

Strategic Thinking for Program Automation

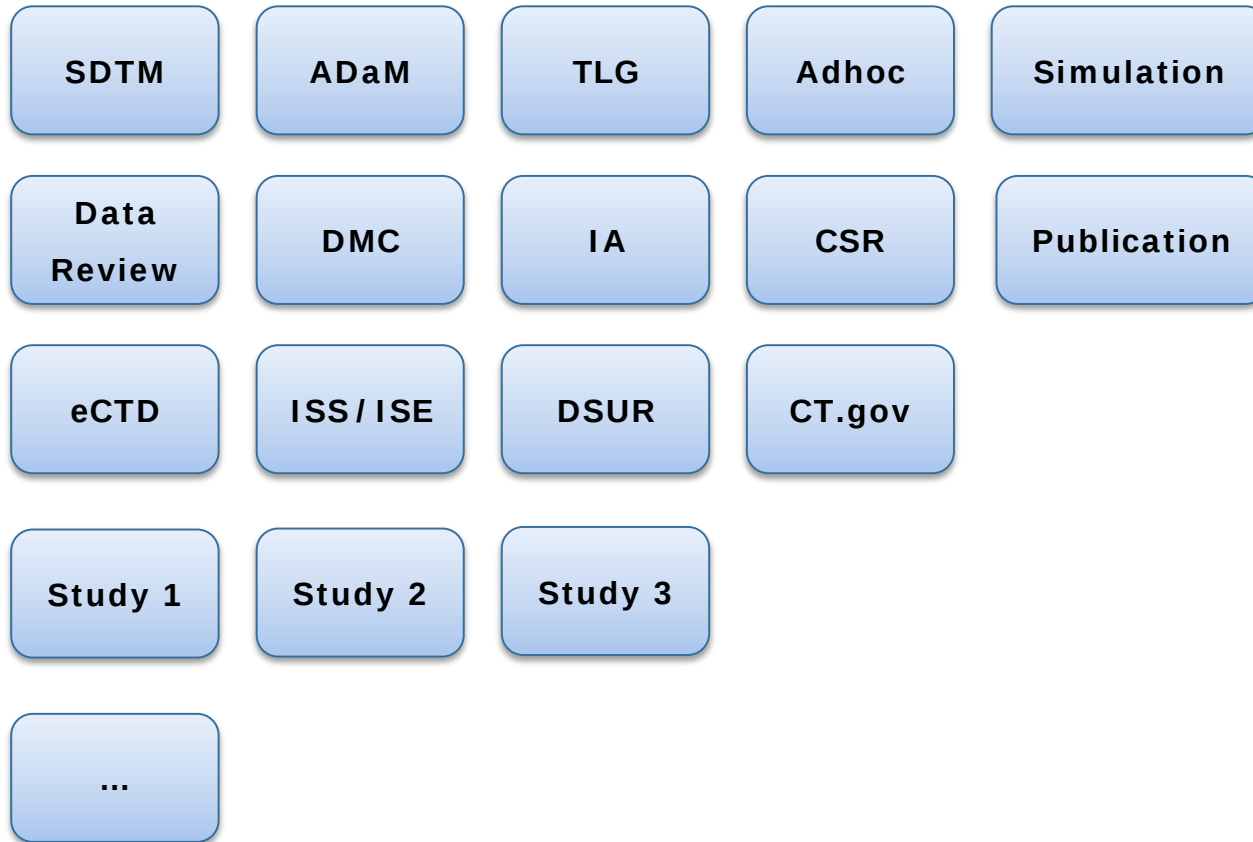
You (Leo) Li

17 May 2019, Beijing



始于信 达于行

Programming in Our Daily Work



Good Programming Practice

Good Programming

- Readable
- Concise
- Simple
- Clear
- Compatible
- Robust
- Efficient
- Automatic

VS

Bad Programming

- Long and Teddy
- Unclear Logic
- Inefficient
- Redundant by-product
- Residual macro variables
- Options change

Important Things



Programming view

Programming thinking

Simple Concept

$$1^3 + 2^3 + 3^3 + 4^3 + 5^3 = ?$$


$$1^3 + 2^3 = ?$$

$$1^3 + 2^3 + 3^3 + 4^3 + 5^3 = ?$$

$$1^3 + 2^3 + 3^3 + 4^3 + 5^3 + \dots + n^3 = ? \Rightarrow$$

Method 1

?

