

Single Day Event

**GenAI-Enabled Automation and
Continuous Traceability Across eCRFs,
SDTM, ADaM and TLFs**
Anette Jakobsen,
Life Science Industry Advisor, SAS

10 June 2026



Anette Dalsgaard Jakobsen

Life Science Industry Advisor, SAS

Anette Dalsgaard Jakobsen is a Senior Industry Advisor in Copenhagen and is part of the Global Health and Life Sciences Practices at SAS. She supports life sciences clients by providing guidance on their digital transformation journey through the use of advanced analytics, Generative AI and agentic AI, especially in areas such as biostatistical programming, productivity, and the responsible adoption of AI in regulated settings.



Pankaj Attri

Pre-Sales Solutions Architect

SAS, Commercial Healthcare & Life Sciences

Pankaj Attri is an AI/ML Solutions Architect focused on leveraging advanced data science and artificial intelligence to support customers in the Life Sciences and Healthcare industries at SAS. He partners with these organizations to design and implement state-of-the-art custom solutions that accelerate innovation across research, development, and clinical operations.

Pankaj brings together a background in Chemical Engineering (B.Tech) from IIT Bombay, India and a Master's in Computer Science from NC State University, U.S.



Agenda

- Why introduce Generative AI in the Clinical Data Flow?
- Example solutions: Generative AI in Clinical Trials Workflows
- Business Impact: Generative AI in Clinical Trial Workflows

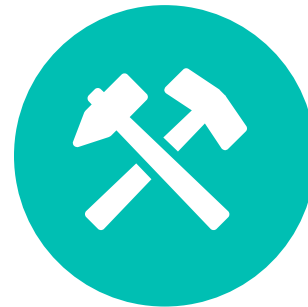


Challenges in Current State of Clinical Trial Biometrics



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

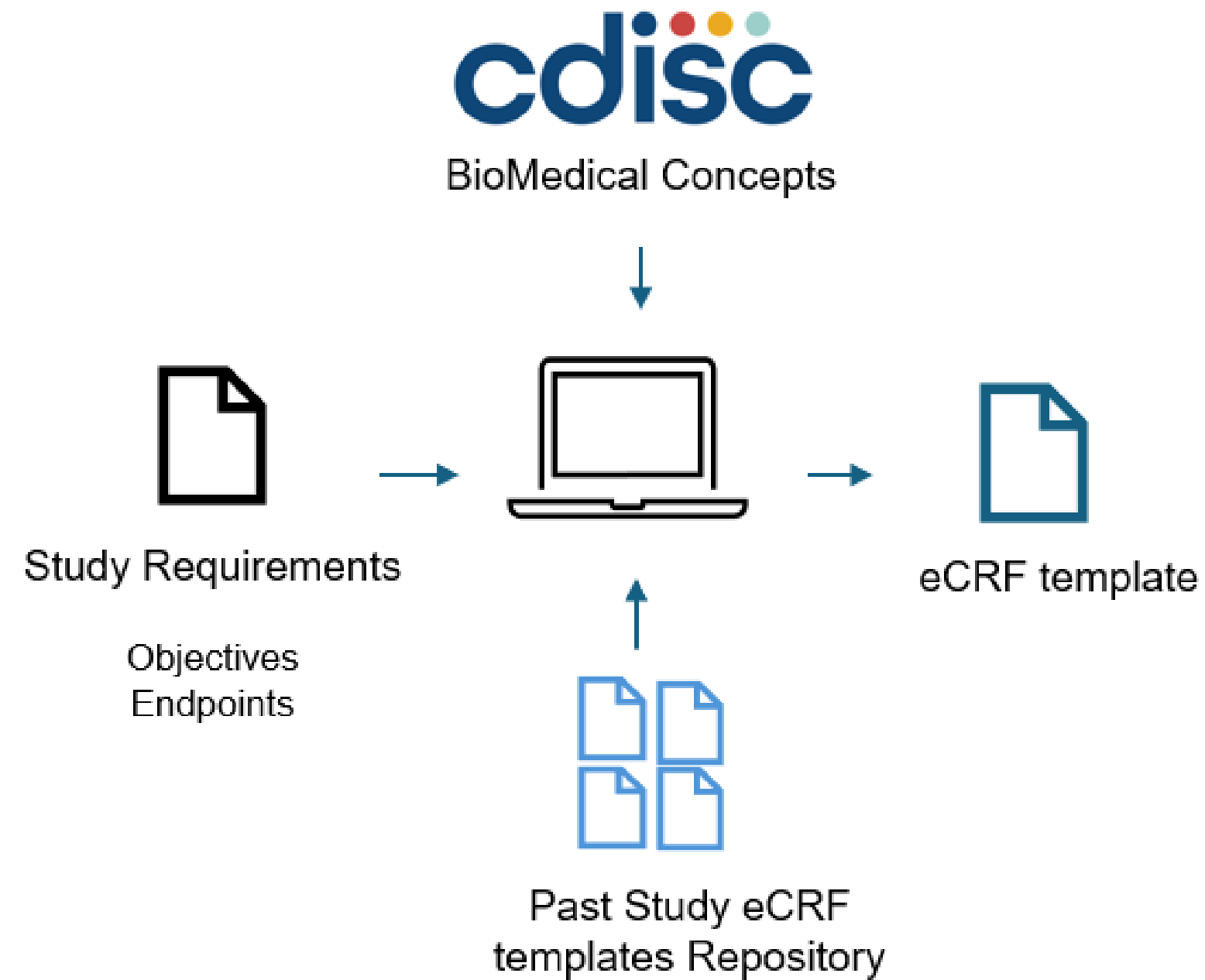
- Repetitive coding tasks
- Impacts submission deadlines

