Paper AD03

Streamlining SEND Dataset Generation: Enhancing Efficiency and Quality with Shiny Dashboards

Madhulika Mishra, Labcorp, Bangaluru, INDIA
Michael Denieu, Labcorp, Madison, WI, USA

ABSTRACT

Data engineering is becoming more important in nonclinical data management because of the Standard for Exchange of Nonclinical Data (SEND) model. The SEND model provides a standardized way to organize and exchange data from nonclinical studies, making it easier to process and analyze. While Laboratory Information Management Systems (LIMS) and SEND software handle most data processing, manual intervention is often needed to ensure conformity. By using data engineering techniques, organizations can streamline the handling of this standardized data, reducing manual work and improving efficiency. In this work, we demonstrate the example of in-house shiny dashboard which enables organization to automate the creation of SEND data and reduce manual intervention and need for QC.

INTRODUCTION

The Standard for Exchange of Nonclinical Data (SEND), developed by the Clinical Data Interchange Standards Consortium (CDISC), is a structured format designed for submitting nonclinical study data to regulatory authorities such as the FDA. Other regulatory agencies like EMA are also investigating the use of SEND as a part of their submission process. Built upon the Study Data Tabulation Model (SDTM), SEND ensures consistency and clarity in data reporting across studies. SEND use is mandated for specific types of nonclinical studies when submitting to the FDA, including single-dose and repeat-dose general toxicology, carcinogenicity, safety pharmacology (covering cardiovascular and respiratory systems), embryo-fetal development (EFD), and studies conducted under the Animal Rule. By standardizing how data is organized and submitted, SEND facilitates efficient regulatory review and supports data integrity and interoperability across the drug development lifecycle.

SEND has a significant impact on drug development by streamlining the way nonclinical data is collected, structured, and submitted to regulatory agencies. By enforcing a standardized format, SEND reduces variability in data presentation, which helps regulatory reviewers quickly interpret and analyze study results. This accelerates the review process, potentially shortening timelines for investigational new drug (IND) applications and new drug approvals. For sponsors, SEND improves data quality and traceability, enabling better internal decision-making and cross-study comparisons. It also facilitates automation in data handling, reducing manual effort, and errors. As drug development becomes increasingly data driven, SEND supports scalability and future integration with advanced analytics, AI tools, and global regulatory frameworks, making it a cornerstone of modern, efficient, and transparent drug development practices.

Generation of SEND-compliant datasets is a technically demanding process that requires precision, consistency, and scalability. As regulatory expectations grow more stringent, and the volume and complexity of nonclinical data increase, traditional manual methods of data preparation are proving insufficient. Manual workflows are not only time-consuming and error-prone but also difficult to reproduce and scale across multiple studies or therapeutic areas.

A data engineering approach addresses these challenges by applying structured, automated, and modular techniques to manage the end-to-end data lifecycle. This includes extracting data from diverse sources such as LIMS, CRO systems, and spreadsheets; transforming it into SEND-compliant formats; applying controlled terminology; and validating it against regulatory standards. Engineering practices such as ETL pipeline design, schema versioning, metadata-driven processing, and automated quality control help ensure that the data is clean, traceable, and submission-ready. Moreover, SEND standards evolve over time, requiring flexible systems that can adapt to new domain structures and controlled terminology updates. Data engineering enables this adaptability while maintaining

compliance and auditability. By shifting from manual curation to automated, rule-based workflows, organizations can reduce operational overhead, improve data integrity, and accelerate the delivery of high-quality SEND datasets—ultimately supporting faster and more reliable drug development.

In this work, we demonstrate the application of a suite of in-house R Shiny dashboards developed to automate and streamline the generation of SEND datasets within a regulatory data engineering framework. Recognizing the increasing complexity of nonclinical data workflows and the limitations of monolithic solutions, we adopted a modular approach—designing multiple dashboards, each tailored to a specific task such as domain mapping, controlled terminology application, time-point edits and other general edits, and metadata generation. This task-specific design not only provides flexibility, scalability, and maintainability but also delivers significant time savings by reducing manual intervention and accelerating routine processes. By distributing responsibilities across focused interfaces, the system enhances data quality, supports efficient quality control, and aligns with evolving regulatory expectations for structured, machine-readable nonclinical submissions. Leveraging R Shiny ensures an interactive, user-friendly environment that empowers data management team to manage SEND workflows with improved transparency, operational efficiency, and measurable productivity gains. As part of this work, we showcase two of these Shiny applications through a case study, showcasing their practical impact on workflow efficiency and data quality.

