Paper SD03

SDTM Navigator: An Agent-Based LLM Framework for Automated Specification and Programming

Benzi Alex Mathew, Saama Technologies, Campbell CA, USA
Jaya Simha Inampudi, Saama Technologies, Campbell CA, USA

ABSTRACT

SDTM Navigator, an AI-powered application, automates the intricate process of creating SDTM specifications and programs. This innovative tool leverages a Large Language Model (LLM) with a Retrieval-Augmented Generation (RAG) backend to streamline the path to submission-ready clinical datasets.

The process commences by ingesting the study protocol and source data. An initial agent analyzes the protocol to discern trial design domains, creating a foundational study map. Subsequently, this information is channeled to a specification agent that correlates the trial design and source metadata to recommend SDTM domains and generate a draft specification.

Throughout this workflow, a critical human-in-the-loop paradigm empowers users to review and modify outputs at each juncture, ensuring digital specifications align with study requirements. Upon receiving the validated output, the final programming agent translates the specifications into executable SAS and R programs.

1. INTRODUCTION

The clinical data lifecycle is frequently bottlenecked by the manual translation of protocol requirements into CDISC-compliant standards. Creating Study Data Tabulation Model (SDTM) specifications and corresponding programs is traditionally labor-intensive, requiring programmers to manually interpret protocols, map source variables, and write repetitive code. This manual approach is often prone to human error and inconsistency.

This paper introduces **SDTM Navigator**, a solution designed to accelerate the delivery of the entire SDTM submission package. By combining Agentic AI with a standards-based approach, the application automates both specification creation and programming, offering a streamlined path from raw protocol to submission-ready datasets.

2. METHODOLOGY: AGENTIC AI WITH HUMAN-IN-THE-LOOP

SDTM Navigator operates on a multi-agent architecture grounded in a **Retrieval-Augmented Generation (RAG)** backend. This ensures that the Large Language Model (LLM) references verified standards (currently supporting CDISC SDTM IG 3.4 and NCI Controlled Terminology) rather than relying solely on generated patterns.

The application follows a four-step sequential workflow, each governed by a specialized AI agent and a **Human-in-the-Loop (HITL)** validation checkpoint:

1. **Auto-Generate Trial Design Domains:**
    - **Input:** Protocol (PDF), CT.Gov/EudraCT data.
    - **Action:** An initial agent analyzes the protocol text to extract trial design elements.
    - **Output:** Draft Trial Design Specifications for user review.

2. **Auto-Identify SDTM Domains:**
    - **Input:** Study Source Data, validated Trial Design specs.
    - **Action:** The system correlates the study design with the raw data structure.
    - **Output:** Recommendations for applicable SDTM domains.

3. **Auto-Generate SDTM Specifications:**
    - **Input:** Historical Specs, Global Macros.
    - **Action:** A specification agent maps source variables to target SDTM variables, leveraging historical study patterns to ensure consistency.
    - **Output:** A complete SDTM Specification.

4. **Auto-Generate Programs:**
    - **Input:** Finalized Specifications.
    - **Action:** The programming agent translates the digital specifications into executable code.
    - **Output:** Native **SAS and R programs**, along with detailed processing notes.

3. DEMONSTRATION

The software demonstration will highlight six core features that distinguish SDTM Navigator from traditional mapping tools:

- **Agentic AI & Standards Automation:** Unlike template-based tools, the system uses autonomous agents to reason through study logic, generating executable programs directly from specifications.

- **Automated Define Spec Generation:** The tool automatically generates **Define specification** based on the final specifications, ensuring "consistency by design" and eliminating duplicate documentation efforts.

- **Standards Management & Reusability:** Users can maintain global SDTM standards and libraries. The system allows teams to "jump-start" new studies by importing historical specifications and verified macros.

- **Trial Design Domains:** By integrating public data sources (like clinicaltrials.gov) with protocol analysis, the tool automates the complex creation of Trial Design (TD) domains.

- **Version Control:** Built-in versioning for specifications ensures that all changes made during the study lifecycle are tracked, managed, and auditable.

- **SAS / R Support:** The platform is language-agnostic, allowing programmers to generate native code in either **SAS or R**, fostering flexibility and preventing vendor lock-in.

4. BUSINESS IMPACT AND BENEFITS

The implementation of SDTM Navigator addresses critical industry pain points regarding speed and traceability.

- **Accelerated Timelines:** The tool has demonstrated a **50% reduction** in the time required to generate specifications and a **69% reduction** in the time to generate and QC first-draft SDTM datasets.

- **Full Traceability:** The application provides a full audit trail and lineage of transformations down to the variable level, supporting rigorous impact assessment.

- **Version Control:** Built-in version control for specifications ensures that changes during the study lifecycle are tracked and managed effectively.

- **Reusability:** By leveraging tested and validated specifications, the system eliminates the need to recreate artifacts from scratch, ensuring cross-study consistency.

5. CONCLUSION

SDTM Navigator represents a significant advancement in clinical data automation. By successfully integrating Agentic AI with established CDISC standards, it provides a robust framework that reduces manual effort while maintaining the

high quality required for regulatory submission.This demonstration will show how the "Human-in-the-Loop" paradigm ensures that automation empowers, rather than replaces, the expertise of clinical programmers.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Jaya Simha Inampudi
Company: Saama Technologies
Email: jaya.inampudi@saama.com

Author Name: Benzi Mathew
Company: Saama Technologies
Email: benzi.mathew@saama.com