Method 2

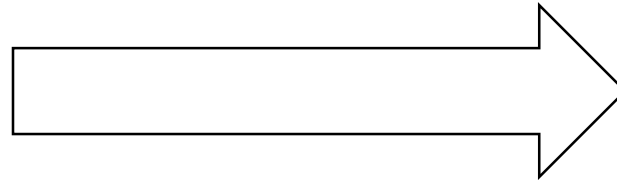
Thinking

View

Case 1

Create the --DTC variables comply with ISO8601 format.

BRTHDAT_RAW
16 JUL 1951
01 AUG 1957
16 AUG 1952
10 JAN 1952
21 MAY 1958
13 JAN 1956
25 SEP 1964
22 DEC 1966
11 SEP 1961
28 OCT 1956
16 AUG 1950
25 DEC 1952



BRTHTDC
1951-07-16
1957-08-01
1952-08-16
1952-01-10
1958-05-21
1956-01-13
1964-09-25
1966-12-22
1961-09-11
1956-10-28
1950-08-16
1952-12-25

Case 1

DM

```
data DM;  
set DM;  
  
dd=input(scan(var, 1, ""), best.);  
mm=input(scan(var, 2, ""), $mmf.);  
yy=input(scan(var, 3, ""), best.);  
  
varDTC = put(mdy(mm, dd, yy), yymmdd10.);  
  
run;
```

Partial Date?

01 JAN 2019
JAN 2019
2019

varDTC=catx("-", yy, mm, dd);

Invalid Date?

31 APR 2019
29 FEB 2019

Case 1

AE

AESTDTC	AEENDTC	AETCDTC
2019-01-18		
2019-01-08	2019-01-08	
2019-01-21	2019-02-13	
2019-04-01		
2019-02-16	2019-03-11	
2019-04-01		
2019-02-07		
2019-02-13	2019-02-18	
2019-02-21		
2019-03-14	2019-03-14	2019-03-14
2019-03-18		
2019-03-11		

```
data AE;  
set AE;  
  
***** AESTDTC *****;  
sssssss;  
  
***** AEENDTC *****;  
sssssss;  
  
***** AEWRDTC *****;  
sssssss;  
  
***** AEDTHDTC *****;  
sssssss;  
  
run;
```

```
data AE;  
set AE;  
  
%dtc(var=AESTDTC) ;  
%dtc(var=AEENDTC) ;  
%dtc(var=AEWRDTC) ;  
%dtc(var=AEDTHDTC) ;  
  
run;
```

Case 1

MH, CM, EX, VS...

Improve the %dtc macro

- **Partial date**
- **Invalid date**
- **Combine time**
- **Different input date format**
- **Add --DT, --DY output**
- **Debug function**
- **Temp var / dbs clean**

Case 1

```
data DM;  
set DM;  
  
%dtt(var=BTHDTC);  
  
run;
```

```
data AE;  
set AE;  
  
%dtt(var=AESTDTC);  
%dtt(var=AEENDTC);  
%dtt(var=AEWRDTC);  
%dtt(var=AEDTHDTC);  
  
run;
```

```
data CM;  
set CM;  
  
%dtt(var=CMSTDTC);  
%dtt(var=CMENDTC);  
  
run;
```

```
data MH;  
set MH;  
  
%dtt(var=MHSTDTC);  
%dtt(var=MHENDTC);  
  
run;
```

```
data VS;  
set VS;  
  
%dtt(var=VSDTC);  
  
run;
```

...