HIGH-LEVEL R SHINY APPLICATION WORKFLOW

To address the complexity of SEND dataset generation and improve operational efficiency, we implemented a modular automation framework using R Shiny. Instead of a single monolithic tool, we developed **multiple R Shiny dashboards**, each dedicated to a specific task within the SEND workflow. Examples include:

- **Domain Mapping Dashboard**: Automates mapping of study data to SEND domains using configurable templates. One of the examples is the SE generation app.
- **Missing Data Handling Dashboard**: Fills missing time-point variables (example --TPT, --TPTNUM, --ELTM, --TPTREF, --RFTDTC) across domains using predefined rules.
- **NSDRG app:** Automates the creation of Nonclinical Study Data Reviewer's Guide by extracting metadata from SEND datasets and populating standardized templates. It ensures regulatory compliance, reduces manual effort, and speeds up submission preparation.

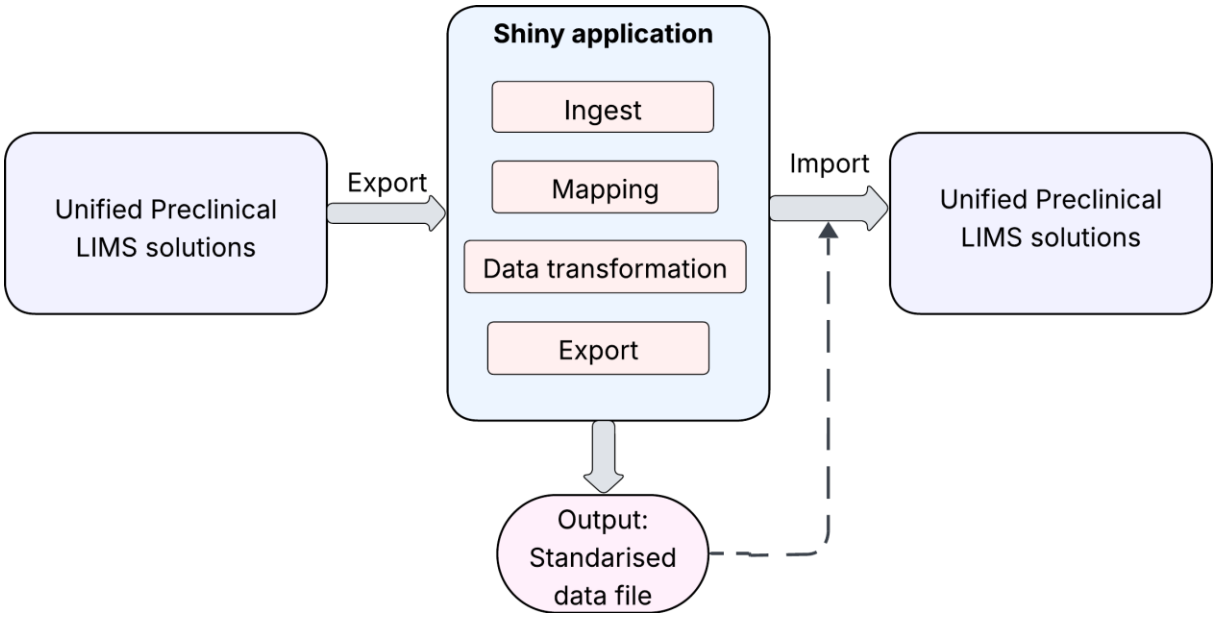

Figure 1: Workflow for data processing using shiny app. The diagram illustrates the end-to-end process, starting with data export from unified preclinical LIMS solutions, ingestion and transformation within the Shiny App, and final upload back to unified preclinical LIMS solutions, incorporating mapping and transformation rules.

