A Guided User Experience for Automated Mapping

Nathan Johnson, eClinical Solutions, Mansfield, MA, USA
Venugopal Mallarapu, eClinical Solutions, Mansfield, MA, USA

ABSTRACT

The hype and promise of fully automated mapping have run counter to the reality of complex, messy source data. In this session, we'll present an approach that aims to bridge that gap – a guided user experience for automated mapping. Rather than being a black box turnkey solution, here the focus is on having a human in the loop expert involved in the most complex aspects, while easing the process for the user with an AI-guided experience. Most attempts at automated SDTM solutions have met the 80/20 rule, capable of automating 80% of mappings. This session will address and consider the challenges and limitations faced by these solutions and present an approach that closes in on the 20% so that we can arrive at a fully automated system that keeps the human in the loop.

INTRODUCTION: AUTOMATED MAPPING AND THE NEED FOR THE HUMAN AT THE HELM

Let's take a moment to think about the USS Enterprise from *Star Trek*. Now, this is a ship equipped with some of the most advanced technology in the galaxy—its computer can calculate warp trajectories, analyze alien life forms, and predict optimal paths through nebulas in seconds. You'd think with all that tech, the crew could just sit back and let the computer fly the ship, right?

But here's the thing: even with this incredibly advanced system, Sulu is still at the helm, hands on the controls, guiding the Enterprise through every twist and turn. You'll see it in episodes like 'The Tholian Web,' where Kirk gives the order, and it's Sulu who navigates them out of a potentially deadly trap. The computer might handle the calculations, but when it comes to real-time decisions—dodging asteroids, avoiding enemy ships—it's the human at the helm who makes the difference.

In clinical data mapping, we face a similar situation. Our AI systems are powerful—they can automate large parts of the process, handling the bulk of the heavy lifting, just like the Enterprise's computer. But when things get complex or the data is messy, we need experts like Sulu, guiding the process and making those critical decisions. Today, I'm going to show you how we keep you, the expert, at the helm, using AI to handle the routine tasks while you make the calls that truly matter.

The journey toward fully automated mapping to the Study Data Tabulation Model (SDTM) has for some time been the goal of many programmers, and the subject of many PHUSE talks. We are reaching a tipping point where it is no longer a question of whether this is possible, but more a question of how we perform this automation in ways that are documented, traceable, reproducible, and auditable.

Currently, mapping data to SDTM standards is a labor-intensive process, relying heavily on standards experts to interpret the SDTM Implementation Guide and deeply understand diverse raw data, and then for programmers to interpret the specification to execute those instructions. This process is not only time-consuming but also prone to inconsistencies.

THE DRIVE AND CHALLENGES OF AUTOMATION

The industry has long recognized the potential of automation in SDTM mapping. By introducing AI and machine learning (ML) to handle repetitive, rule-based tasks, we could significantly reduce cycle times, cut down on manual intervention, and improve overall consistency in data transformation. Automation promises faster timelines, allowing clinical trials to move from data collection to analysis and submission in a fraction of the time it currently takes. For stakeholders, this means getting insights faster, accelerating drug development, and ultimately benefiting patients.

However, the promise of fully automated SDTM mapping remains challenging to realize for several reasons:

1. COMPLEXITY OF CLINICAL DATA

Clinical trial data is notoriously complex and varied. Each study has unique variables, data structures, and terminologies that don't always fit neatly into predefined SDTM standards. For example, data from specialized labs, new therapeutic areas, and novel digital health devices often arrives in formats that resist standardization. This complexity means that even the most advanced AI models can struggle to accurately interpret and map data without human intervention. Further, the interpretation of the standards is difficult even for human experts with many therapeutic area-specific or study-specific nuances.

2. LACK OF TRUST IN AUTOMATION

Beyond the technical challenges, there's a more fundamental barrier to automation: trust. Many data managers and clinical programmers (and nearly all regulatory stakeholders!) are hesitant to rely on automated systems for SDTM mapping because of concerns over the "black box" nature of AI. Without visibility into how an AI model makes its decisions, users are reluctant to fully adopt automation, especially when regulatory compliance and data accuracy are at stake. Mistrust can arise from the fear that automated mappings may miss crucial context or create errors that are difficult to trace back and correct.

