



cdisc® 2026 Europe Interchange

THE FUTURE IS CONNECTED: STANDARDS AND AI POWERING DIGITAL TRANSFORMATION



MILAN, ITALY | MAIN CONFERENCE: 20-21 MAY | TRAININGS & WORKSHOPS: 18, 19, & 22 MAY

LLM-powered end-to-end clinical data automation: standardized eCRFs to SDTM, ADaM, and TLF with full traceability

Anja Jazbec, Standards Developer, Novo Nordisk
Pankaj Attri, Pre-Sales Solutions Architect, Global Health and Life Sciences Advisory, SAS
Anette Jakobsen, Principal Industry Advisor, Global Health and Life Sciences Advisory, SAS

Speakers



Anja Jazbec
Standards Developer
Novo Nordisk



Anette Dalsgaard Jakobsen
Life Science Senior Industry Advisor
SAS



Pankaj Attri
Solutions Architect
SAS



Agenda

1. Introduction

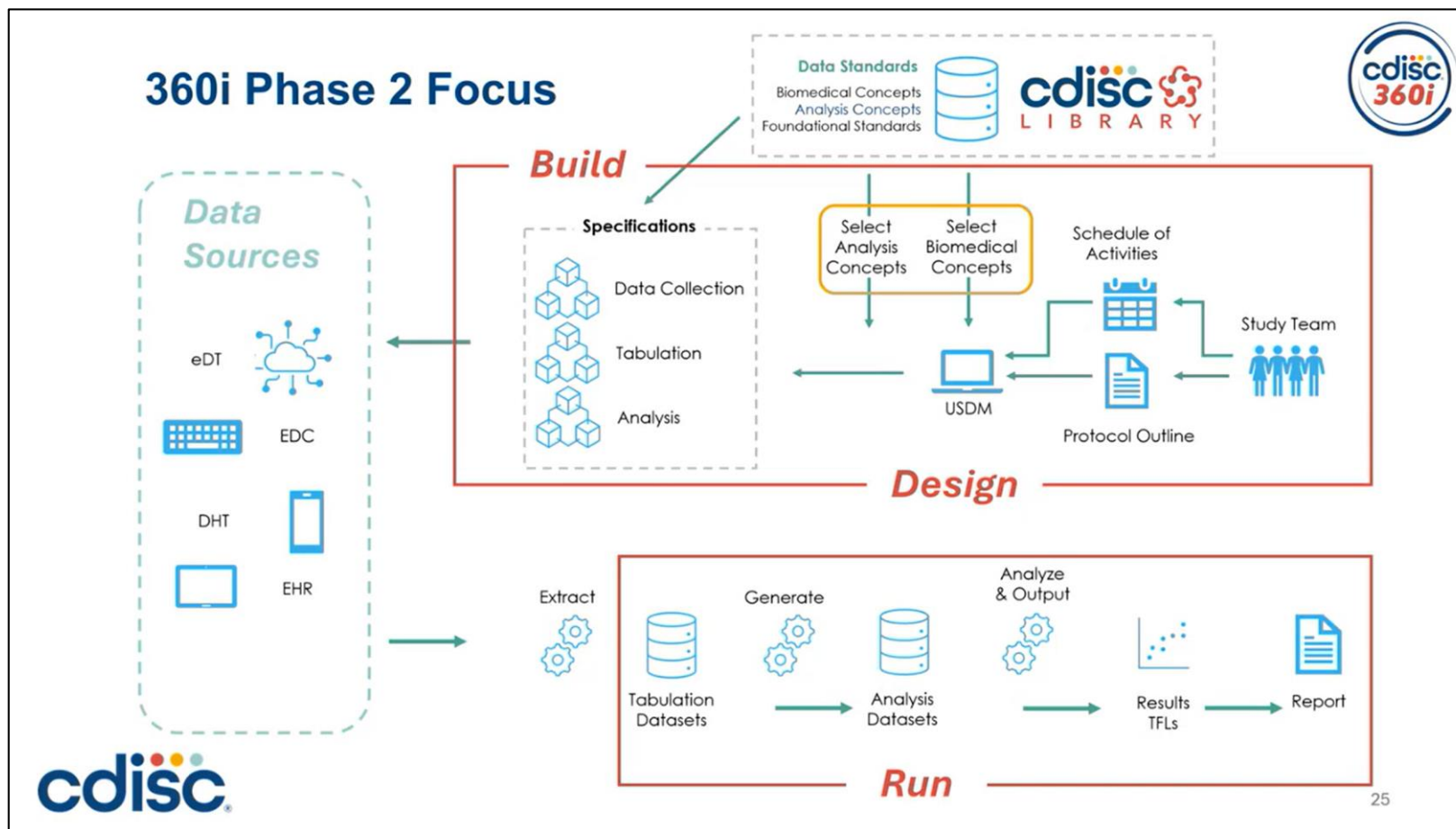
Automation in Clinical Trial workflows

2. Solutions

Generative AI in Clinical Trial workflows

3. Business Impact

Operationalizing 360i vision



Accelerated, standards-driven eCRF design with prequalified metadata and controlled terminology, ensuring data are “SDTM-ready” by design for the RAW→SDTM pipeline.

Automated transformation from RAW to SDTM, ADaM, and TLFs with end-to-end lineage for audit-ready traceability.

