

Auto-mapping of CDISC standards using Agentic AI

Nithyanandan Thangavelu, Lifio.ai, Bangalore, India

Rohit Bailur, Lifio.ai, Bangalore, India

ABSTRACT

The increasing complexity and volume of clinical trial data necessitate more efficient methods for transforming raw data into standardized formats such as CDISC's SDTM and ADaM. Manual mapping is time-consuming, error-prone, and resource-intensive. This abstract introduces an innovative approach leveraging Agentic AI—an autonomous, goal-driven artificial intelligence paradigm—to automate the mapping of source clinical data to CDISC standards. Agentic AI systems combine large language models (LLMs), reasoning capabilities, and task execution agents to learn from historical mapping rules, metadata, and controlled terminologies (e.g., CDASH, SDTMIG, ADaMIG). The system dynamically interprets EDC data structures, applies contextual reasoning, and generates traceable, auditable mappings with minimal human intervention. It also supports iterative learning, continuously refining its logic based on user feedback and validation results. Pilot implementations have demonstrated significant gains in mapping speed (up to 60% faster), improved consistency, and early error detection across therapeutic areas. This automation framework has the potential to transform clinical data standardization by enabling scalable, intelligent, and regulatory-compliant data transformation, aligning with the goals of faster clinical development and submission readiness.

INTRODUCTION

Standardization of clinical trial data into CDISC-compliant SDTM and ADaM datasets is essential for regulatory submissions, statistical analysis, and cross-study integration. Mapping raw electronic Case Report Form (eCRF) data to these standards is complex due to variability in CRF structures, metadata, terminology, and derivation requirements. Traditional manual and rule-based approaches are often resource-intensive and difficult to scale. Agentic AI, based on multiple autonomous but coordinated agents, provides a structured approach to support consistent, explainable, and auditable CDISC mapping workflows.

AIM

To implement an Agentic AI-based framework that:

- Automates extraction of CDISC specifications and study metadata
- Supports syntactic and semantic mapping to SDTM and ADaM
- Aligns mappings across CDASH, SDTM, and ADaM
- Continuously improves through human feedback and prior study learnings

METHODS

The solution is implemented as a multi-agent system, with each agent responsible for a specific stage of the CDISC mapping lifecycle.

Agentic AI Framework

1. Extraction Agent

- Extracts domains, variables, and specifications from SDTMIG and ADaMIG
- Extracts CRFs and relevant metadata from study protocols
- Structures source and standards metadata for downstream processing

2. Mapping Agent

- Performs syntactic and semantic similarity-based mapping between source variables and CDISC standards
- Supports one-to-one, one-to-many, direct, and indirect mappings
- Identifies cases requiring derived variables when direct mappings are not available
- Triggers mapping algorithms to derive SDTM or ADaM variables as needed

3. Standardization Agent

- Aligns mappings consistently across CDASH → SDTM → ADaM
- Ensures harmonized interpretation of variable intent and standard usage

4. Feedback Agent

- Incorporates human reviewer feedback into mapping decisions
- Refines mapping outputs based on corrections and approvals
- Supports controlled human-in-the-loop validation

5. Reinforcement Learning Agent

- Learns from previous study mappings and feedback outcomes
- Improves future mapping accuracy and consistency
- Enables continuous system improvement across studies

RESULTS

Initial applications of the framework indicate:

- Reduced manual effort in SDTM and ADaM mapping activities
- Pilot implementations have demonstrated significant gains in mapping speed (up to 60% faster)
- Improved consistency in domain mapping and variable derivations
- Better handling of indirect and derived mappings
- Enhanced traceability of mapping decisions across studies

CONCLUSION

A focused Agentic AI framework, built on well-defined and coordinated agents, can effectively support standardized mapping to CDISC SDTM and ADaM while maintaining transparency, traceability, and human oversight. By integrating structured metadata extraction, syntactic and semantic mapping, standards alignment, and continuous learning from prior studies, the framework complements existing clinical data workflows. This approach supports scalable and inspection-ready CDISC data standardization and has the potential to improve efficiency and consistency in clinical data preparation for regulatory submission.

CONTACT INFORMATION

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

Author Name: Nithyanandan Thangavelu

Company: Lifio.ai

Address: Bangalore, India

Work Phone: +91 97399 75504

Email: Nithyanandan.thangavelu@lifio.ai

Website: www.lifio.ai