Type	Vendor - Method	Possible Results	Vendor Conventional units	CSR Analysis Flag? (Y/N)
R_CAT	R_TESTCD	R_TEST	R_SPEC	R_METHOD	R_ORRES	R_ORRESU	IS_MAPPED
GPC3 IHC	CYTOMEM1+	% GPC3 CYTO-MEM >= 1+	Tumor tissue	IHC	0-100 <blank>	%	Y
GPC3 IHC	IHCSTATUS	GPC3 IHC Status	Tumor tissue	IHC	Positive Negative Unevaluable		Y
GPC3 IHC	CYTOMEM0	CYTOMEM 0	Tumor tissue	IHC	0-100	%	Y
GPC3 IHC	CYTOMEM1+	CYTOMEM 1+	Tumor tissue	IHC	0-100	%	Y
GPC3 IHC	CYTOMEM2+	CYTOMEM 2+	Tumor tissue	IHC	0-100	%	Y
GPC3 IHC	CYTOMEM3+	CYTOMEM 3+	Tumor tissue	IHC	0-100	%	Y
GPC3 IHC	CYTO/S	CYTOMEM H SCORE	Tumor tissue	IHC	0-300		Y
GPC3 IHC	GPCCOMM	GPC3 Comments	Tumor tissue	IHC	Freetext		N
GPC3 IHC	SCODAT	GPC3 Score Date	Tumor tissue	IHC	DDMMMYYYY		N

Standardized SDTM mappings used for analysis and regulatory submissions

Domain Identifier	Short Name of Measurement, Test, or Exam	Name of Measurement, Test, or Exam	Measurement, Test, or Examination Detail	Method of Test or Examination	Specimen Material Type	Original Units	Standard Units	Subcellular Location
DOMAIN	TESTCD	TEST	TSTDTL	METHOD	SPEC	ORRESU	STRESU	SCELOC
MI	GPC3	Glypican-3	PERCENT POSITIVE CELL WITH STAINING INTENSITY	IHC	TUMOR TISSUE	%	%	CYTOPLASM AND CELL MEMBRANE
MI	GPC3	Glypican-3	OVERALL STATUS	IHC	TUMOR TISSUE			CYTOPLASM AND CELL MEMBRANE
MI	GPC3	Glypican-3	STAINING INTENSITY 0	IHC	TUMOR TISSUE	%	%	CYTOPLASM AND CELL MEMBRANE
MI	GPC3	Glypican-3	STAINING INTENSITY 1+	IHC	TUMOR TISSUE	%	%	CYTOPLASM AND CELL MEMBRANE
MI	GPC3	Glypican-3	STAINING INTENSITY 2+	IHC	TUMOR TISSUE	%	%	CYTOPLASM AND CELL MEMBRANE
MI	GPC3	Glypican-3	STAINING INTENSITY 3+	IHC	TUMOR TISSUE	%	%	CYTOPLASM AND CELL MEMBRANE
MI	GPC3	Glypican-3	H SCORE TOTAL SCORE	IHC	TUMOR TISSUE			CYTOPLASM AND CELL MEMBRANE

Example 2: Genomic Vendor Data to into Reusable CDISC SDTM standards

Vendor-provided source metadata

Panel Name / Category	Vendor Test code	Vendor Test Name	Vendor - Specimen Type	Vendor - Method	Possible Results	Vendor Conventional units	CSR Analysis Flag? (Y/N)
R_CAT	R_TESTCD	R_TEST	R_SPEC	R_METHOD	R_ORRES	R_ORRESU	IS_MAPPED
DEL17	DEL17PAN	del17p total nuclei analysed	Blood	FISH	0-200		N
DEL11	DEL11QAN	del11q total nuclei analysed	Blood	FISH	0-200		N
DEL17	DEL17PP	Percent of del17p abnormal nuclei	Blood	FISH	0-100	%	N
DEL11	DEL11QP	Percent of del11q abnormal nuclei	Blood	FISH	0-100	%	N
DEL17	DEL17PR	del17p results	Blood	FISH	Normal, Abnormal, Indeterminate		Y
DEL11	DEL11QR	del11q results	Blood	FISH	Normal, Abnormal, Indeterminate		Y