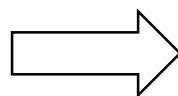
Shell we do the same thing in next study?

Case 1

Do we know which variables should be handled by %dte in one study?

PROC Contents

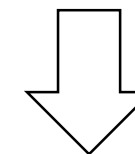
MEMNAME	NAME
AE_H	AEENDAT_RAW
AE_H	AESTDAT_RAW
AE_H	AETCDAT_RAW
BM_H	LBDAT_RAW
CM_H	CMENDAT_RAW
CM_H	CMRLPDAT_RAW
CM_H	CMSTDAT_RAW
DD_H	DDDAT_RAW
DM_H	BRTHDAT_RAW
DS_H	DSSTDAT_RAW
DS_H	SRVCDAT_RAW
EG_H	EGDAT_RAW
EX_H	EXSTDAT_RAW
FAMH_H	MHSTDAT_RAW
IE_H	IEPVDAT_RAW
LB_H	LBDAT_RAW
MH_H	MHENDAT_RAW
MH_H	MHSTDAT_RAW
PE_H	PEDAT_RAW
PK_H	PKSTDAT_RAW
PR_H	PRENDAT_RAW
PR_H	PRSTDAT_RAW
PS_H	PFMDAT_RAW
QS_H	QSDAT_RAW
SU_H	SUENDAT_RAW
TR_H	IOTADAT_RAW
TU_H	TMADAT_RAW
VS_H	VSDAT_RAW



MEMNAME	NAME	dte
AE_H	AEENDAT_RAW	%dte(var=AEENDTC);
AE_H	AESTDAT_RAW	%dte(var=AESTDTC);
AE_H	AETCDAT_RAW	%dte(var=AETCDTC);
BM_H	LBDAT_RAW	%dte(var=LBDTC);
CM_H	CMENDAT_RAW	%dte(var=CMENDTC);
CM_H	CMRLPDAT_RAW	%dte(var=CMRLPDTC);
CM_H	CMSTDAT_RAW	%dte(var=CMSTDTC);
DD_H	DDDAT_RAW	%dte(var=DDDTTC);
DM_H	BRTHDAT_RAW	%dte(var=BRTHDTC);
DS_H	DSSTDAT_RAW	%dte(var=DSSTDTC);
DS_H	SRVCDAT_RAW	%dte(var=SRVCDTC);
EG_H	EGDAT_RAW	%dte(var=EGDTC);
EX_H	EXSTDAT_RAW	%dte(var=EXSTDTC);
FAMH_H	MHSTDAT_RAW	%dte(var=MHSTDTC);
IE_H	IEPVDAT_RAW	%dte(var=IEPVDTC);
LB_H	LBDAT_RAW	%dte(var=LBDTC);
MH_H	MHENDAT_RAW	%dte(var=MHENDTC);
MH_H	MHSTDAT_RAW	%dte(var=MHSTDTC);
PE_H	PEDAT_RAW	%dte(var=PEDTC);
PK_H	PKSTDAT_RAW	%dte(var=PKSTDTC);
PR_H	PRENDAT_RAW	%dte(var=PRENDTC);
PR_H	PRSTDAT_RAW	%dte(var=PRSTDTC);
PS_H	PFMDAT_RAW	%dte(var=PFMDTC);
QS_H	QSDAT_RAW	%dte(var=QSDTC);
SU_H	SUENDAT_RAW	%dte(var=SUENDTC);
TR_H	IOTADAT_RAW	%dte(var=IOTADTC);
TU_H	TMADAT_RAW	%dte(var=TMADTC);
VS_H	VSDAT_RAW	%dte(var=VSDTC);

} &ae_dtc

```
%dte(var=AEENDTC);  
%dte(var=AESTDTC);  
%dte(var=AETCDTC);
```



```
data AE;  
set AE;  
&ae_dtc. ;  
run;
```

Case 1

Automation???

```
% do i=1 %to &domain_num.  
  call symput("domain", xxx);  
  sql; select vvv into :domain._dtc  
data &domain.  
  set &domain.  
  &&domain._dtc.;  
run;  
% end
```

←

←

Case 2

ADSL

Label	Age (Yrs)	Age Group	Sex	Race	BMI (kg/m^2)	...
Var	AGE	AGESRP	SEX	RACE	BMI	...

