PharmaSUG China 2024 - Paper AA-10053

An Automatic SDTM Platform: Create Your SDTM Since Case Report Form is Ready!

Mingyue Zhu, Wei Wang, and Mingfeng Zhou, BeiGene

ABSTRACT

The industry trend towards integrated platforms for standardized SAS deliveries, including SDTM, TFL, and ADaM, has prompted the development of an innovative SDTM Platform at BeiGene. This platform leverages a Python-powered backend and a JavaScript-driven frontend to provide SAS programmers with a streamlined method to generate SDTM datasets. Upon uploading Case Report Forms (CRFs) and their corresponding Architect Loader Specification (ALS) files, the platform automates the SDTM process efficiently, regardless of whether actual raw data is present.

Initially, the platform generates an annotated CRF (aCRF) and specifications tailored for SDTM programming and Trial Design Model (TDM). These outputs are presented through an intuitive, user-friendly interface, leveraging study-specific ALS files and global accumulated SDTM annotation library. Additionally, the platform offers study self-adaptive derived variables for SDTM, simplifying the process for users.

Programmer can easily trigger backend SAS macros via an interface button to generate SDTM individual programs and datasets. After reviewing the TDM, aCRF, and study self-adaptive SDTM derived rules, programmers can make necessary adjustments. Subsequently, the final SDTM datasets are generated and updated seamlessly.

INTRODUCTION

In the clinical programming industry, the generation of SDTM datasets traditionally involves SAS programmers manually creating annotated CRFs (aCRF) and SDTM programming specifications, followed by developing individual SAS programs for each domain based on these documents. While there are existing automated methods that enhance efficiency in certain steps, such as automating aCRF generation, this paper introduces an intelligent platform that revolutionizes the SDTM creation process.

This platform offers SAS programmers a comprehensive approach to SDTM generation, encompassing the creation of aCRF, Trial Design Model (TDM), individual programs, and SDTM datasets. By significantly reducing manual and repetitive programming work, the platform eliminates the need for programmers to write code for each domain from scratch.

The simplified workflow involves:

1. Uploading study CRFs, ALS files, and Study Edit Check Specifications to the platform, which manages these files in the backend database.
2. Reviewing generated content such as TDM, a preview of the aCRF (referred to as aCRF hereafter), SDTM specifications (used for annotation and SDTM-like data conversion, referred to as SDTM minus specification or minus spec hereafter), and study-specific SDTM derived specifications.
3. Making necessary adjustments through the platform's interface to tailor these configurations.
4. Clicking a button on the platform to automatically generate SDTM programs and datasets in a predefined location. Programmers can then update individual programs as needed for specific study requirements.

This intelligent platform marks a great advancement, offering a streamlined, user-friendly solution for SDTM creation.

SDTM PLATFORM FLOWCHART

The SDTM Platform operates by handling uploaded files within a backend database using Python. Upon upload, the Study Blank CRF and ALS files are merged with the company's comprehensive SDTM

annotation library to generate an aCRF and Minus Spec. Additionally, the Study Edit Check Specification, containing protocol inclusion/exclusion criteria, is merged with ALS to initialize the Trial Design Model (TDM). All backend-generated files are formatted in JSON data.

Subsequently, the backend transfers this JSON data to the platform interface via an Application Programming Interface (API). The platform interface, powered by JavaScript, receives, and displays the JSON data in a user-friendly manner. Moreover, the interface also utilizes JavaScript to facilitate the creation and execution of SAS programs in a predefined path.

This functionality allows users to interact with the platform interface, where they can click buttons to utilize the JSON data as source data. Users can then execute the company's global SDTM macros, which generate SDTM-Like datasets from raw data, Study Derive Specification, SDTM programs, and SDTM datasets. A detailed flowchart illustrating this process is depicted in Figure 1.

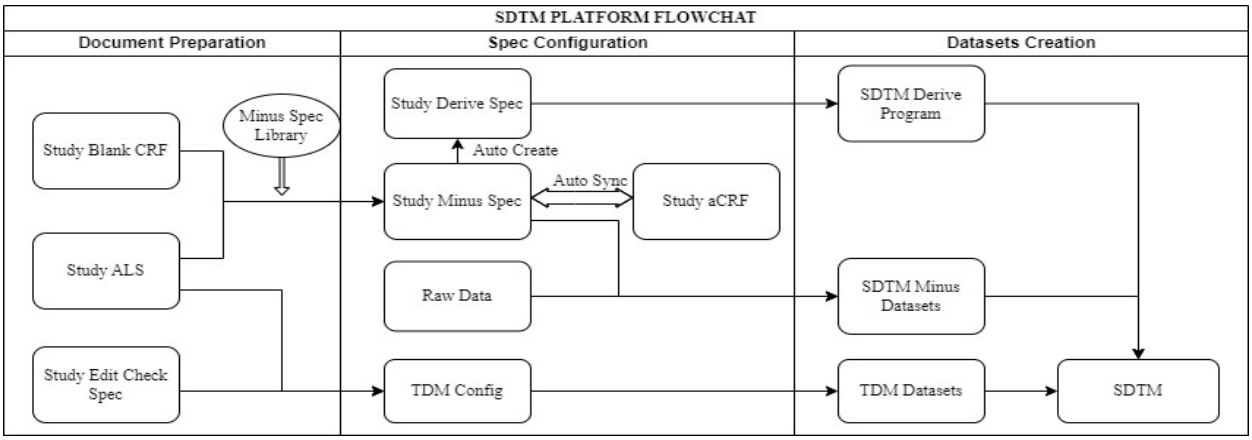

Figure 1. SDTM Platform Flowchart

INITIAL SETUP AND UPLOAD FILES TO DATABASE

To begin setting up a clinical study on the platform, programmers can define a corresponding working folder where the produced products will be exported. This ensures organization and accessibility throughout the study lifecycle.

HOME PANEL

Upon logging into the platform, users are directed to the "Home" panel, illustrated in Figure 2. This panel is one of six panels available, each serving distinct functions detailed further in the accompanying diagram.

