phuse·share
connect·share·advance

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

phuse.global

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 an 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

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

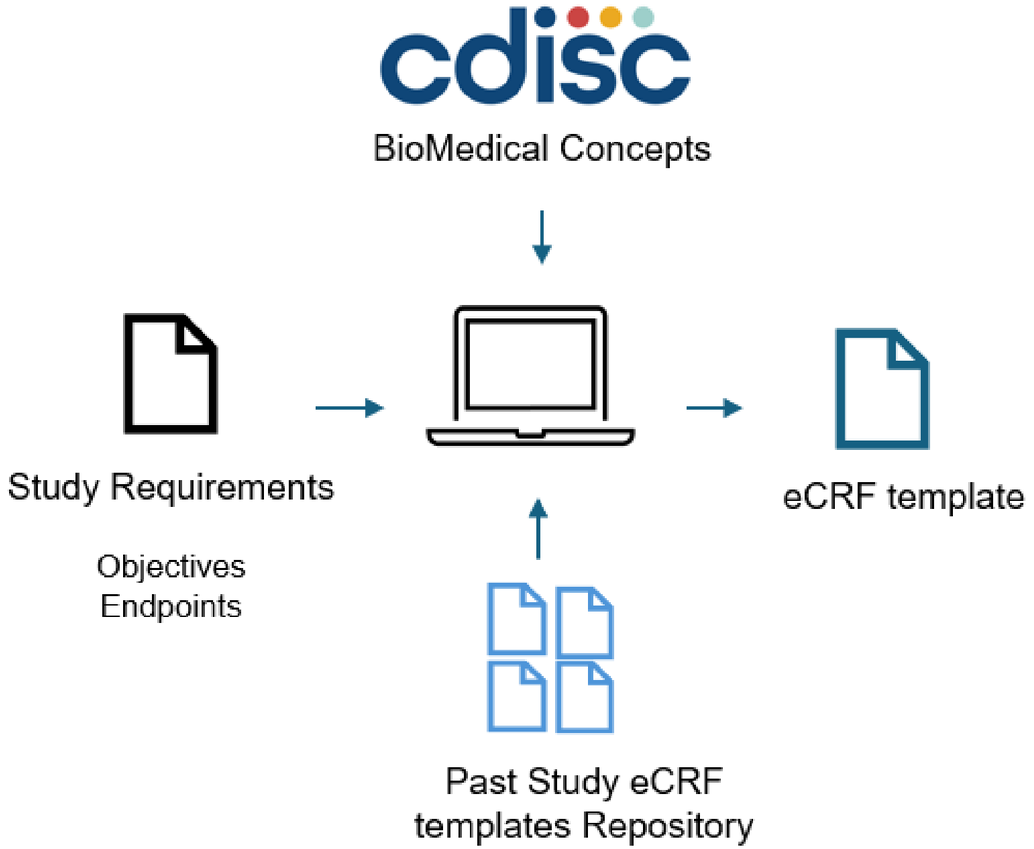
Double Programming

