PharmaSUG China 2023 - Paper AP-130

Trial Design Model: Digitalization and Automation

Han Wang, Beigene (Beijing) Co., Ltd.

ABSTRACT

Digital transformation has been impacting more and more industries, and the pharmaceutical industry is at the forefront. As a next-generation biopharmaceutical company, BeiGene has always been a pioneer of digitalization. From a programming perspective, our transformation strategy involves digitalization, automation, and intelligence, aiming to develop an intelligent programming platform that could offer high-quality digitalization solutions to our daily work.

In this paper, we will share our innovative effort on digitalizing and automating the Trial Design Model (TDM) domains generation approach with Python. We will provide our design ideas, current achievements, limitations of this approach, and prospects we look forward to. We hope to showcase our digital transformation implementation by sharing our experience in developing it. And hope that this paper could be “throwing a brick to attract jade” and provide valuable insights and direction for programmers that are experiencing the process of digital transformation.

INTRODUCTION

The trend of digitalization in the pharma industry is expected to continue in the coming years, driven by a range of factors such as increasing demand for personalized medicine, rising healthcare costs, and the need to accelerate drug development and improve patient outcomes.

Starting from the programmer's perspective, digitalization is seen as a necessary step for development. We are working to construct a fully digital and intelligent platform that initiates from raw data and goes all the way to outputs and even submission packages. In this process, SDTM, ADaM, and TFLs can achieve a high degree of automation with the support of standardized CRFs and shells, which reduces the programmer's repetitive work.

TDM domains are always found to require extensive manual effort, deep understanding, and extra expertise. So, we developed a tool with Python to automate the TDM domains creation process by extracting relevant data from trial information and filling it into a standard structure. With this automation process, TA, TE, and TV will be automatically derived and TI, TS will be more easily built. The tool also enables programmers to modify and update the data sets quickly and provides real-time feedback on the impact of changes on the trial design.

Besides, through this tool, this article will introduce ideas of BeiGene's programming team and design on digitalization, with the hope of inspiring industry peers who share the same ideas to evolve towards digitalization together.

WHAT: THE CREATION OF THE TDM DOMAINS

As an essential component of the SDTM, the TDM datasets provide a standardized way of representing the trial design information in clinical trial data (Figure 1), which facilitates data sharing and analysis across different studies and sponsors. They are special purpose data sets and do not contain subject-level data.

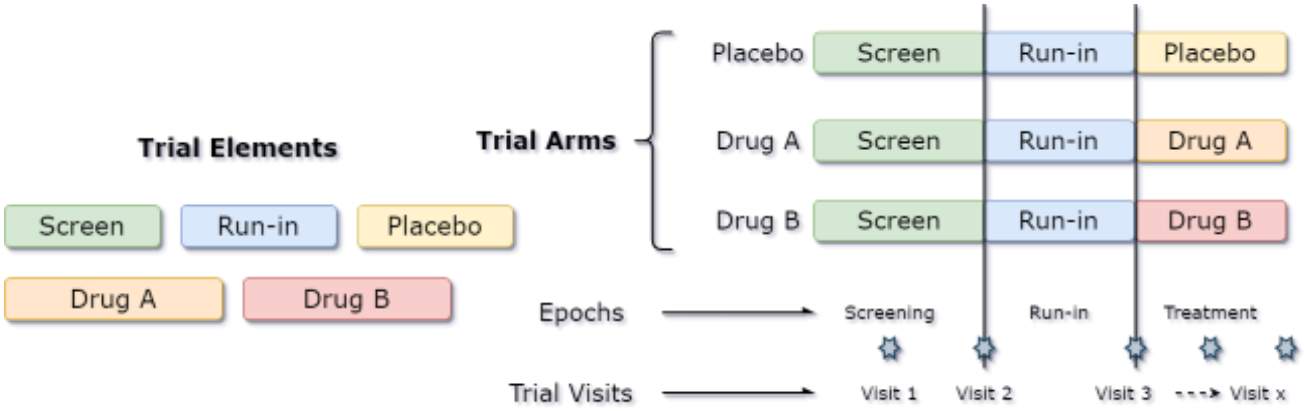

Figure 1. Trial Design Basic Concepts

Usually, the process of TDM data sets includes (Figure 2):

1. Capturing trial design information by identifying the relevant details, such as study design, study arms, treatment groups, and planned analyses.
2. Determining the appropriate variables and their attributes to represent the trial design per SDTM specifications.
3. Creating the TDM dataset using SAS.
4. Populating the dataset with accurate, consistent, and complete trial design information; validating the dataset.
5. Updating the TDM dataset as needed throughout the trial to reflect any changes in the trial design or planned analyses.