These factors have led to our current state in which, despite the enormous potential of AI, organizations still rely heavily on manual processes. The industry needs an approach that combines the efficiency of automation with the transparency and adaptability of human expertise.

In this paper, we'll discuss how a guided user experience—with AI acting as a support tool rather than a replacement—can address these challenges. By not only having a "human in the loop," but truly having a "human at the helm," just like Sulu on the Enterprise, we can build an automated mapping solution that is both efficient and trustworthy.

CURRENT STATE AND LIMITATIONS OF AUTOMATED MAPPING

TRADITIONAL APPROACH TO SDTM MAPPING

In today's SDTM mapping process, clinical programmers and standards experts follow a meticulous, multi-step workflow to ensure data is accurately transformed and mapped for regulatory submission:

1. Specification Development: Standards experts interpret the SDTM Implementation Guide (SDTM-IG), study data, and the study protocols to create a detailed mapping specification. This document provides a detailed blueprint for how raw study data should be transformed into standardized SDTM variables. Developing these specifications is labor-intensive, requiring deep knowledge of both SDTM standards and the unique data structures of each study.

2. Programming the Transformation: Once specifications are finalized, clinical programmers write custom code to execute the transformations, translating the detailed specifications into SAS® or R® scripts that map data according to SDTM requirements. Even with organizational eCRF standards, each study still contains unique variables and data nuances, and thus mapping code is often highly customized.

3. Review and Quality Control (QC): Each step undergoes thorough review to validate accuracy. The specification undergoes thorough review and quality control (QC). Then all code requires its own quality control and output datasets require extensive validation to ensure accuracy and compliance. This process is time-intensive and requires close collaboration across programming, data management, and quality assurance teams.

LIMITATIONS OF CURRENT AUTOMATED APPROACHES

While automation holds promise for reducing manual effort, current automated approaches face two significant limitations:

1. Partial Automation: Only Programming is Automated

Some automated approaches focus primarily on the coding step, providing metadata-driven tools that can generate transformation scripts based on pre-written specifications. This strategy can speed up programming but doesn't address the challenge of manually authoring accurate, detailed specifications. With such approaches programming is not eliminated, as some complex derivations cannot be adequately expressed in even the most robust specifications.

Increased Specification Complexity: Because the automated system needs precise inputs to function, the specification must now include detailed technical information—such as complex data logic and transformation rules—that wasn't necessary when programming was performed by humans who could interpret context. In the worst case the machine-readable specification can become its own proxy programming language. This additional detail demands more time from standards experts and introduces new layers of review, potentially offsetting the time savings achieved by automated programming.

Limited Flexibility: Automating only the programming step can restrict flexibility, especially when unexpected data issues arise. Since the system lacks contextual knowledge, it may not adapt well to data anomalies or changes in mapping requirements, requiring manual intervention to adjust specifications or troubleshoot issues.

2. End-to-end Automation, but with Black Box Limitations

More recent advanced approaches aim to automate the entire SDTM mapping process, from specification to transformation and validation. While ambitious, these solutions often function as “black box” systems, where the internal decision-making processes aren’t transparent to the users. This lack of transparency presents significant challenges for validation, traceability, and, ultimately, trust.

Traceability and Compliance Risks: In regulatory submissions, every transformation must not only be validated as compliant but must also be documented and auditable. When a system’s decision-making process is opaque, validating these automated mappings becomes difficult. Traceability is critical for regulatory compliance, and black box systems make it difficult to document transformation decisions accurately, which can hinder traceability. When users can’t see the logic behind each mapping choice, it becomes challenging to create the Define document and audit trail that regulatory bodies require.

Erosion of Trust: Ultimately, these black box systems erode user trust. When users can’t verify how or why the AI made specific decisions, they become less likely to adopt the technology fully, particularly for complex or high-stakes data transformations where human judgment is essential.

These limitations underscore why fully automated solutions, while promising, haven’t yet replaced traditional approaches. For automation to be viable, it needs to go beyond simply generating code or producing black box outputs; it must offer transparency, adaptability, and human oversight. In the next section, we’ll explore how a guided user experience—keeping “the human at the helm”—can address these gaps by balancing automation with expert involvement, ensuring both efficiency and trust in automated SDTM mapping.