Demographic

		Group A	Group B
Age (Yrs)	N	xxx	xxx
	Mean (STD)	xxx.x (xxx.x)	xxx.x (xxx.x)
	Median	xxx.x	xxx.x
	Min - Max	xx-xx	xx-xx
Age Group, n(%)	18-<55 yrs	xxx (xx.x%)	xxx (xx.x%)
	55-<65 yrs	xxx (xx.x%)	xxx (xx.x%)
	>=65 yrs	xxx (xx.x%)	xxx (xx.x%)
Sex, n (%)	Male	xxx (xx.x%)	xxx (xx.x%)
	Female	xxx (xx.x%)	xxx (xx.x%)

Case 2

Demographic

		Group A	Group B
Age (Yrs)	N	xxx	xxx
	Mean (STD)	xxx.x (xxx.x)	xxx.x (xxx.x)
	Median	xxx.x	xxx.x
	Min - Max	xx-xx	xx-xx
Age Group, n(%)	18-<55 yrs	xxx (xx.x%)	xxx (xx.x%)
	55-<65 yrs	xxx (xx.x%)	xxx (xx.x%)
	>=65 yrs	xxx (xx.x%)	xxx (xx.x%)
Sex, n (%)	Male	xxx (xx.x%)	xxx (xx.x%)
	Female	xxx (xx.x%)	xxx (xx.x%)

% mean(var = age, ...)

% freq(var = agegrp, ...)

% freq(var = sex, ...)

Case 2

PROC Contents

Domain	Var	Label
ADSL	AGE	Age (Yrs)
ADSL	AGEGRP	Age Group
ADSL	SEX	Sex
ADSL	RACE	Race
ADSL	BMI	BMI (kg/m^2)
ADSL	...	...

Case 2

PROC Contents

Add Component

Domain	Var	Label	Digital	Method	CT
ADSL	AGE	Age (Yrs)	0	Mean	
ADSL	AGEGRP	Age Group		Freq	AGEGRP
ADSL	SEX	Sex		Freq	SEX
ADSL	RACE	Race		Freq	RACE
ADSL	BMI	BMI (kg/m^2)	1	Mean	
ADSL	...	...			

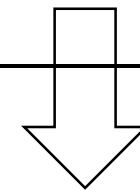
Case 2

PROC Contents

Add Component

Compile Macro

Domain	Var	Label	Digital	Method	CT	Analysis
ADSL	AGE	Age (Yrs)	0	Mean		%mean(var=,label=,digit=,...)
ADSL	AGEGRP	Age Group		Freq	AGEGRP	%freq(var=,label=,ct=,...)
ADSL	SEX	Sex		Freq	SEX	%freq(var=,label=,ct=,...)
ADSL	RACE	Race		Freq	RACE	%freq(var=,label=,ct=,...)
ADSL	BMI	BMI (kg/m^2)	1	Mean		%mean(var=,label=,digit=,...)
ADSL	...	...				