Figure 2 Trial Design Basic Concepts

And within this process, several documents and information are required including Protocol, eCRF, Inclusion/Exclusion criteria and so on. The protocol is the main source for TA, TE and TV domain. Since the TI domain must contain the inclusion and exclusion criteria for the trial, a summary of the Inclusion/Exclusion criteria changes (if any) including all the wording updates and protocol version number where they were changed need to be prepared per protocol. For TS domain, it will require that each PARAMCD described in the SDTM IG appendix is filled with the study information present in the protocol.

WHY: CHALLENGES IN A REGULAR PROCESS

As demonstrated above, the structure of TDM domains datasets, while not complicated, differs significantly from the mapping process and workflow of our conventional SDTM datasets. Consequently, the generation of TDM domains by programmers presents many challenges.

Firstly, careful reading and comprehension of the relevant files such as protocol and aCRF for clinical trials are required, and effective information such as arms, elements, and inclusion/exclusion criteria must be collated.

Secondly, understanding the relevant rules for TDM as specified in the SDTM IG, variable settings, and required mandatory data is necessary.

Additionally, updates to TDM domains are required when there are changes to the protocol or inclusion/exclusion criteria, which can result in repetitive work.

For junior programmers or those generating TDM domains for the first time, relative experience is necessary, as well as a deep understanding of clinical trials and the SDTM IG, which results in a relatively steep learning curve. For senior programmers, a significant amount of manual work is required not only to develop TDM domains but to validate them as well. Moreover, if there are changes to the protocol or other experimental designs later, time must be spent updating and performing repeat work. Thus, TDM domains represent a relatively challenging aspect of SDTM data sets.

HOW: DIGITALIZATION AND AUTOMATION

GENERAL IDEA

Our design aims to minimize the challenges in the TDM generation process by providing tools to assist programmers in TDM development. These tools guide users in collecting and providing trial design information, centralizing management, and reducing repetitive work. The automation process is divided into two parts: Excel and Python (Figure 3). The Excel part involves collecting and centralizing trial design information and presenting the corresponding domain dataset. The Python part automates the generation of TA, TE, and TV, and centralizes the management and maintenance of TS and TI domains.

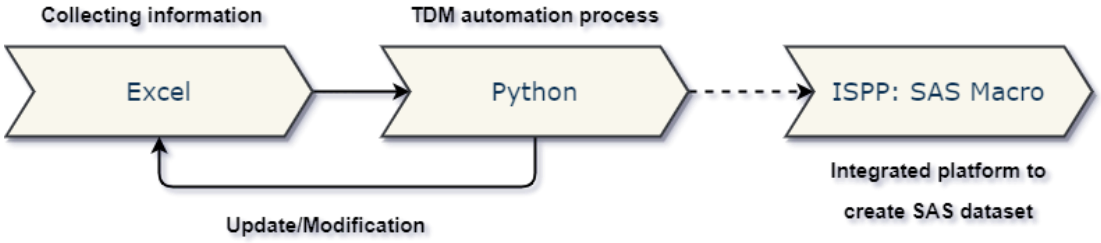

Figure 3 Automation Process

EXCEL PART

We have set up three main functions in Excel (Figure 4): tool-related sheets, trial design information collection sheets, and domain dataset presentation sheets.

The tool-related sheets include version control and instructions. In the version update sheet, users can view updated information, ensure they are using the latest version, and understand the specific content of the update. In the instruction sheet, users can learn about the required variables for TA, TE, TV, TI, and TS domains, the corresponding columns in domain sheets, and the corresponding rules and examples for trial design information filling.