- Repetitive coding tasks
- Impacts submission deadlines

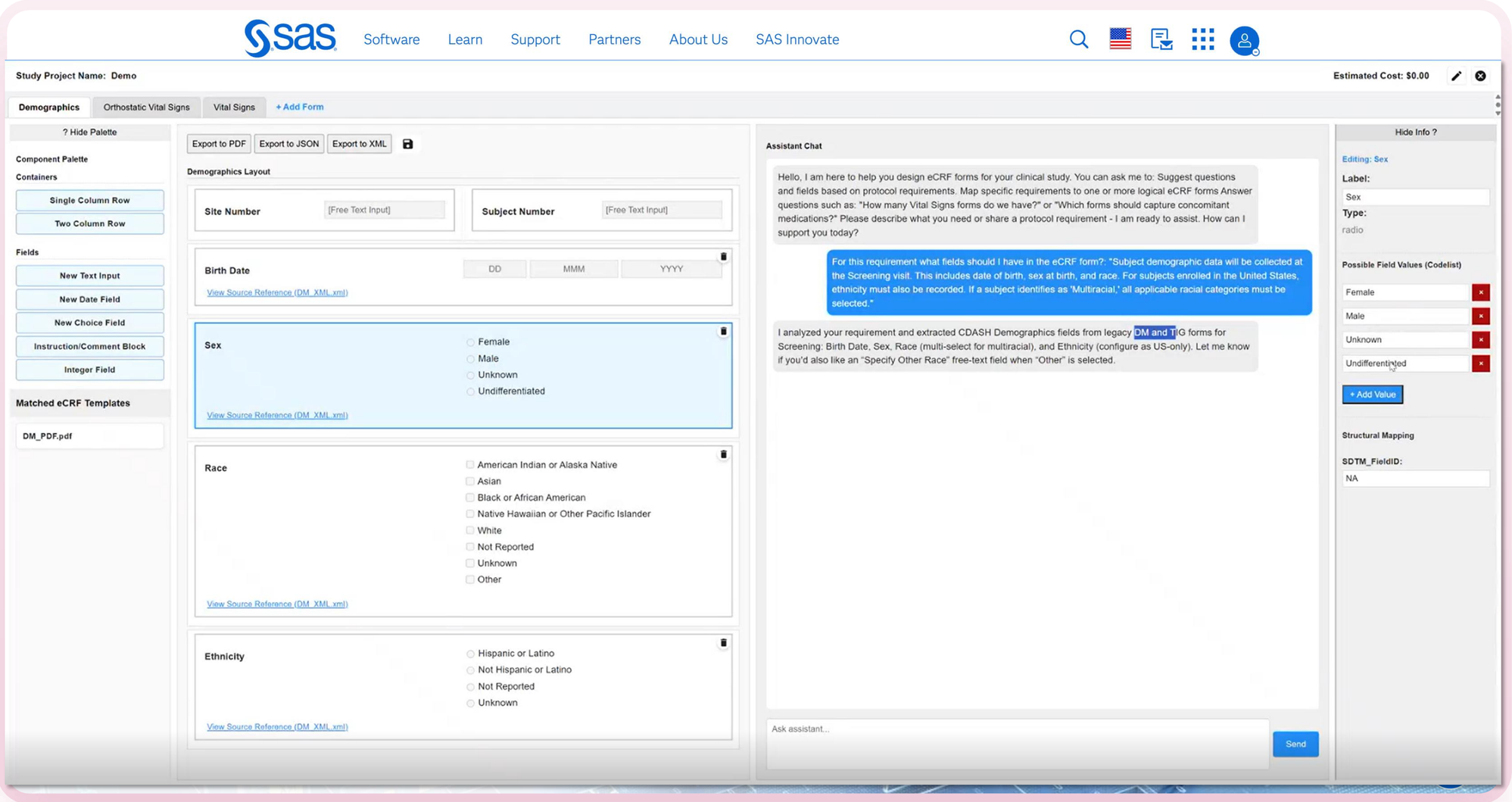
LLM-powered eCRF generation

Example prototype of a generative AI solution

AI Assisted eCRF form Creation



AI Assisted eCRF Creation - Prototype

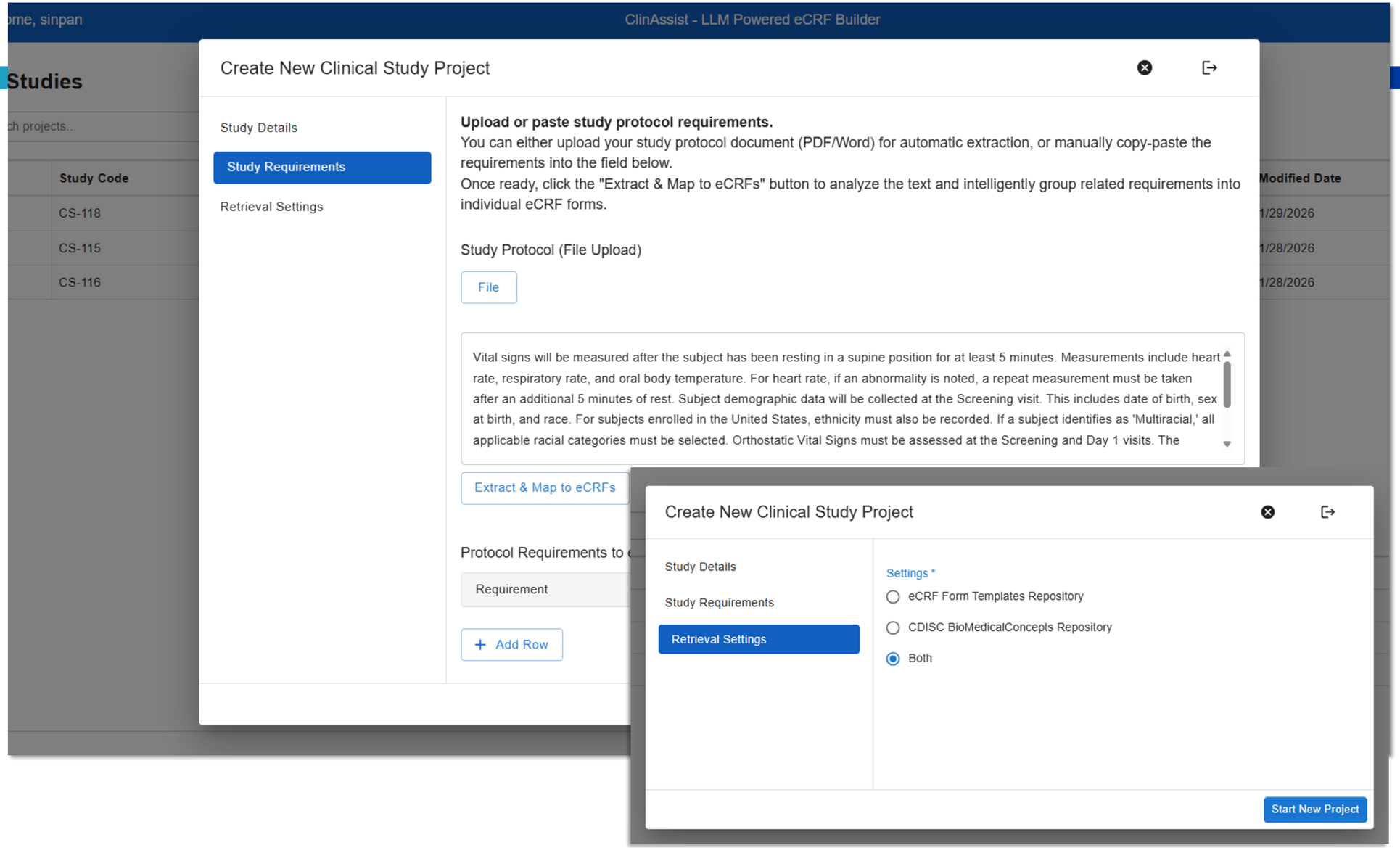


Clinical Trials eCRF Build Automation