Our development process begins with requirement gathering and specification clarification through detailed discussions with Subject Matter Experts (SMEs). These interactions ensure accurate and compliant data handling by defining transformation rules, mapping logic, and output expectations. Our suite of Shiny applications is internally hosted enabling secure and scalable deployment. In **figure 1**, we describe the overall workflow which follows a three-step process:

1. Data Export from unified preclinical LIMS solutions.

2. Data Upload and Transformation via Shiny dashboards for standardization and compliance. We ensure that the dashboards have interactive interfaces requiring minimal user input, enabling study teams to complete tasks efficiently with reduced manual effort which makes these applications user friendly.

3. Output Integration back into the unified preclinical LIMS environment.

Case Study Demonstration

We present two Shiny applications as part of this work. The first one is the **SE Domain Mapping Dashboard** which demonstrates transformation of raw study data into SEND-compliant domains. The next one is the **EDITCSV Dashboard** where we illustrate an automated population of missing timepoint variables, nominal label for a timepoint or collection event for various domains with minimal user input. Through these case studies we achieve measurable improvements in processing time, data quality, and regulatory compliance compared to traditional manual workflows. Time savings were quantified by comparing manual versus automated processes across multiple studies. Using this structured, modular approach, we ensure consistency, traceability, and adherence to SEND standards throughout the data lifecycle, significantly reducing complexity and improving efficiency.

SE DOMAIN GENERATOR

The Subject Elements (SE) domain defines the chronological sequence of trial elements—such as treatment periods, washout phases, and recovery phases—that each subject experiences during a study. It establishes a critical link between individual subjects and trial elements specified in the TE (Trial Elements) domain, while also providing precise timing information for each element. Traditionally, manual generation of SEND domains like SE is labor-intensive and prone to errors. To address this, we developed SE Domain Generator, an R Shiny-based tool that automates SE domain creation by integrating data across key SEND domains (DM, TA, TS, and EX). This automation significantly reduced manual effort, enhanced data integrity, and promoted reproducibility and regulatory compliance in our organization. We describe the SE domain generation workflow for general toxicology studies below (High-level pseudocode workflow is shown in **figure 2, page 4**).

UI section: File Upload & Validation

1. Start: Wait for user to upload CSV files.
2. Validate Input:
    a) Check if required domains (DM, BW, DS, EX, TA and TS) are present.
    b) If any domain is missing, display an error message.
3. Load CSVs: read files into a reactive list of tibbles.
4. Validation Check:
    a) If validation fails, disable the GO button.
    b) If validation passes, enable the GO button.

Server (Part-I): SE Domain Construction

5. On GO Button Click: Retrieve domain names from uploaded files.
6. Prepare SE Structure: Use DM and TA domain to create frame of SE domain
    a. Extract ARMCD values from dm.
    b. Initialize empty lists for TAETORD, ETCD, ELEMENT, USUBJID.
7. Loop Through ARMCD: For each ARMCD, filter ta and dm domains. Repeat TA elements for each subject in dm and Populate lists for SE domain.
8. Combine Lists: Merge lists into a single SE dataframe. And add columns: STUDYID, DOMAIN, USUBJID, SESEQ, ETCD, ELEMENT, etc.
9. Populate timing variables for Predose elements
    a) TS Domain Check: If TS exists, set SESTDTC for predose using EXPSTDTC.

b) Join DM: Merge RFXSTDTC from dm to update SEENDTC.

