Spec Helper：AI助力CDISC元数据处理

江苏恒瑞医药股份有限公司

JIANGSU HENGRUI PHARMACEUTICALS CO., LTD.

刘慧

2025年12月5日

CONTENTS

目录 | Contents

01 临床数据元数据管理的现状与挑战

当前痛点

define-xml为监管要求的数据递交中最重要的文件之一，相关metadata梳理耗时耗力

- 人工编写耗时长
- 元数据一致性难以保证
- 项目相似部分重复编写

核心诉求

打通metadata间关联，基于项目数据收集情况自动产生metadata文件

- 基于元数据间关联性自动产生
- 确保元数据逻辑一致性
- 维护部门模板，项目相关内容自己动匹配

02 解决方案概述——工具核心价值

Metadata间依赖关系

AE转归

- 死亡
- 未恢复/未解决
- 恢复/解决
- 恢复/解决有后遗症
- 恢复中
- 未知

EDC

不良事件AE转归收集了6个转归状态

aCRF

基于标准库，根据“AE转归”这个自断自动匹配到AE.AEOUT

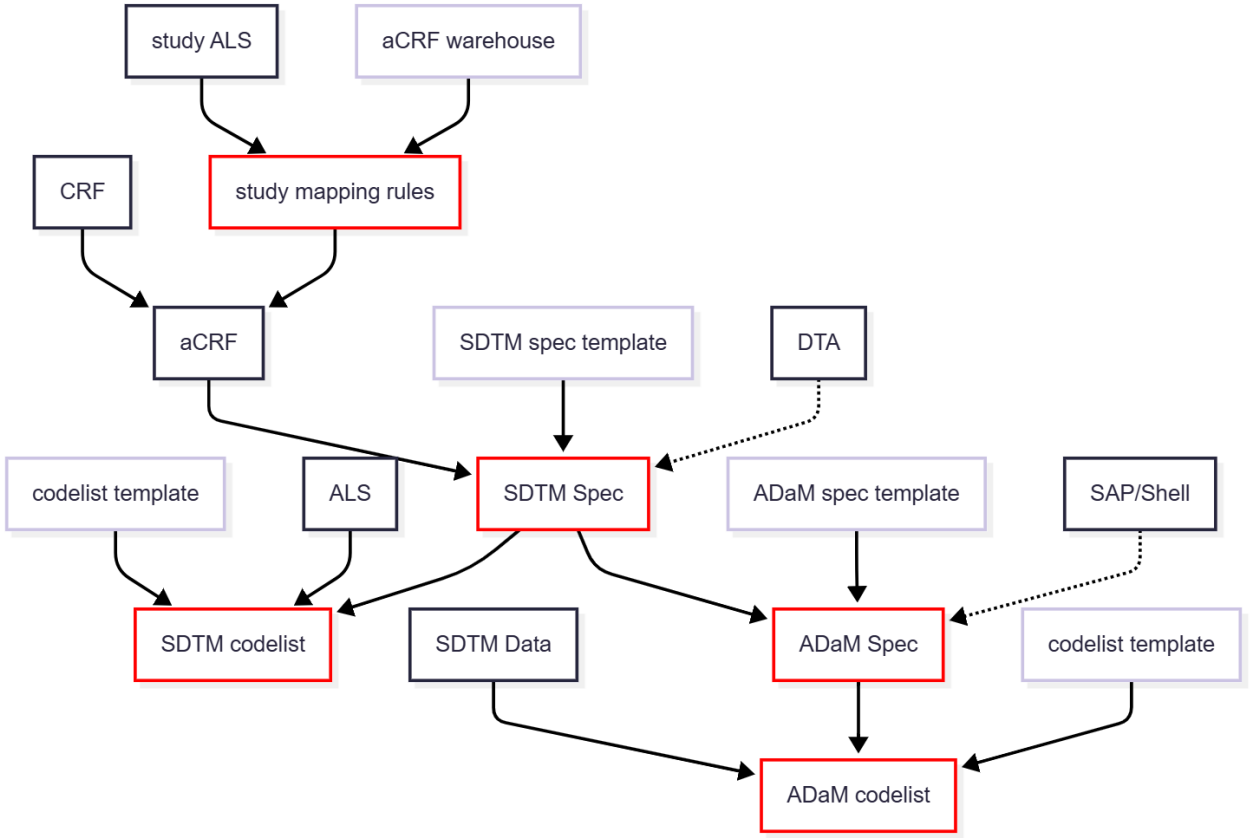
SDTM

基于aCRF, SDTM增加AEOUT这边变量，取值可通过EDC收集值和CDISC CT匹配

ADaM

将SDTM变量AEOUT同步平移到ADAE

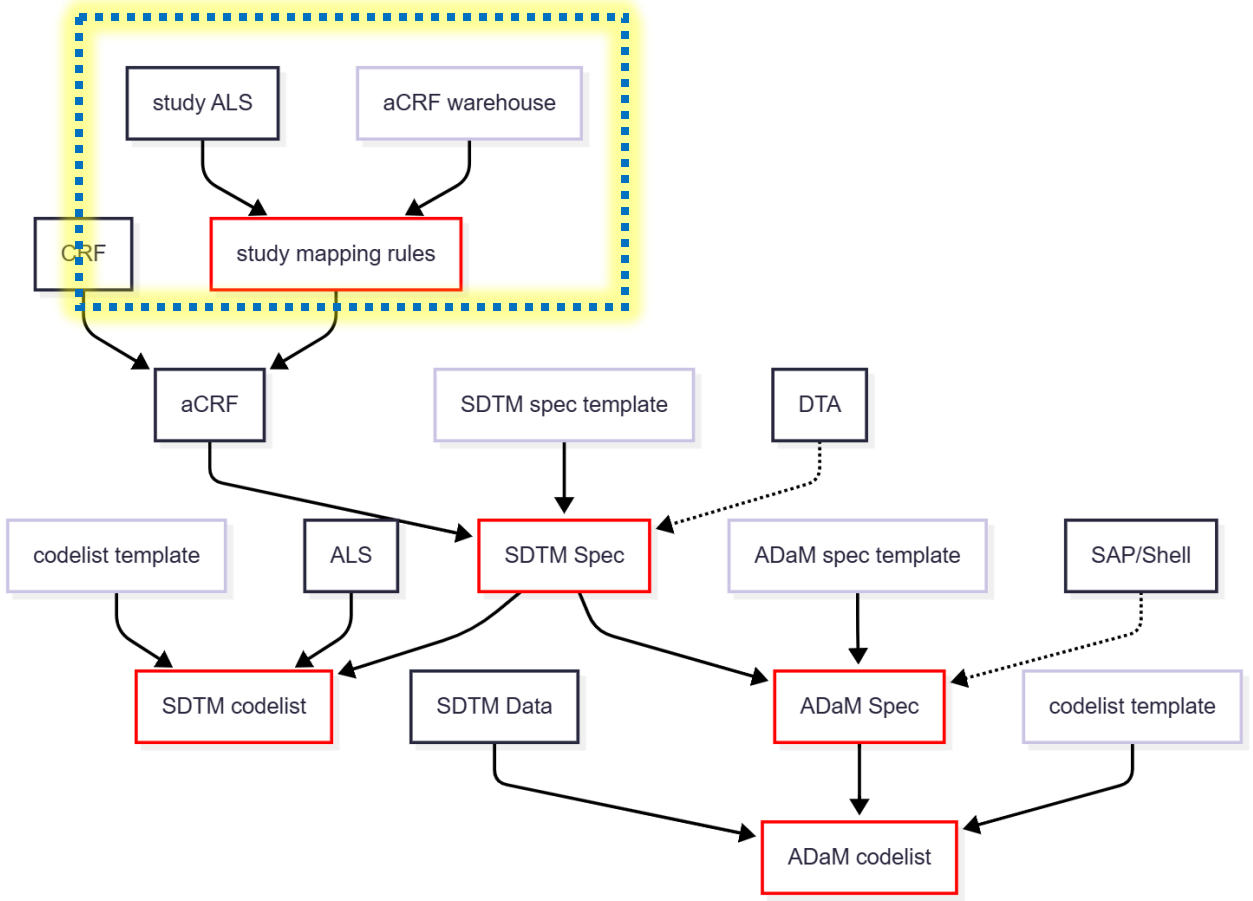
Metadata间依赖关系



03 工具功能演示

aCRF Mapping

基于项目ALS和标准库自动匹配项目相关字段对应的SDTM变量，80%以上可实现自动化，人工补充项目特异性部分，产生mapping rules后自动使用工具产生aCRF



03 工具功能演示

aCRF Mapping

基于项目ALS和标准库自动匹配项目相关字段对应的SDTM变量，80%以上可实现自动化，人工补充项目特异性部分，产生mapping rules后自动使用工具产生aCRF