The "Home" panel provides an overview of all CRF versions associated with the selected study. Different CRF IDs are linked to various CRF versions or owners. By selecting Button 7, users can seamlessly proceed to the subsequent processes outlined within the platform.

Figure 2. Platform Home Panel

SETTING PANEL

In the Setting panel, users can upload necessary documents to initialize all SDTM metadata. This includes specifying the default annotation reference order to merge annotations from all historical libraries. For instance, if a CRF undergoes a new migration, the platform automatically sets the latest version of the CRF annotation as the highest priority in the annotation reference order. The default rule covers most of the practical need. User can specify priority reference order overriding the default setting as needed.

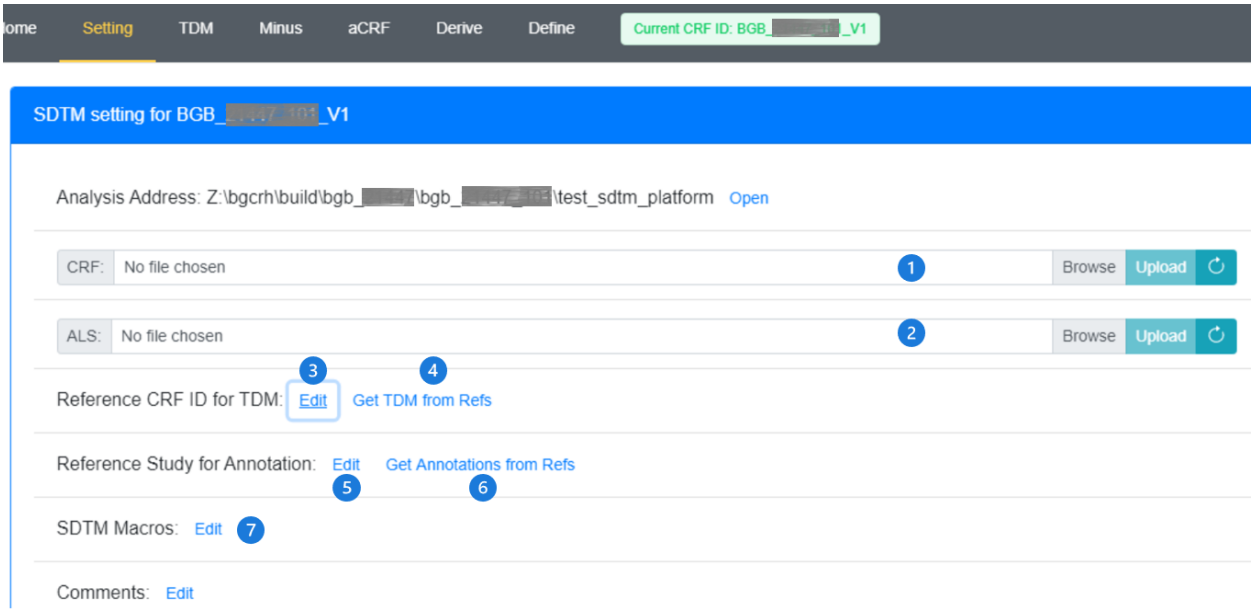

Figure 3. Platform Setting Panel

General settings for SDTM macros can be configured as shown in Figure 4. It includes macro package versions and macro parameters. And it also defines default paths such as raw data, sdtm datasets and programs and so on.

