phuse

phuse·share
connect·share·advance

Single Day Event

Modernizing Study Metadata Management with AI-Enabled Web Applications Development

PhUSE SDE Presentation, Shanghai, China

Stanley Wei

Novartis

phuse.global

November 7, 2025

Disclaimer

⚠️ Important Notice

The views and opinions expressed in this presentation are solely those of the author(s) and do not necessarily represent the official policy or position of Novartis.

Novartis makes no representations or warranties regarding the accuracy, completeness, or reliability of the content presented.

⚖️ Regulatory Compliance

All methodologies and approaches discussed should be evaluated for compliance with applicable regulatory requirements and organizational policies before implementation.

Agenda

1. **The Critical Role of Metadata in Clinical Trials**
   Understanding metadata as the backbone of clinical data operations
2. **Limitations of Traditional Metadata Tools**
   Challenges with Excel-based tools and manual processes
3. **Blueprint for Modern Web-Based Platform**
   Architecture principles and core functionalities
4. **Essential Features & Compliance**
   Regulatory requirements and validation frameworks
5. **AI-Powered Enhancements & Automation**
   Intelligent mapping, validation, and metadata-driven programming
6. **Future Outlook & Implementation**
   Vision, benefits, and next steps for modernization

Why Study Metadata Matters

1

Foundation of Clinical Data Operations

Provides essential structure and context, establishing the groundwork for everything that follows

2

Critical for Statistical Programming

Essential for SDTM, ADaM, and TLF programming activities in compliance with regulatory standards and process

3

Ensures Data Traceability

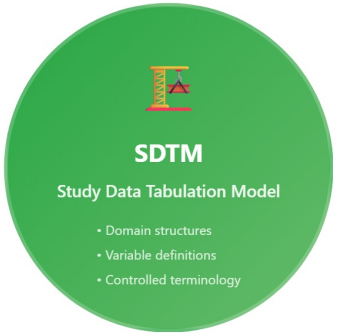
Maintains complete lineage from source data to final outputs, enabling full audit trails and compliance

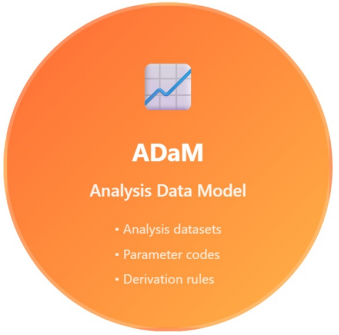
4

Enables Metadata-Driven Automation

Powers automated processes across the entire data lifecycle, reducing manual effort and ensuring consistency

Metadata in Clinical Trial Data Operations





Comprehensive metadata ecosystem supporting clinical data operations and analysis

Challenges with Excel-Based Metadata Tool

Why traditional spreadsheet approaches fall short in modern clinical environments

Poor User Experience

HIGH LEARNING CURVE

Difficult for first-time users with no intuitive help text or tooltips, creating steep learning curves and adoption barriers

No Concurrent Collaboration

TEAM BOTTLENECKS

Only one user can work at a time, creating bottlenecks and limiting team productivity in fast-paced environments

Version Control Nightmare

LOST PROGRESS

Difficult to track changes from old versions, leading to lost work, confusion, and potential compliance issues

No Quality Controls

ERROR-PRONE

Hard to debug with no edit checks or validation rules implemented to prevent mis-configurations and errors

Time-Consuming Manual Processes

LOW EFFICIENCY

Multiple manual steps required for simple tasks like adding/dropping variables, reducing efficiency and increasing risk

Why Metadata-Driven Automation Matters

Traditional Process

Manual Documents Review

Read protocol, SAP, etc. and extract metadata manually

1-2 weeks

Excel-Based Mapping

Create SDTM/ADaM mappings using spreadsheets and VBA macros

2-3 weeks

Manual Code Generation

Write SAS/R programs based on specifications

3-4 weeks

Multiple QC Rounds

Extensive quality control due to manual errors

2-3 weeks

Total Time: 8-12 weeks

High error risk, inconsistent quality

AI-Powered Automation

Metadata-Driven Process

AI Document Analysis

Automated extraction of metadata and key information from protocol, SAP, etc.