3. THE GUIDED USER EXPERIENCE APPROACH

HUMAN-IN-THE-LOOP DESIGN FOR BALANCED CONTROL AND FLEXIBILITY

A guided user experience leverages human expertise at critical junctures in the mapping process. Rather than requiring exhaustive specifications or generating black box decisions, this approach integrates human input to navigate ambiguities. For instance, when an AI model is uncertain or encounters data that deviates from expected patterns, it can flag these cases for expert review, enabling quick adjustments and informed decisions without needing highly detailed specifications. Rather than the model making one pass and leaving the human reviewer to completely take care of the remainder, an interactive experience allows for the model and the human expert to inform one another throughout the process.

Rather than making specifications overly complex to guide automated programming, the guided approach allows for simpler, high-level specifications that the AI can interpret dynamically, seeking human clarification as needed. This minimizes the specification burden and ensures that the system remains adaptable to different study requirements and data types, letting experts focus on refining mappings only where necessary.

AI GUIDED EXPERIENCE FOR TRANSPARENCY AND DECISION SUPPORT

In a guided user experience, AI acts as a supportive tool that provides interactive suggestions rather than prescriptive outputs. For instance, when mapping to SDTM domains or variables, the AI can suggest appropriate mappings based on historical mappings and similarity analyses, surfacing possible options but leaving the final decision to the user. This approach maintains transparency, because ultimately the human expert can accept or adjust suggestions based on their judgment.

Instead of making “black box” decisions, the AI presents confidence levels for its mappings and transformation suggestions, allowing users to easily identify which recommendations are high-confidence and which may require further scrutiny. By presenting these indicators clearly, the guided user experience allows experts to use a risk-based approach to validate AI-generated mappings as they go, streamlining the QC process and ensuring that all decisions are documented and traceable.

The guided user interface can log each interaction, capturing both AI-generated suggestions and user adjustments. This provides a full audit trail, ensuring that each transformation is transparent and meets regulatory requirements for traceability. When users accept or modify a mapping, these changes are documented automatically, enabling a clear record of decision-making.

BALANCING AUTOMATION WITH USER CONTROL TO ADDRESS COMPLEXITY AND BUILD TRUST

Enhanced Transparency for Trust and Adoption: By involving users actively in complex mappings, the guided approach mitigates the “black box” concern and fosters trust. Users are more likely to adopt automation when they can understand and validate each decision, especially in high-stakes environments like clinical research. Seeing the logic behind each transformation builds confidence in the system, creating a collaborative AI environment where automation supports human expertise without replacing it.

User-centric Design to Increase Adoption: The guided approach must provide a user experience that is intuitive and simple enough that the user could complete an entire specification manually at least as easily as they could in previous Excel-based solutions. Accelerators such as pick-lists, intelligent auto-complete, and derivation copilots, make the user interface an acceptable solution upon which the AI completion has a multiplying affect.

Learning from Expert Input: Each human-in-the-loop intervention provides valuable training data for the AI, enabling it to improve over time. As users adjust, the system can learn from these changes, gradually building a knowledge base that reduces the need for future interventions. With human experts able to refine mappings in real time, the system can adapt to new domains or atypical data structures This feedback loop is essential for achieving higher levels of automation without compromising accuracy or traceability.

4. CREATING A ROADMAP TO FULL AUTOMATION WITH HUMAN-AT-THE-HELM AI

A. INCREMENTAL STEPS TOWARD AUTOMATION: BUILDING TRUST THROUGH ITERATION

Rather than aiming for a fully autonomous system from the outset, the guided user experience provides a framework for incremental automation, where each human interaction serves as a training opportunity for the AI. As users make the adjustments and decisions, the AI learns from these inputs, refining its accuracy over time. This iterative process not only builds a stronger AI model but also fosters trust, as users see the system's performance improve through their input.

By incorporating real-time user feedback, the system can identify patterns in complex mappings and gradually automate more of these edge cases. Over time, as the AI captures and applies these patterns, the volume of necessary human interventions decreases, moving closer to a more autonomous mapping solution without sacrificing quality.

B. DEVELOPING TRANSPARENT AI CAPABILITIES FOR END-TO-END TRACEABILITY

Future development much expand the AI's contextual understanding so that it can make more nuanced decisions. By building an AI that can recognize when additional input is needed (such as when data structures are highly nuanced or when new standards emerge), the system becomes capable of handling a wider range of cases autonomously without risking traceability.