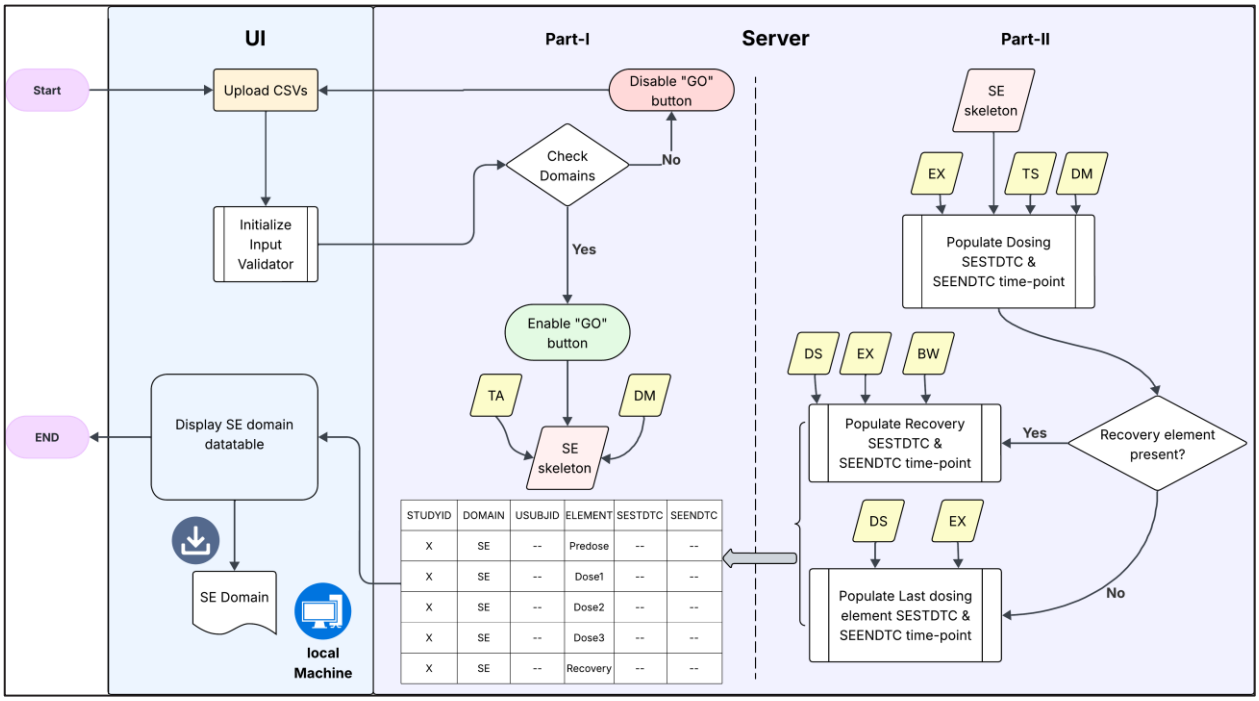

Figure 2. SE Application Process Overview. The SE application follows a structured workflow starting with input validation and CSV upload for domain verification. The "GO" button remains disabled until all domains are validated. Using DM and TA domain data, the app constructs the initial SE domain skeleton, excluding SESTDTC and SEENDTC. In the second phase, dosing-phase start and end dates are inferred from EX, DM, and TS domains. If recovery elements are present, the app determines recovery details and final dosing end date using DS, EX, and BW domains. If the recovery element is not present, the app determines the last SESTDTC and SEENDTC of last dosing element using the DS and EX domain. Once the SE domain is fully assembled, it is displayed in the UI for visual review and can be downloaded by the user.

Server (Part-II): Final Adjustment to populate timing variables for dosing and recovery elements

10. Modify EX Domain:
    a) Convert EXSTDTC to date, EXDOSE to numeric.
    b) Group by USUBJID and summarize first dose date.
    c) Add the next-row column for ENDTC.
11. Merge EX with SE: Update SESTDTC and SEENDTC using EX data.
12. Join DS Domain: Merge DSSTDTC for subjects without recovery phase.
13. Recovery Phase Logic: If recovery exists:
    a) Use the BW domain to compute the recovery start date.
    b) Update SESTDTC and SEENDTC for recovery elements.
14. Finalize SE Dataset: Apply conditional logic for recovery.
15. Render Table: Display SE domain in UI.

How to Run SE Generation App

1. Access the App

- We are hosting an application internally, and users need to select **"SE"** from their content page.
- Below is a screenshot of the initial application interface (**figure3**):