Standardized SDTM mappings used for analysis and regulatory submissions

Domain Identifier	Short Name of Measurement, Test, or Exam	Name of Measurement, Test, or Exam	Measurement, Test, or Examination Detail	Category	Chromosome Identifier	Method of Test or Examination	Specimen Material Type	Original Units	Standard Units	Collected Specimen
DOMA	TESTCD	TEST	TSTDTL	CAT	GFCHROI	METHOD	SPEC	ORRESU	STRESU	CLSPEC
GF	DELETN	Deletion	OVERALL STATUS	CYTOGENETICS	17p	FISH	DNA			BLOOD
GF	DELETN	Deletion	OVERALL STATUS	CYTOGENETICS	11q	FISH	DNA			BLOOD

Example 3: Flowcytometry Vendor Data to into Reusable CDISC SDTM standards

Vendor-provided source metadata

Laboratory / Vendor Name	Panel Name / Category	Vendor Test code	Vendor Test Name	Vendor - Specimen Type	Vendor - Method	Possible Results	Vendor Conventional units	CSR Analysis Flag? (Y/N)
NAM	R_CAT	R_TESTCD	R_TEST	R_SPEC	R_METHOD	R_ORRES	R_ORRESU	IS_MAPPED
Q2 Solutions	A167 Panel	A167-1	CD3+	Whole Blood (Cyto-Chex BCT)	Flow Cytometry		%	Y
Q2 Solutions	A167 Panel	A167-12	CD4:CD8 Ratio	Whole Blood (Cyto-Chex BCT)	Flow Cytometry		Ratio	Y
Q2 Solutions	A167 Panel	A167-15	CD3+CD8+ ABS	Whole Blood (Cyto-Chex BCT)	Flow Cytometry		cells/ul	Y
Q2 Solutions	A167 Panel	A167-16	CD3-CD19+ ABS	Whole Blood (Cyto-Chex BCT)	Flow Cytometry		cells/ul	Y
Q2 Solutions	A167 Panel	A167-17	CD3-CD56+/CD16+ ABS	Whole Blood (Cyto-Chex BCT)	Flow Cytometry		cells/ul	Y
Q2 Solutions	A167 Panel	A167-18	CD14+ ABS	Whole Blood (Cyto-Chex BCT)	Flow Cytometry		cells/ul	N
Q2 Solutions	A167 Panel	A167-19	Lymph ABS	Whole Blood (Cyto-Chex BCT)	Flow Cytometry		cells/ul	N

Standardized SDTM mappings used for analysis and regulatory submissions

Domain Identifier	Short Name of Measurement, Test, or Exam	Name of Measurement, Test, or Exam	Category	Marker String	Method of Test or Examination	Specimen Material Type	Original Units	Standard Units
DOMAIN	TESTCD	TEST	CAT	CPMRKSTR	METHOD	SPEC	ORRESU	STRESU
CP	TLYLY	TLym/Lym	IMMUNOPHENOTYPING	CD3+	FLOW CYTOMETRY	BLOOD	%	%
CP	TLYHTLYC	TLym Help/TLym Cytx	IMMUNOPHENOTYPING	CD4+/CD8+	FLOW CYTOMETRY	BLOOD	RATIO	RATIO
CP	TLC	TLym Cytx	IMMUNOPHENOTYPING	CD3+CD8+	FLOW CYTOMETRY	BLOOD	10^6/L	10^6/L
CP	BLYCE	B-Lymphocytes	IMMUNOPHENOTYPING	CD3-CD19+	FLOW CYTOMETRY	BLOOD	10^6/L	10^6/L
CP	NKCE	Natural Killer Cells	IMMUNOPHENOTYPING	CD3-CD56+CD16+	FLOW CYTOMETRY	BLOOD	10^6/L	10^6/L

Key Takeaways: Why This Matters

- The biomarker technology may change, but the standardization challenge is often the same.
- Standardized DTS requirements create a common foundation for data transfer and integration.
- Reusable mapping metadata allow vendor-specific concepts to be standardized once and reused many times.
- Early governance improves traceability, consistency, and regulatory readiness.
- Standardization is not only about producing SDTM datasets today; it also creates the structured metadata and reusable rules needed for AI-assisted automation tomorrow.

Standardize Once, Reuse Across Studies and Enable AI-assisted Automation

cdisc®

Questions?