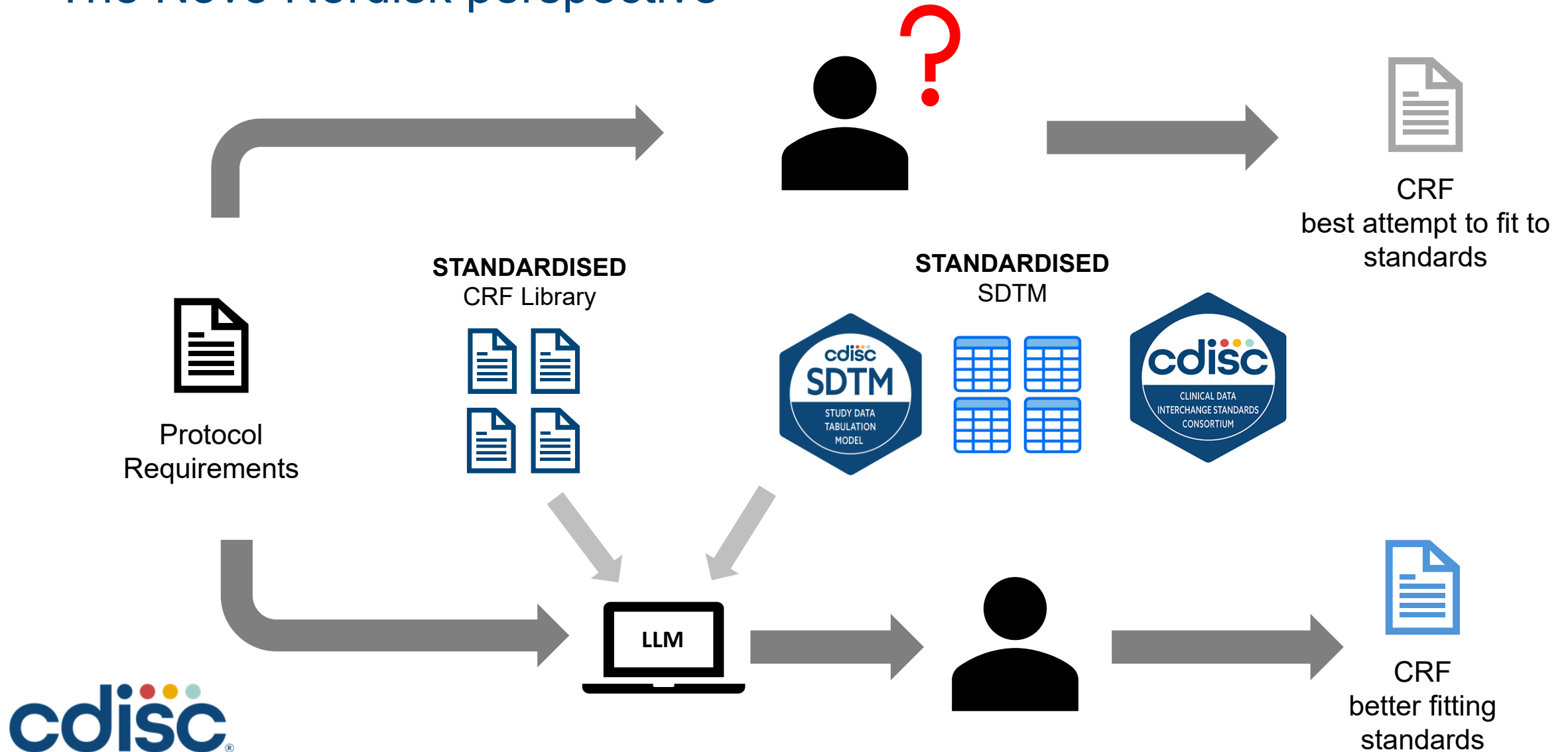


LLM Powered eCRF Builder

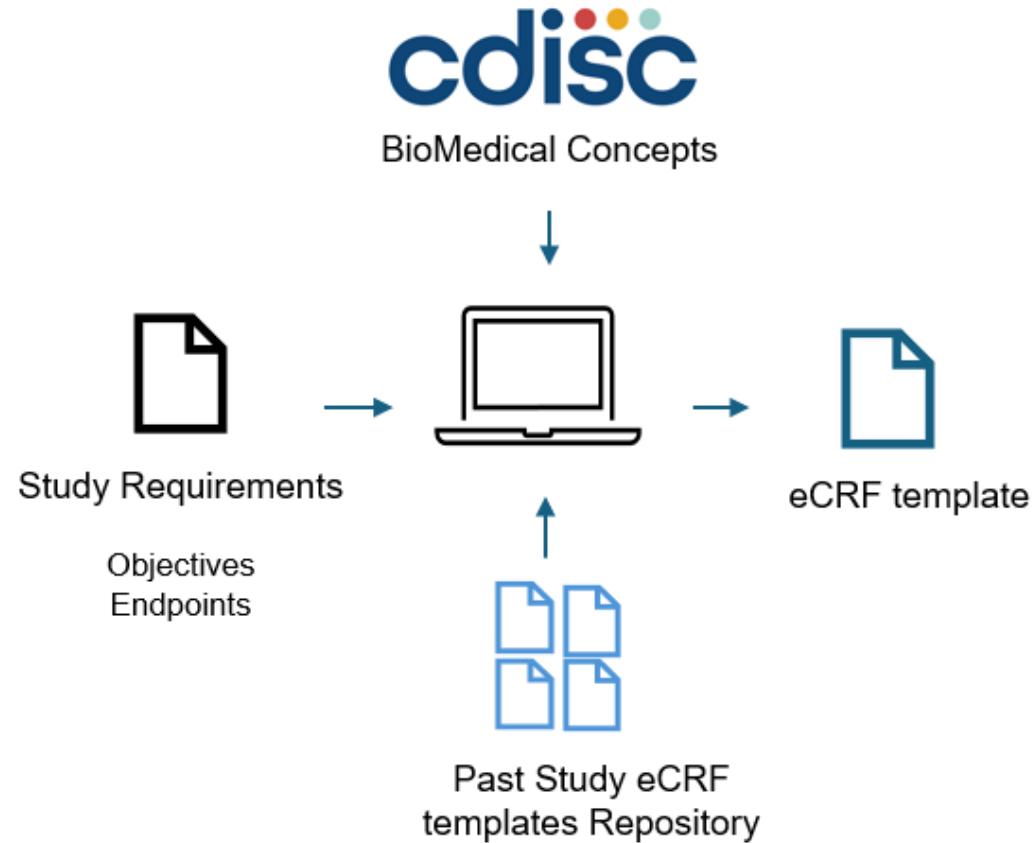
Using eCRF Templates and CDISC BioMedical Concepts Repositories