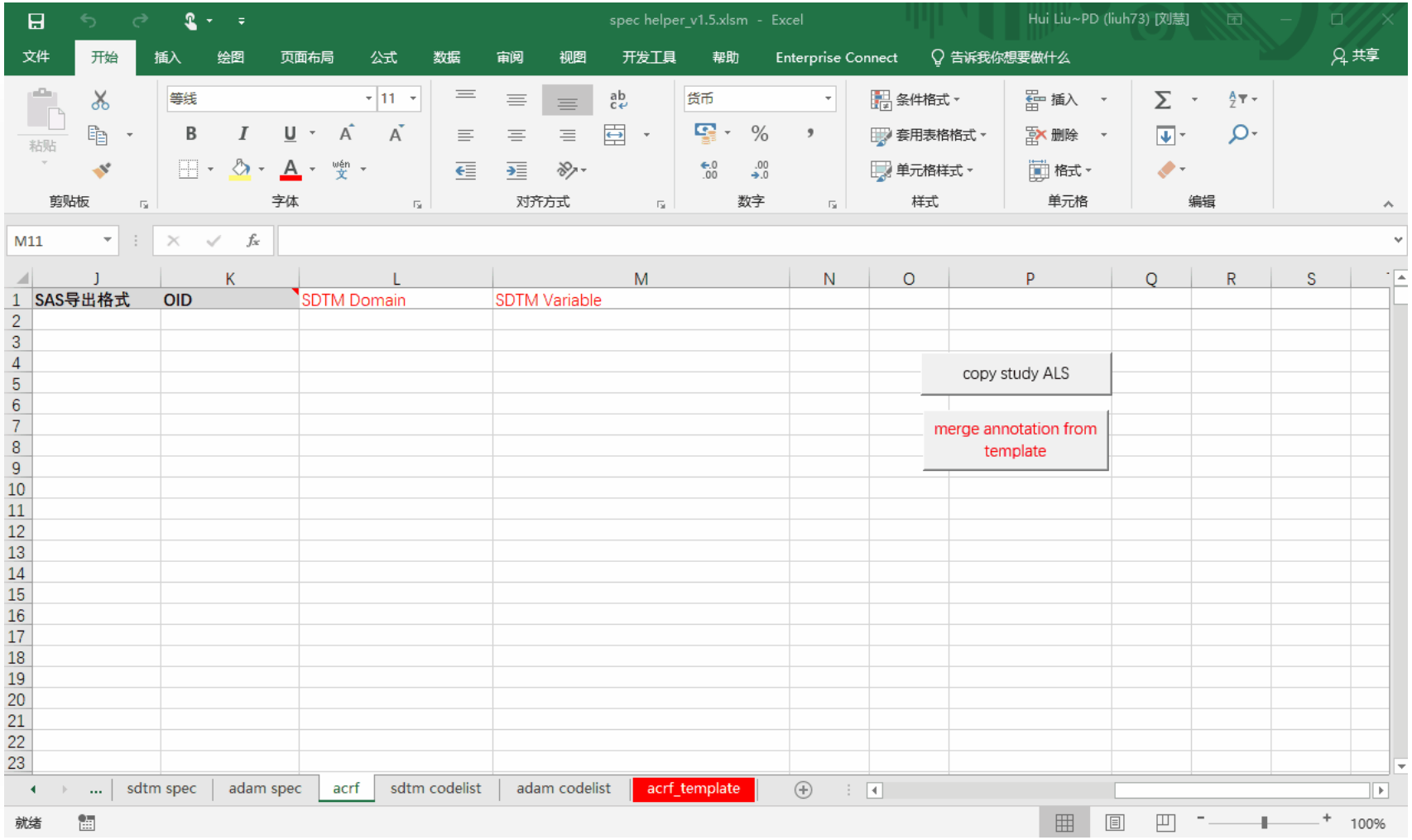
	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	表名	编号	变量名	变量	变量类型	格式	编码名称	编码排列方式	SAS导出格式	SDTM Domain	SDTM Variable	Comments		
2	访视日期	1	是否进行访视?	SVYN	单选		YN	单行Radio	¥5.00	SV (受试者访视)	SVOCCUR			
3	访视日期	2	未进行访视的原因	SVREAS	单选		NDREAS	单行Radio	¥5.00	SV (受试者访视)	SVREASOC			
4	访视日期	4	访视日期	VISDAT	日期				yymmdd10.	SV (受试者访视)	SVSTDTC			
5	人口学资料	7	出生年份	BRTHYY	数值	6			¥10.00	DC (人口学收集)	BRTHDTC			
6	人口学资料	8	年龄	AGE	数值	3			¥10.00	DC (人口学收集)	AGE			
7	人口学资料	11	民族	ETHNIC	单选		ETHNIC	多行Radio	¥7.00	DC (人口学收集)	ETHNIC			
8	人口学资料	12	种族	RACE	单选		RACE	多行Radio	¥7.00	DC (人口学收集)	RACE			
9	人口学资料	16	知情同意签署日期	RFICDAT	日期				yymmdd10.	DC (人口学收集)	RFICDTC			
10	人口学资料	17	出生日期	BRTHDAT	数值	4			¥10.00	DC (人口学收集)	BRTHDTC			
11	人口学资料	18	性别	SEX	单选		SEX	单行Radio	¥5.00	DC (人口学收集)	SEX			
12	人口学资料	19	是否有生育能力	PREGYN	单选		YN	单行Radio	¥5.00	RP (生殖系统发现)	RPORRES when RPTESTCD = CHILDPOT			
13	人口学资料	20	身高	HEIGHT	字符数值	6.3			¥10.00	VS (生命体征)	VSORRES when VSTESTCD = HEIGHT			
14	人口学资料	21	复筛前参与者代码	PSUBJID	字符				¥8.00	DC (人口学收集)	SUBJID			
15	人口学资料	22	身高单位	HEIGHTU	单选		HEIGHTU	单行Radio	¥5.00	VS (生命体征)	VSORRESU			
16	既往病史	23	疾病名称/症状	MHTERM	字符				¥200.00	MH (病史)	MHTERM			