1

Map Requirements to eCRFs

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

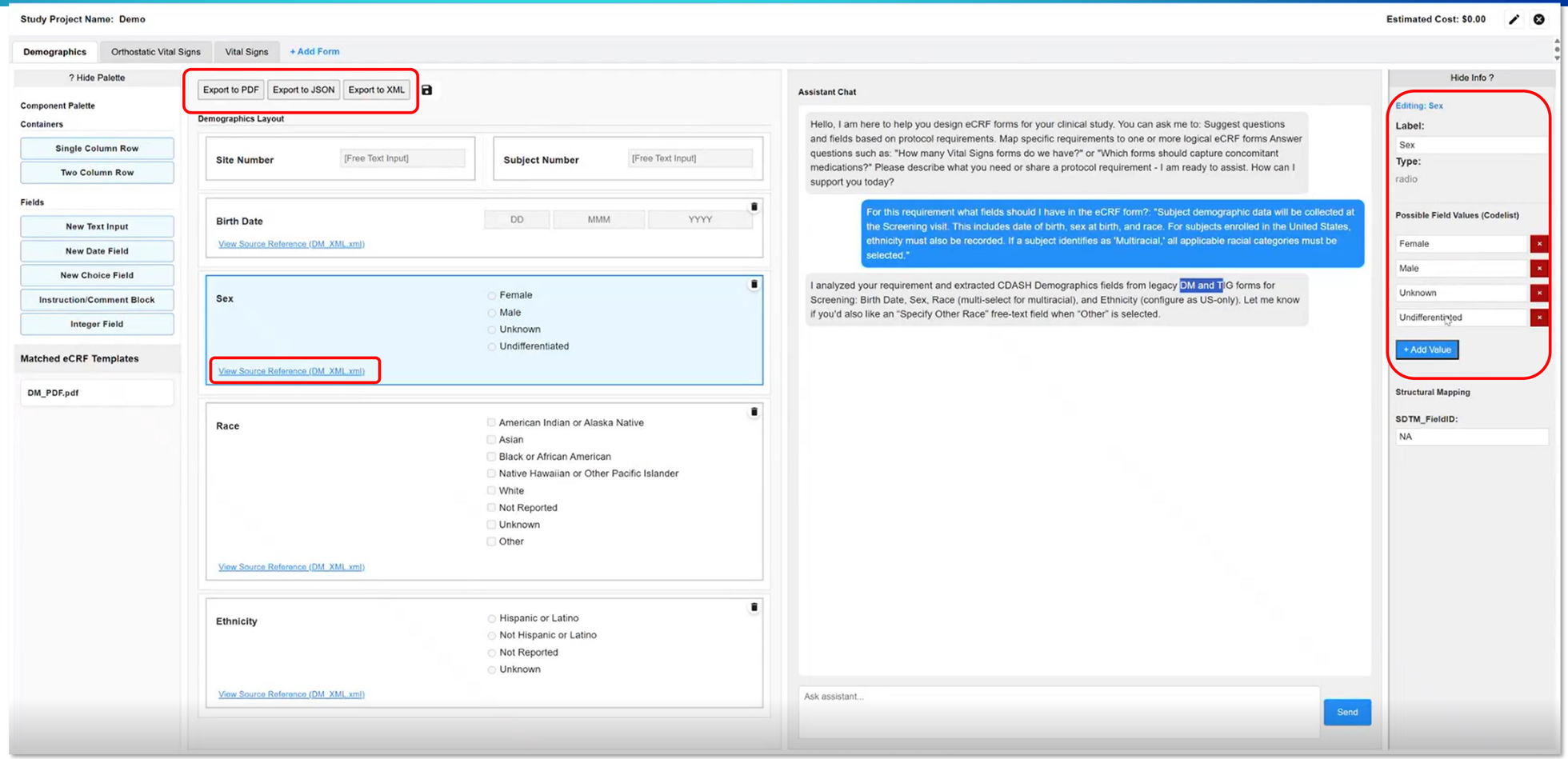


Clinical Trials eCRF Build Automation

2

Edit and Export eCRFs

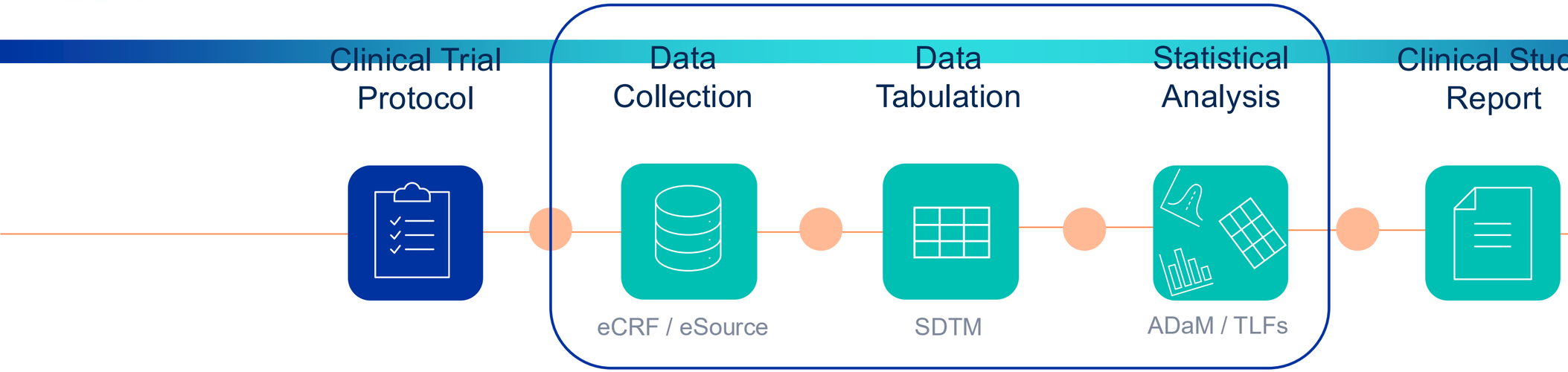
Edit metadata, add/edit new fields to eCRF form, export to PDF/JSON/XML.