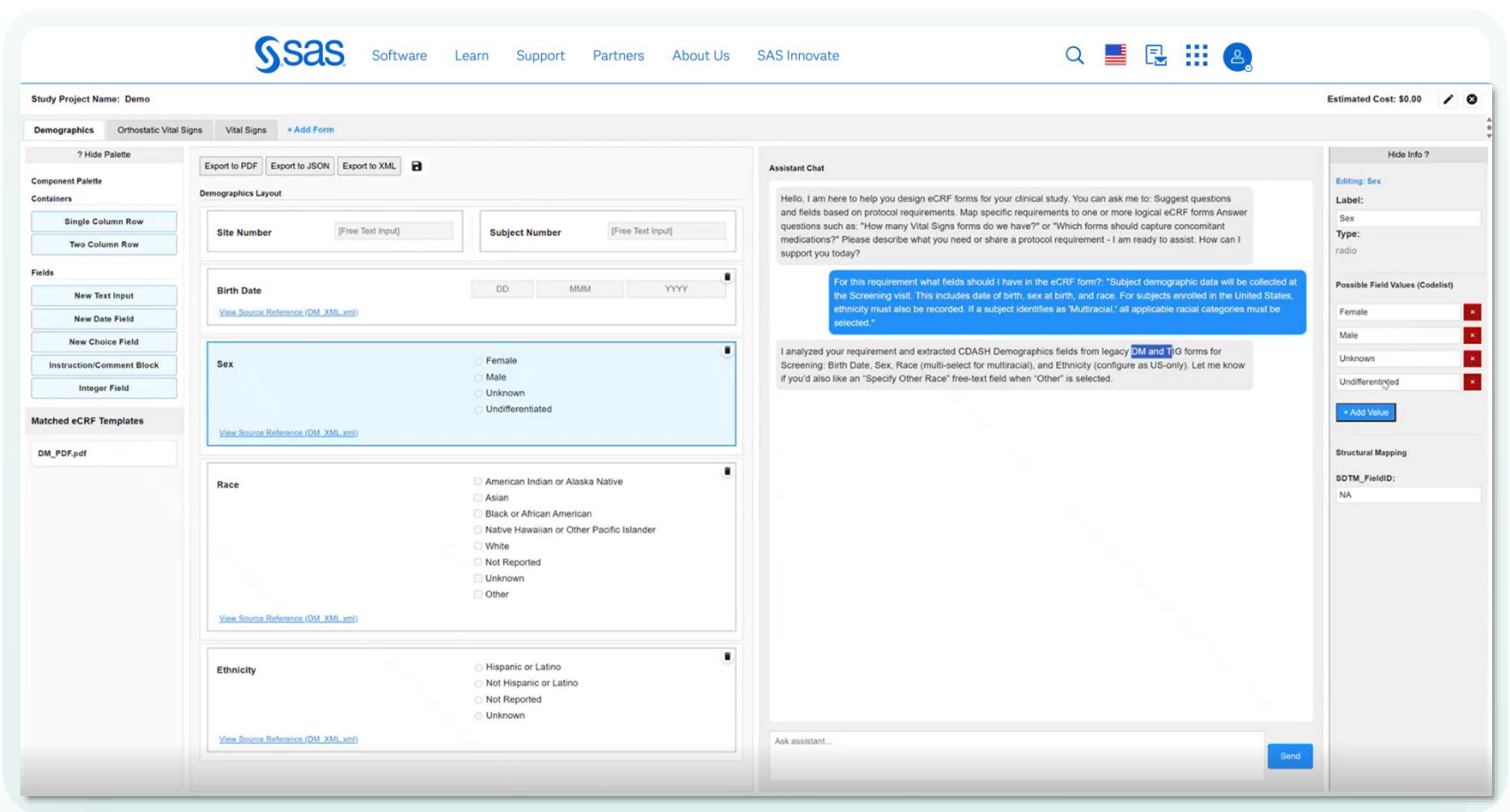
Challenges of standardisation

The Novo Nordisk perspective



AI Assisted eCRF form Creation





AI Assisted eCRF Creation - Prototype

Clinical Trials eCRF Build Automation

1

Map Requirements to eCRFs

Upload Protocol, other digital documents, OR copy/paste study requirements.

ome, sinpan

ClinAssist - LLM Powered eCRF Builder

Studies

Study Code	Modified Date
CS-118	1/29/2026
CS-115	1/28/2026
CS-116	1/28/2026

Create New Clinical Study Project

Study Details

Study Requirements

Retrieval Settings

Upload or paste study protocol requirements.
You can either upload your study protocol document (PDF/Word) for automatic extraction, or manually copy-paste the requirements into the field below.
Once ready, click the "Extract & Map to eCRFs" button to analyze the text and intelligently group related requirements into individual eCRF forms.

Study Protocol (File Upload)

File

Vital signs will be measured after the subject has been resting in a supine position for at least 5 minutes. Measurements include heart rate, respiratory rate, and oral body temperature. For heart rate, if an abnormality is noted, a repeat measurement must be taken after an additional 5 minutes of rest. Subject demographic data will be collected at the Screening visit. This includes date of birth, sex at birth, and race. For subjects enrolled in the United States, ethnicity must also be recorded. If a subject identifies as 'Multiracial,' all applicable racial categories must be selected. Orthostatic Vital Signs must be assessed at the Screening and Day 1 visits. The

Extract & Map to eCRFs

Protocol Requirements to eCRF form match

Requirement

+ Add Row

Settings *

☐ eCRF Form Templates Repository

☐ CDISC BioMedicalConcepts Repository

☒ Both

Start New Project

Clinical Trials eCRF Build Automation

2

Edit and Export eCRFs
Edit metadata, add/edit new fields to eCRF form, export to PDF/JSON/XML.

The screenshot displays the CDISC eCRF Build Automation interface for a study named "Demo". The top navigation bar includes tabs for "Demographics", "Orthostatic Vital Signs", and "Vital Signs", along with a "+ Add Form" button. The "Demographics" tab is active, showing a form layout with fields for "Site Number", "Subject Number", "Birth Date", "Sex", "Race", and "Ethnicity". The "Sex" field is highlighted with a red box, and its "Possible Field Values (Codelist)" is shown on the right, including "Female", "Male", "Unknown", and "Undifferentiated". The "Export to PDF", "Export to JSON", and "Export to XML" buttons are also highlighted with a red box. The "Assistant Chat" on the right provides guidance on form design and data collection. The "Structural Mapping" section on the far right shows the "SDTM_FieldID" as "NA".

Study Project Name: Demo Estimated Cost: \$0.00

Demographics Orthostatic Vital Signs Vital Signs + Add Form

? Hide Palette

Component Palette

Containers

- Single Column Row
- Two Column Row

Fields

- New Text Input
- New Date Field
- New Choice Field
- Instruction/Comment Block
- Integer Field

Matched eCRF Templates

- DM_PDF.pdf

Demographics Layout

Site Number [Free Text Input] Subject Number [Free Text Input]

Birth Date DD MMM YYYY [View Source Reference \(DM XML.xml\)](#)

Sex ☐ Female ☐ Male ☐ Unknown ☐ Undifferentiated [View Source Reference \(DM XML.xml\)](#)

Race ☐ American Indian or Alaska Native ☐ Asian ☐ Black or African American ☐ Native Hawaiian or Other Pacific Islander ☐ White ☐ Not Reported ☐ Unknown ☐ Other [View Source Reference \(DM XML.xml\)](#)

Ethnicity ☐ Hispanic or Latino ☐ Not Hispanic or Latino ☐ Not Reported ☐ Unknown [View Source Reference \(DM XML.xml\)](#)

Assistant Chat

Hello, I am here to help you design eCRF forms for your clinical study. You can ask me to: Suggest questions and fields based on protocol requirements. Map specific requirements to one or more logical eCRF forms Answer questions such as: "How many Vital Signs forms do we have?" or "Which forms should capture concomitant medications?" Please describe what you need or share a protocol requirement - I am ready to assist. How can I support you today?