As the AI continues to learn from expert inputs, it also improves in generating audit trails for all decisions, allowing users to see precisely how and why specific mappings were chosen. This is essential for regulatory compliance and transparency, enabling teams to rely on automated mappings with confidence.

C. TARGETED AUTOMATION FOR SPECIFIC, HIGH-IMPACT MAPPING AREAS

Prioritizing High-Frequency and High-Burden Mappings: The roadmap focuses on automating mappings that occur frequently or involve particularly repetitive tasks. This targeted approach maximizes efficiency while freeing up experts to focus on high-value work, like novel or complex data elements that need special attention.

Expanding to Emerging Data Types and Standards: As AI models continue to evolve, the system's adaptability can extend to new therapeutic areas or data types not yet covered by SDTM standards. By including human experts in the loop initially, the system becomes robust enough to integrate these new data sources smoothly, setting the groundwork for broader future automation.

D. THE LONG-TERM VISION: A BALANCED ECOSYSTEM OF AUTOMATION AND HUMAN EXPERTISE

Achieving Higher Automation with Confidence: The roadmap isn't about fully eliminating the human role but rather about automating the routine while keeping humans engaged for complex or unpredictable cases. This balanced ecosystem allows organizations to handle data with greater speed and accuracy while ensuring that the human touch remains at initially at all, then at most, and later only at critical decision points.

Empowering Users as System Supervisors: Over time, as AI takes on more of the mapping workload, human experts shift into a supervisory role, overseeing quality and providing guidance only where necessary. This approach reduces cycle times while maintaining high data quality and compliance, making the system a trusted partner in clinical research.

5. CASE STUDY: EARLY WIREFRAMES OF A GUIDED USER EXPERIENCE

As we've discussed, the path to effective automation in SDTM mapping requires an approach that balances AI-driven automation with human expertise. Now, let's turn from theory to practice with some early wireframes that illustrate this guided user experience in action.

These initial designs reflect the principles we've covered: enabling real-time AI suggestions while keeping the human 'at the helm' to make complex decisions, ensuring transparency for each transformation, and simplifying the specification burden. The goal here isn't to create a black box but rather a collaborative interface where users have full control over mappings and can guide the AI, much like Sulu navigating with Kirk's direction on the Enterprise.

In the following wireframes, you'll see how the user interface is structured to present AI-driven recommendations while allowing users to review, adjust, and document each choice. Each design element is intended to build trust, simplify the workflow, and increase mapping efficiency, setting a clear path for future enhancements in automation without sacrificing user oversight. This is not meant to represent any specific software implementation, but to provide a vision of how the principles described in this paper could be put into practice.

In Figure 1, the specification process begins either by running the model on all raw data to determine what SDTM domains are present, or by the user selecting specific domains and working on them one-by-one.

User can kick off the automatic mapping model, or it could be set to automatically run.