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.

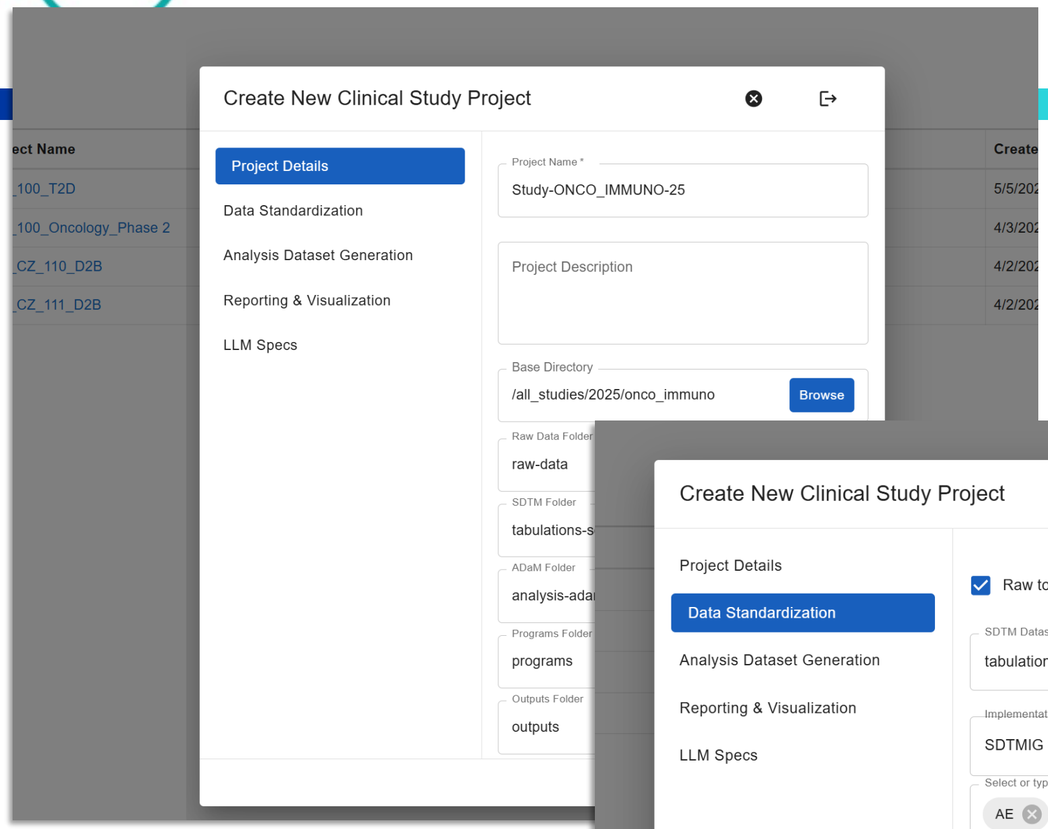
Code Generation

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

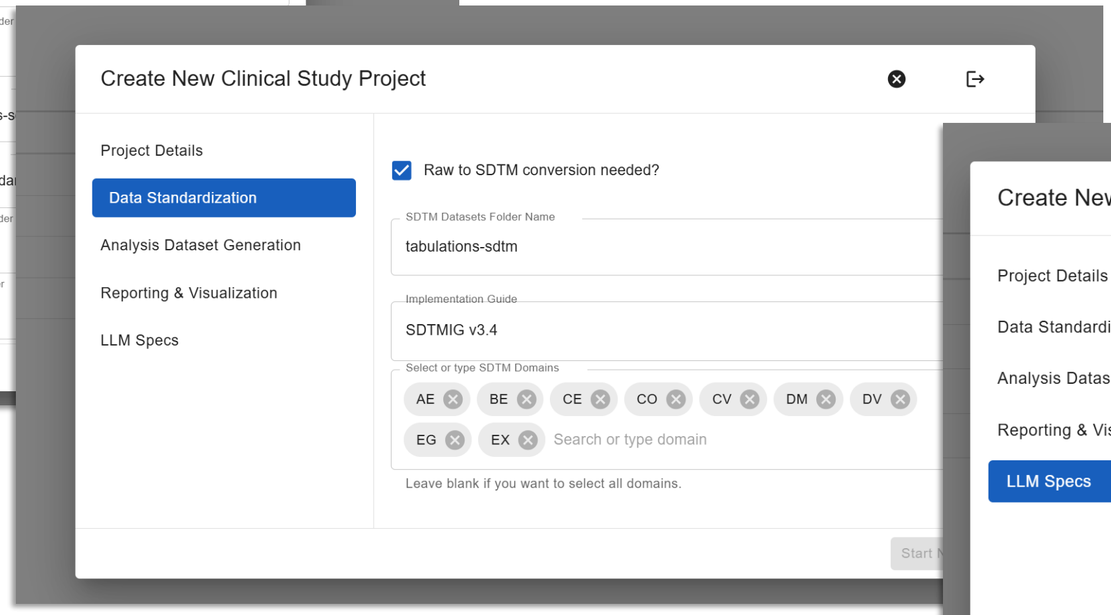
Traceability

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

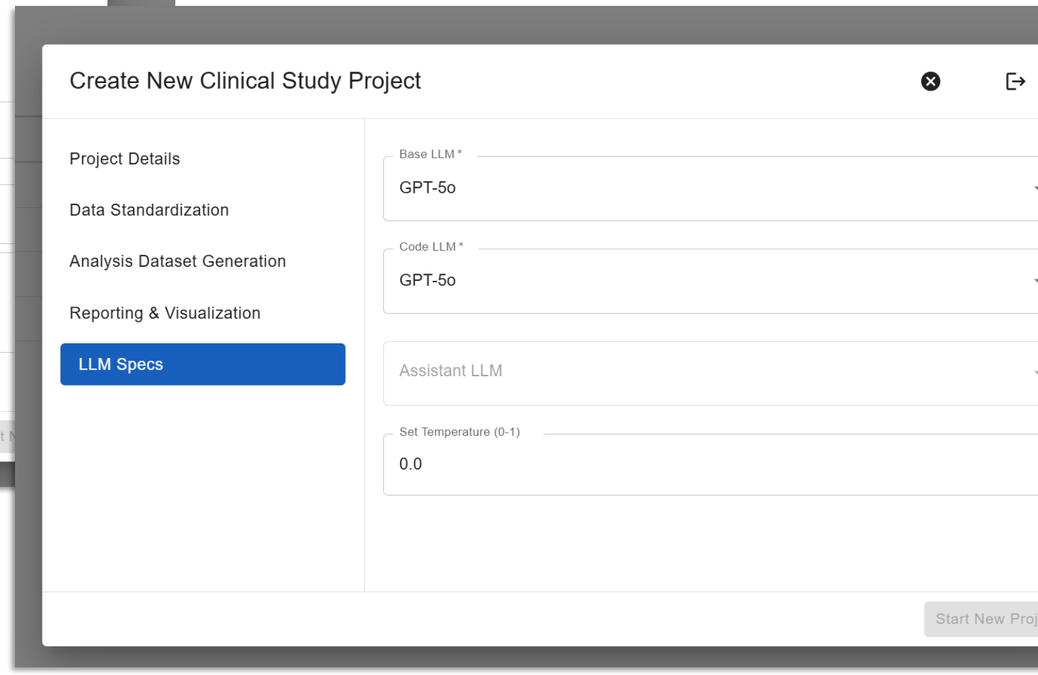
Clinical Data Flow - Create Study Project



