PharmaSUG China 2023 – Paper AP-103

Meta-programming in Table Automation

Chunpeng Zhao, BeiGene

ABSTRACT

Meta-programming takes information in table programming code as data; it digitalizes and modularizes the information into Table Block Digital Capsules (TBDCs). TBDCs are machine-readable metadata, which make programming code information recyclable and reproducible per table block. The direct outputs of table metaprogramming are neither data sets nor tables, but ADaM annotated Table Shells, Metadata TBDCs and executable table creation computer programs. Meta-programming makes table creation coding-free and table creation computer programs more reader friendly. For example, table creation SAS programs can be macro-free.

INTRODUCTION

Real-time SDTM is realized in BeiGene through metaprogramming (Chunpeng Zhao etc., 2022). Using metaprogramming, SDTM transformation cost is reduced by 90% compared with outsourcing model. The paper is exploring how to make table creation coding-free using metaprogramming. The rationality of table automation is that table computer programs are not data driven. Everything in table creation computer programs can be pre-defined. To make table automation happen, we need to make ADaM annotated Table shells and Table creation computer programs ready before ADaM data are ready. Then tables can be created immediately once ADaM data are available. The paper will introduce how to achieve the goal making everything ready before ADaM data are available.

DIGITALIZATION & MODULARIZATION

Table shells determine what tables look like. Annotations per table block determine how tables are created. Annotations always include a table block corresponding ADaM dataset names, ADaM variable names, and ADaM variable values. So, annotations on table blocks are ADaM annotations. Table shells with ADaM annotations are ADaM annotated Table Shells. ADaM annotated Table Shells include all information that table creation computer programs need.

If ADaM annotated Table Shells are designed in a structured interface, e.g., Excel, ADaM annotated Table Shells can be automatically decomposed and modularized into Table-block-digital-capsules (TBDCs) by machine. TBDCs are digitalized ADaM annotated Table Shells. TBDCs can be reverted to ADaM annotated Table Shells. TBDCs and ADaM annotated Table Shells are equivalent on information.

A TBDC include all information that the table block corresponding programming code need. TBDCs are machine-readable and can be transformed to programming code automatically. TBDCs and table creation computer programs are equivalent on information. TBDC is a data model. It defines a standard structure for digitalization and modularization of ADaM annotated Table Shells. TBDC variables can be categorized into 4 classes:

- Identifier variables, which identify the study, the report task (e.g., CSR), the table, the table shell, and the table block. They also include corresponding reference identifiers. TBDCs can be borrowed from previous studies. These reference identifiers can be used for traceability and reference prioritizing. For example, TBDCs from standard table shells can be prioritized for reference.
- Data-in qualifiers, which define which ADaM datasets, variables and values are used in the table blocks.
- Calculation qualifiers, which specify which statistics (e.g., mean, standard deviation, percentage, etc.) are calculated, and how to calculate denominator for percentage.
- Display qualifiers, which specify what calculation results in table blocks look like (e.g., block order, block label, block values, table column number and pattern of each calculation result, etc.)

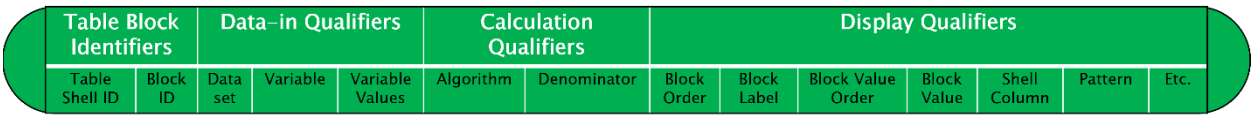

Figure 1. Table-Block-Digital-Capsule (TBDC) Model

TBDCs are universal data model for all types of Tables. The data model has two features.

1. ADaM Annotated Table Shells and their corresponding TBDCs can be auto transformed to each other.
   a) TBDCs can be auto derived from ADaM Annotated Table Shells.
   b) ADaM Annotated Table Shells can be auto restored from their TBDCs.
2. Executable table creation computer programs can be auto generated from TBDCs.

