Paper AD13

Balancing Automation and Flexibility in Programming Processes

Punitha Philip, Cmed Clinical Services, Horsham, UK
Helen Nicholson, Cmed Clinical Services, Horsham, UK

ABSTRACT

Many technical solutions and systems these days are built around end-to-end standardization, generating efficiency through consistency from one study to the next. However, this is not always completely feasible for CROs with small to mid-sized clients. In this case there may be distinct requirements or interpretation of standards, not always fully utilizing CDASH. This can result in challenges in completely standardizing metadata and hence the production of automated systems. This paper describes some of the methods Cmed have used to build adaptable solutions, striking the balance between automation and flexibility, avoiding the restrictions of a rigid system. Examples include making effective use of SAS® to semi-automate program generation for SDTM and ADaM datasets.

INTRODUCTION

There is a certain amount of repetition in the process of mapping data to standards across different projects. However, each project is slightly different depending on a number of varying factors. The challenge lies in finding the delicate balance between trying to fit every trial into the same template thereby making the programming fully automatic, versus analyzing each trial to identify which aspects are different and require bespoke programming.

VARYING REQUIREMENTS

As target deadlines are getting shorter, we need to generate the programs faster, validate the outputs more quickly, and enhance the efficiency of the end-to-end process. There is an ongoing aim within the industry to improve the efficiency in programming processes, leading to quicker mapping of data to standards, thereby contributing to getting new drugs to market as quickly as possible without compromising the quality of deliverables produced.

An efficient and fast program creation process that is able to produce datasets of high quality would significantly reduce validation time by decreasing the number of feedback cycles and findings. The most obvious way of achieving this is to build a fully validated application that assists all parties in getting their work done better, faster and to a higher standard.

However, the challenge of streamlining the process for different sponsors plays a vital role in determining the level of automation possible in a standardized system (**Figure 1**).

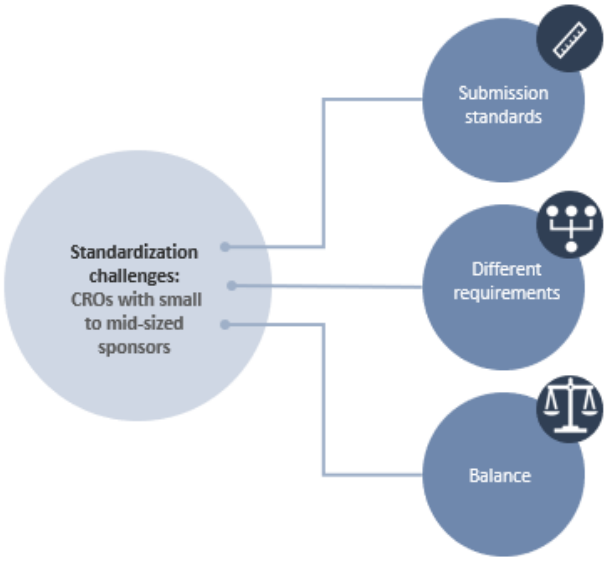

Figure 1

Submission standards need to be met regardless of the varied nature of trials and their requirements. There is flexibility in how submission standards are interpreted and followed by different sponsors, and they may approach the standards in different ways whilst still following the same core standards and their prerequisites.

As a CRO, there are not only the standards of our company to meet, but often each sponsor has their own standards, which may differ from one another yet still meet the submission standards. The sponsor may also have their own requirements for the electronic CRF; for example, the format the date is collected in could be different between sponsors.

Different study indications may lead to the collection of many varied data points due to specific protocol requirements (**Figure 2**). Also, the phase of the clinical trial will dictate what sort of information is being collected; first-in-man trials will probably have a lot more laboratory and vital signs data compared to a post-marketing or real-world studies.