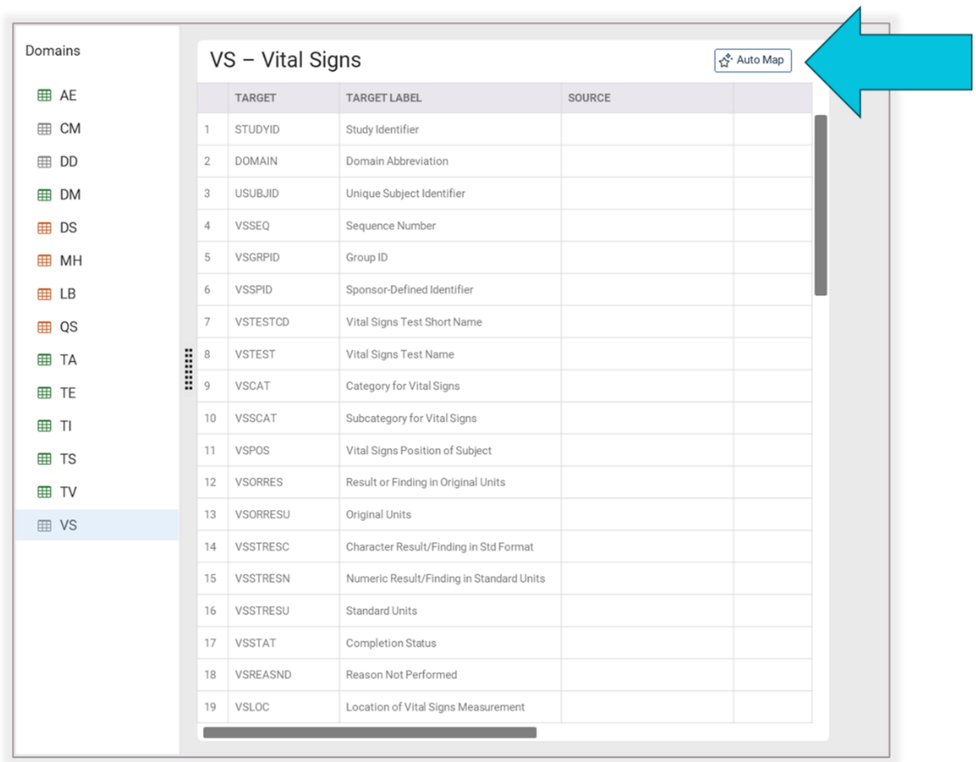

Fig. 1

In Figure 2, results are returned in the UI, with a confidence score associated with each mapping. User is guided to those SDTM variables that need the most attention.

The cutoffs for confidence level could be configured based on the user's tolerance for accepting model decisions.

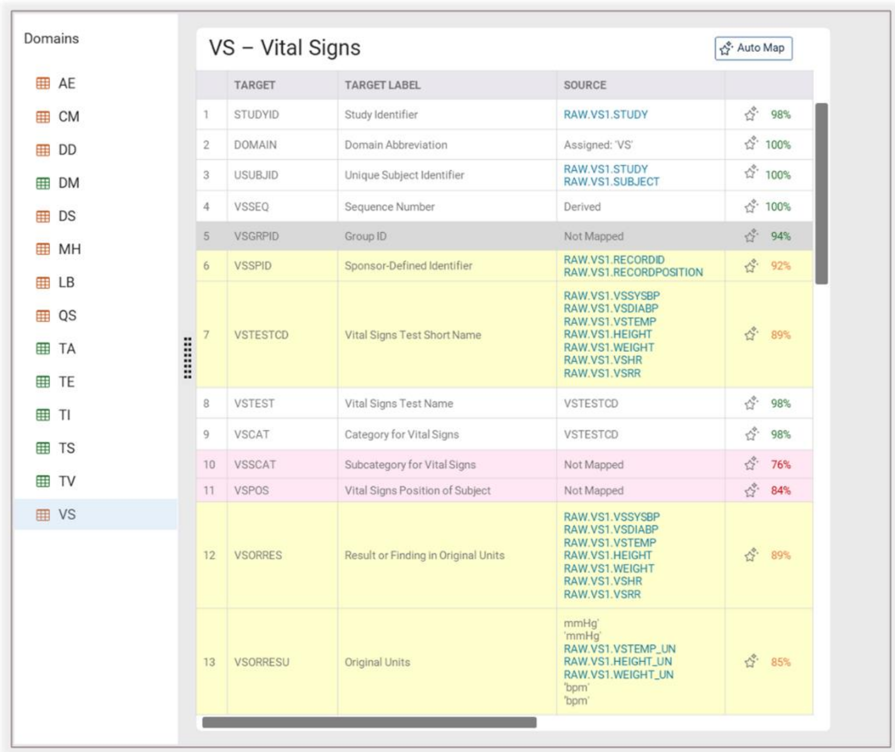

Fig. 2

In Figure 3, If the user disagrees with the model decision, they can update the mapping details. The UI indicates that the mapping is based on user input.