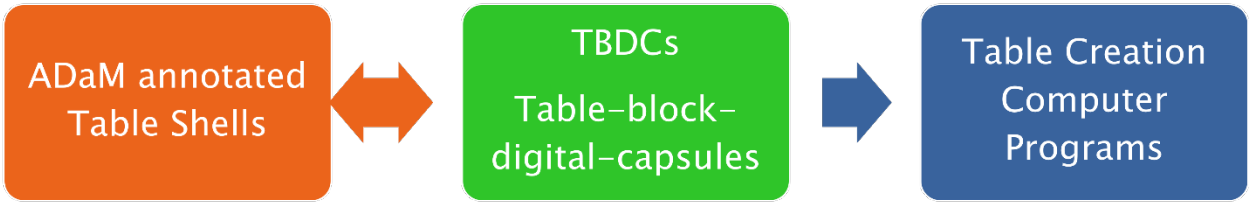

Figure 2. Flowchart of Table Metaprogramming

RECYCLING & INTEGRATION

All TBDCs are uploaded to a shared cloud folder automatically when we build up TBDCs in each study. A subsequent study's Table Shells are mainly integrated by existing Table Blocks which are defined in other various studies, with only limited additional, newly designed Table Blocks added. For a subsequent study, only the newly designed TBDCs need to be created.