17	既往病史	24	确诊/开始日期	MHSTDAT	字符日期				¥20.00	MH (病史)	MHSTDTC			
18	既往病史	25	是否持续	MHONGO	单选		YN	单行Radio	¥5.00	MH (病史)	MHENRTPT= 仍持续			
19	既往病史	26	结束日期	MHENDAT	字符日期				¥20.00	MH (病史)	MHENDTC			
20	饮酒史	28	酒龄	SUCDUR	字符数值	6.3			¥10.00	SU (嗜好品使用)	SUDUR			
21	饮酒史	29	单位	SUCDURU	单选		SUCDURU	单行Radio	¥5.00	SU (嗜好品使用)	SUDUR			
22	饮酒史	30	戒酒日期	SUENDAT	字符日期				¥20.00	SU (嗜好品使用)	SUENDTC			
23	饮酒史	31	平均每日饮酒量单位	SUDSTXTU	单选		SUDSTXTU_ALCO	单行Radio	¥5.00	SU (嗜好品使用)	SUDOSU			
24	饮酒史	32	平均每日饮酒量	SUDSTXT	字符数值	6.3			¥10.00	SU (嗜好品使用)	SUDOSE			
25	饮酒史	33	饮酒状况	SUNCF	单选		SUNCF	多行Radio	¥7.00	SU (嗜好品使用)	SUNCF in SUPPSU			
26	吸烟史	35	烟龄	SUCDUR	字符数值	6.3			¥10.00	SU (嗜好品使用)	SUDUR			
27	吸烟史	36	单位	SUCDURU	单选		SUCDURU	单行Radio	¥5.00	SU (嗜好品使用)	SUDUR			
28	吸烟史	37	戒烟日期	SUENDAT	字符日期				¥20.00	SU (嗜好品使用)	SUENDTC			
29	吸烟史	38	平均每日吸烟量	SUDSTXT	字符数值	6.3			¥10.00	SU (嗜好品使用)	SUDOSE			
30	吸烟史	39	平均每日吸烟量单位	SUDSTXTU	单选		SUDSTXTU_CIGR	下拉菜单	¥5.00	SU (嗜好品使用)	SUDOSU			
31	吸烟史	40	吸烟状况	SUNCF	单选		SUNCF	多行Radio	¥7.00	SU (嗜好品使用)	SUNCF in SUPPSU			
32	生命体征	42	计划检查时间点	VSTPT	单选		VSTPT	下拉菜单	¥6.00	VS (生命体征)	VSTPT			
33	生命体征	44	检查时间	VSTIM	时间				time5.	VS (生命体征)	VSDTC			
34	生命体征明细	48	临床意义	VSCLSIG	单选		CLSIG	下拉菜单	¥6.00	VS (生命体征)	VSCLSIG			