Next is the Trial Design Matrix for information collection (Display 1), which is the core of our Excel-based tool. In this section, users are required to collect and organize relevant information related to the trial design based on the template provided. At the top, users need to provide the study ID, whether it is a randomized study, cycle duration, and ALS files. Below are the treatment-related details, including ARM/ARMCD, ELEMENT/ETCD, and so on. Based on the information provided by the user in this section, we will proceed with the development of TDM domains.

Finally, there are domain dataset sheets for each domain. TA, TE, and TV are automatically generated, while for TI and TS, we provide templates that users can fill in according to their specific trial design. Although these two domains cannot be fully automated at present, with all the information integrated and templates provided, users can develop and maintain more conveniently.

Figure 4 Excel Sheets Setting

TRIAL DESIGN MATRIX:								
Please fill in below trial design matrix and item #1 to 5 firstly:								
1. STUDYID:	XXXXX-666							
2. Randomized Study (Yes/ No):	No							
3. Study ALS xlsx File Path:	C:\Users\han.wang\Documents\ALS file							
4. Cycle Duration (Weeks):	3							
5. VISIT by ARM (Yes/ No):	No							
	ARMCD, ARM - Fill in Dark Blue cells							
	EPOCH - Fill in Light Green cells							
	ETCD - Fill in Yellow cells							
	VISIT to be excluded for each ARM							
ARMCD	ARM	PRE-SCREENING	SCREENING	TREATMENT	*TREATMENT 2*	*TREATMENT 3*	FOLLOW-UP	If VISIT differs between ARMs, please list VISIT to be excluded separated by ', ' per ARM (blank by default):
ArmACD	ArmA	PRESCRN	SCRN	ARMAETCD			FU	
ArmBCD	ArmB	PRESCRN	SCRN	ARMBETCD			FU	

Display 1 Trial Design Matrix

PYTHON PART

In the backend logic part, we mainly use the *openpyxl* package to read and write Excel information. In the processing of TA and TE domains, we use for loops to generate them row by row based on trial design matrices to obtain trial arm and element information. For the TV domain, we read visit-related information from the ALS file's Folder sheet and use regular expressions to identify visit texts. We then separately handle special visit pages such as response assessment and end of treatment according to different features and write them into the TV domain sheet layer by layer with loops to obtain the final dataset. Although we cannot fully automate the acquisition of TI and TS domains at present, some fixed information such as study ID and SDTM IG version will be automatically filled in. (Figure 5)

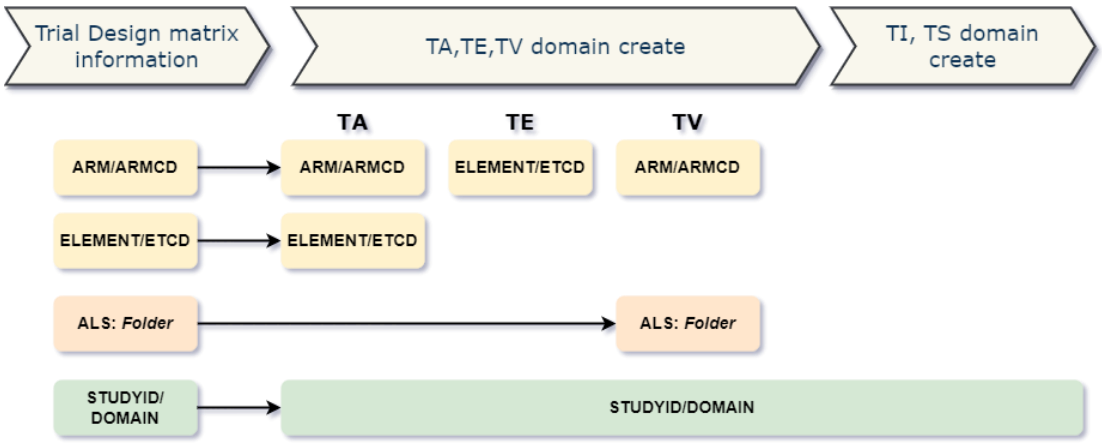

Figure 5 General Mapping Rules

SUMMARY

This automation process simplifies the TDM domains development process compared to traditional methods. It presents the necessary information to developers in a straightforward manner while achieving centralized management. Automation of the TA, TE, and TV domains can be achieved with a simple click of a button after filling out the necessary information (Figure 6), which improves initial

development efficiency and saves subsequent update time and workload. Additionally, the tool can assist programmers in some work for the TI and TS domains, providing a more convenient way for programmers to manage and maintain these two data sets. Since the tool is based on Excel, users can freely modify all information based on their study settings when dealing with complex clinical trial designs and generate datasets based on the modified content. This process also provides a gentler learning curve to guide programmers in understanding the relevant concepts of the TDM process.