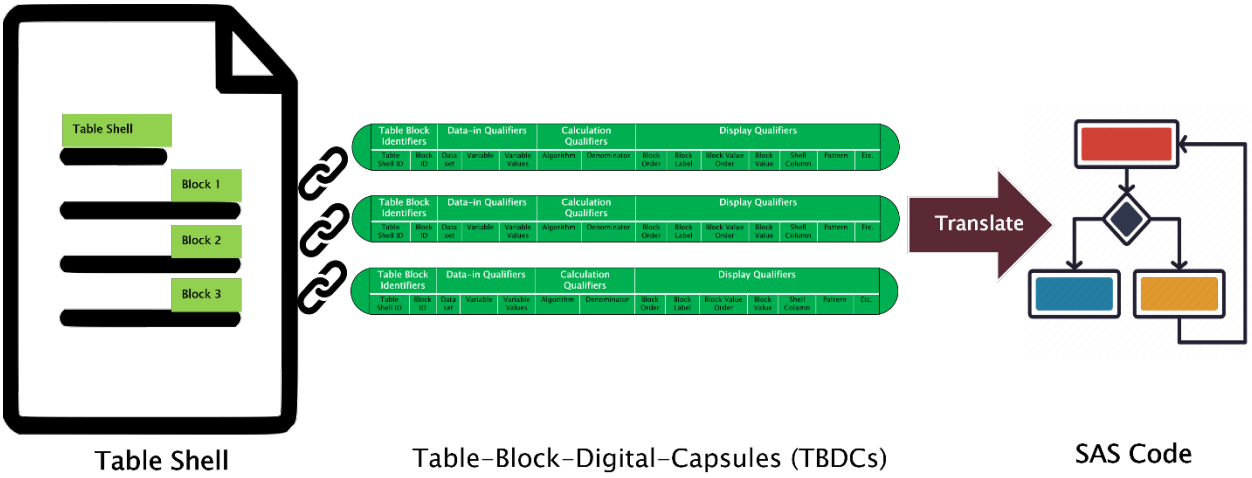

Figure 3. Table-shell-driven Coding-free Table Creation Process

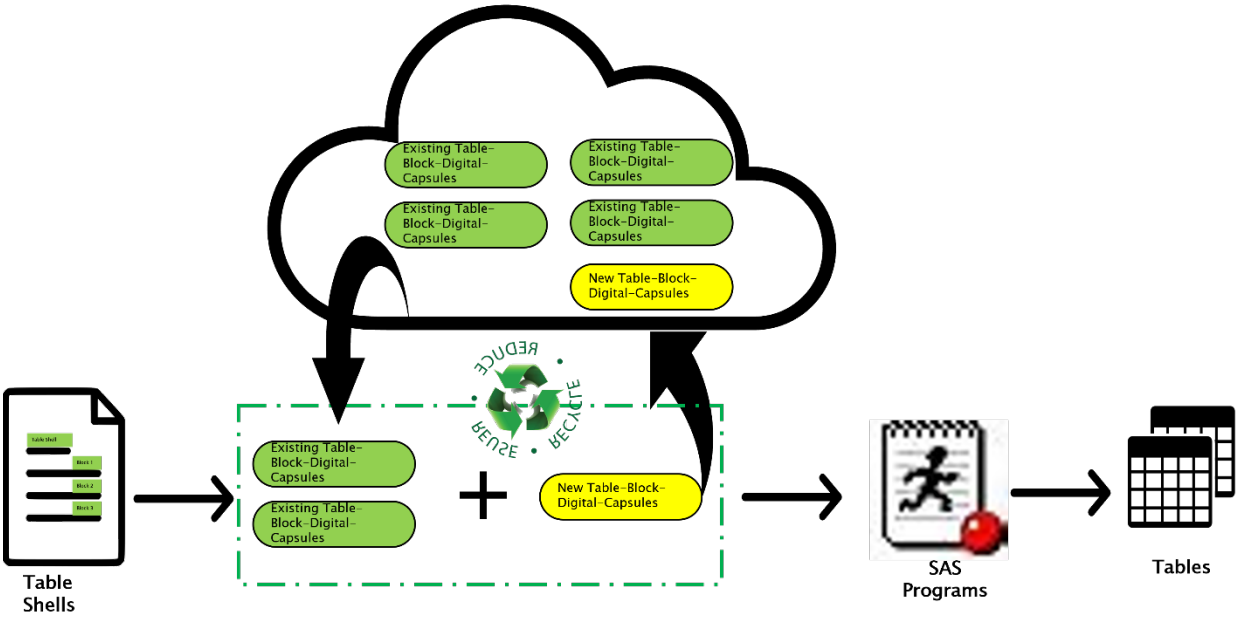

Figure 4. Flowchart of Table Creation

CENTRALIZATION & FRONT-LOADING

All studies' newly designed TBDCs are created centrally by the Table Shell Design and Annotation Centralization Group at the Table Shell design stage. ADaM annotated Table Shells and Table creation computer programs can be ready once Table Shells are done. Once ADaM data are available, Tables can be created immediately.

META-PROGRAMMING

Meta-programming takes programming code as data; Table meta-programming takes ADaM annotated Table Shells equivalent to programming code; it digitalizes and modularizes ADaM annotated Table Shells into machine-readable Metadata TBDCs, which are recyclable and reproducible per table block. The direct outputs of meta-programming are not tables, but ADaM annotated Table Shells, Metadata TBDCs and table creation computer programs. TBDCs from various studies are automatically recycled and recombined according to a study's Table Shells and converted to executable table creation computer programs. Meta-programming makes table creation coding-free.

There are three parts on table meta-programming.

- Program to create Study ADaM annotated Table Shells based on TBDC library.
- Program to create TBDCs based on Study ADaM annotated Table Shells, then auto upload to TBDC library.
- Program to create executable table creation computer programs based on TBDCs.

The key to meta-programming is Metadata TBDCs. TBDC is a universal data model, which is applicable to all types of Table Shells. TBDCs include all information how to create Tables from ADaM datasets. Machine-readable TBDCs have two meanings. First, TBDCs can be recycled and reused automatically by machine. Second, table creation executable computer programs can be generated based on TBDCs through programming instead of manually, because all information which executable Table creation computer programs need, can be machine-readable in TBDCs.

Meta-programming creates a common tool applicable to all studies. From Metadata TBDCs to executable table computer programs, it can be taken as a compilation process. In table creation, only one compiler is needed, if metadata model TBDC is robust enough that can be applicable for all types of tables. With the complier, table creation is coding-free. It means we can create executable computer programs on a User Interface ADaM Annotated Table Shell Designer, instead of typing computer code directly by hands. Since table computer code is generated by machine automatically, we can make computer code more reader friendly without consideration of typing efficiency. For example, table creation SAS programs can be macro-free. There are three levels of programming philosophy. Creating programs by programming, i.e., writing programs to handle text streams, is the third level, because that is a universal interface (Erik van Baaren, The Unix philosophy).

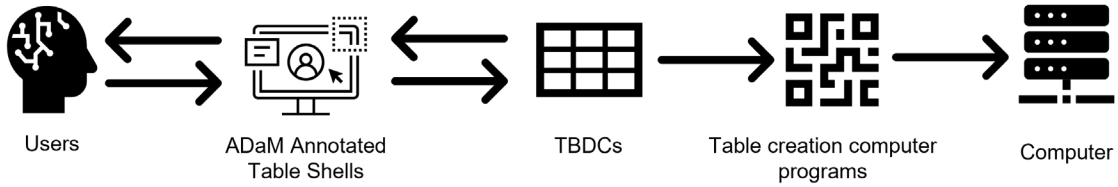

Figure 5. Two interfaces of TBDCs: User Interface (ADaM Annotated Table Shells) and Computer Interface (Table creation computer programs)

For example, in Table 1, a typical demographic table shell is created in Column ShellCol1, ShellCol2, ShellCol3 and ShellCol4. Identifiers are added in Column TFLID, ShellID and BlockID per Table Block. And ADaM annotations are added in Column “ADaM Annotation” per table block. Table 1 can be taken as a User Interface. Table Shell ADaM Annotations can be done during Table Shells design. The user interface can be called ADaM Annotated Table Shell Designer. With ADaM Annotated Table Shell Designer, all information in ADaM Annotated Table Shell can be read directly by machine automatically. This information is decomposed and assembled to TBDCs, e.g., Table 2. An executable table creation computer program can be auto generated based on Table 2 by a compilation process. Table TBDC compilation process is very similar with CRF-field-digital-capsules (CRF-FDCs) (Chunpeng Zhao etc., 2022). Finally, the demographic table can be created by the executable table creation computer program.