1-2 days

Intelligent Mapping

AI-suggested SDTM/ADaM mapping specifications based on extracted metadata and standards

2-3 days

Auto-Code Generation

Automatically metadata-driven code generation for datasets and TLFs

3-5 days

Automated Validation

Built-in quality checks and consistency validation

1-2 days

Total Time: 1-2 weeks

Consistent quality, reduced errors

Modern Study Metadata Platform Blueprint

Comprehensive solution addressing all Excel limitations

Enhanced User Experience

- Concurrent access and collaborative editing capabilities
- Friendly edit, filter, query, and highlight functions
- Integrated help text, tooltips, and online manual
- Lower SME involvement required for routine tasks

Data Quality & Validation

- In-line validations implemented when applicable
- Controlled codelist columns to reduce typos and unexpected values
- Built-in cross-table checks, queries, and duplicate avoidance
- Extensible rules engine with validation reports

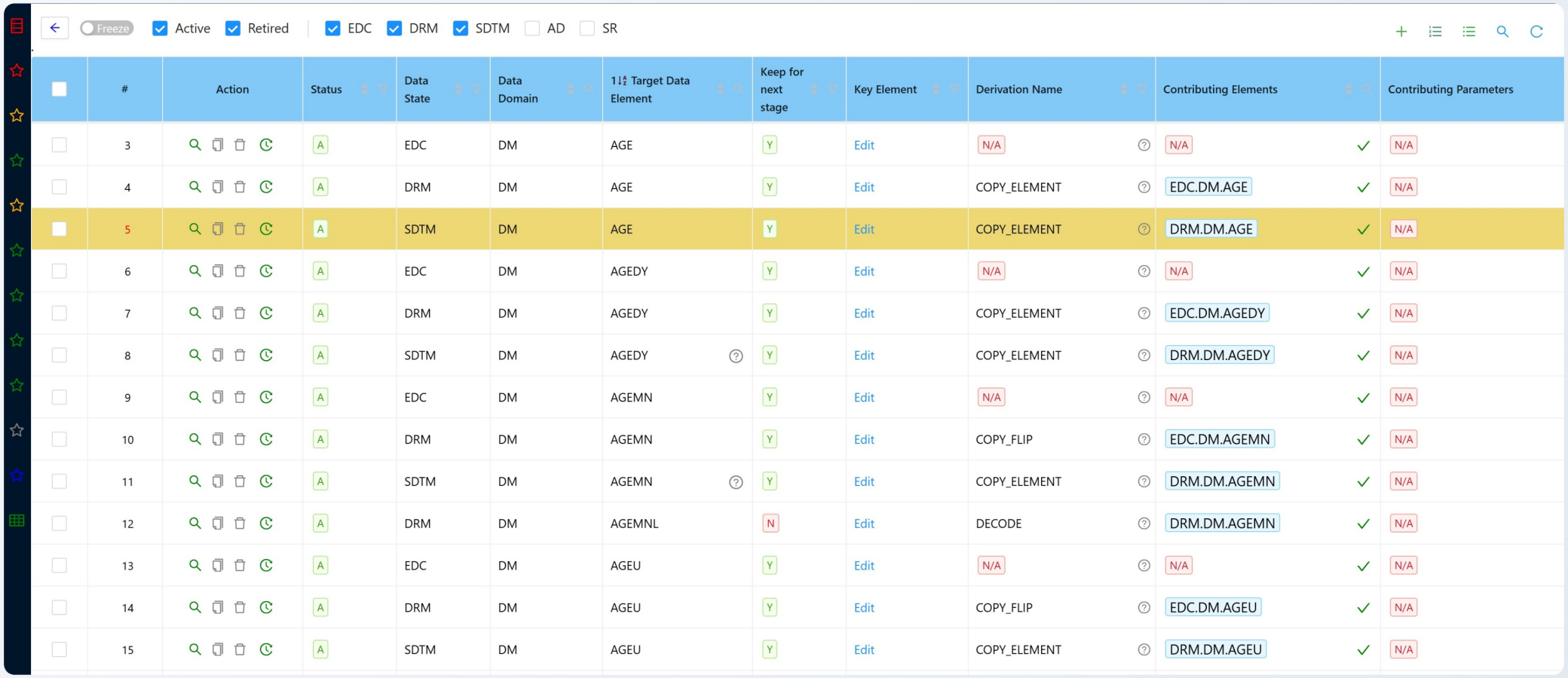
Governance & Control

- Centralized study metadata repository
- Role-based edit modes with granular permissions
- Built-in cell-level track changes and audit trails
- Reduces configuration traps identified only after SMT execution

Lifecycle Management

- Row-level lifecycle: Initialized → Ongoing/Draft → Under Review
- Lock/Final and Archive states with proper transitions
- Summary reports of changes and status tracking
- Email notifications for status changes and approvals