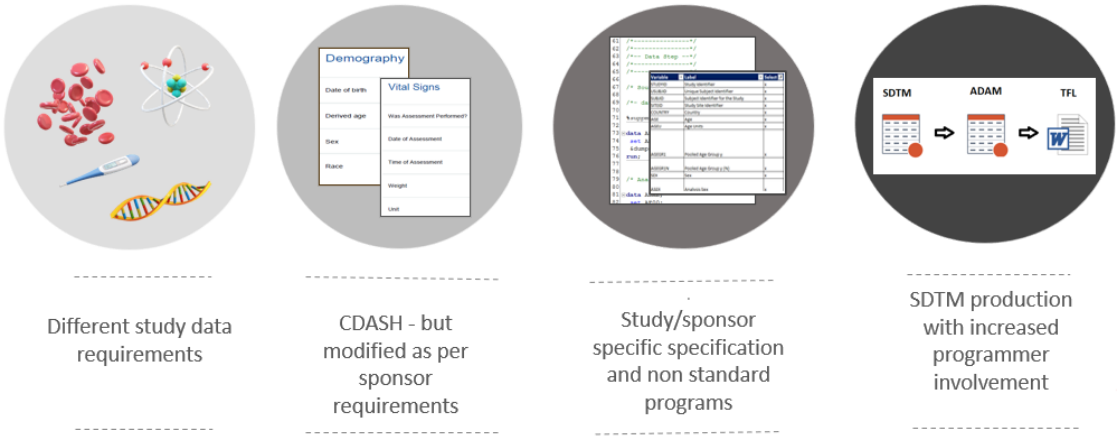

Figure 2

The ideal solution would be to standardize the CRF pages in accordance with CDASH. This would result in standard variable names, which would provide the ability to standardize the specification template, create a standard program suite and streamline and automate the dataset production process.

The statistician would be able to easily specify which variables they need to analyze when they review the CRF, as the pages are standard, thereby enabling the ADaM and TFL production to be easily streamlined.

However, it is rarely possible to achieve absolute consistency across all studies. A rigid system built entirely around the assumption that database design will remain exactly the same from study to study may not have the capacity to easily work around differences during the dataset programming. Also, for deliveries when the study is ongoing, the data may have open queries and not be fully clean. Such deliveries would require a degree of flexibility that a rigid system may not be able to provide.

All of this can make the programming process relatively difficult to standardize which increases programmer involvement in editing and finalizing the programs.

This impact on standardization carries through SDTM into the ADaM and TFL production, increasing the programmer involvement at each stage and impacting efficiency.

Therefore, we have developed a system that uses automation where practical, possible and sensible, balanced with flexibility that takes into account each study's individual requirements, is simple to use and reduces the overall time scale.

SOLUTION

Our solution combines the best of both worlds, flexible programming and automation. The initial program for each SDTM domain is automatically generated for the programmer, who can then edit for the unique requirements and finalize the code (**Figure 3**).

The concept is based on using dataset specifications and standard programs as building blocks which, through automation, create study specific code.

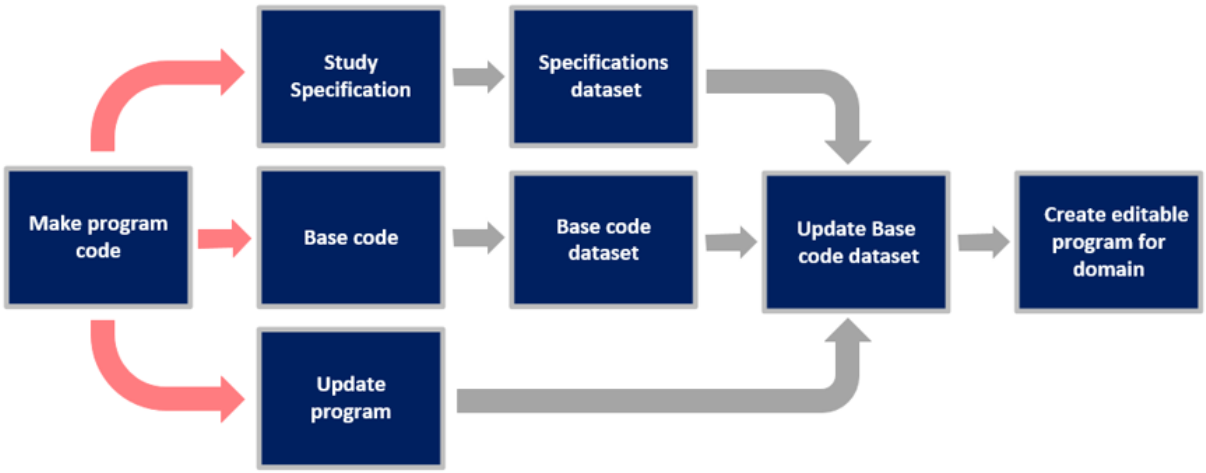

Figure 3

The programmer provides key information to a SAS® program, which in turn picks up the base code, updates it based on what is in the dataset specification and outputs a program ready for the programmer to perform any necessary editing.

SPECIFICATION

As the specification forms the basis for all downstream programming, it needs to be completed according to the guidelines. The better the quality of the specifications, the smoother all the downstream processes will be.

The specification needs to be generated with no ambiguity in any of the wording to ensure interpretation would be consistent.

All of the variables need to be present, each with its label, derivation and all the information that will be required for both the SDTM dataset programming and also the define.xml production.

Where code lists are used, they should be checked against the controlled terminology. The programmer would need to ensure that all the values required for the study are present in the dataset specifications and remove any codes that are not needed.

The specifications are used to produce the editable program via SAS® programs, therefore the quality of the specifications has a direct impact on the success of the process to follow.

PROGRAM PRODUCTION

Regardless of the domain being worked on there will be some derivations that will be required for that dataset in all cases. These derivations have been added to a base or starter program; one for each domain.