META-PROGRAMMING EXAMPLE

Table 1 Sample of ADaM Annotated Table Shell

TFLID	ShellID	BlockID	ShellCol1	ShellCol2	ShellCol3	ShellCol4	ADaM Annotation
Table 14.1.2.1	T-DM	Title	Demographics and Baseline Characteristics				
Table 14.1.2.1	T-DM	Population	Enrolled Analysis Set				ADSL.ENRLFL="Y"
Table 14.1.2.1	T-DM	ColumnHeader					
Table 14.1.2.1	T-DM	ColumnHeader		Treatment A	Treatment B	Total	TRT01P="Treatment A"|"Treatment B"|("Treatment A", "Treatment B"){Width=40 20 20 20}
Table 14.1.2.1	T-DM	ColumnHeader		(N = XX)	(N = XX)	(N = XX)	
Table 14.1.2.1	T-DM	ColumnHeader					
Table 14.1.2.1	T-DM	Sex					
Table 14.1.2.1	T-DM	Sex	Sex, n (%)				SEX
Table 14.1.2.1	T-DM	Sex	Male	xx (xx.x)	xx (xx.x)	xx (xx.x)	M
Table 14.1.2.1	T-DM	Sex	Female	xx (xx.x)	xx (xx.x)	xx (xx.x)	F
Table 14.1.2.1	T-DM	AGE					
Table 14.1.2.1	T-DM	AGE	Age (years)				ADSL.AGE
Table 14.1.2.1	T-DM	AGE	n	xx	xx	xx	
Table 14.1.2.1	T-DM	AGE	Mean (SD)	xx.xx (xx.xxx)	xx.xx (xx.xxx)	xx.xx (xx.xxx)	
Table 14.1.2.1	T-DM	AGE	Median	xx.xx	xx.xx	xx.xx	
Table 14.1.2.1	T-DM	AGE	Q1, Q3	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x	
Table 14.1.2.1	T-DM	AGE	Min, Max	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x	
Table 14.1.2.1	T-DM	AGEGR1					
Table 14.1.2.1	T-DM	AGEGR1	Age Group, n (%)				AGEGR1
Table 14.1.2.1	T-DM	AGEGR1	< 65 years	xx (xx.x)	xx (xx.x)	xx (xx.x)	
Table 14.1.2.1	T-DM	AGEGR1	>= 65 years	xx (xx.x)	xx (xx.x)	xx (xx.x)	
Table 14.1.2.1	T-DM	Footnotes					
Table 14.1.2.1	T-DM	ProgrmmingNotes					