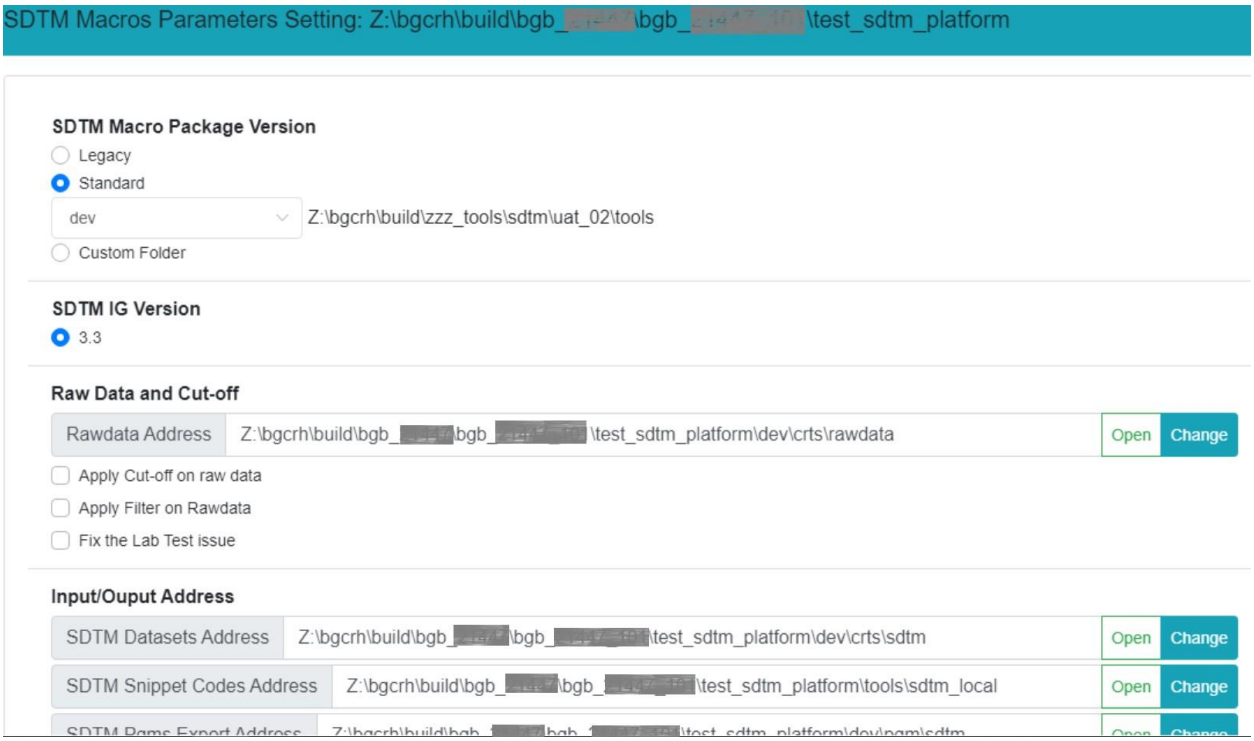

Figure 4. SDTM Macro Setting

REVIEW AND UPDATE SDTM FILES IN PLATFORM

TDM PANEL

Once the initial setup and file uploads are complete, programmers proceed to review, check, and update the SDTM-related files displayed in the platform interface. The TDM panel plays a crucial role in this phase. Here, the uploaded ALS and Edit Check Spec files initialize the Trial Design Model (TDM) with accuracy and efficiency. The panel features components like "TDM Setting" and "TDM Preview," which are further detailed below.

TDM Setting

The “TDM Setting" component allows users to review and update default values retrieved from the database, automatically refreshing the final TDM datasets (see Figure 5). For instance:

- In the "Visit" tab, the "Foldername" value is imported entirely from the ALS file. The "Visit_Type" categorizes visit types into three options: "Image" for scheduled image scan visits (e.g., Response Assessment Week 52), "Non-Image" for general scheduled visits (e.g., Screening, Cycle 2 Day 5), and "N/A" for others (e.g., Adverse Event, Unscheduled Visit). Users can modify these values as needed, adjusting the visit type can exclude certain visits (“N/A”) from the final TV dataset. The database includes built-in algorithms for default VISIT/VISITNUM mappings based on common foldername values in TV, which users can also customize.
- In the "Trial Element" tab, values are initially populated from the database but can be updated according to the study protocol. Icons in the "Action" column facilitate operations such as row deletion, insertion, copying, and pasting. The "EPOCH" column value automatically transposes to corresponding columns in the "Trial Arm and Visit" tab for seamless data management.

This interactive functionality enhances user control and accuracy in configuring the TDM, ensuring alignment with study-specific requirements and protocols.

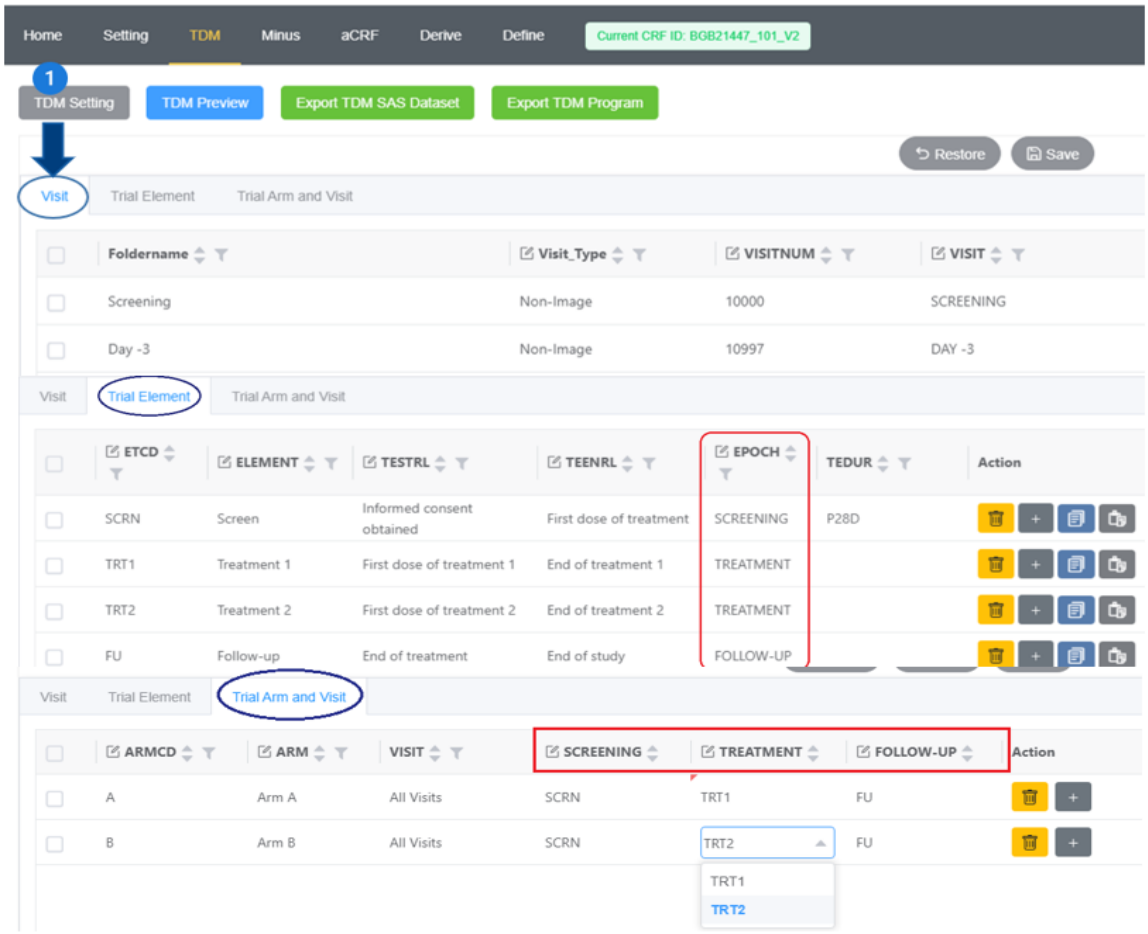

Figure 5. TDM Setting Panel

In the "Trial Arm and Visit" tab shown in Figure 5, users can set arms, visits, and epoch mappings. For each arm, the values in the SCREENING, TREATMENT, and FOLLOW-UP columns automatically synchronize with the "ETCD" of the corresponding epoch in the "Trial Element" tab. Users need to select the appropriate values from the "ETCD" list. For some studies, different arms may experience different visits, and users can select the specified visits for each arm, as illustrated in Figure 6.