For this requirement what fields should I have in the eCRF form?: "Subject demographic data will be collected at the Screening visit. This includes date of birth, sex at birth, and race. For subjects enrolled in the United States, ethnicity must also be recorded. If a subject identifies as 'Multiracial,' all applicable racial categories must be selected."

I analyzed your requirement and extracted CDASH Demographics fields from legacy DM and TIG forms for Screening: Birth Date, Sex, Race (multi-select for multiracial), and Ethnicity (configure as US-only). Let me know if you'd also like an "Specify Other Race" free-text field when "Other" is selected.

Ask assistant... Send

Hide Info ?

Editing: Sex

Label: Sex

Type: radio

Possible Field Values (Codelist)

- Female
- Male
- Unknown
- Undifferentiated

+ Add Value

Structural Mapping

SDTM_FieldID: NA



Introducing Generative AI in Clinical Data Flow

Challenges in Current State of Clinical Trial Biometrics



Complex Clinical trials

- More trials globally
- Intricate data sources
- Traceability requirements



Manual data mapping coding

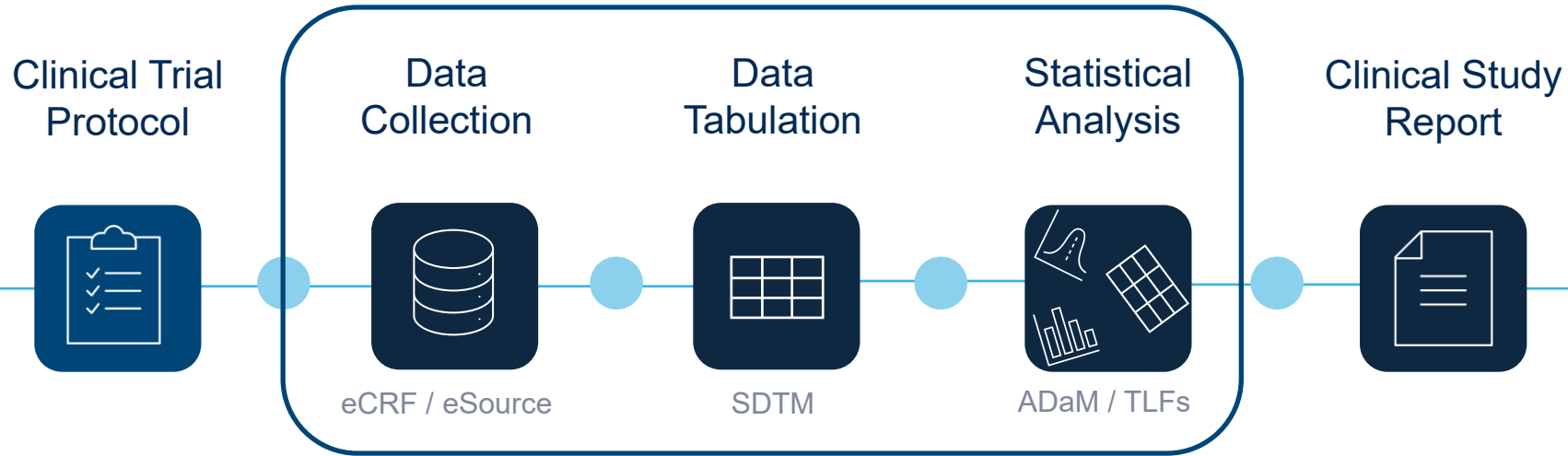
- Even with CDISC standards
- Time consuming
- Resource intensive



Double Programming

- Repetitive coding tasks
- Impacts submission deadlines

End to End Clinical Data Flow



Raw data to SDTM format

SDTM to ADaM

Creation of define.xml

Creation of TLFs



Clinical Data Flow (CDF) Prototype

- Will be part of SAS Clinical Acceleration

Features



Data Mapping

- Auto generate data transformation rules. (Raw -> SDTM -> ADaM)
- Investigate documents such as Study Protocols, SAP, TLF Shells to generate mapping specifications.



Code Generation

- SAS code generation with error feedback loop.
- Ensure Cyclic dependencies are taken care of when generating datasets.



Traceability

- Lineage view to track source to target data transformation.
- Source data coverage, missed variables reports.

CDF Prototype - Create Project for Study

Create Project for a Clinical Study

- Select data , program, output folder paths
- Select
 - Implementation Guide version
 - Domains
- Choose LLMs

This screenshot shows the 'Project Details' tab of the 'Create New Clinical Study Project' dialog. The 'Project Name' field is filled with 'Study-ONCO_IMMUNO-25'. The 'Project Description' field is empty. The 'Base Directory' field is filled with '/all_studies/2025/onco_immuno'. The 'Raw Data Folder' field is filled with 'raw-data'. The 'SDTM Folder' field is filled with 'tabulations-s'. The 'ADaM Folder' field is filled with 'analysis-ad'. The 'Programs Folder' field is filled with 'programs'. The 'Outputs Folder' field is filled with 'outputs'. The 'LLM Specs' tab is selected.

This screenshot shows the 'Data Standardization' tab of the 'Create New Clinical Study Project' dialog. The 'Raw to SDTM conversion needed?' checkbox is checked. The 'SDTM Datasets Folder Name' field is filled with 'tabulations-sdtm'. The 'Implementation Guide' field is filled with 'SDTMIG v3.4'. The 'Select or type SDTM Domains' section shows a grid of domain buttons: AE, BE, CE, CO, CV, DM, DV, EG, and EX. The 'Search or type domain' field is empty. The 'Leave blank if you want to select all domains.' text is present.