By having a base or starter program for each domain, the code for the derivations and variables that are always required are added accompanied by markers to indicate optional derivations or variables that may be required depending on the study specifications.

Before the base program is read into SAS® there are some programmed checks in the calling program. This ensures that any pre-existing programs are not overwritten, that the information required is available, and that the specifications for the domain to be produced are available.

If there is no base code for the specific domain, the calling program uses a generic base code starting point which provides the structure of the program only.

As well as there being a base program for each domain there is also an update program which is run at the time of domain program creation. The update program uses the information from the specifications to work out which variables are present in the raw data and updates the base program code for the domain.

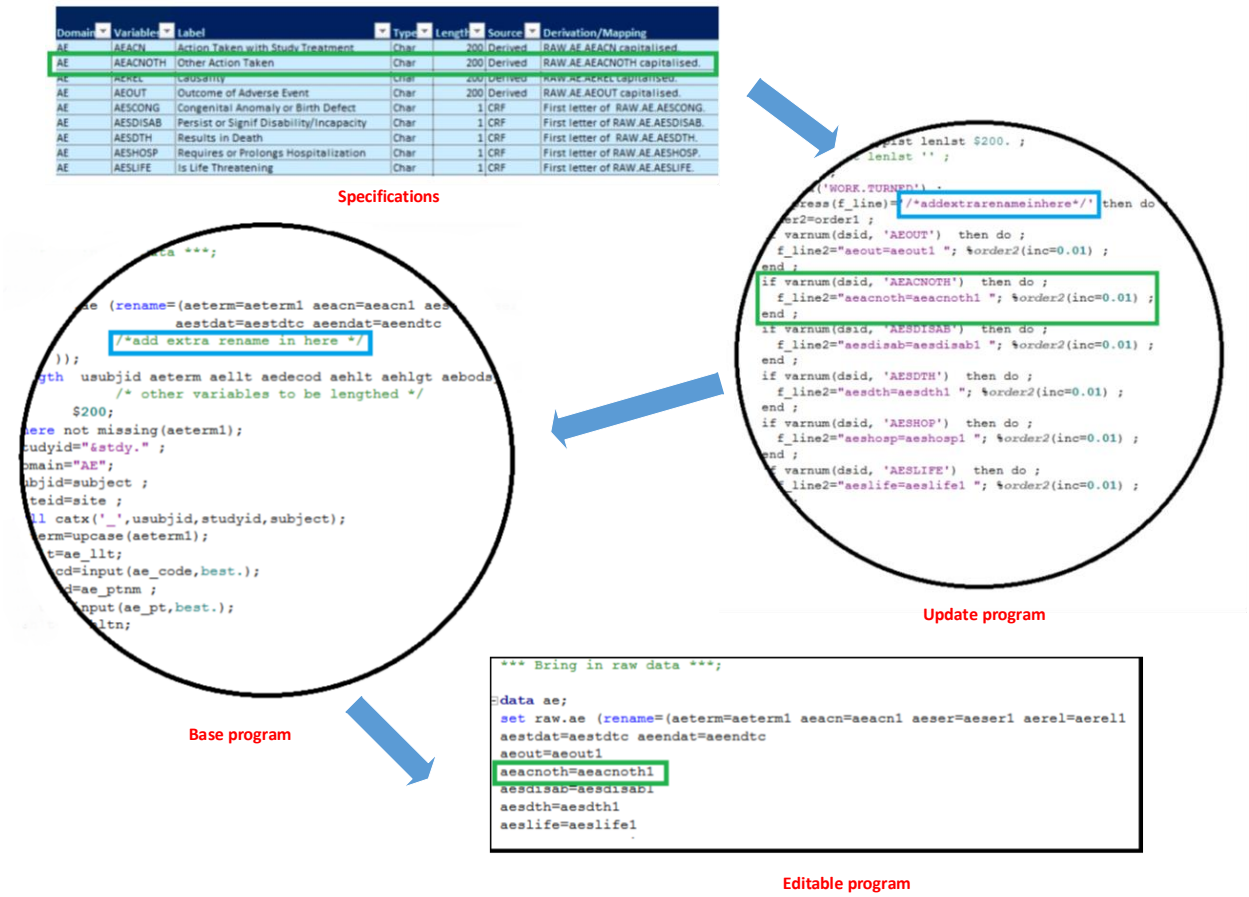

Figure 4

The program checks for markers in the code and using these markers adapts the base program as per the information the specifications identify. For example, in an SDTM Adverse Event domain it may add in extra rename statements if certain variables are present in the specifications when the raw data is read in and add the actual derivation of the variable further down. This is illustrated in **Figure 4** above; the blue boxes indicate the marker in the program; the green boxes indicate the information from the specifications which is added to the base program in the form of additional code by the update program to create the editable program.