Table 2 Sample of ADaM Annotated Table Shell Metadata

Block Order	BlockID	Denominator Block	Block Label	Stat	BlockValueOrder	BlockValue	Block Value Width	Shell Col	BlockCDISC Dataset	BlockCDISC Variable	BlockCDISCVarVal	Pattern
1	Title				1	Demographics and Baseline Characteristics						
2	Population				1	Enrolled Analysis Set			ADSL	ENRLFL	Y	
3	ColumnHeader				1		40	Shell Col1	ADSL	TRT01P		
3	ColumnHeader				2	Treatment A~(N = XX)	20	Shell Col2	ADSL	TRT01P	Treatment A	
3	ColumnHeader				3	Treatment B~(N = XX)	20	Shell Col3	ADSL	TRT01P	Treatment B	
3	ColumnHeader				4	Total~(N = XX)	20	Shell Col4	ADSL	TRT01P	Treatment A	
3	ColumnHeader				4	Total~(N = XX)	20	Shell Col4	ADSL	TRT01P	Treatment B	
4	Sex	ColumnHeader	Sex, n (%)	Freq	1	Male		Shell Col2	ADSL	SEX	M	xx (xx.x)
4	Sex	ColumnHeader	Sex, n (%)	Freq	1	Male		Shell Col3	ADSL	SEX	M	xx (xx.x)
4	Sex	ColumnHeader	Sex, n (%)	Freq	1	Male		Shell Col4	ADSL	SEX	M	xx (xx.x)
4	Sex	ColumnHeader	Sex, n (%)	Freq	2	Female		Shell Col2	ADSL	SEX	F	xx (xx.x)
4	Sex	ColumnHeader	Sex, n (%)	Freq	2	Female		Shell Col3	ADSL	SEX	F	xx (xx.x)
4	Sex	ColumnHeader	Sex, n (%)	Freq	2	Female		Shell Col4	ADSL	SEX	F	xx (xx.x)
6	AGE		Age (years)	Sum	1	n		Shell Col2	ADSL	AGE		xx
6	AGE		Age (years)	Sum	1	n		Shell Col3	ADSL	AGE		xx
6	AGE		Age (years)	Sum	1	n		Shell Col4	ADSL	AGE		xx
6	AGE		Age (years)	Sum	2	Mean (SD)		Shell Col2	ADSL	AGE		xx.xx (xx.xxx)
6	AGE		Age (years)	Sum	2	Mean (SD)		Shell Col3	ADSL	AGE		xx.xx (xx.xxx)
6	AGE		Age (years)	Sum	2	Mean (SD)		Shell Col4	ADSL	AGE		xx.xx (xx.xxx)

Block Order	BlockID	Denominator Block	Block Label	Stat	BlockValueOrder	BlockValue	Block Value Width	Shell Col	BlockCDISC Dataset	BlockCDISCVariable	BlockCDISCVarVal	Pattern
6	AGE		Age (years)	Sum	3	Median		Shell Col2	ADSL	AGE		xx.xx
6	AGE		Age (years)	Sum	3	Median		Shell Col3	ADSL	AGE		xx.xx
6	AGE		Age (years)	Sum	3	Median		Shell Col4	ADSL	AGE		xx.xx
6	AGE		Age (years)	Sum	4	Q1, Q3		Shell Col2	ADSL	AGE		xx.x, xx.x
6	AGE		Age (years)	Sum	4	Q1, Q3		Shell Col3	ADSL	AGE		xx.x, xx.x
6	AGE		Age (years)	Sum	4	Q1, Q3		Shell Col4	ADSL	AGE		xx.x, xx.x
6	AGE		Age (years)	Sum	5	Min, Max		Shell Col2	ADSL	AGE		xx.x, xx.x
6	AGE		Age (years)	Sum	5	Min, Max		Shell Col3	ADSL	AGE		xx.x, xx.x
6	AGE		Age (years)	Sum	5	Min, Max		Shell Col4	ADSL	AGE		xx.x, xx.x
7	AGEGR1	ColumnHeader	Age Group, n (%)	Freq	1	< 65 years		Shell Col2	ADSL	AGEGR1		xx (xx.x)
7	AGEGR1	ColumnHeader	Age Group, n (%)	Freq	1	< 65 years		Shell Col3	ADSL	AGEGR1		xx (xx.x)
7	AGEGR1	ColumnHeader	Age Group, n (%)	Freq	1	< 65 years		Shell Col4	ADSL	AGEGR1		xx (xx.x)
7	AGEGR1	ColumnHeader	Age Group, n (%)	Freq	2	>= 65 years		Shell Col2	ADSL	AGEGR1		xx (xx.x)
7	AGEGR1	ColumnHeader	Age Group, n (%)	Freq	2	>= 65 years		Shell Col3	ADSL	AGEGR1		xx (xx.x)
7	AGEGR1	ColumnHeader	Age Group, n (%)	Freq	2	>= 65 years		Shell Col4	ADSL	AGEGR1		xx (xx.x)
11	Footnotes											
12	ProgrmmingNotes											

CONCLUSION

Table creation program is not data driven. With Digitalization & Modularization, Recycling & Integration, Centralization & Front-Loading and Meta-programming, we make ADaM annotated Table Shells and executable table creation computer programs ready before First Patient First Visit. Once ADaM data are available, tables can be created immediately. Executable table creation computer programs can be generated with coding-free.

REFERENCES

Chunpeng Zhao etc., 2022. “Real-time SDTM” https://www.lexjansen.com/pharmasug-cn/2022/AT/Pharmasug-China-2022-AT105.pdf

Erik van Baaren, The Unix philosophy. “The Unix philosophy” https://python.land/the-unix-shell

CONTACT INFORMATION

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

Chunpeng Zhao
BeiGene
Chunpeng.zhao@beigene.com