This screenshot shows the 'LLM Specs' tab of the 'Create New Clinical Study Project' dialog. The 'Base LLM' field is filled with 'GPT-5o'. The 'Code LLM' field is filled with 'GPT-5o'. The 'Assistant LLM' field is empty. The 'Set Temperature (0-1)' field is filled with '0.0'. The 'Start New Project' button is visible at the bottom right.

CDF Prototype - Project Details Screen

Welcome, sinpan

Clinical Data Flow AI Assistant

Project Name: Study-ONCO_IMMUNO-2025

Estimated Cost: \$0.00

Clinical Data Transformation Steps

Convert Raw Data to SDTM standards

Start here as the first step. You can choose to upload a Study Design document. Once done press Start next to 'Identify..' to get SDTM datasets, variables, and mapping logic. Once satisfied, click Start next to 'Generate..' to produce SDTM Datasets & SAS code.

Study Design:

Choose File

No file chosen

Instructions:

Optional instructions for the study protocol...

Identify SDTM Datasets:

Start

Macros File:

Choose File

No file chosen

Macros instructions:

Optional instructions for the SDTM macros file...

Generate SDTM Datasets:

Start

State:

In-Process

Extract TLF shells logic

Upload the TLF shells document, and press Start to extract TLF shell details. When content is satisfactory, select "Approved" in status to proceed to the next step.

TLF Shells:

Choose File

No file chosen

Instructions:

Optional instructions for TLF shells...

Extract

Agent outputs

SDTM Datasets

SDTM Code

ADaM Datasets

ADaM Code

TLF Logic

TLF Code

Domain

Variable

Type

Description

Mapping Rule

Data Source

Confidence

Status

No data available

Clinical Data Flow Steps

No SDTM datasets available

Project details

- Steps on the Left
- Agent (LLM) outputs on the right

CDF Prototype – SDTM Mapping

Agent Outputs

SDTM Datasets SDTM Code ADaM Datasets ADaM Code TLF Logic TLF Code

🔍 ✎ ⬇ ↺

Domain	Variable	Type	Description	Mapping Rule	Data Source	Confidence	Status
AE	AEACN	Expected	Describes changes to the study treatment as a result of the event. (Examples: 'DRUG WITHDRAWN', 'DOSE NOT CHANGED')	Map from ADVERSE_EVENTS.ACTION_TAKEN. ACTION_TAKEN appears to represent action taken with study treatment; convert to standard AEACN	ADVERSE_EVENTS.ACTION_TAKEN	Medium	⌚
AE	AEACNDEV	Permissible	An action taken with a device as the result of the event.	No device-action variable present in raw data. Set to Null.	Not available in raw data. Set to Null.	Low	⌚
AE	AEACNOTH	Permissible	Describes other actions taken as a result of the event that are unrelated to dose adjustments of study treatment.	No explicit 'other action' free-text variable present in raw ADVERSE_EVENTS. If ADVERSE_EVENTS.ACTION_TAKEN includes free	ADVERSE_EVENTS.ACTION_TAKEN (if contains non-dose-	Low	⌚
AE	AEBODSYS	Expected	Dictionary derived Body System or Organ Class used by the sponsor from the coding dictionary.	Map from ADVERSE_EVENTS.AE_BODSYS (sponsor-level body system/SOC).	ADVERSE_EVENTS.AE_BODSYS	High	⌚
AE	AEDECOD	Required	Dictionary-derived text description of AETERM (e.g., MedDRA PT).	Map directly from ADVERSE_EVENTS.AE_DECOD (dictionary-derived term). Sponsor to provide dictionary name/version in Define-XML.	ADVERSE_EVENTS.AE_DECOD	High	⌚
AE	AEDUR	Permissible	Collected duration and unit of an adverse event (only if collected on CRF and not derived).	No explicit duration collected field in ADVERSE_EVENTS. Derive duration from ADVERSE_EVENTS.START_DATE and END_DATE	ADVERSE_EVENTS.START_DATE, ADVERSE_EVENTS.END_DATE	Medium	⌚
AE	AEENDTC	Expected	End date/time of the adverse event represented in ISO 8601 character format.	Map from ADVERSE_EVENTS.END_DATE. Convert to ISO 8601 if needed. If blank in raw data, set to Null.	ADVERSE_EVENTS.END_DATE	High	⌚
AE	AEENDY	Permissible	Study day of end of event relative to the sponsor-defined RFSTDTC.	Calculate as (date(AEENDTC) - date(DM.RFSTDTC)) + 1. Use ADVERSE_EVENTS.END_DATE for AEENDTC and DM.RFSTDTC as reference. If	ADVERSE_EVENTS.END_DATE, DM.RFSTDTC	Medium	⌚
AE	AEENRF	Permissible	Describes the end of the event relative to the sponsor-defined reference period.	Not explicitly provided in raw ADVERSE_EVENTS. Could be derived if sponsor defines RFSTDTC/RFENDTC and event end date; default to	ADVERSE_EVENTS.END_DATE, DM.RFSTDTC, DM.RFENDTC	Low	⌚