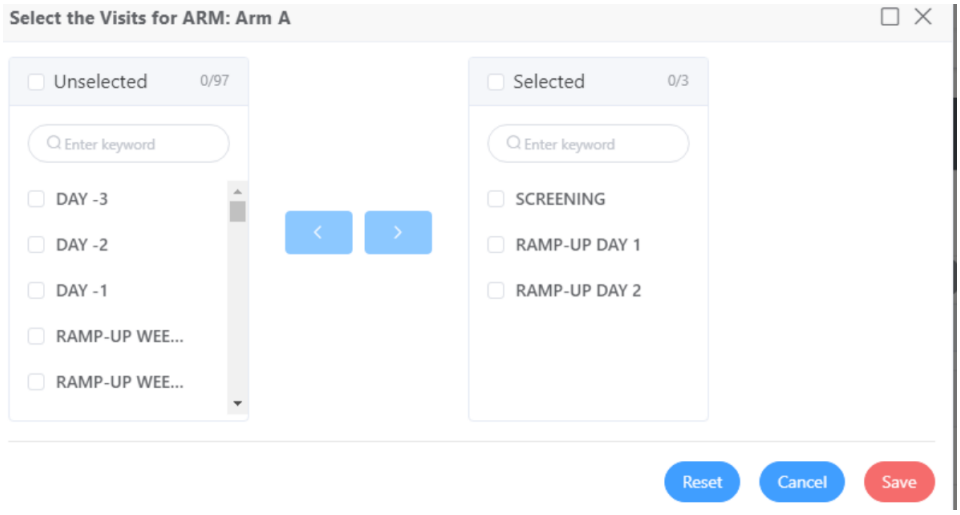

Figure 6. ARM Visits Selection Interface

TDM Preview

Once everything in the "TDM Setting" is configured, click the "TDM Preview" button to review the final TDM files, as shown in Figure 7. The values in the TA/TE/TV tabs automatically synchronize with the "TDM Setting" and are uneditable. Users can update the "TDM Setting" to resolve any issues found while reviewing these tabs.

For the "TI" tab, users need to obtain the Study Edit Check Specification and upload this file through the "Upload" button. The platform will then locate the related contents in this specification and export the inclusion/exclusion criteria, as illustrated in Figure 7. If the length of IETEST exceeds 200 characters, a notification will prompt the user to rewrite the criteria.

After reviewing all the tabs, users can export TDM SAS datasets by clicking the "Export TDM SAS Dataset" button. The backend will silently download TDM files to a default location, and the interface will create and autorun a SAS program to convert TDM files into SAS datasets in a predefined path. Additionally, the backend includes a checklist to identify any issues in the TDM files that could cause the export to fail. If any issues are found, a pop-up window will appear, as shown in Figure 8. Users can also click the "Export TDM Program" button to export the TDM SAS program to a specified location.