... | Instruction | sdtm spec | adam spec | acrf | sdtm codelist | adam codelist | acrf_template

03 工具功能演示

aCRF Mapping

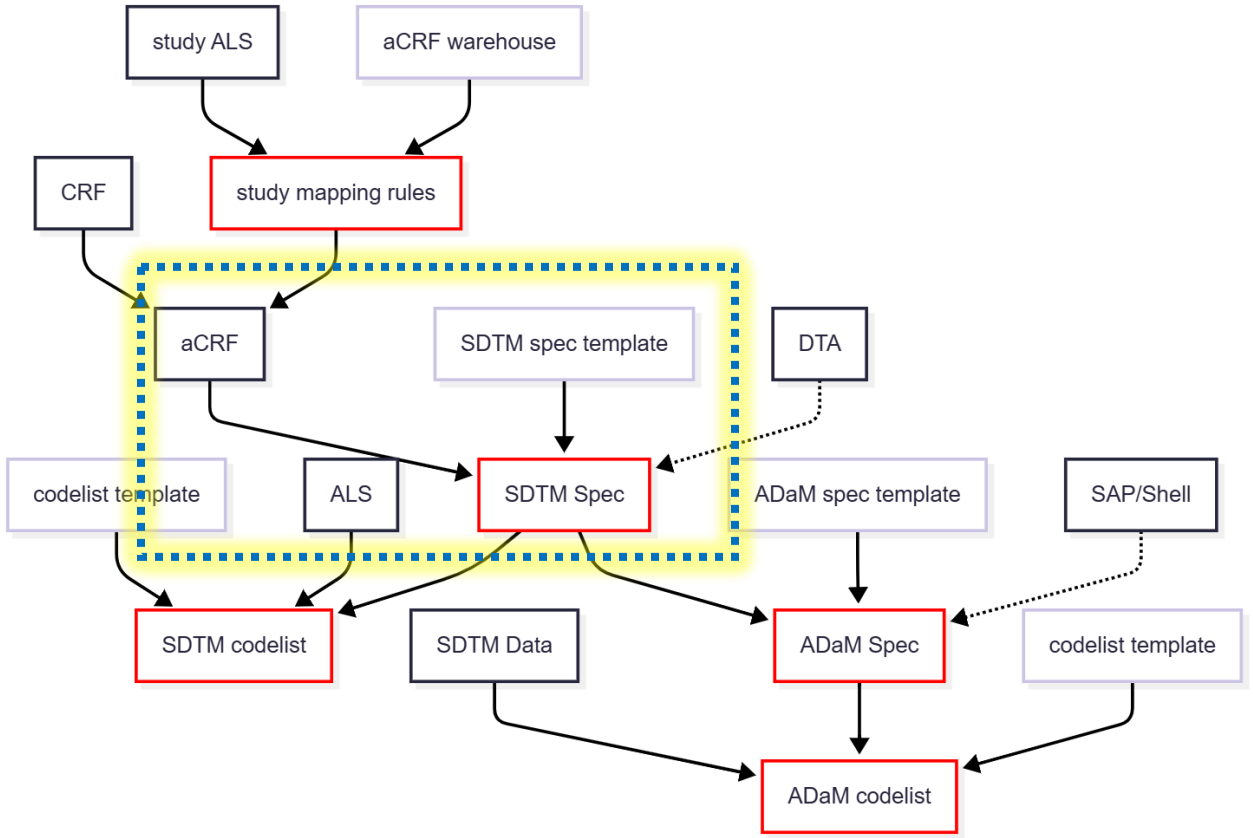
基于项目ALS和标准库自动匹配项目相关字段对应的SDTM变量，80%以上可实现自动化，人工补充项目特异性部分，产生mapping rules后自动使用工具产生aCRF