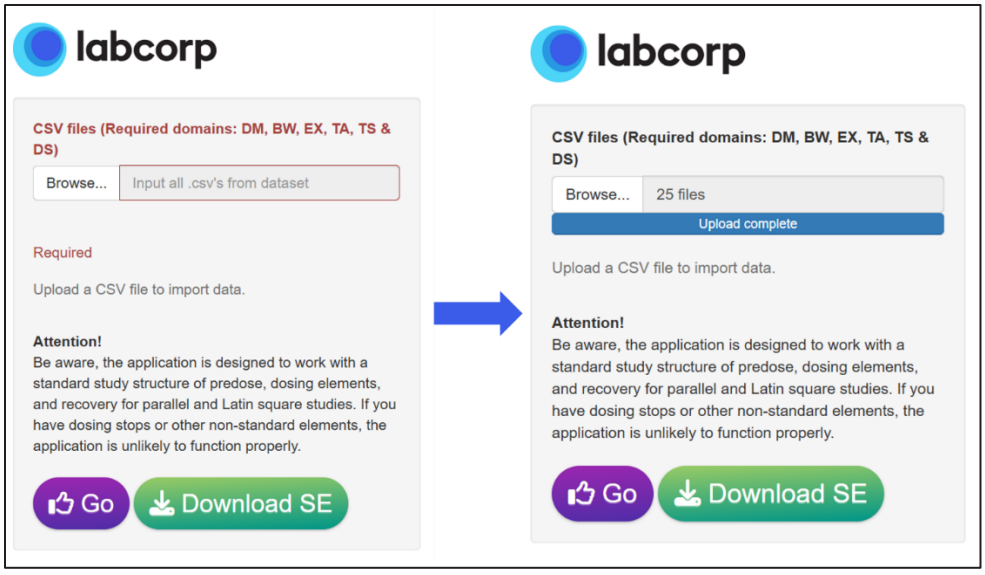

Figure 3. SE Application UI Overview.

2. Upload Required Domain Files

- User needs to use the interface to upload all required domain files.

Note:

- If any required domain is missing, the app displays an error and prevents further steps.
- When all files are uploaded successfully, the app shows **"Upload complete"**.

3. Generate SE Domain

Click **GO** to generate the SE domain.

4. Download the Output

- Use the **Download SE** button to download the generated file. Example output is shown in **figure 4**(page 6).

Output

- SE domain generated for the given study type.
- Downloadable as **CSV**.
- Displayed in-app for a quick sanity check.

Impact of SE application

While a standard conversion study may take around 20 minutes, the SE app completes the task in under 10 seconds. This dramatic reduction in time highlights its efficiency, even when accounting for staggered starts or added complexity.

EDITCSV APPLICATION

Preparing nonclinical study data for regulatory submission requires strict adherence to the SEND Implementation Guide **(SENDIG)** standards. While preclinical LIMS solutions provide raw exports of study data, these outputs often lack certain critical fields and structural conformance, necessitating manual intervention. Traditionally, these edits are time-consuming and prone to errors. We developed EDITCSV application to bridge this gap, offering an interactive, automated solution for transforming raw LIMS outputs into SENDIG-compliant datasets. EDITCSV application streamlines SEND data preparation by performing complex data transformations, validations, and automated data completion, across multiple domains. In the next section, we describe the core capabilities of the EDITCSV application.

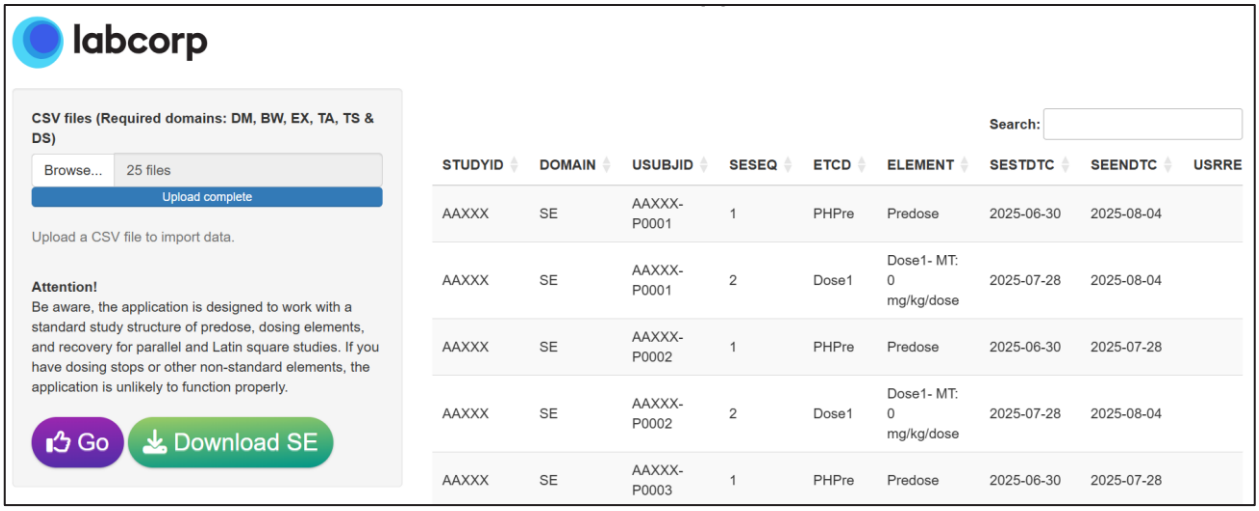

Figure 4. Example output generated using SE application.