SDTM Mappings

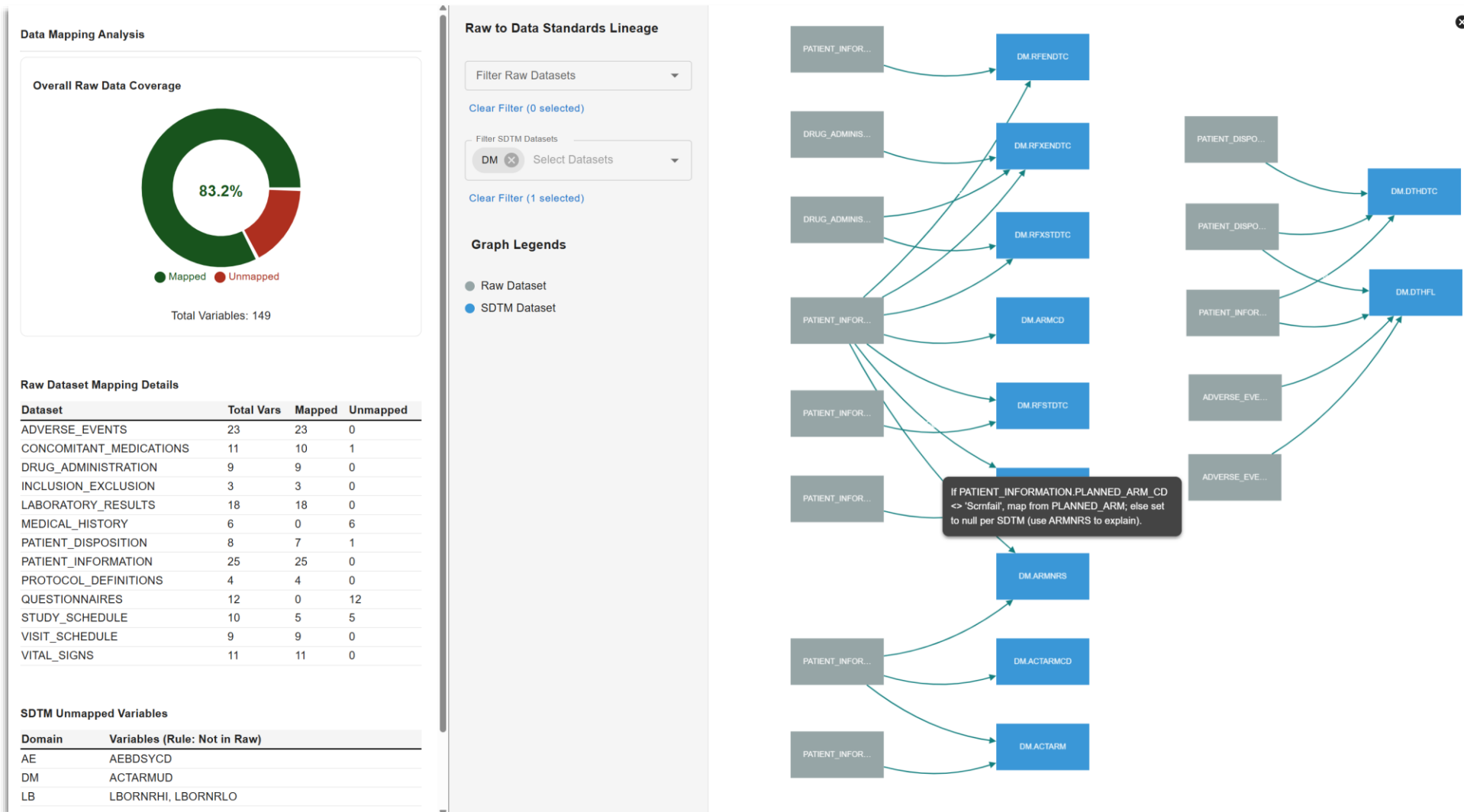
Inputs:

- Raw Data Schema
- Study Design Documents
- SDTMIG/Custom

Outputs:

- Editable Field Level mappings

CDF prototype - Traceability and Data Coverage



Lineage View

- Raw Data coverage reports.
- Variables with missing mappings.
- Graph view to track lineage.

CDF prototype – SDTM Code

The screenshot shows the 'Agent Outputs' section of the CDF prototype. At the top, there are tabs for 'SDTM Datasets', 'SDTM Code' (selected), 'ADaM Datasets', 'ADaM Code', 'TLF Logic', and 'TLF Code'. Below these are sub-tabs for various domains: AE, CM, DM, DS, EX, LB, RELREC, SC, SE, SUPPAE, SUPPDM, SUPPDS, SUPPLB, SV, TA, TE, TI, TS, TV, and VS. The 'AE' tab is selected. Below the sub-tabs are three main sections: 'Code' (selected), 'Last Run Log', and 'Last Run Output'. In the 'Code' section, there are buttons for 'Run Code' and 'Save Code', and a 'Last Run Status: Success' indicator. Red arrows point from the 'SDTM Domain Tabs' label to the 'AE' tab, from the 'Edit, Re-generate, Run/Save code' label to the 'Run Code' and 'Save Code' buttons, and from the 'Generated SAS Code, Logs, and Outputs' label to the 'Last Run Log' and 'Last Run Output' sections. The main area displays SAS code for creating the SDTM AE dataset.

Agent Outputs

SDTM Domain Tabs

SDTM Datasets SDTM Code ADaM Datasets ADaM Code TLF Logic TLF Code

AE CM DM DS EX LB RELREC SC SE SUPPAE SUPPDM SUPPDS SUPPLB SV TA TE TI TS TV VS

Code Last Run Log Last Run Output

Run Code Save Code Last Run Status: Success

Edit, Re-generate, Run/Save code

Generated SAS Code, Logs, and Outputs

```
/* ***** */
/* PROGRAM:      create_sdtm_ae.sas */
/* DESCRIPTION:   Creates the SDTM Adverse Events (AE) domain from RAW data per specifications. */
/* */
/* INPUT:        RAW.ADVERSE_EVENTS, RAW.PATIENT_INFORMATION, RAW.VISIT_SCHEDULE, */
/*              RAW.PROTOCOL_DEFINITIONS */
/* */
/* OUTPUT:       SDTM.AE */
/* */
/* AUTHOR:       Clinical Data Programmer AI */
/* DATE:         06 May 2026 */
/* */
/* NOTES:        - This program adheres strictly to the provided SDTM AE mapping specifications.*/
/*              - Only RAW datasets are used to derive SDTM AE. */
/* ***** */

options nocenter ls=132 ps=64;

/*-----*/
/* Define LIBNAMEs for source Raw data and target SDTM datasets. */
/*-----*/
libname RAW  "/mnt/viya-share/data/CDF_DEMO/raw-data";
libname SDTM "/mnt/viya-share/data/CDF_DEMO/tabulations-sdtm";

/*-----*/
/* 0. STUDYID: Obtain Study Identifier from Protocol Definitions (RAW.PROTOCOL_DEFINITIONS). */
/*-----*/
proc sql noprint;
  select min(study_id) into :_studyid trimmed
  from raw.protocol_definitions
  where not missing(study_id);
quit;

/*-----*/
```

SDTM Code

- Code generated with error feedback loop.
- Edit, re-generate, re-run code form this screen.

CDF prototype – ADaM Mappings

Clinical Data Transformation Steps

Upload the TLF shells document, and press start to extract TLF shell details. When content is satisfactory, select "Approved" in status to proceed to the next step.

TLF Shells: Choose File No file chosen Uploaded: CDISCPILLOT_TLF_Shells_Small.pdf

Instructions:

Extract

State: Approved

Identify and Generate ADaM Datasets

Upload the SAP document and press Start next to 'Identify..' to get ADaM datasets, variables, and mapping logic. Once satisfied, click Start next to 'Generate..' to produce ADaM Datasets & SAS code.