03 工具功能演示

SDTM Specification

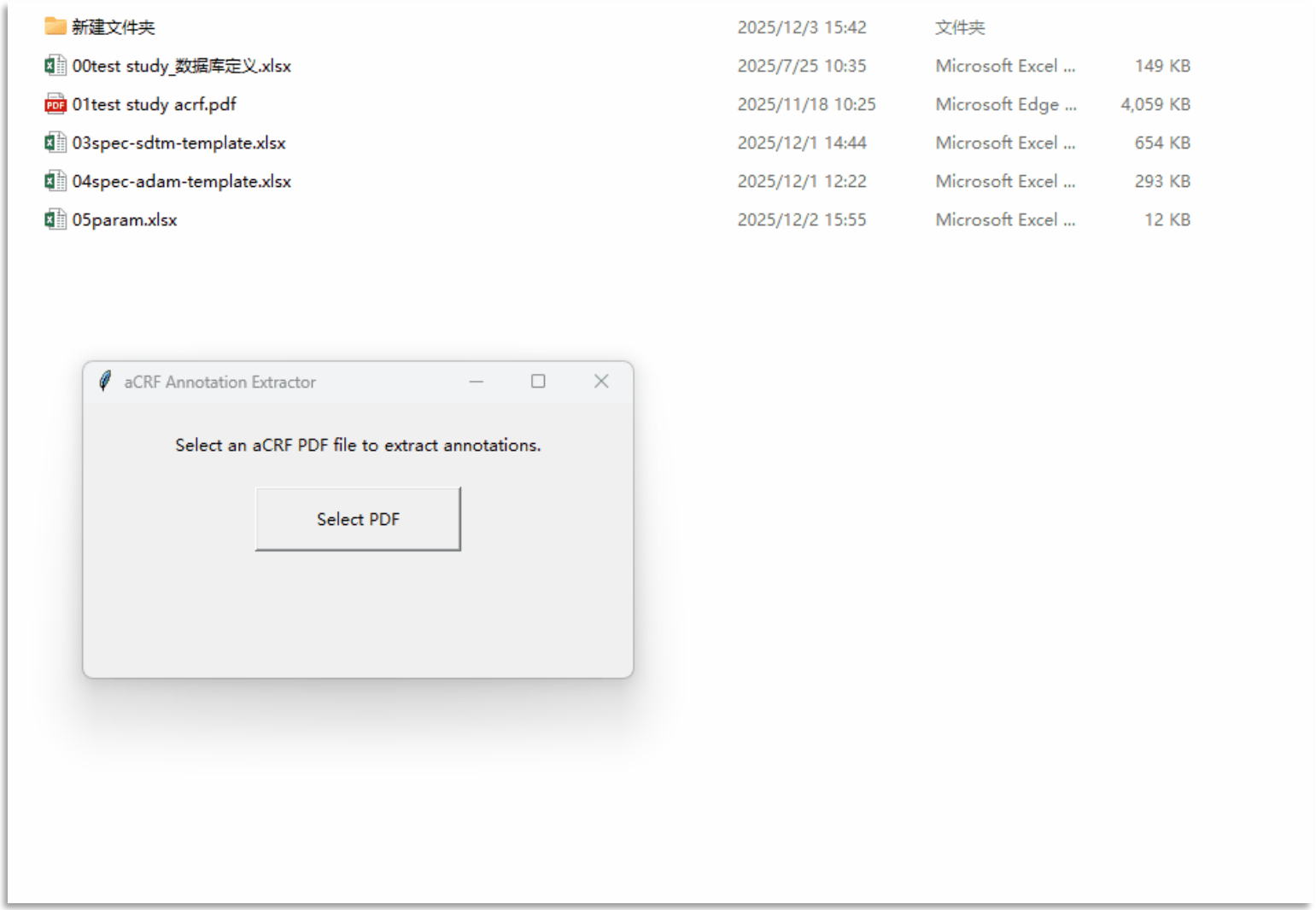
从项目aCRF中抓取相关SDTM变量信息，页码，来源；读取SDTM spec模板，自动产生项目SDTM spec