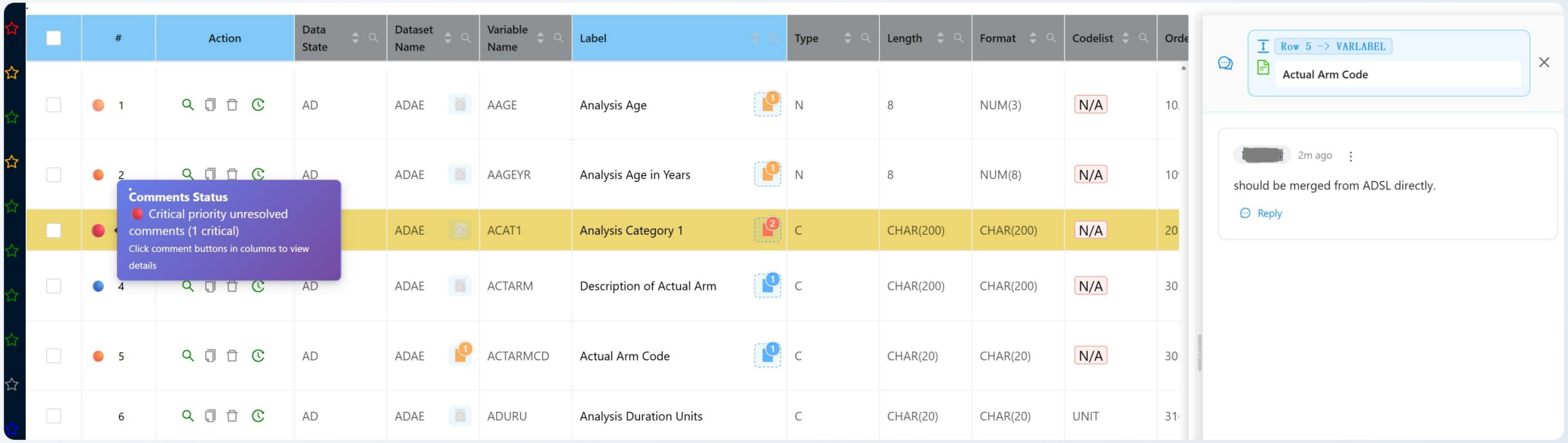
Centralized Metadata Repository View



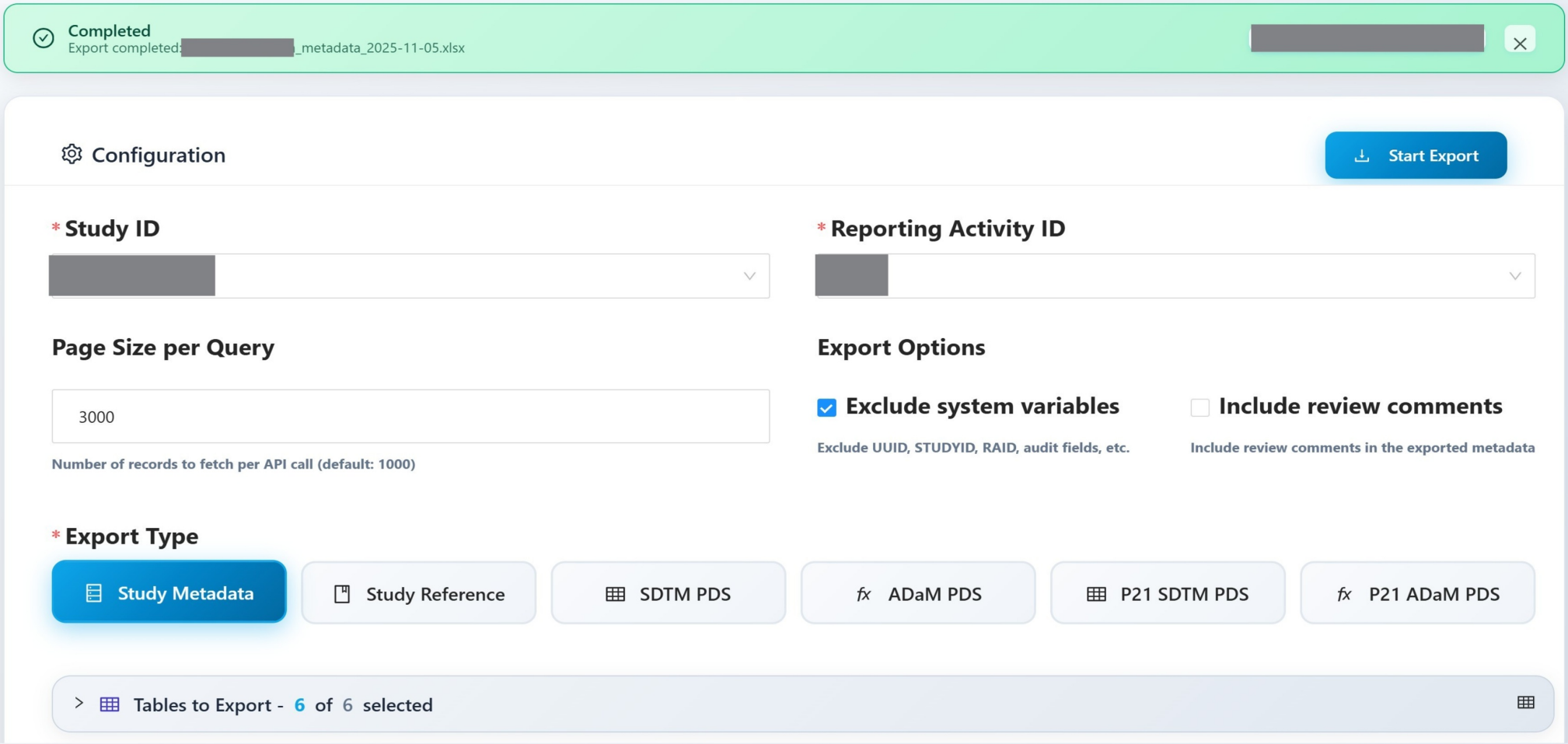
Freeze | Active | Retired | EDC | DRM | SDTM | AD | SR

#	Action	Status	Data State	Data Domain	Target Data Element	Keep for next stage	Key Element	Derivation Name	Contributing Elements	Contributing Parameters
3		A	EDC	DM	AGE	Y	Edit	N/A	N/A	N/A
4		A	DRM	DM	AGE	Y	Edit	COPY_ELEMENT	EDC.DM.AGE	N/A
5		A	SDTM	DM	AGE	Y	Edit	COPY_ELEMENT	DRM.DM.AGE	N/A
6		A	EDC	DM	AGEDY	Y	Edit	N/A	N/A	N/A
7		A	DRM	DM	AGEDY	Y	Edit	COPY_ELEMENT	EDC.DM.AGEDY	N/A
8		A	SDTM	DM	AGEDY	Y	Edit	COPY_ELEMENT	DRM.DM.AGEDY	N/A
9		A	EDC	DM	AGEMN	Y	Edit	N/A	N/A	N/A
10		A	DRM	DM	AGEMN	Y	Edit	COPY_FLIP	EDC.DM.AGEMN	N/A
11		A	SDTM	DM	AGEMN	Y	Edit	COPY_ELEMENT	DRM.DM.AGEMN	N/A
12		A	DRM	DM	AGEMNL	N	Edit	DECODE	DRM.DM.AGEMN	N/A
13		A	EDC	DM	AGEU	Y	Edit	N/A	N/A	N/A
14		A	DRM	DM	AGEU	Y	Edit	COPY_FLIP	EDC.DM.AGEU	N/A
15		A	SDTM	DM	AGEU	Y	Edit	COPY_ELEMENT	DRM.DM.AGEU	N/A