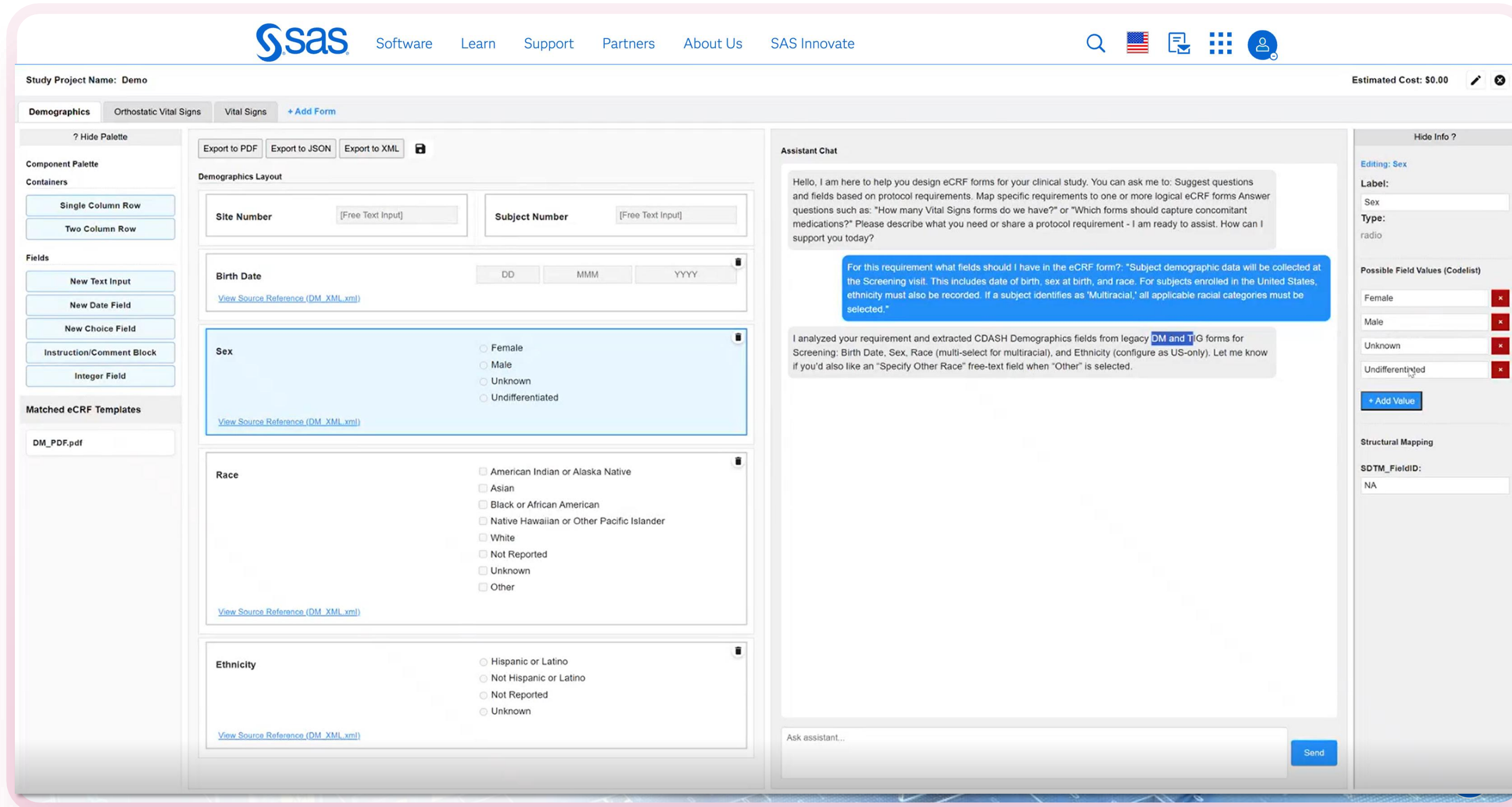
LLM-powered eCRF generation

Example prototype of a generative AI
solution



AI Assisted eCRF form Creation







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 e

Requirement
+ Add Row

Study Details

Study Requirements

Retrieval Settings

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.

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

[Export to PDF](#) [Export to JSON](#) [Export to XML](#)