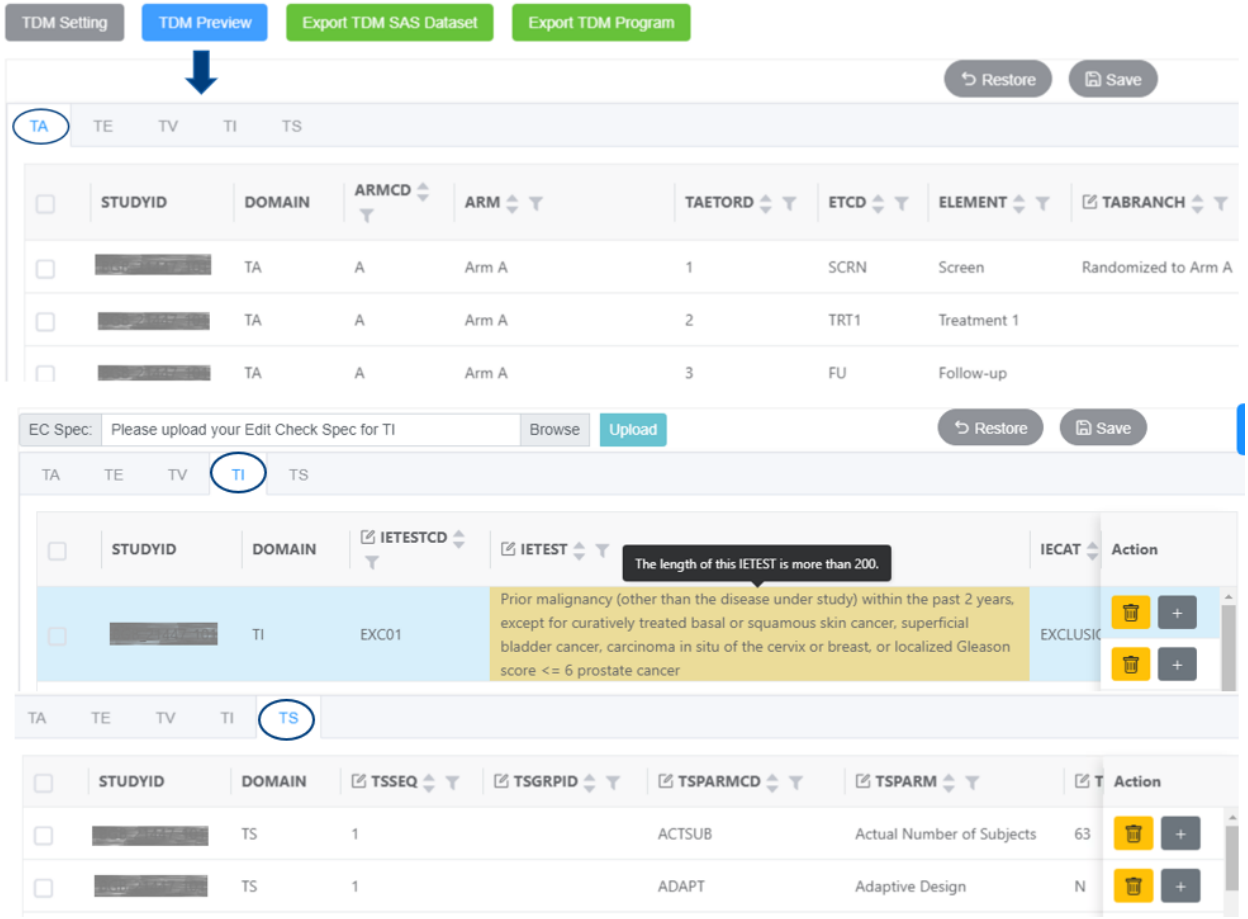

Figure 7. TDM Preview Panel

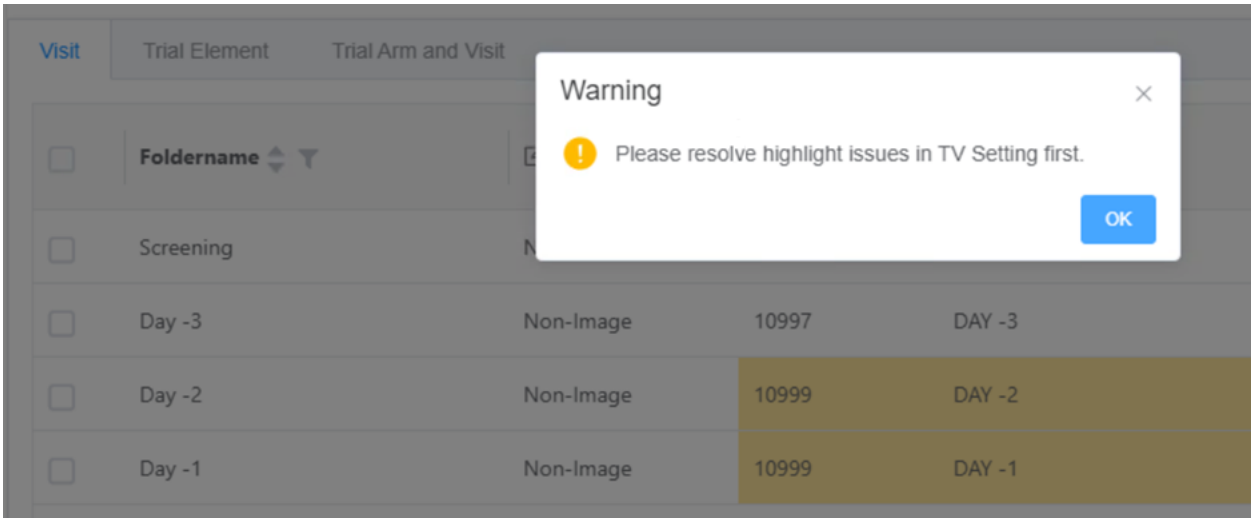

Figure 8. TDM Export Warning

MINUS PANEL

When the blank CRF and ALS are successfully uploaded in the "Setting" panel, the minus specification and aCRF are automatically created in the database. By clicking the "Minus" panel, users can review and update the minus spec, as shown in Figure 9. The minus spec can be interpreted as aCRF equivalent data in Excel format, containing information such as Form, Field, Field Value, reference source, annotation position coordinates, and corresponding SDTM variable mappings (e.g., domain, variable, label, and variable value). Most importantly, it includes the algorithm to process this field for SDTM conversion. It serves as SDTM metadata tailored for minus dataset conversion during SDTM programming. The minus dataset is an SDTM-like dataset, almost directly converted from the EDC raw data per the defined algorithm. By clicking the "Create Minus Datasets" button, the interface will use the minus spec and predefined raw data as source data to run the company's SAS macro and generate minus datasets.

On the left side of the "Minus" tab, users can click the bookmark to activate the corresponding CRF form in red color. There is a built-in data checklist indicating which annotations require further confirmation, with an exclamation mark icon next to the bookmark reminding users to pay attention to highlighted annotation comments. This allows users to easily identify which field’s annotations need careful review.

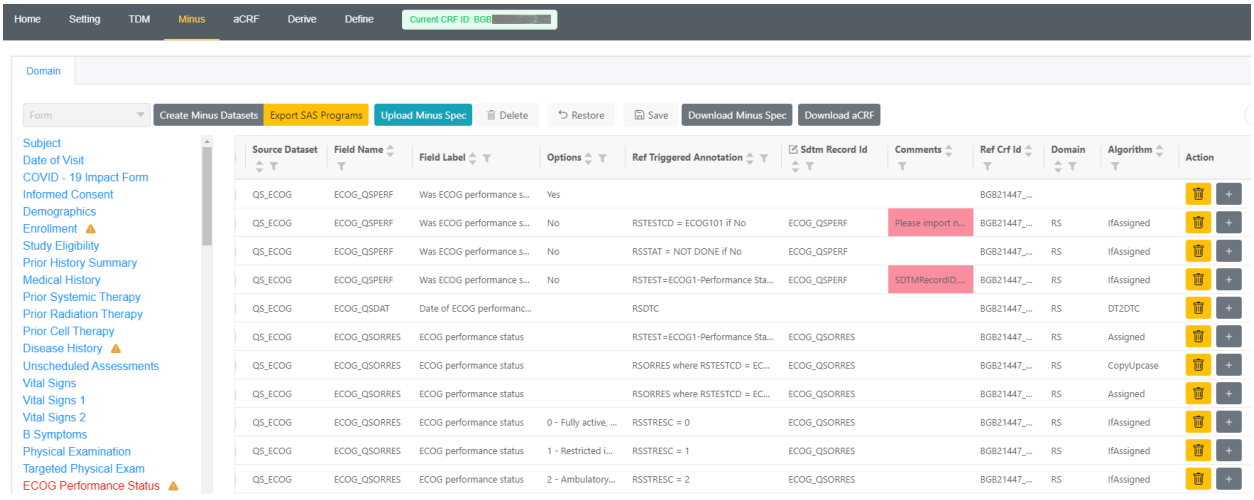

Figure 9. Minus Panel

To edit the current annotation, users can double-click the “Ref Triggered Annotation” column to trigger a pop-up window, as shown in Figure 10. The annotation is decomposed into multiple boxes, allowing users to easily update the annotation in the corresponding box. A list of input suggestions will appear for users to choose from, helping to avoid common input mistakes. The values in the "SDTM Domain/Variable" table will be automatically updated according to the changes made to the annotation boxes mentioned above. The “Algorithm” value explains the derivation method for the current annotation and will be used in the SAS macro to generate minus datasets.

