

SDTM Specification Automation Using an ML Approach to Provide Mappings

Benzi Alex Mathew, Jaya Simha Inampudi
(Saama Technologies)



US Connect 2026

The Bottleneck in Clinical Data Lifecycle

- **Manual Translation:** Converting protocol requirements to CDISC standards is labor-intensive.
- **Repetitive Tasks:** Programmers manually interpret protocols, map variables, and write code.
- **Risk:** Manual approaches are prone to human error and inconsistency.

Our Goal: Accelerate delivery of the entire SDTM submission package without sacrificing quality.



Introducing SDTM Navigator

An Agentic AI Solution

Core Concept: Combines **Agentic AI** with a **Standards-Based Approach**.

Technology Stack:

- Large Language Model (LLM).
- Retrieval-Augmented Generation (RAG) backend.

Compliance: References verified standards rather than relying solely on generated patterns.

- *Currently supports CDISC SDTM IG 3.4 and NCI Controlled Terminology.*



From Protocol to Program

The 4-Step Agentic Workflow

- **Step 1: Auto-Generate Trial Design**
 - *Input:* Protocol (PDF), CT.Gov data.
 - *Action:* Extract trial design elements.
- **Step 2: Auto-Identify Domains**
 - *Input:* Source Data, Trial Design Specs.
 - *Output:* Recommended SDTM Domains.
- **Step 3: Auto-Generate Specifications**
 - *Input:* Historical Specs, Global Macros.
 - *Action:* Map source to target variables using historical patterns.
- **Step 4: Auto-Generate Programs**
 - *Output:* Executable **SAS** and **R** programs.



Human-in-the-Loop (HITL) Paradigm

Empowering the Clinical Programmer

The Paradigm: Automation empowers, rather than replaces, expertise.

Validation Checkpoints:

- Users review and modify outputs at **each critical juncture**.
- Ensures digital specifications align with unique study requirements *before* programming begins.

Result: High-quality, validated output ready for regulatory submission.

Key Features



- **Multilingual Support:** Generates native code in **SAS or R**, preventing vendor lock-in.
- **Reasoning & Traceability:** Provides rationale for the mapping recommendation.
- **Trial Design Integration:** Integrates public sources (ClinicalTrials.gov) to automate complex Trial Design domain creation.
- **Standards Management:** "Jump-start" studies using historical specifications and global libraries.



Business Impact & Metrics

Quantifiable Efficiency Gains

- **Time Reduction:**
 - **50% reduction** in time to generate Specifications.
 - **69% reduction** in time to Generate & QC first draft SDTM datasets.
- **Traceability:** Full audit trail of transformations at the variable level.
- **Reusability:** Leverage tested artifacts to eliminate "recreating from scratch".

Summary



- **Paradigm Shift:** Moves from manual transcription to Agentic Automation.
- **Robust Framework:** Integrates AI with established CDISC standards.
- **Future Ready:** Flexible, language-agnostic, and scalable.



Benzi Alex Mathew

Head,
Biometrics & Clinical Solutions (BCS)
VP, Saama
benzi.mathew@saama.com

Jaya Simha Inampudi

Sr Director,
Biometrics & Clinical Solutions (BCS), Saama
jaya.inampudi@saama.com