SQL

select analysis

into :demo separated by “;”

Case 2

```
1 data adsl;  
  set ADAM.adsl;  
  where .....;  
run;
```

```
&demo. ;
```

```
1 data report;  
  set &demo_out. ;  
run;
```

```
1 proc report data=..... ;  
run;
```

```
%mean(var=,label=,digit=,...)
```

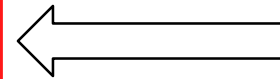
```
%freq(var=,label=,ct=,...)
```

```
%freq(var=,label=,ct=,...)
```

```
%freq(var=,label=,ct=,...)
```

```
%mean(var=,label=,digit=,...)
```

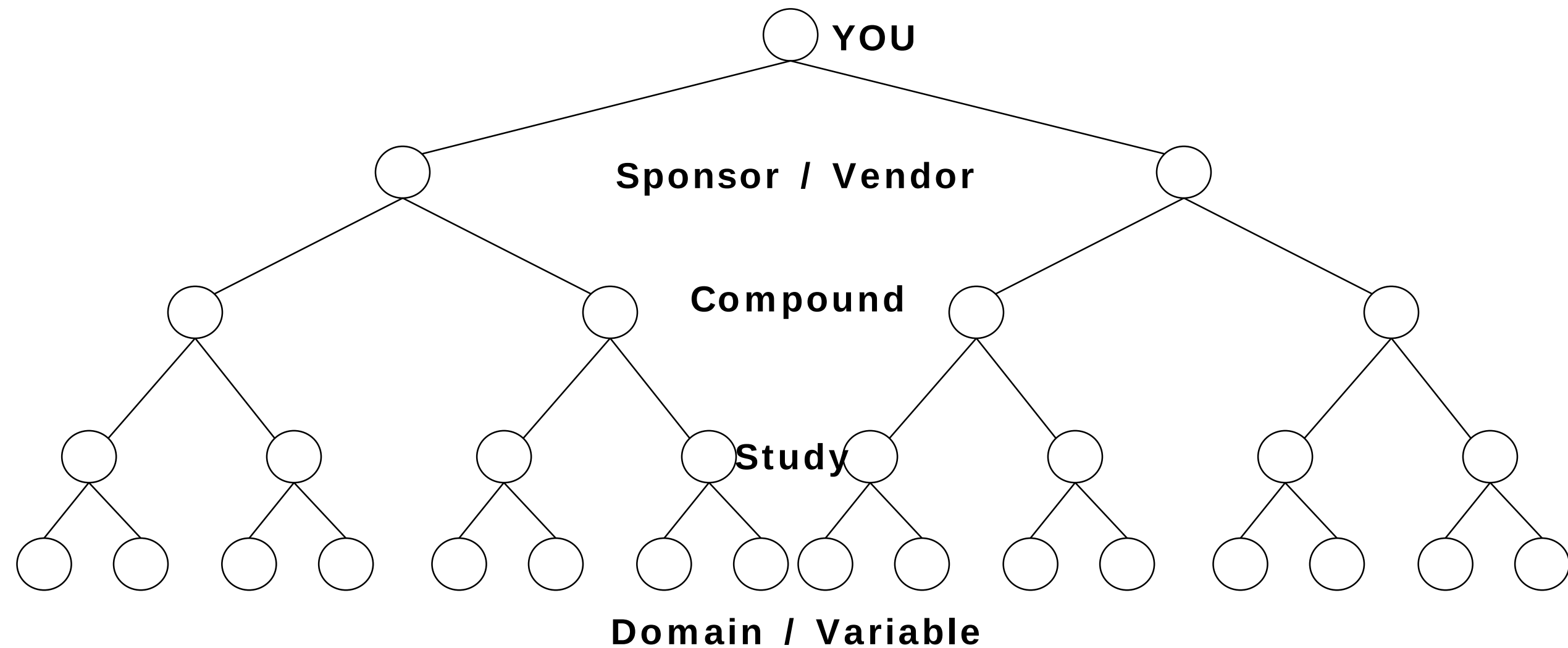
```
...
```



Summary

- **PROC contents to get the variables need to be handled**
- **Compile the macro statement**
- **Count the domain number**
- **With a DOTO loop**
 - **Retrieve the domain**
 - **Execute the macro**
- **Automatically complete all dataset / variable one-by-one**

Summary



Thank You

You Li

you.li@innoventbio.com

Innovent
信达生物制药

始于信 达于行