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

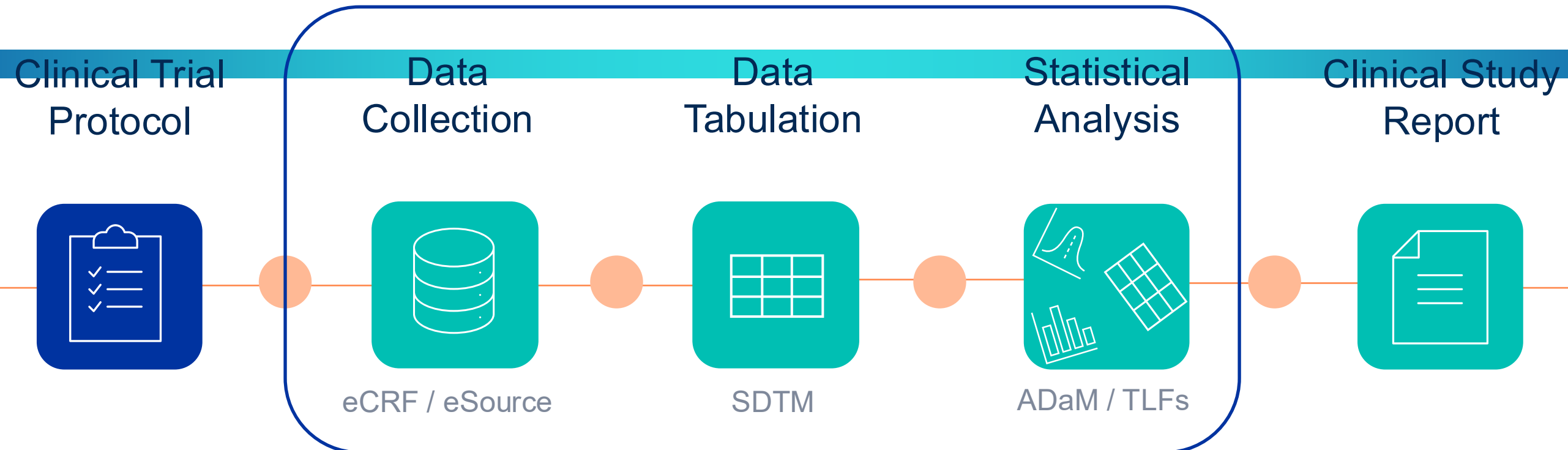
SDTM_FieldID: NA

Clinical Data Flow (CDF) Prototype

- *Will be part of [SAS Clinical Acceleration](#)*



End to End Clinical Data Flow



Raw data to SDTM format

SDTM to ADaM

Creation of define.xml

Creation of TLFs

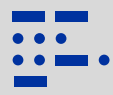


Features



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.