Core capabilities of EDITCSV application

The application's core capabilities span three pillars: multidomain automation, data completion, and validation & review (**figure 5**). It enables real time editing of timepoint variables --TPT, --TPTNUM, and --ELTM—across applicable domains and automatically populates the --USCHFL flag to distinguish scheduled from unscheduled collections. Where reference timing is required, app can derive --TPTREF and update --RFTDTC to keep elapsed time calculations consistent. For incomplete records, the app performs automatic completion of missing --DTC and --DY values, aligning date/time and relative day with study references only for a few domains. Finally, it supports quality oversight by displaying read-only exception tables whenever --STAT/--REASND are incorrectly populated, ensuring reviewers can quickly spot and resolve issues while preserving an audit friendly, noneditable view of the discrepancies. We showcase two examples of the EDITCSV application in detail below.

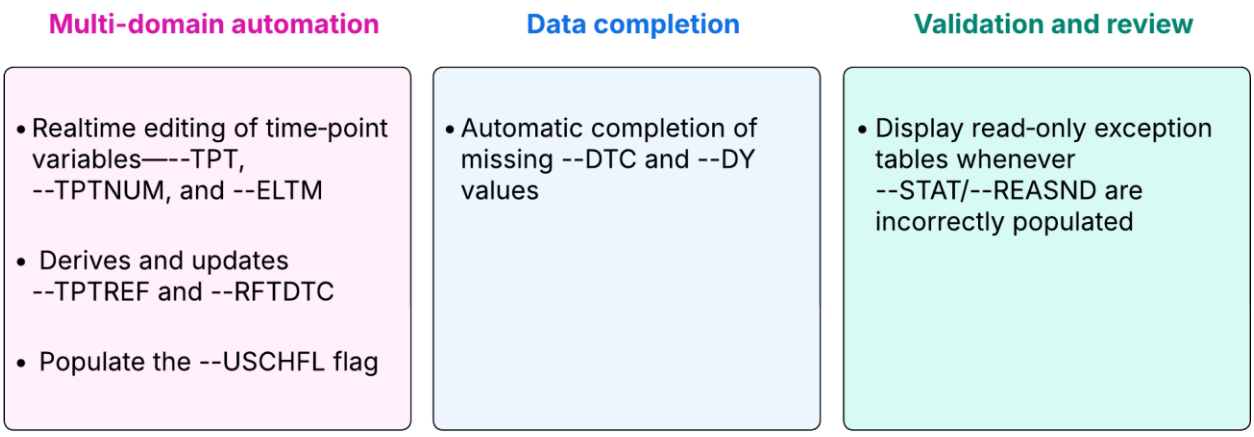

Figure 5. Core features of EDITCSV application: multi-domain automation, data completion, and validation for SEND compliance

Case Study 1: Real-time timepoint edit feature

One of the key functionalities of the EDITCSV application is its ability to handle time point (TPT) edits in SEND-compliant datasets. In **figure 6**, we demonstrate this capability using the Clinical Observations (CL) domain as an example. In SEND standards, CLTPT (Time Point Name), CLTPTNUM (Numeric Identifier), and CLELTM (Elapsed Time) are three important variables that together ensure proper sequencing of observations.

The workflow begins with the user uploading domain-specific CSV files and selecting configuration parameters such as dosing reference points (e.g., EXSTDTC for start of dosing). The application then validates the dataset against CDISC SEND rules, checking for consistency among these variables. If errors such as duplication or certain TPT/TPTNUM/ELTM combinations are detected, an alert is displayed (TPT/TPTNUM/ELTM error), prompting corrective action.