This method is more beneficial than having standard template code for each domain for programming, as it adapts the code automatically as per the study requirements using the specifications and creates study specific code for each domain with minimal programmer input.

PROGRAMMER UPDATES AND FINALIZATION

Once generated, the editable program is saved to the required location. The file is a text file, with the correct extension for a SAS® program, enabling the programmer to edit the program in SAS® to make any study-specific changes that are not covered by the update program.

This editing process will involve addition of one-off study-specific definitions and inclusion of non-standard input datasets. Once edited and saved the program format is no different to any other program.

At the start of each program a macro is called which picks up the list of variables and labels from the specifications, so if the labels change then the programmer does not have to update or recreate the program.

FUTURE DEVELOPMENT AND CONCLUSION

This approach to the SDTM dataset programming process can be used on several different deliverables produced by programming.

The same concept can be adapted for any dataset programming process and is not limited to SDTM/ADaMs. This process can also be further extended to tables, listing and figures using the Statistical Analysis Plan shells as the specifications.

Over time as more and more domains are programmed, the code will be developed further, allowing more domains to

be added to the specification template and base/update programs, and for more details and derivations to be added to the update programs. As the code is refined through an ongoing feedback process, it will reduce the programmer involvement in editing code for standard domains and derivations.

In conclusion, there are significant advantages to using this process (**Figure 5**).

The metadata approach often seen in dataset production solutions is replaced in this model with a metadata-built specification template paired with meta-programming.

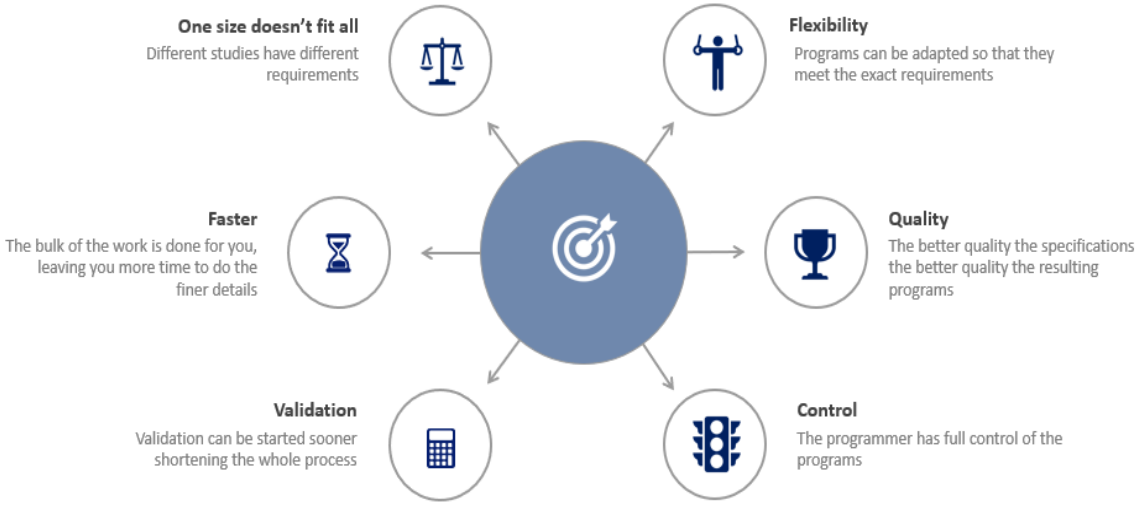

Figure 5

The system allows each program to be unique for the domains and variables present in a trial, without starting from scratch each time. Good quality specifications allow production of well-tailored programs best suited for each set of variables in a domain whilst allowing the programmer to still have full control of the study programs. With the standard derivations already implemented for a trial, it frees up programming resource for more complex derivations and study specific requirements. The process quickens the production of final programs enabling validation to be started sooner and given high quality programs results in a reduction in the number of feedback cycles during validation the deliverables are finalized earlier overall.

Faster SDTM program production in turn leads to quicker production of SDTM datasets, and this enables ADaM and TFLs to be produced quicker for a trial, ultimately helping reduce the time taken to get a drug to market.

CONTACT INFORMATION

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

Author Name: **Punitha Philip**
Company: **Cmed Clinical Services**
Address: Cmed (Clinical Research Services) Limited
Ashurst, Broadlands Business Campus
Langhurstwood Road
Horsham, West Sussex, RH12 4QP, United Kingdom
Email: **pphilip@cmedresearch.com**

Author Name: **Helen Nicholson**
Company: **Cmed Clinical Services**
Address: Cmed (Clinical Research Services) Limited
Ashurst, Broadlands Business Campus
Langhurstwood Road
Horsham, West Sussex, RH12 4QP, United Kingdom
Email: **hnicholson@cmedresearch.com**