Clinical Data Flow - Create Study Project

Project Name

Study-ONCO_IMMUNO-25

Data Standardization

Analysis Dataset Generation

Reporting & Visualization

LLM Specs

Project Name *

Study-ONCO_IMMUNO-25

Project Description

Base Directory

/all_studies/2025/onco_immuno

Browse

Raw Data Folder

raw-data

SDTM Folder

tabulations-s

ADaM Folder

analysis-adam

Programs Folder

programs

Outputs Folder

outputs

Project Details

Data Standardization

Analysis Dataset Generation

Reporting & Visualization

LLM Specs

☒ Raw to SDTM conversion needed?

SDTM Datasets Folder Name

tabulations-sdtm

Implementation Guide

SDTMIG v3.4

Select or type SDTM Domains

AE BE CE CO CV DM DV EG EX Search or type domain

Leave blank if you want to select all domains.

Project Details

Data Standardization

Analysis Dataset Generation

Reporting & Visualization

LLM Specs

Base LLM *

GPT-5o

Code LLM *

GPT-5o

Assistant LLM

Set Temperature (0-1)

0.0

Select data, programs, outputs folders

Select Implementation Guide and LLMs



CDF Prototype - Project Details Screen

Welcome, sinpan

Clinical Data Flow AI Assistant

Project Name: Study-ONCO_IMMUNO-2025

Agent outputs

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



Mapping Rules Generation - SDTM

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.E	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	

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

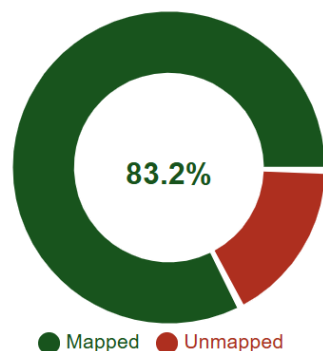
- Outputs:
- Editable SDTM mapping rules



Traceability and Lineage Reports

Data Mapping Analysis

Overall Raw Data Coverage



Total Variables: 149

Raw Dataset Mapping Details

Dataset	Total Vars	Mapped	Unmapped
ADVERSE_EVENTS	23	23	0
CONCOMITANT_MEDICATIONS	11	10	1
DRUG_ADMINISTRATION	9	9	0
INCLUSION_EXCLUSION	3	3	0
LABORATORY_RESULTS	18	18	0
MEDICAL_HISTORY	6	0	6
PATIENT_DISPOSITION	8	7	1
PATIENT_INFORMATION	25	25	0
PROTOCOL_DEFINITIONS	4	4	0
QUESTIONNAIRES	12	0	12
STUDY_SCHEDULE	10	5	5
VISIT_SCHEDULE	9	9	0
VITAL_SIGNS	11	11	0

SDTM Unmapped Variables

Domain	Variables (Rule: Not in Raw)
AE	AEBDSYCD
DM	ACTARMUD
LB	LBORNRHI, LBORNRLO

Raw to Data Standards Lineage

Filter Raw Datasets

Clear Filter (0 selected)