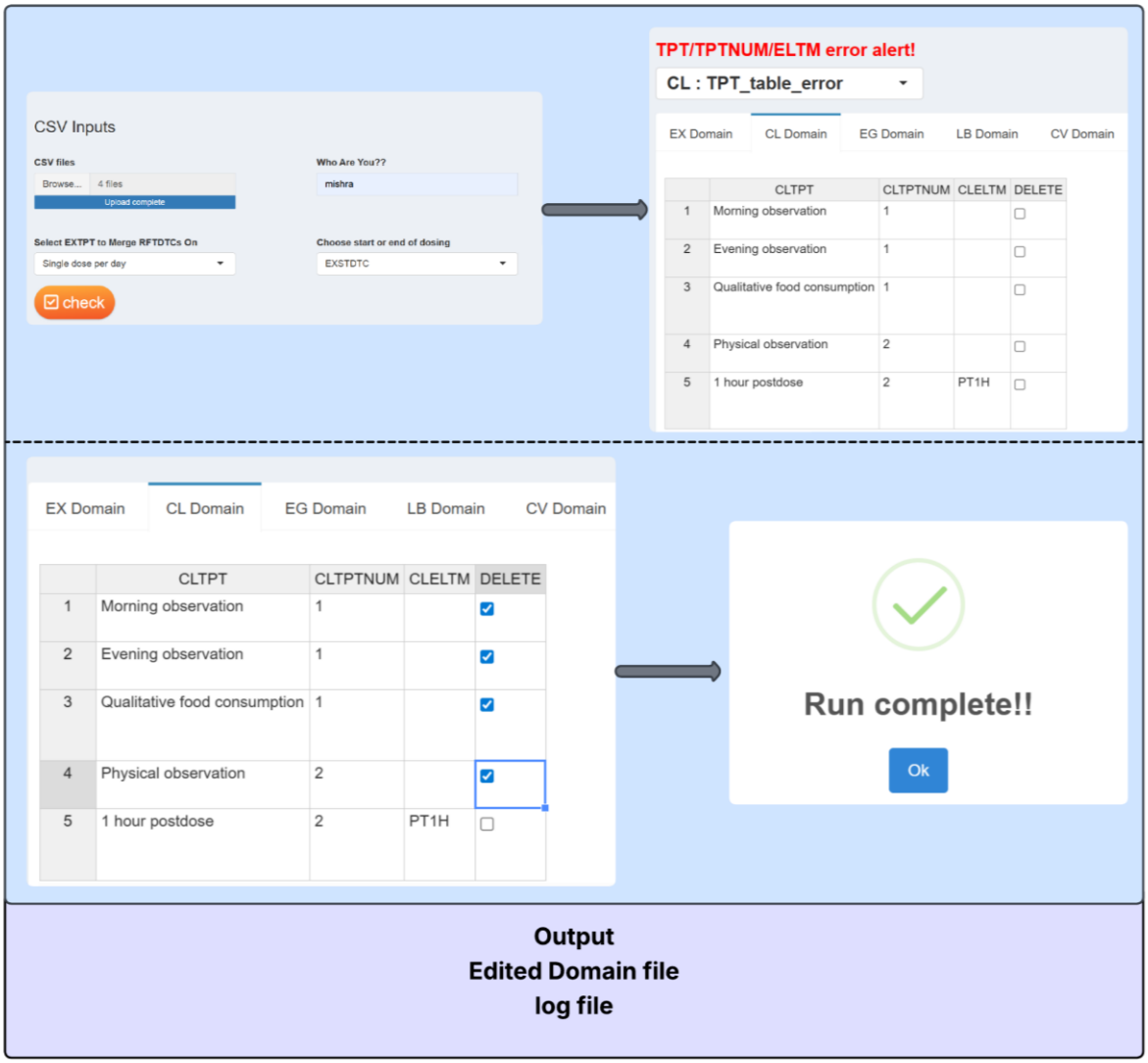

Figure 6. Example of time point edits in clinical observations domain (CL) Using EDITCSV application. Users can edit or append TPT values directly in the CL domain table, and these changes propagate across the entire dataset to maintain referential integrity. Once edits are finalized, the cleaning process is executed, confirmed by the "Run complete!!" message. The output includes an updated domain file and a log file documenting all modifications for traceability. This example highlights how EDITCSV automates complex SEND compliance checks and supports efficient, scalable data standardization for nonclinical studies.