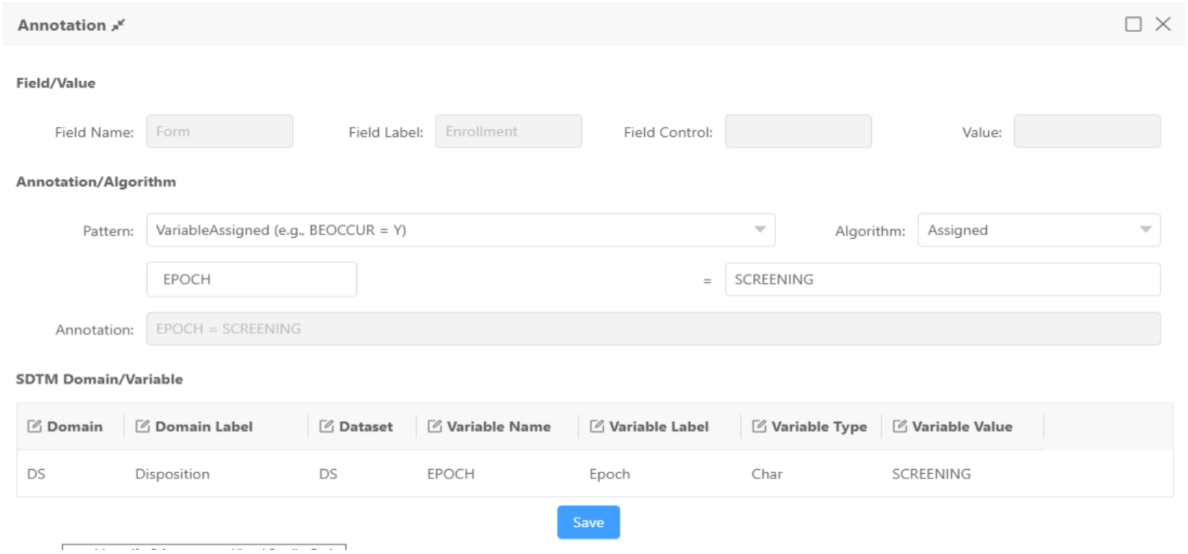

Figure 10. Annotation Pop-up Window

The platform can also analyze the potential annotation patterns for a field based on the input data type of the current field and the annotation level (form level, field value, or value level). The pattern dropdown list in Figure 11 displays all allowed annotation types for the field, and users can select one to fill in the annotation boxes. These features help prevent users from using improper annotations.

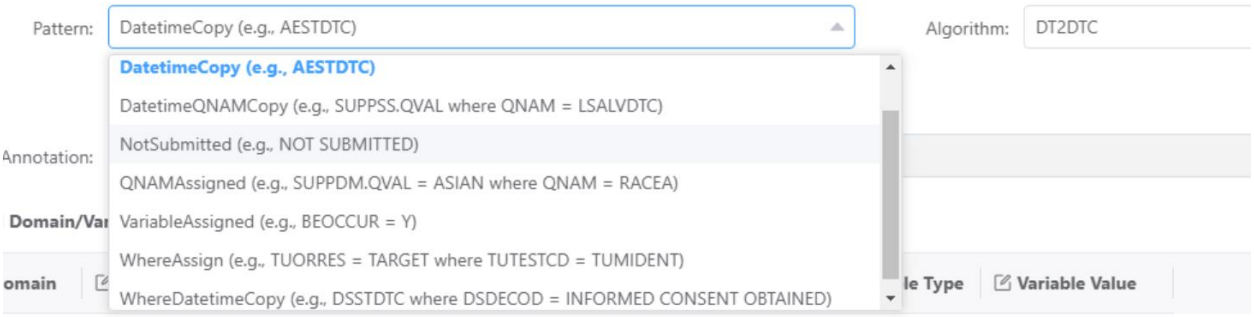

Figure 11. Example of Annotation Pattern

ACRF PANEL

The "aCRF" panel provides users with a real-time, bilaterally synchronized aCRF renderer, as shown in Figure 12. The aCRF renderer and minus spec are essentially the same data in different formats. Updates in either one will synchronize with its counterpart simultaneously. For newly created annotation entries in the minus spec, the box automatically fits the size of the text content, colors the background, and places itself in the blank area of the aCRF renderer.