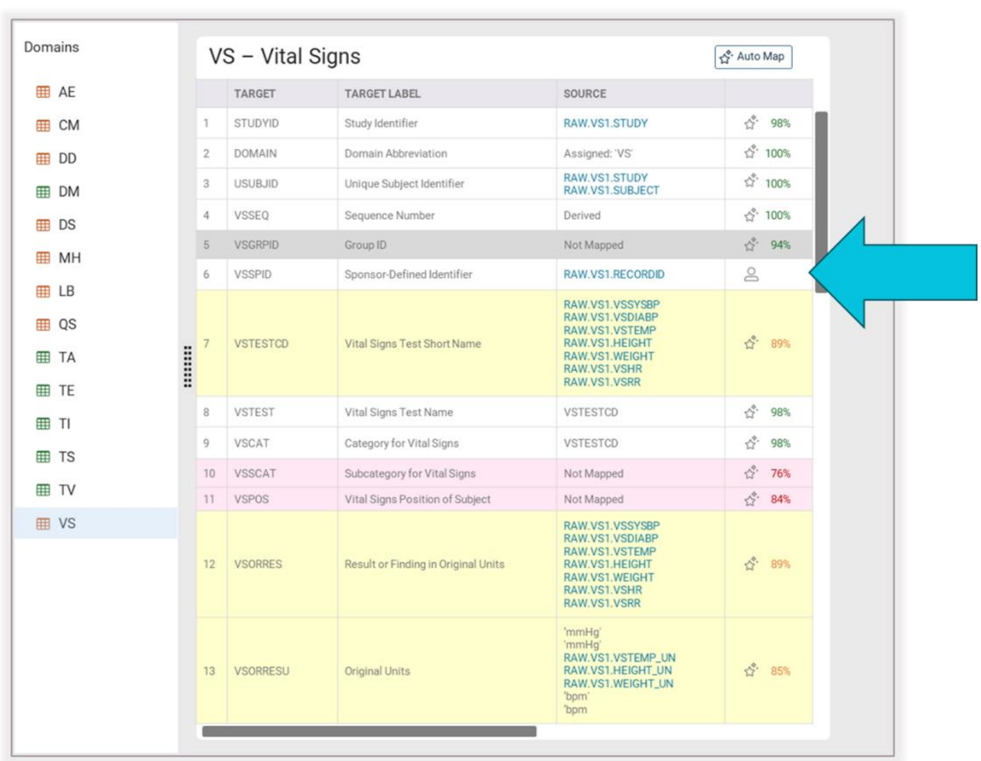

Fig. 3

In Figure 4, if everything looks ok, the user can accept the model decision, and the UI indicates that it has been manually adjudicated.

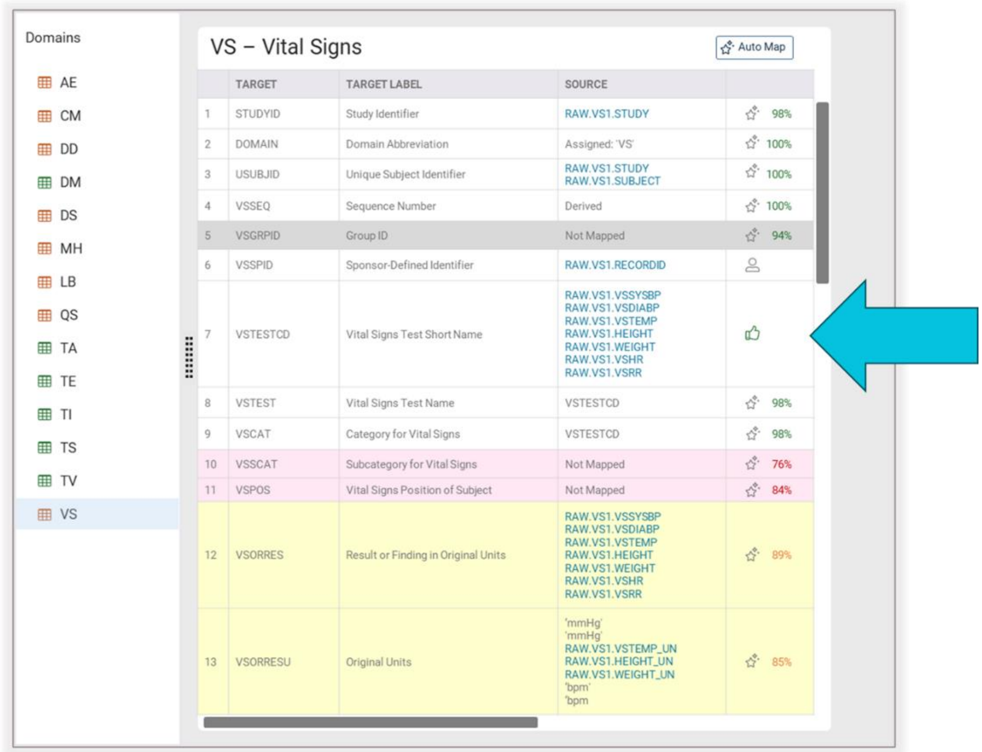

Fig. 4

As shown in Figure 5, this adjudication and review process continues until the entire domain is complete. This is a much simpler process compared to manual specification authoring.