Case Study 2: Populating the Unscheduled Flag for SEND compliance using EDITCSV

Unscheduled data provides critical insight into events that occur outside the planned study schedule, such as adverse events, protocol deviations, or additional safety assessments. Regulatory reviewers rely on this information to evaluate the integrity of the study, assess subject safety, and ensure that any deviations are properly documented and justified. Proper identification and flagging of unscheduled observations help maintain transparency and compliance with SEND standards, reducing ambiguity during data review.

In SEND datasets, the Unscheduled Flag (--USCHFL) plays a critical role in distinguishing observations that occur outside the planned schedule from those that follow the study protocol. This separation ensures accurate interpretation of protocol-driven vs. ad hoc data and is required for compliance with SEND standards to maintain traceability. Having this flag helps reviewers clearly identify unscheduled events and supports downstream analysis.

The --USCHFL variable is commonly used in domains with scheduled timepoints, such as Clinical Observations (CL), Vital Signs (VS), Laboratory Tests (LB), and others.

Examples of --USCHFL usages-

Domain	--NOMLBL	Description	--USCHFL
CL	Week 4	Observation per protocol	N
CL	Unscheduled	Extra observation due to AE	Y
VS	Pre-dose	Measured at planned timepoint	N
VS	*(blank)*	Ad hoc measurement	Y

Table 1: Examples of scenarios of --USCHFL usage. Allowed values are- **Y**→ Record is unscheduled (outside planned nominal schedule), **N** → Record is scheduled (matches planned schedule), and **Blank** → Not applicable or not determined (should be minimized when --NOMLBL exists).

The EDITCSV application automatically detects and populates USCHFL if missing. An example is shown below-

Before edit

USUBJID	CLTESTCD	CLNOMLBL	CLUSCHFL
SUBJ-001	HGB	Visit 2	
SUBJ-001	HGB	Unscheduled	
SUBJ-002	HGB	UNSCHEDULED	

SUBJ-003	WBC	Visit 5 (unscheduled)	

After edit

USUBJID	CLTESTCD	CLNOMLBL	CLUSCHFL
SUBJ-001	HGB	Visit 2	
SUBJ-001	HGB	Unscheduled	Y
SUBJ-002	HGB	UNSCHEDULED	Y
SUBJ-003	WBC	Visit 5 (unscheduled)	Y

Impact of EDITCSV application

The EDITCSV application completes the task ~10sec which can take up to 35-40 min per study, saving significant time and ensures the SEND compliance.

Conclusion

The adoption of modular, automation-driven solutions such as R Shiny dashboards represents a transformative leap in SEND-compliant dataset generation. By replacing manual, error-prone workflows with interactive, rule-based applications, organizations can achieve gains in efficiency, accuracy, and scalability across nonclinical data preparation. Our implementation of tools like the SE Domain Generator and EDITCSV application demonstrates how automation can compress processing times from minutes to mere seconds, while ensuring rigorous compliance with CDISC SEND standards and maintaining full traceability throughout the data lifecycle.

Automating critical tasks—such as domain mapping, timepoint edits, and population of compliance flags like --USCHFL—not only accelerating data readiness but also delivers clarity and transparency for regulatory reviewers. These innovations eliminate ambiguity, enforce consistency across domains, and embed robust quality controls, ultimately strengthening the reliability and integrity of regulatory submissions.

Looking ahead, expanding this modular framework to include advanced validation checks, metadata automation, and domain-specific editors will further enhance compliance and operational agility. By embracing data engineering principles and leveraging interactive technologies, organizations can transform SEND dataset generation into a repeatable, auditable, and high-velocity process—a cornerstone for faster, smarter, and more reliable drug development.

Libraries and tools utilized in shiny framework

User Interface and Dashboard: shiny, shinythemes, shinyWidgets, shinyvalidate, shinyjs, rhandsontable
Data Handling and Manipulation: Dplyr, Purrr, Readr, Stringr, Tools, lubridate
Tables and Interactive Components: DT

CONTACT INFORMATION

Madhulika Mishra
Company: Labcorp, Bangaluru, INDIA
Address: MSR Vaishnavi, No 29 Union Street, Bangalore, INDIA
email: madhulika.mishra@labcorp.com

Michael Denieu
Company: Labcorp, Madison, WI, USA
Address: Madison - 3301 Kinsman Boulevard, WI, USA
email: michael.denieu@labcorp.com