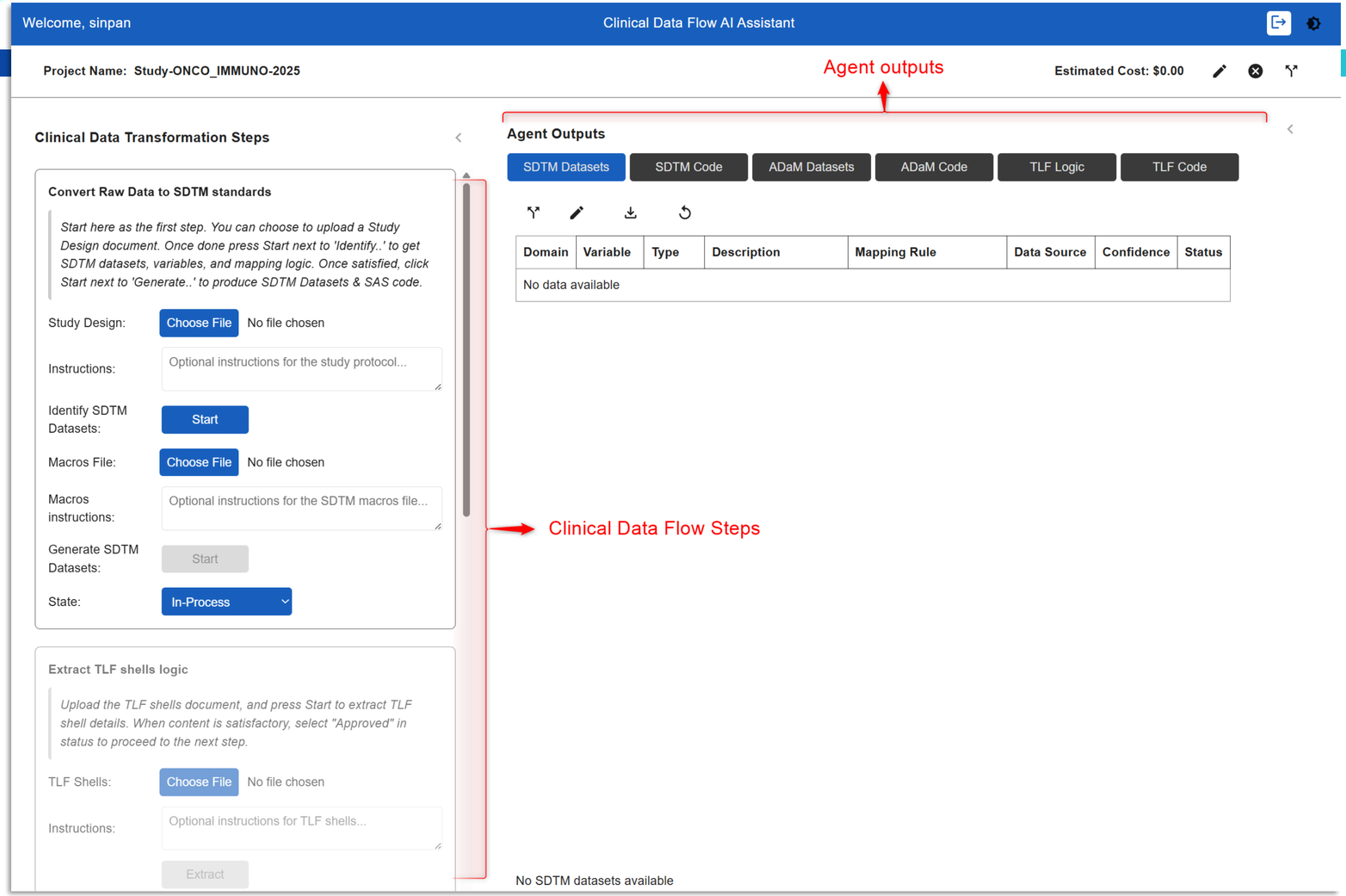
Select data, programs, outputs folders



Select Implementation Guide a LLMs



CDF Prototype - Project Details Screen