SAP: Choose File No file chosen Uploaded: CDISCPILLOT_SAP.pdf ☐ No SAP available

Instructions:

Identify ADaM Datasets: Start

Macros File: Choose File No file chosen

Instructions:

Generate ADaM Datasets: Start

State: In-Process

Agent Outputs

SDTM Datasets SDTM Code ADaM Datasets ADaM Code TLF Logic TLF Code

Dataset Name	Variable	Type	Description	Mapping Rule	Source	Data Source	Confidence	Status
ADADAS	ABLFL	conditional	Baseline Record Flag. Flag set to "Y" for the single record	For each USUBJID and PARAMCD: identify candidate source records	ADaM IG; SAP/TL	COG.C OGDTC, COG.C	Medium	
ADADAS	ADT	conditional	Analysis Date. Date of the assessment (numeric date)	Convert the source SDTM assessment date (e.g., COG.COGDT or	ADaM IG; SAP/TL	COG.C OGDTC, COG.C	High	
ADADAS	ADTM	conditional	Analysis Datetime. Datetime of the assessment derived	Convert source SDTM datetime (COG.COGDT/COG.COG	ADaM IG	COG.C OGDTC, COG.C	High	
ADADAS	ADY	conditional	Analysis Relative Day. Day relative to anchor date	Calculate ADY = (ADT - anchor_date) where anchor_date =	ADaM IG; SAP/TL	ADSLR FSTDTC, ADSL	High	
ADADAS	ANLADAFI	conditional	Analysis Flag for ADAS. An example record-level analysis	Apply SAP-specified analysis-selection logic to mark the single record per	ADaM IG; SAP/TL	ADADA S.AVISIT, ADADA	Medium	
ADADAS	ASEQ	permissible	Analysis Sequence Number. Sequence number generated to	Generate ASEQ as incremental integer per USUBJID ordered by	ADaM IG	ADSL.U SUBJID, ADADA	High	
ADADAS	AVAL	required	Analysis Value. Numeric analysis value derived from	Direct mapping from SDTM source assessment numeric	ADaM IG; SAP/TL	COG.C OGSTR ESN,C	High	
ADADAS	AVALLOC	conditional	LOCF-imputed Analysis Value. Study-specific field	If AVAL is missing for a postbaseline targeted visit, apply LOCF: select	SAP/TL F; Section	ADADA S.AVAL, ADADA	Medium	
			Analysis Visit.	Assign AVISIT by	ADaM	ADADA		

ADaM Mappings

Inputs:

- TLF Shells Logic
- SAP
- SDTM datasets schema

Outputs:

- Editable Field Level mappings

CDF prototype - End to end traceability

Details

Study Name: Safety and Efficacy of the Xanomeline Transdermal Therapeutic System (TTS)

Study Objectives:

Primary Objective 1 (Efficacy): To determine if there is a statistically significant relationship between the change in both the ADAS-Cog (11) and CIBIC+ scores, and drug dose. (SAP Section 3.1.1: 'To determine if there is a statistically significant relationship ... between the change in both the ADAS-Cog (11) and CIBIC+ scores, and drug dose')

Primary Objective 2 (Safety): To document the safety profile of the xanomeline TTS. (SAP Section 3.1.1: 'To document the safety profile of the xanomeline TTS.')

Secondary Objective (Efficacy): To assess the dose-dependent improvement in behavior, indicated by improved scores on the Revised Neuropsychiatric Inventory (NPI-X). (SAP Section 3.1.2: 'To assess the dose-dependent improvement in behavior.')

Primary Endpoint 1: Alzheimer's Disease Assessment Scale - Cognitive Subscale, total of 11 items [ADAS-Cog (11)] at Week 24. (SAP Section 3.2.1)

Primary Endpoint 2: Video-referenced Clinician's Interview-based Impression of Change (CIBIC+) at Week 24. (SAP Section 3.2.1)

Secondary Endpoint 1: ADAS-Cog (11) at Weeks 8 and 16. (SAP Section 3.2.2)

Secondary Endpoint 2: CIBIC+ at Weeks 8 and 16. (SAP Section 3.2.2)

Secondary Endpoint 3: Mean Revised Neuropsychiatric Inventory (NPI-X) from Week 4 to Week 24. (SAP Section 3.2.2)

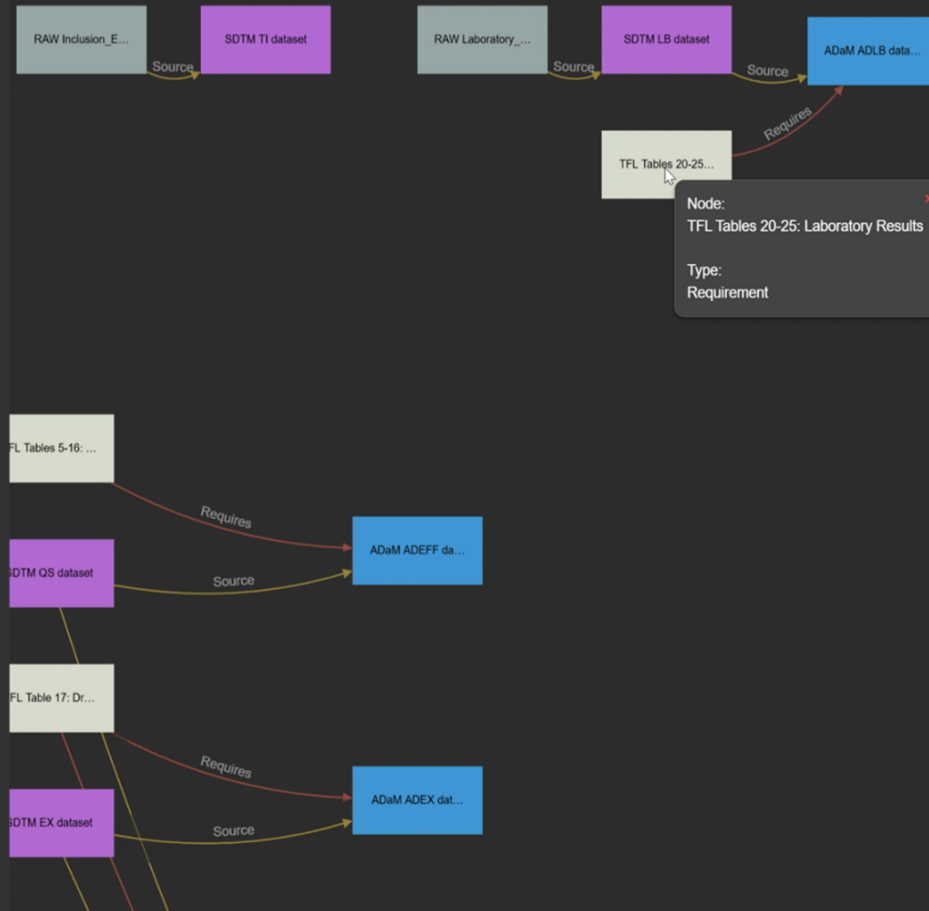
Safety Endpoints: Adverse events, vital signs (weight, blood pressure, heart rate), and laboratory evaluations. (SAP Section 3.2.2)