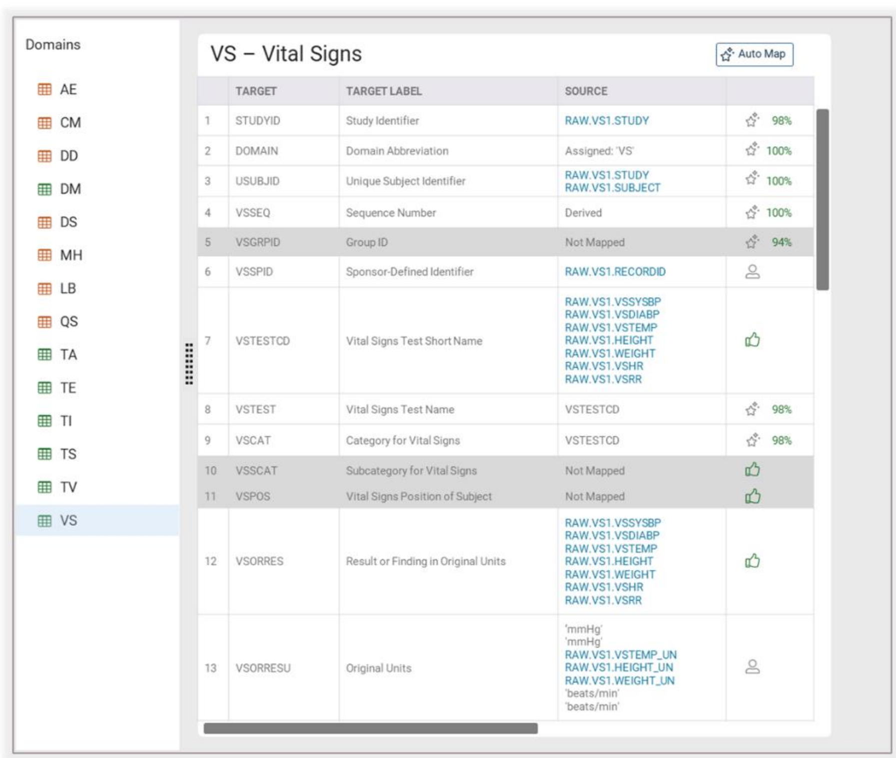

Fig. 5

As shown by Figure 6, this all can be wrapped into a workflow that includes not only SDTM mapping but the overall flow of data from source to target.