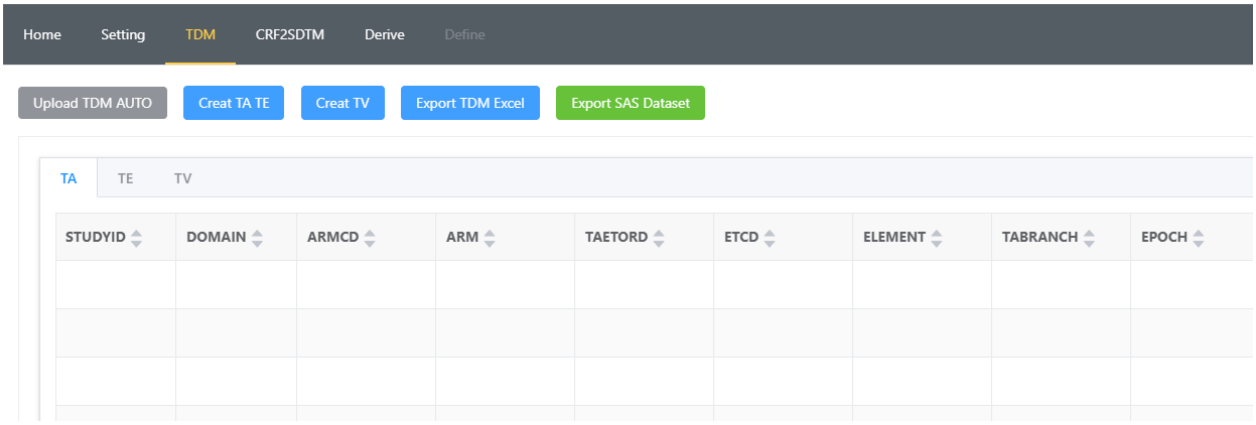

Figure 6 TDM Development Interface Example

LIMITATIONS

As a tool, it also has its limitations. Since we have only tested oncology-related studies, it may not be possible to automate the process for non-oncology or more complex and special trial designs, and manual work may still be required. Additionally, after developing TDM domains, we currently do not have an automated validation tool to better ensure our work quality, so manual verification is still required.

CONCLUSION

Although the tool can help programmers better understand the TDM development process, a thorough understanding still requires careful reading of the SDTM IG. And programmers still need to take responsibility for their work instead of relying entirely on the tool. By acknowledging its limitations and using it appropriately, programmers can leverage this tool to achieve more efficient and streamlined TDM domains development.

FUTURE EXPECTATION

TDM (Trial Design Model) domains are just parts of the SDTM mapping process. Currently, we are devoting ourselves to achieving full digitalization of the whole procedure, from CRF annotation to the generation of SDTM draft programs and datasets, batch runs, validation, and preparation of the submission package. In the future, this part will be integrated into our own ISPP (Integrated SAS Programming Platform) application (Figure 7), forming a fully digitalized platform that includes all the necessary work need to be done by programmers and a more streamlined and efficient workflow.

Besides, as digital technology and AI continue to evolve and become more prevalent in the pharma industry, our digitalization process is expected to become increasingly intelligent, significantly improving the accuracy and speed of programmers' daily work, reducing manual errors, and enhancing patients' data safety. We also aim to dedicate efforts towards building the ISPP server in the future and expanding its bidirectional communication with databases and other servers to further enhance the performance of ISPP (Figure 8). There is also a growing demand for real-time monitoring of clinical trials in some projects. So, our future work also hopes to encapsulate this function to enable early detection of potential issues and faster decision-making.

With the integration of TDM domains into Beigene's own ISPP platform, we can expect to see a fully digitalized platform that includes other work done by programmers, forming a more streamlined and efficient workflow.