03 工具功能演示

SDTM Specification

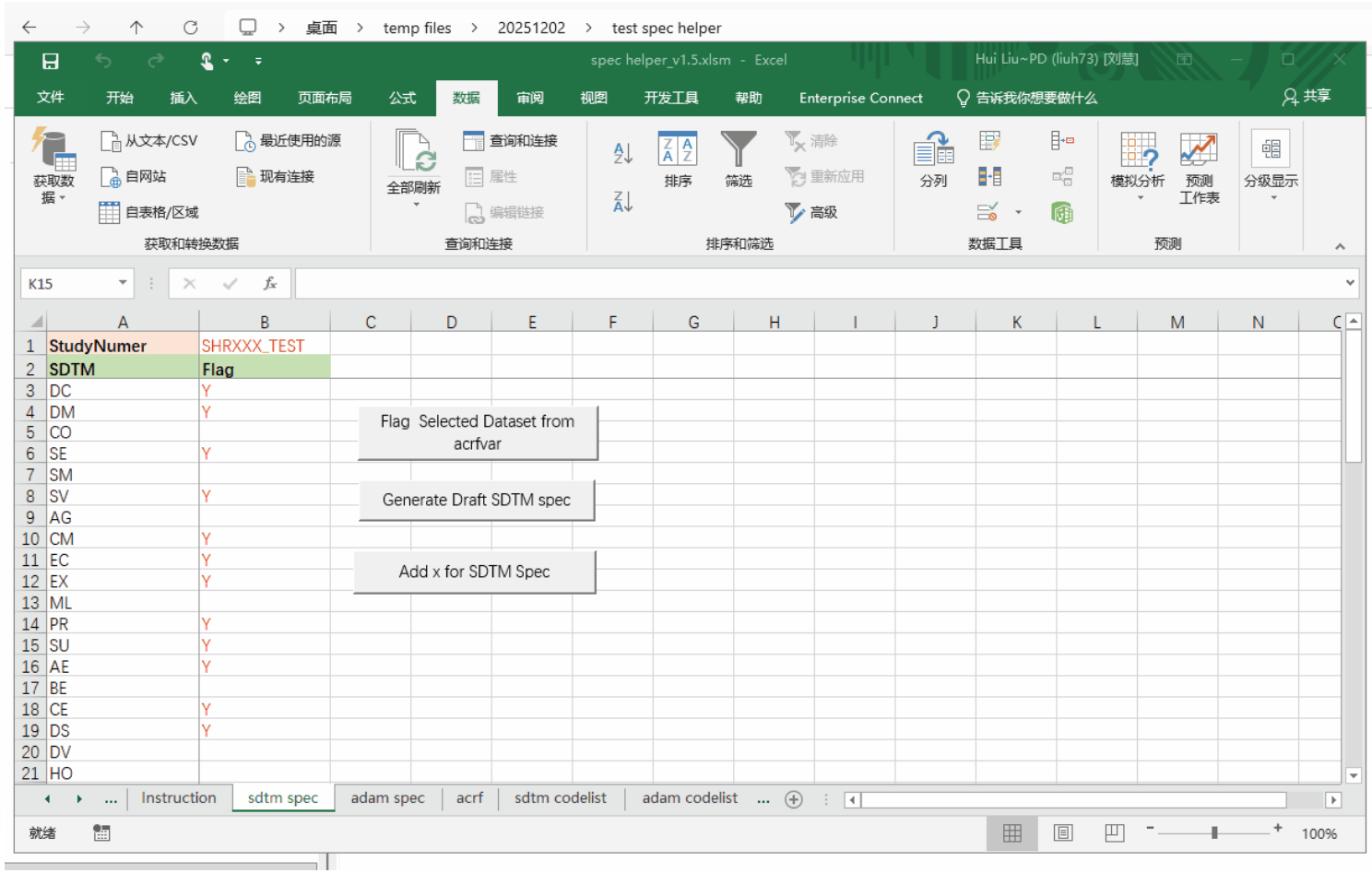
从项目aCRF中抓取相关SDTM变量信息，页码，来源；读取SDTM spec模板，自动产生项目SDTM spec