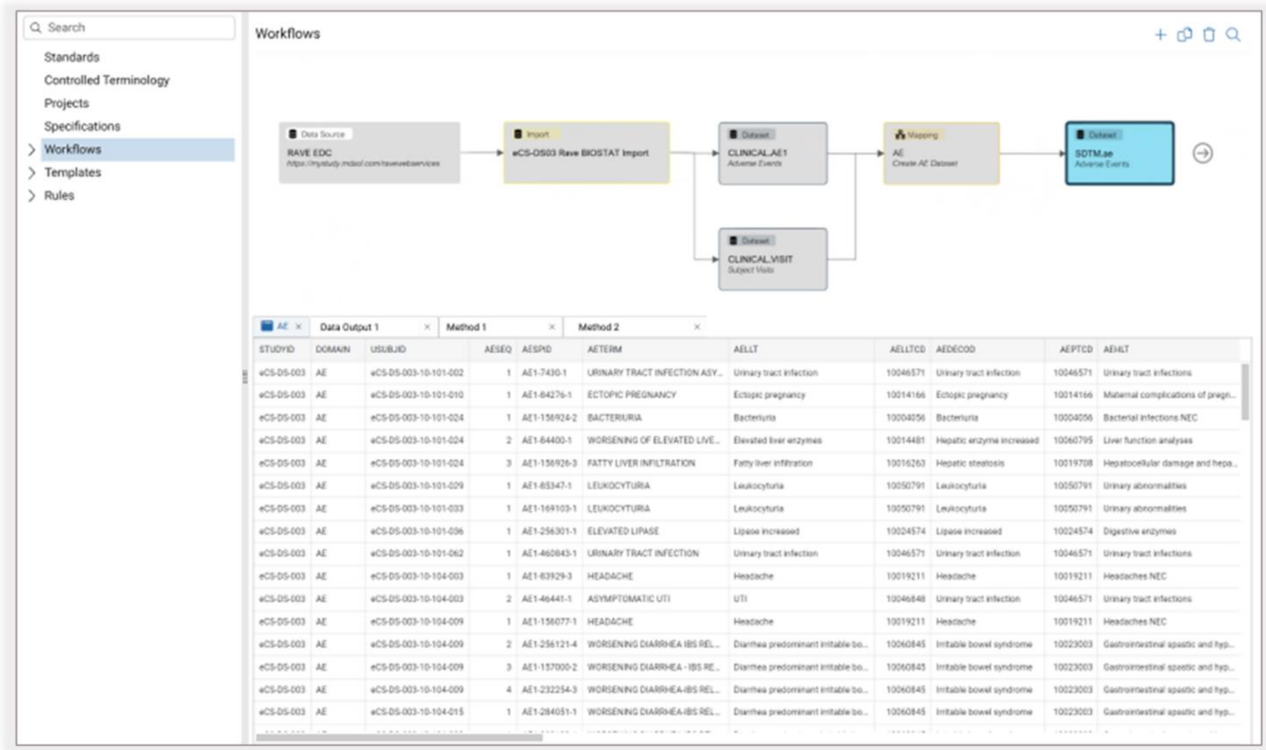

Fig. 6

6. CONCLUSION: TOWARDS A FULLY AUTOMATED, USER CENTRIC FUTURE

Today, we've journeyed through the complexities and challenges of SDTM mapping, examining how traditional manual methods are giving way to a future where AI and human expertise work side by side. We started by recognizing the industry's need for an automated solution to reduce cycle times in SDTM mapping—yet saw how the intricacies of clinical trial data and a lack of trust in black-box models often hinder full automation.

We then delved into the limitations of current automated approaches. Automated coding alone doesn't solve the specification burden, and fully automated systems are too opaque to foster trust or traceability. This led us to the core solution—a guided user experience that retains humans in the loop, bringing transparency and control to the mapping process. In this approach, AI becomes a supportive tool, suggesting mappings and surfacing potential areas for review, but always allowing the expert to steer and refine.

Through the wireframes we reviewed, we saw how this guided experience is structured to bring together efficiency and flexibility. The AI provides a first pass, handling repetitive tasks and highlighting areas of uncertainty, while the human expert brings the context and judgment that only experience can offer. Each user interaction enhances the AI's knowledge, building a progressively more accurate model that continues to improve over time.

The vision, ultimately, is not a world without human experts but a balanced ecosystem where automation handles routine tasks and humans focus on the nuances that make data mapping both art and science. By incrementally increasing automation, guided by continuous human oversight, we move toward a fully transparent, adaptable mapping process that meets compliance standards without sacrificing control or traceability.

In closing, this approach transforms SDTM mapping from a time-consuming, manual task into a collaborative experience with AI—a tool designed to support, not replace, human expertise. By keeping 'the human at the helm' like Sulu on the Enterprise, we're not only building trust in automation but also charting a path to a future where data transformation is faster and more accurate, leading to faster insights. Tomorrow's breakthroughs are out there, and patients are waiting. "Warp speed, Mister Sulu."

CONTACT INFORMATION

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

Author Name: Nathan Johnson
Company: eClinical Solutions LLC
Address: 603 West St., Mansfield MA, USA 02048
Work Phone: +1 216 245 5166
Email: njohnson@eclinicalsol.com
Website: www.eclinicalsol.com

Author Name: Venugopal Mallarapu
Company: eClinical Solutions LLC
Address: 603 West St., Mansfield MA, USA 02048
Work Phone: +1 919 599 3012
Email: vmallarapu@eclinicalsol.com
Website: www.eclinicalsol.com