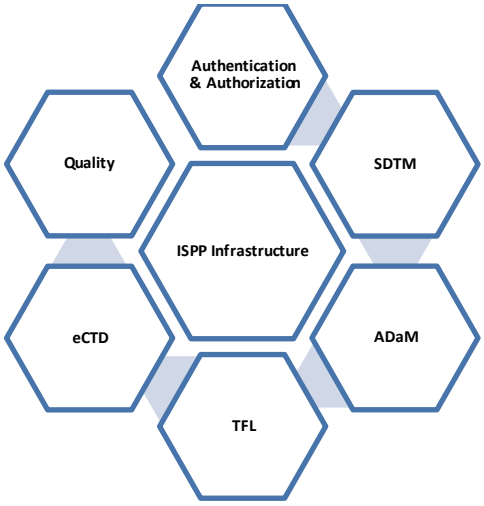

Figure 7 ISPP Infrastructure

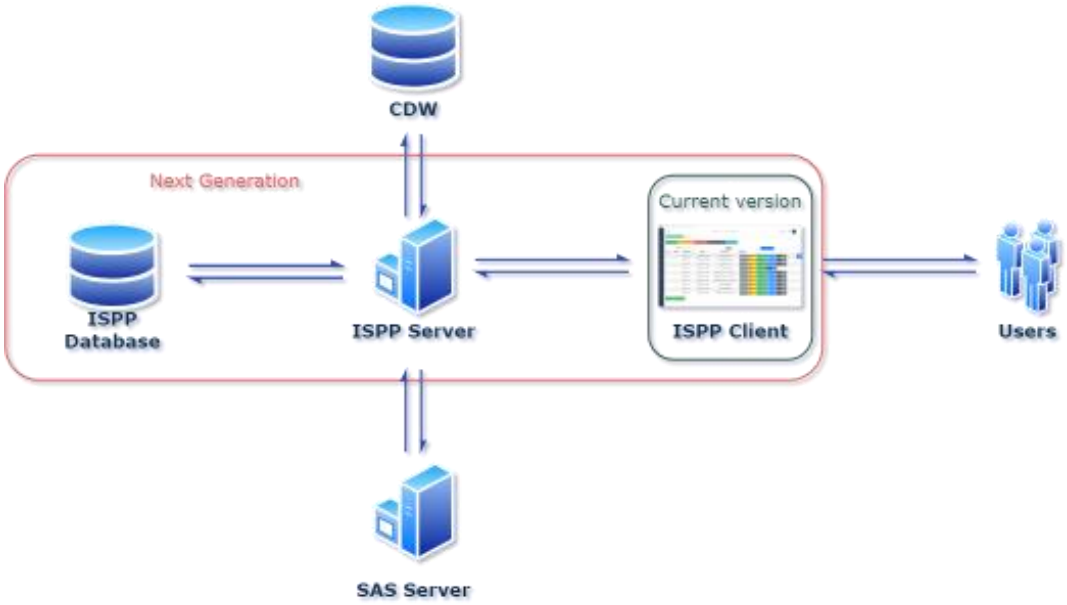

Figure 8 Future Generation

REFERENCES

Jerry Salyers, Richard Lewis, Kim Minkalis, Fred Wood. “Trials and Tribulations of Trial Design” *PharmaSUG 2014 Proceedings*

Study Data Tabulation Model. Clinical Data Interchange Standards Consortium (CDISC) Submission Data

Standards (SDS) Team, Version 1.4, December 2013.

Study Data Tabulation Model Implementation Guide: Human Clinical Trials-Version 3.3 (Final). https://www.cdisc.org/standards/foundational/sdtmig/sdtmig-v3-3/html#Trial+Design+Model+Datasets

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my boss Wei Wei, and my colleague Liming Xie for their invaluable support and guidance throughout this paper. Their mentorship, encouragement, and expertise have been instrumental in my personal and professional growth, and I am grateful for the opportunities they have provided me to contribute to the company's digitalization. I have learned so much from their leadership and teamwork, and I am honored to be a part of such a dynamic and innovative organization.

CONTACT INFORMATION

Han Wang
Beigene (Beijing) Co., Ltd.
+86 15201591465
han.wang@beigene.com