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,	Low	

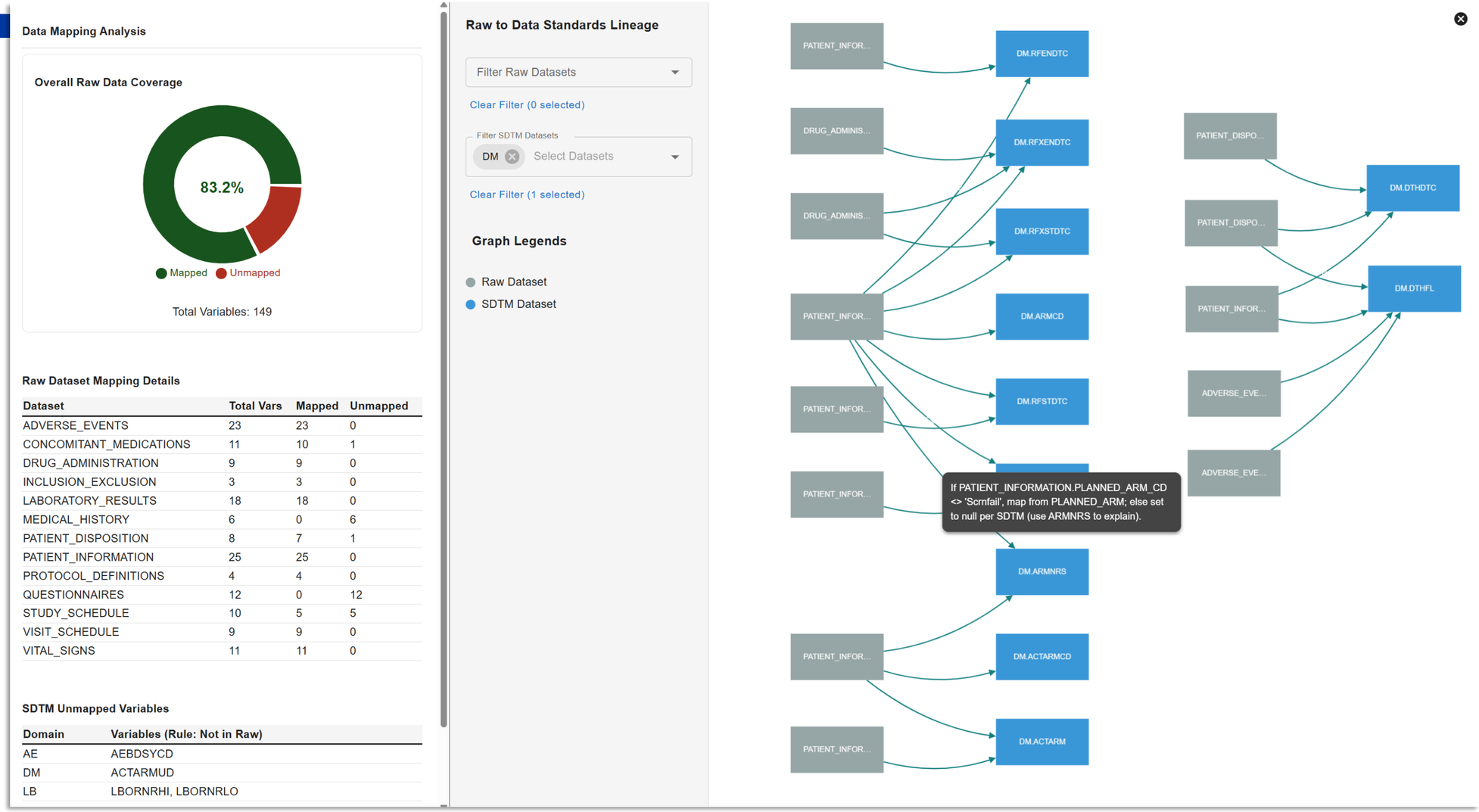
Inputs:

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

Outputs:

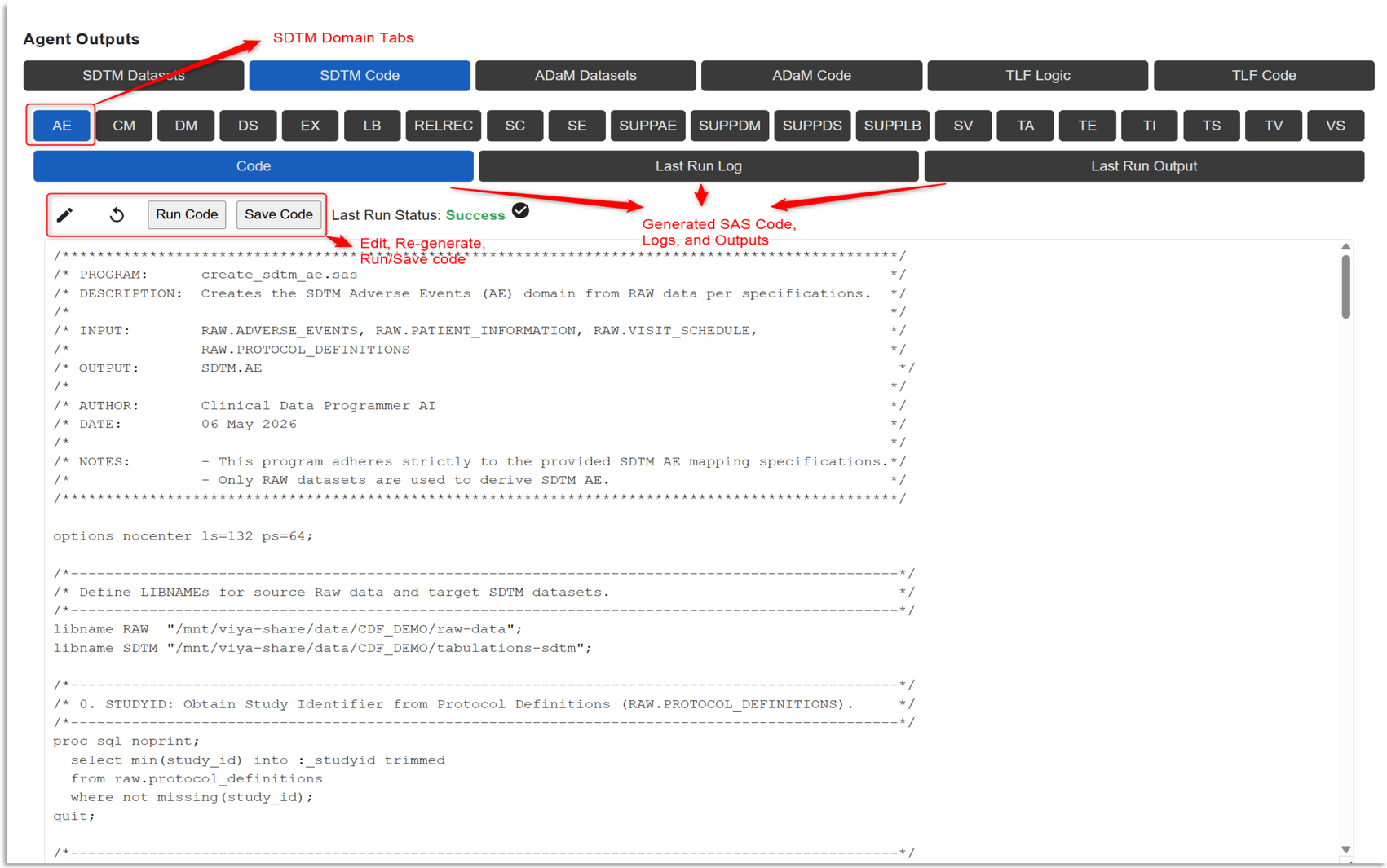
- Editable SDTM mapping rules

Traceability and Lineage Reports



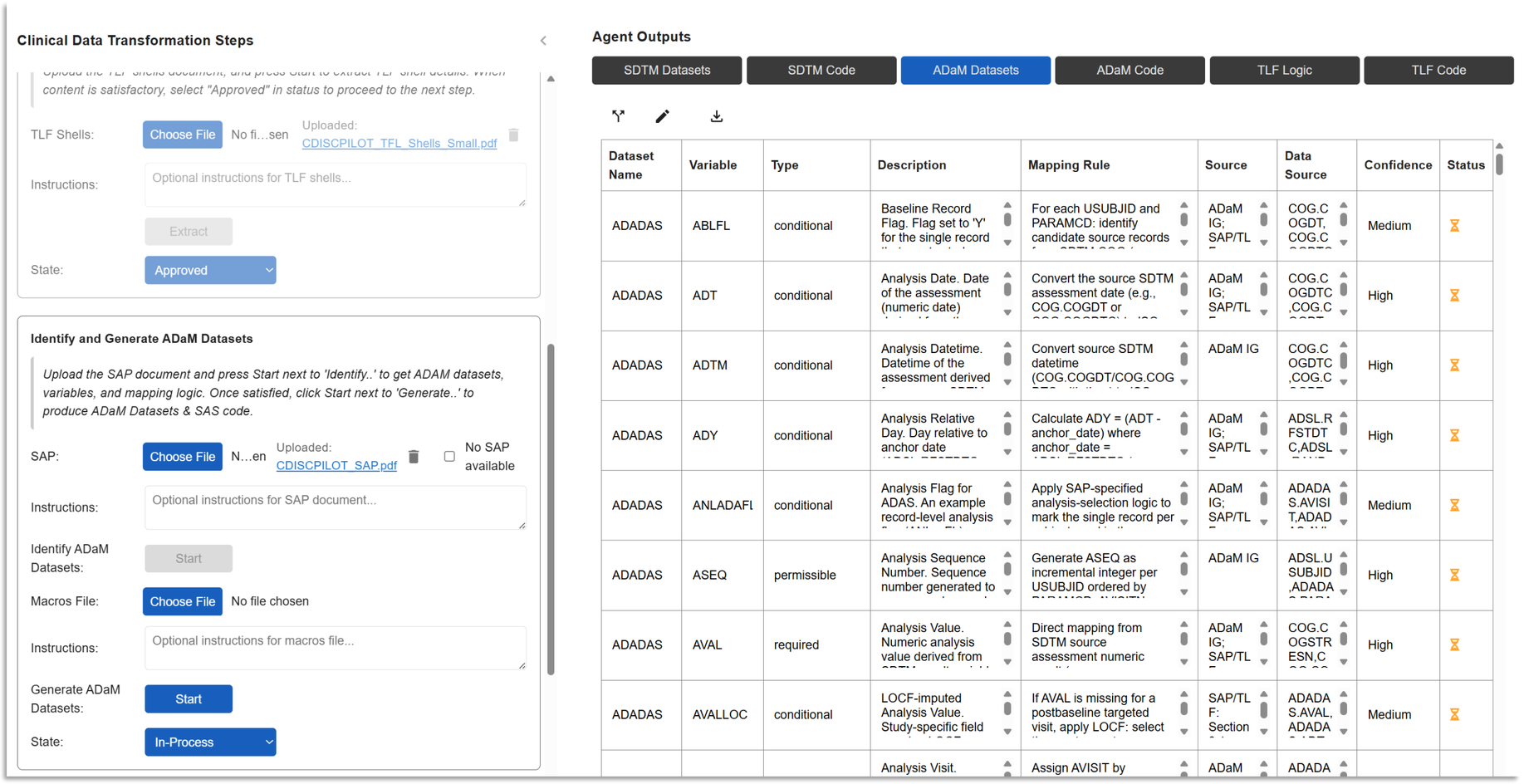
- Raw Data coverage reports.
- Variables with missing mappings.
- Graph view to track lineage.