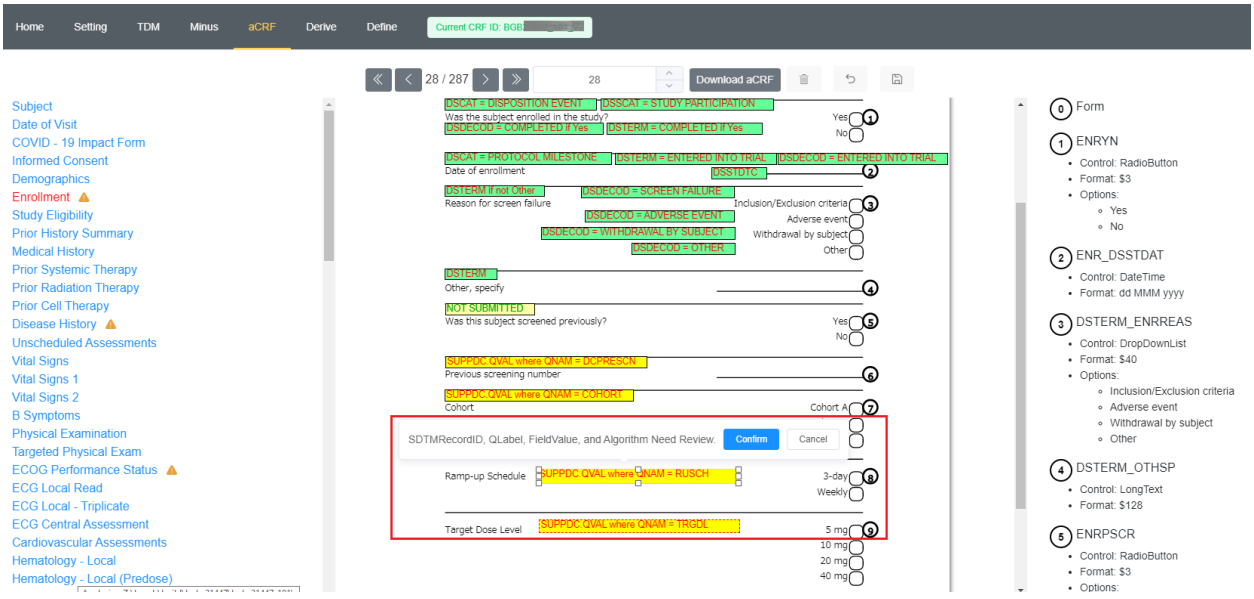

Figure 12. aCRF Renderer Panel

The "aCRF" panel consists of three parts. On the left side, the bookmark can be used to select a CRF form and indicate further review reminders, similar to the "Minus" panel. On the right side, a list of all fields in the current form is provided. Users can drag a relevant field to the middle section, and an annotation pop-up window, like in Figure 10, will appear once the drag is complete. Users can add a new annotation in the pop-up window and click save to see an annotation box with automatic size adjustment and background coloring at the drop position.

The middle section is the main body of the aCRF renderer(or called preview). Users can double-click the box to edit the annotation in a pop-up window like in Figure 10. They can also smoothly move and adjust the size of the annotation box within the aCRF renderer.

The aCRF renderer includes several features to enhance the user experience and streamline operations. If further confirmation and review of an annotation are required, dashed lines will appear around the box, as shown in the red rectangular box in Figure 12. When users click the dashed annotation box, a reminder box with a detailed review checklist will pop up. Additionally, there may be duplicate annotations for repeat fields within one form across multiple pages, such as annotations for the form header area or for the same fields at multiple time points (e.g., VSTPT). In traditional aCRF annotation processes, programmers need to manually add these repeat annotations. However, the platform synchronizes these repeat annotations with the first occurrence origin, so users only need to edit or move the first origin annotation. For convenience, the repeat annotations will be half-transparent and unable to be edited or moved, as shown in Figure 13.

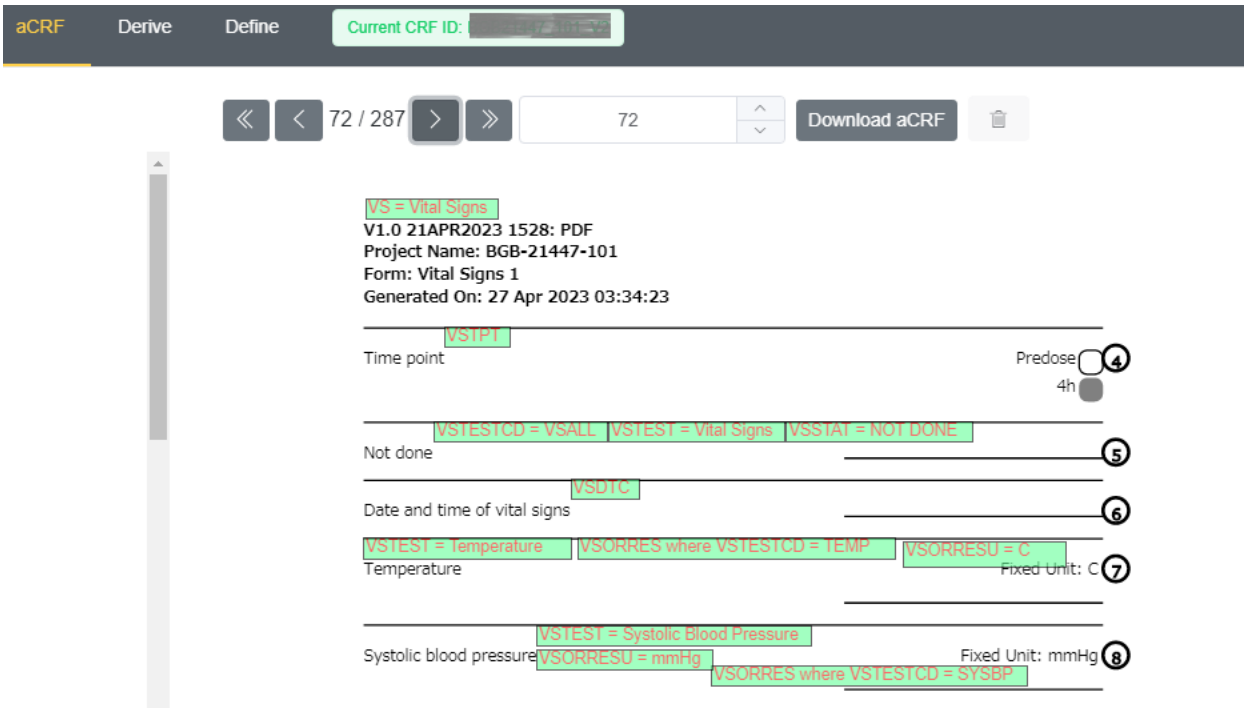

Figure 13. Duplicate Annotations Format in aCRF Renderer

The aCRF is ready to download when users have completed the review and updates of minus spec or aCRF renderer. The downloadable aCRF is totally identical to aCRF renderer. Therefore, the movement and updates in minus spec or aCRF renderer will be automatically reflected in the downloaded aCRF. Additionally, it's very easy to export aCRF by clicking the button "Downloading aCRF" in "aCRF" panel or "Minus" panel, and it only takes less than 15 seconds to download one complete aCRF in pdf format to predefined path.