Filter SDTM Datasets

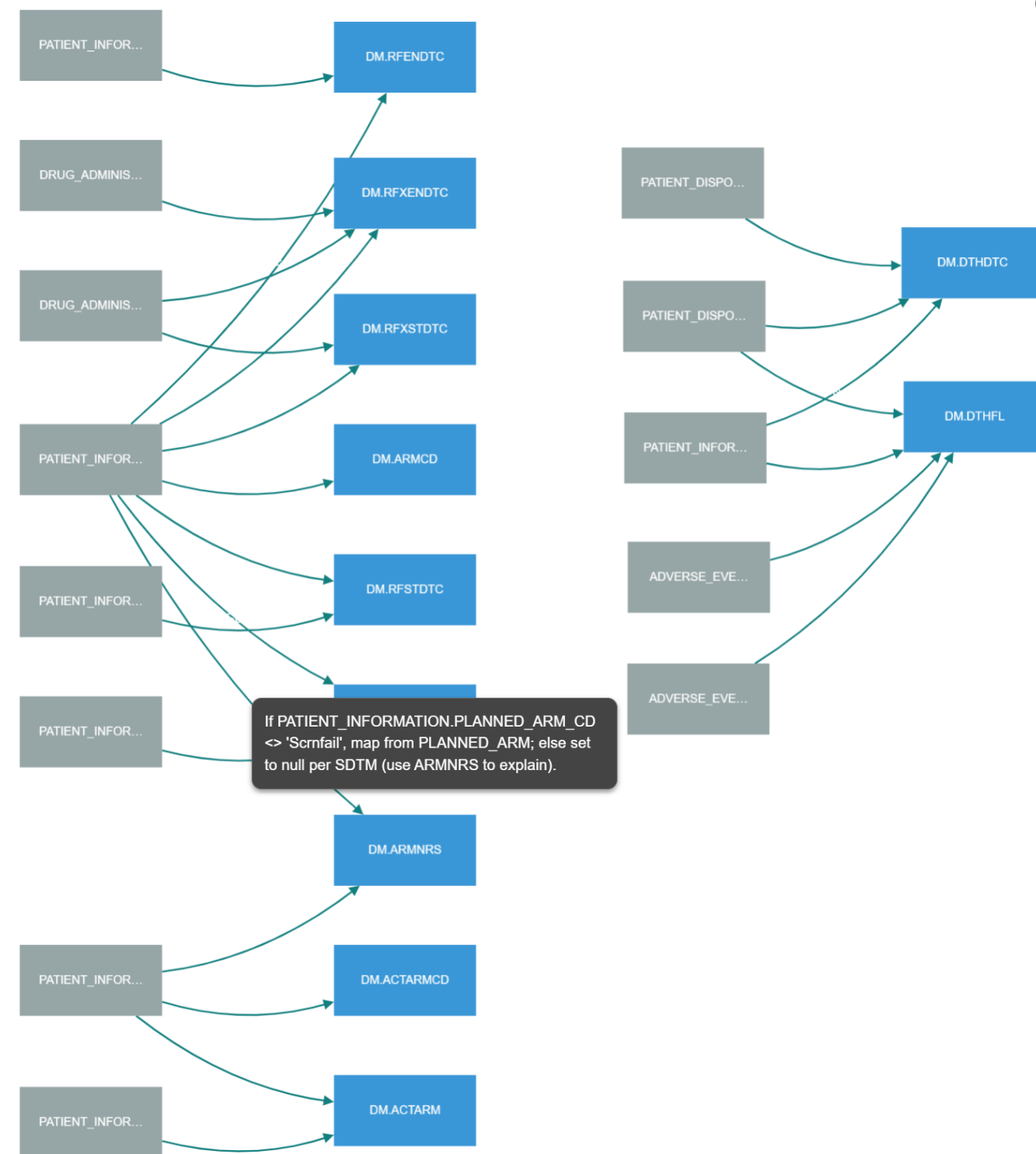
DM X Select Datasets

Clear Filter (1 selected)

Graph Legends

● Raw Dataset

● SDTM Dataset



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



Code Generation - SDTM

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

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



Mapping Rules Generation – ADaM

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 selected

Uploaded:

CDISCPILOT_TLF_Shells_Small.pdf

Instructions:

Optional instructions for TLF shells...

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 selected

Uploaded:

CDISCPILOT_SAP.pdf

No SAP available

Instructions:

Optional instructions for SAP document...

Identify ADaM Datasets:

Start

Macros File:

Choose File

No file chosen

Instructions:

Optional instructions for macros file...

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	ADSL.R FSTDT C,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		

Inputs:

- TLF Shells Logic
- SAP
- SDTM datasets schema

Outputs:

- Editable SDTM – ADaM mapping rules



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)