Graph Node Legend

- Requirement as specified in SAP or TFL
- ADaM Dataset
- SDTM Dataset
- RAW Dataset

Graph Edge Legend

- Source
- Requires



Tracing SAP/TLFs requirements from Raw data to SDTM and ADaM datasets



Business Impact

Automated mapping & code generation

Faster builds, lower cost

AI AS ASSISTANT

LLMs enhance rather than replace human programmers, transforming routines into streamlined processes.

COMPRESS TIMELINES

Mapping, SDTM/ADaM, and TLF generation shrink from months to days.

JUMP-START BUT NOT TURNKEY

Automate bulk of the upstream process to create a robust baseline; humans finalize for submission readiness.

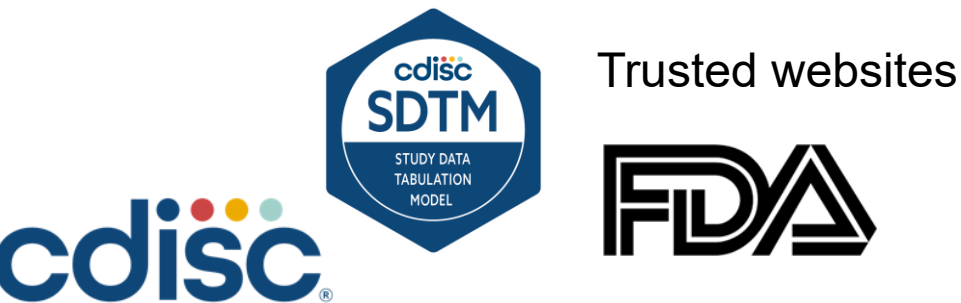
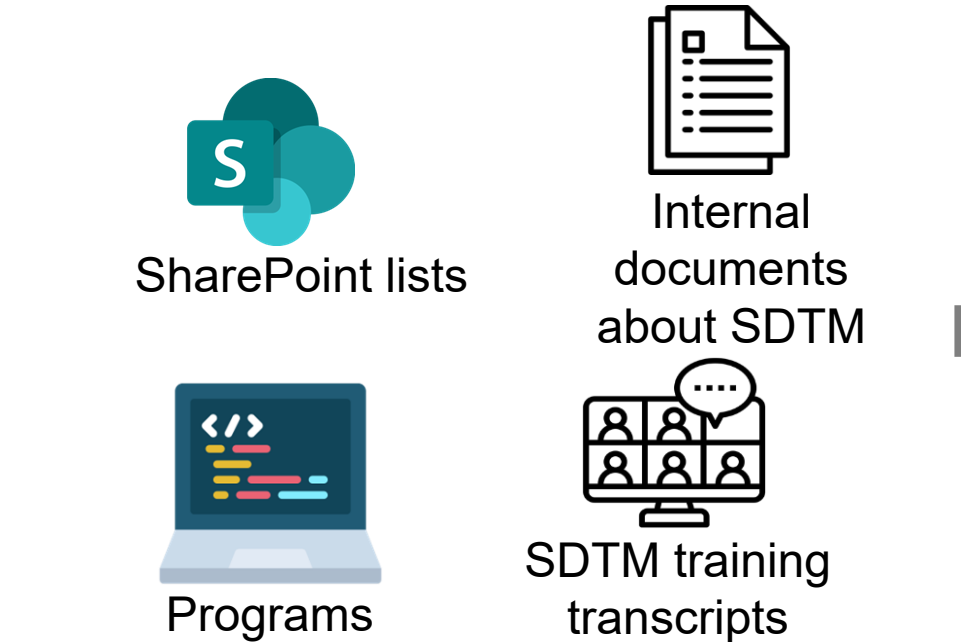
REALLOCATE EXPERTISE

Reallocate expertise to high-value work (e.g., complex efficacy datasets), high-level validation, scientific analyses.

Other AI tools supporting Standards Development in Novo Nordisk



Chatbot with domain knowledge





What is CRF?

CRF stands for Case Report Form. It is a printed, optical, or electronic document designed to record all required information to be reported to the sponsor for each trial subject.

SDTMIG-MD v1.1taug-diabetesv101CRF&defineSDTMIG_v3.3_FINALDefine-XML v2.1SDTMIG-AP v1.0

Unsatisfied with the answer?

Provide your feedback about the answer

Type your feedback here

Clear conversation history & reset document source categories

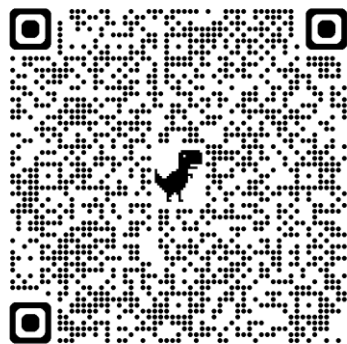
Type your question here...

PapillonsChat is here to help you find information about the **SDTM generation framework** and can retrieve relevant solutions from the **SDTM Super Users Help Desk**. Just describe your issue, and it will guide you to solutions from similar past cases.

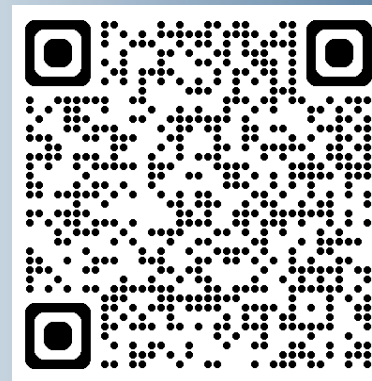
Before you create a new SDTMSU Help Desk ticket, please try to ask the PapillonsChat about the issue.

Please keep in mind:

While PapillonsChat can point you to helpful information, it **does not** replace the original sources. Always verify the information by consulting the original source documents.



Anette Jakobsen
Anette.DalsgaardJakobsen@sas.com



Anja Jazbec
ajvr@novonordisk.com