Code Generation - SDTM



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

Mapping Rules Generation – ADaM



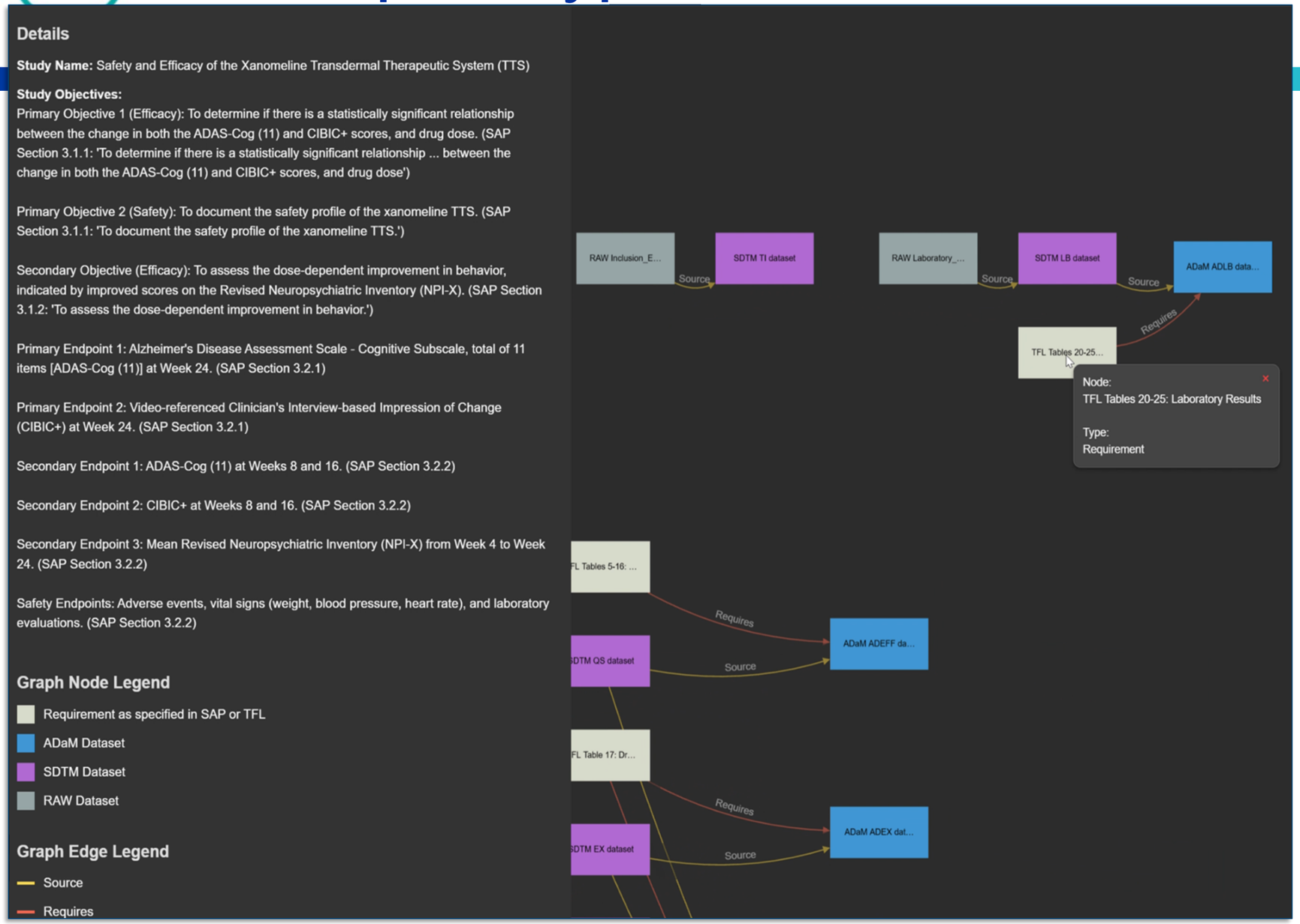
Inputs:

- TLF Shells Logic
- SAP
- SDTM datasets schema

Outputs:

- Editable SDTM – ADaM mapping rules

CDF prototype - End to end traceability



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/