Comments & Collaboration View



Export & Download View



Metadata-Driven Dummy Data Factory

A concrete example of how structured metadata enables intelligent dummy data generation

Variable	Label	Type	Length	Controlled Terms
STUDYID	Study Identifier	Char	12	e.g., "ABC-123-001"
DOMAIN	Domain Abbreviation	Char	2	"AE"
USUBJID	Unique Subject Identifier	Char	20	e.g., "ABC-123-001-001"
AESEQ	Sequence Number	Num	8	1, 2, 3...
AETERM	Reported Term for AE	Char	200	Free text (realistic symptoms)
AEDECOD	Dictionary-Derived Term	Char	200	MedDRA PT terms
AESEV	Severity/Intensity	Char	20	MILD, MODERATE, SEVERE
AESER	Serious Event	Char	1	Y, N
AESTDTC	Start Date/Time	Char	19	ISO 8601 format
AEENDTC	End Date/Time	Char	19	ISO 8601 format

Metadata-Driven Dummy Data Factory

Key Insight: Metadata from multiple domains (DM, EX, MH) drives intelligent, medically-realistic data generation with proper CDISC compliance and MedDRA coding

Metadata Repository

Demographics (DM)

- Age: 68 years
- Sex: Male
- Race: Caucasian

Concomitant Medication (CM)

- Drug Class: Cardiovascular
- Duration: 180 days
- Dose: 50 mg QD

Medical History (MH)

- Hypertension
- Type 2 Diabetes

Metadata-Driven Algorithm

1. Age 68 → ↑2.5x Serious AE Risk
2. Male → ↑1.0x MI Risk
3. CV Drug → ↑2.0x Hypertension
4. Duration 180d → Chronic Events
5. MedDRA Mapping Applied

Generated AE Data

SERIOUS AE — Cardiac Disorders

AETERM:	Myocardial Infarction
AESEV:	SEVERE
AESER:	Y (Hospitalization)
AESTDTC:	2024-03-15
AEDECOD:	Myocardial infarction
AEBODSYS:	Cardiac disorders
Duration:	21 days (90% resolution)

Metadata-Driven Dummy Data Factory

DM - Demographics × AE - Adverse Events × VS - Vital Signs × CM - Concomitant Medications × MH - Medical History × LB - Laboratory Test Results ×

30

▶ Generate

Select domains and click Generate to create CDISC SDTM data

AI-Powered Solutions

Natural Language

Built-in LLM chatbot to ask questions about metadata and get instant answers, instructions and even authorized actions

Intelligent Mapping & Logic Derivation

AI suggests optimal variable mappings & logics based on historical data patterns and study context

Auto-Generation

Generate SDTM/ADaM specifications and dummy data with various format from study metadata automatically

Smart Validation & Compliance

Built-in compliance checks and prevents common metadata errors in a batch mode, ensuring high data quality

Technology Stack Overview

Frontend Framework

Vue 3.x + Ant Design Responsive Design
Composition API

TypeScript
Type Safety

Styling & Animation

CSS3
Animations & Grid

Canvas API
Dynamic Effects

Backend Technology

Python + MySQL/MariaDB
Relational Database

Redis
In-Memory Cache

Build Tools

Vite
Fast Build Tool

Node.js
Runtime Environment

Q&A and Discussion

Stanley Wei

Novartis

Associate Director, Statistical Programming, AQS

zhijun.wei@novartis.com

CT-FUN

Scan to follow CT-FUN

Thank You!

Questions and Comments Welcome