OBTAIN SDTM DATASETS AND PROGRAMS

DERIVE PANEL

Once all the previous panels are prepared, the final step to generate individual SDTM programs and datasets is to add SDTM-derived variables to the minus datasets and convert them to the final SDTM datasets. This can be accomplished in the "Derive" panel, as illustrated in Figure 14.

Clicking on the "Derive" panel triggers the interface to dynamically generate a programming specification (also called derive spec) containing the derived rules for SDTM-derived variables, such as VISIT, xxSEQ, xxDY, xxBLFL, and so on. We call it 'dynamically' because only the required and expected derived variables will be added not as long as they are not present in SDTM minus. The derive spec will be displayed in the interface under two tabs: "Domain" and "Variables." Users can click the arrow button at the end of each row in the "Domain" tab to go to the "Variable" tab to review derive rules for the corresponding domain.

The "Variable" tab lists detailed programming methods, macros used for the variable, and macro algorithm descriptions. Users can review and update the macro parameters as needed for the study and generate all SDTM programs by clicking the "Create All Dev SDTM pgms" button, using the updated derive spec as source data. Users can then batch run all SDTM programs to generate SDTM datasets. Additionally, if users only want to update the program for a certain domain, they can update the derive rules in the "Variable" tab and return to the "Domain" tab to click the "Dev" or "QC" button to re-generate the corresponding SDTM program for development or validation.

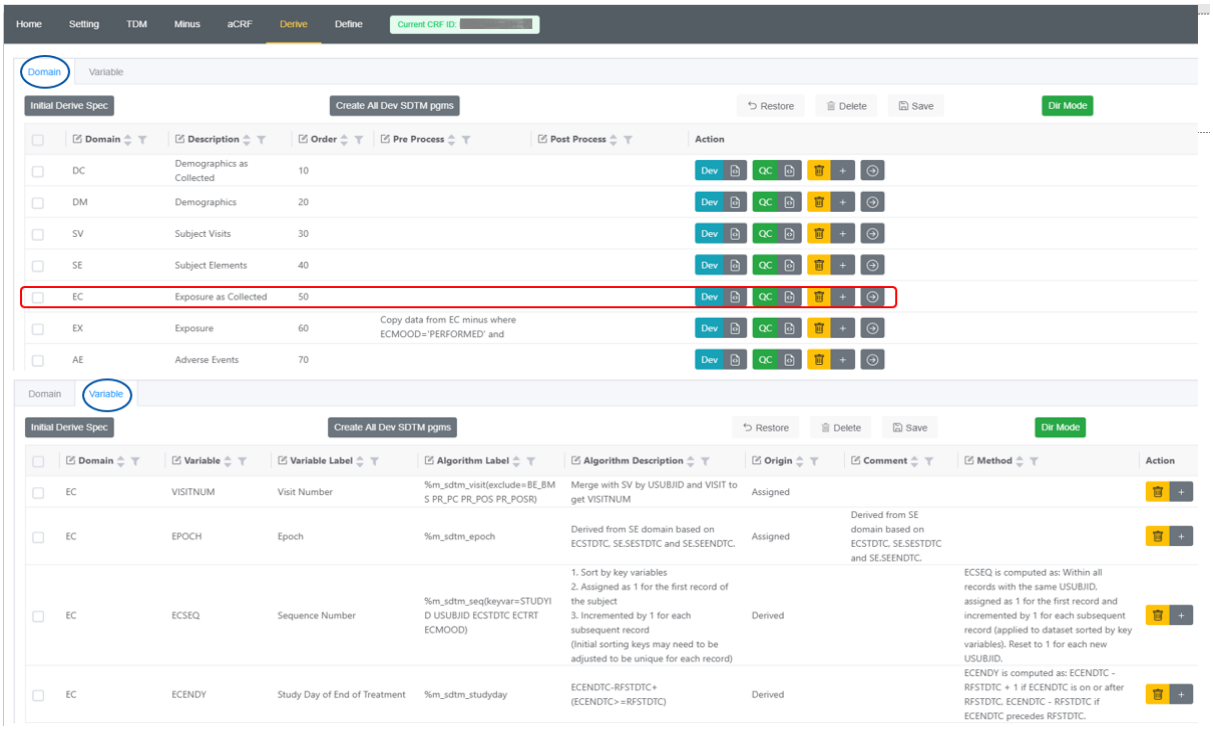

Figure 14. Domain and Variable Tab in Derive Panel

CONCLUSION

This SDTM platform provides programmers with an intelligent, more convenient, and time-saving method to generate SDTM datasets, in terms of saving on tedious and repetitive coding so as to put more focus on reviewing and refining annotation rules to ensure quality. The platform facilitates the SDTM process with the following advantages:

1. Reduced coding requirements.
2. Easy-to-use, straightforward, and self-explanatory UI buttons.
3. Convenient sharing and referencing of study metadata.

Additionally, the functionalities for creating the SDTM submission package are handled in "Define" panel, which is not covered in this paper.

Platformization and innovation are popular trends in the clinical industry. This platform design allows for flexible addition of smart functions in the future such as embed AI application at designated points. Furthermore, it can define input patterns and data checks to guide users in filling data more conveniently with minimized mistakes. This brings use a great of conveniency and efficiency gain.

ACKNOWLEDGMENTS

Heartful thanks to SIR team (Jinling Li, Mijun Hu, Wei Wang, Mingfeng Zhou, Jiapeng He and Xuejie Zhou) and really treasure the amazing and memorable cooperation during exploring the SDTM Platform automation. And I would like to express my great appreciation to Mijun Hu for his meticulous guidance and invaluable insights throughout this paper.

CONTACT INFORMATION

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

Mingyue Zhu
BeiGene
Mingyue.zhu@beigene.com