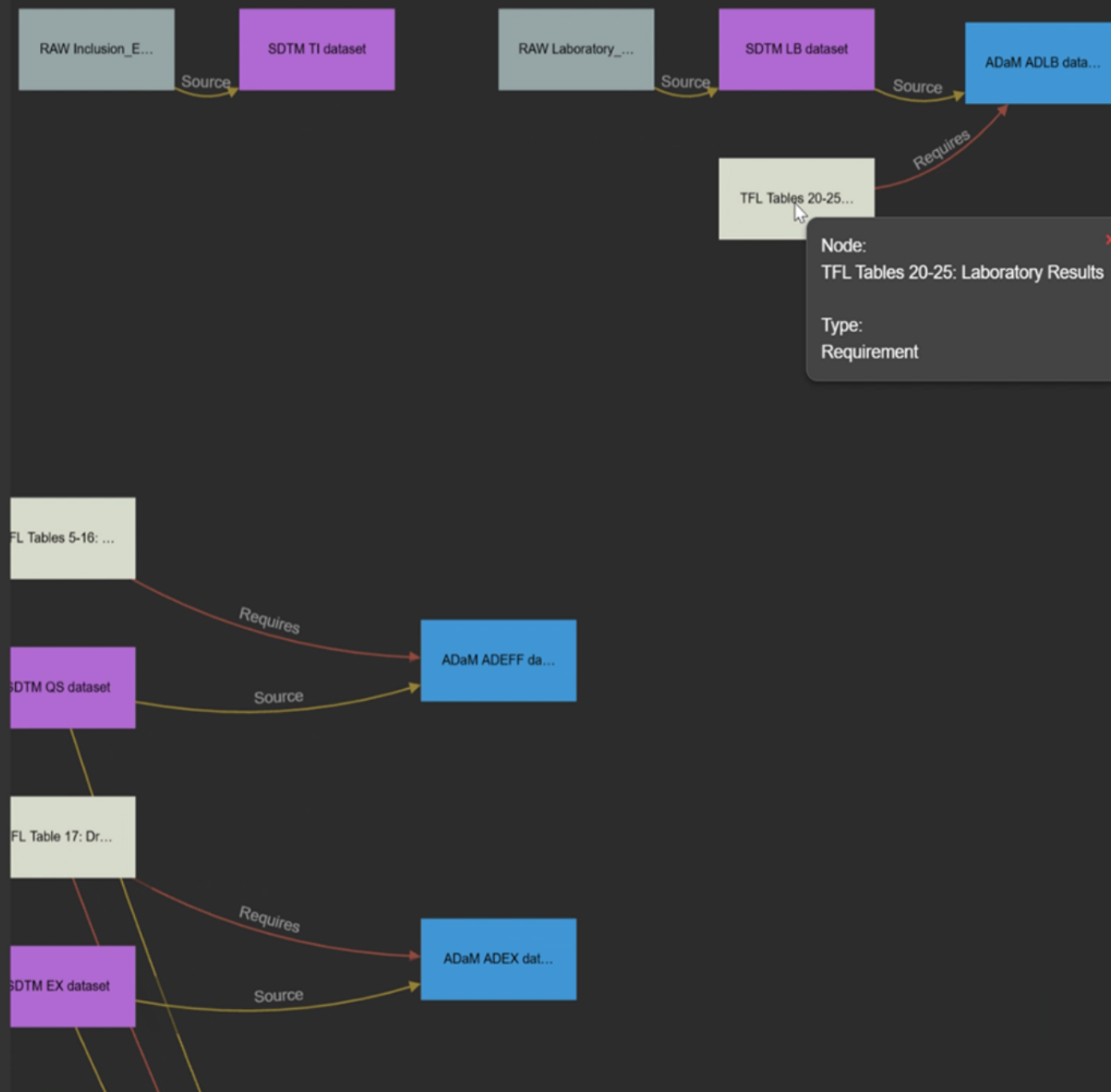
Safety Endpoints: Adverse events, vital signs (weight, blood pressure, heart rate), and laboratory evaluations. (SAP Section 3.2.2)

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

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.



CDISC in Milan May 2026



- **Automating** Define-XML Generation Using a New Metadata Specification Model in the CDISC 360i Program
Pierre Dostie, UCB
- **Using Open-Source Tools to Drive Standards-Based Automation:** From SDTM Setup to CDISC-Compliant Submission Katarzyna Bzymek, SGS Pharma
- **AI-Assisted** Workflows for Enhanced Consistency and Auditability of SDTM generation
Sapna More and Sarah Jamal, Oracle
- **Deterministic Automation Meets Generative AI: ADaM Programming with the Mighty Framework Gives the Best of Both Worlds**
Matthew Phelps, Novo Nordisk
- **LLM-Powered End-to-End Clinical Data Automation:** Standardized eCRFs to SDTM, ADaM, and TLF with Full Traceability
Anja Jazbec, Novo Nordisk; and Anette Dalsgaard, SAS



Questions or Comments?



Anette Dalsgaard Jakobsen
anette.dalsgaardjakobsen@sas.com
[Anette Jakobsen/](#)