03 工具功能演示

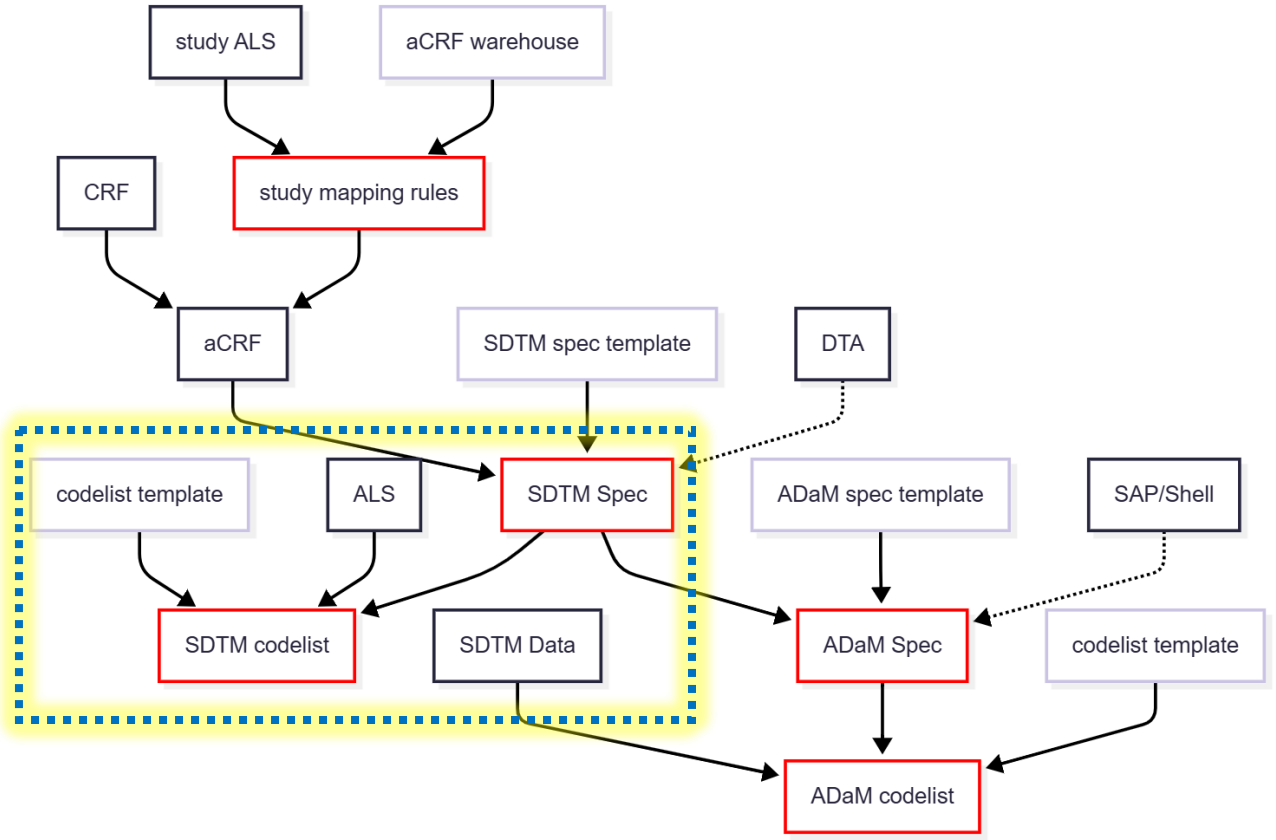
SDTM Specification

从项目aCRF中抓取相关SDTM变量信息，页码，来源；读取SDTM spec模板，自动产生项目SDTM spec



SDTM Codelist

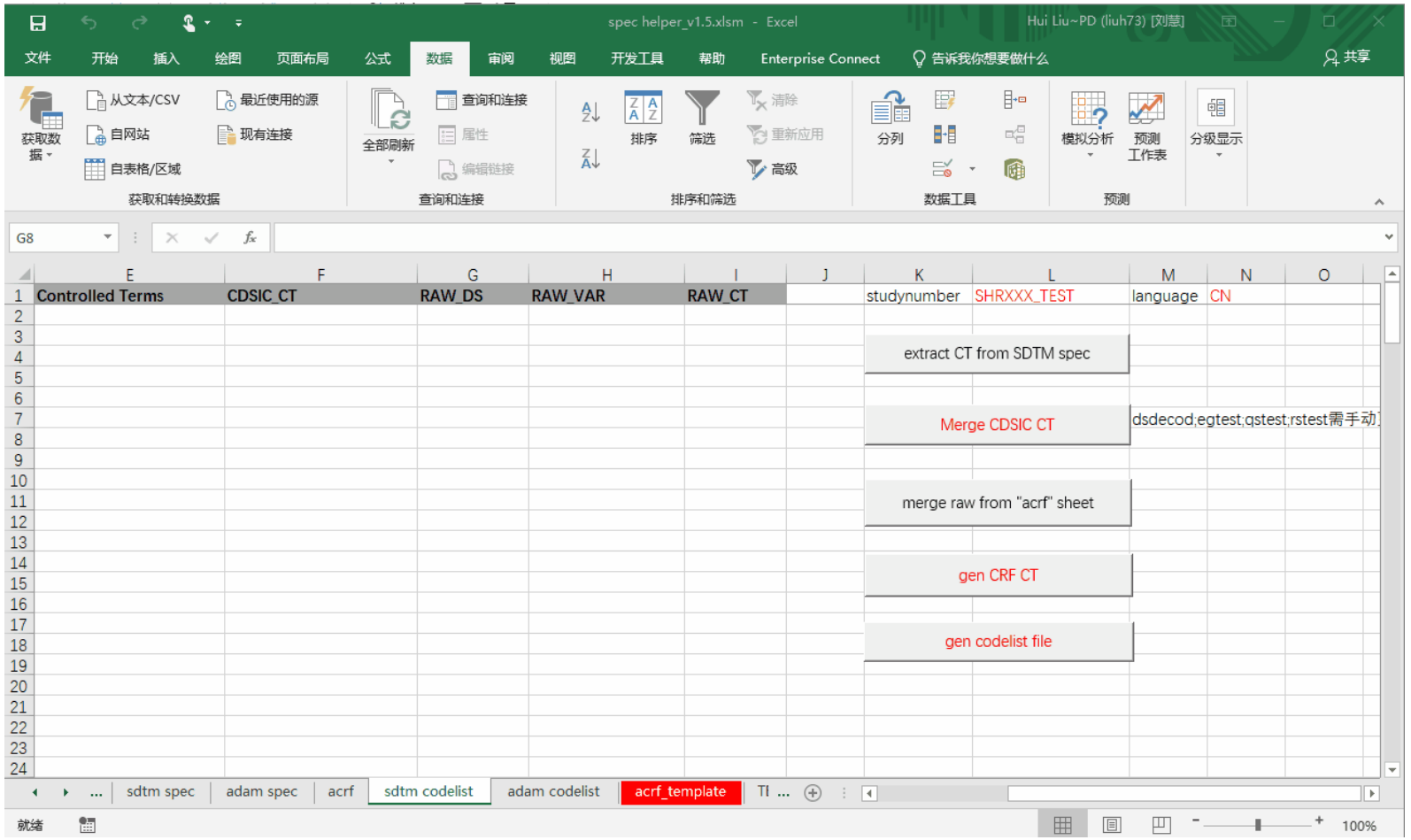
从SDTM spec中抓取CT name；与CDISC controlled terminology匹配获得对应CDSIC CT；从mapping rules中抓取对应raw variable；产生SDTM codelist。



03 工具功能演示

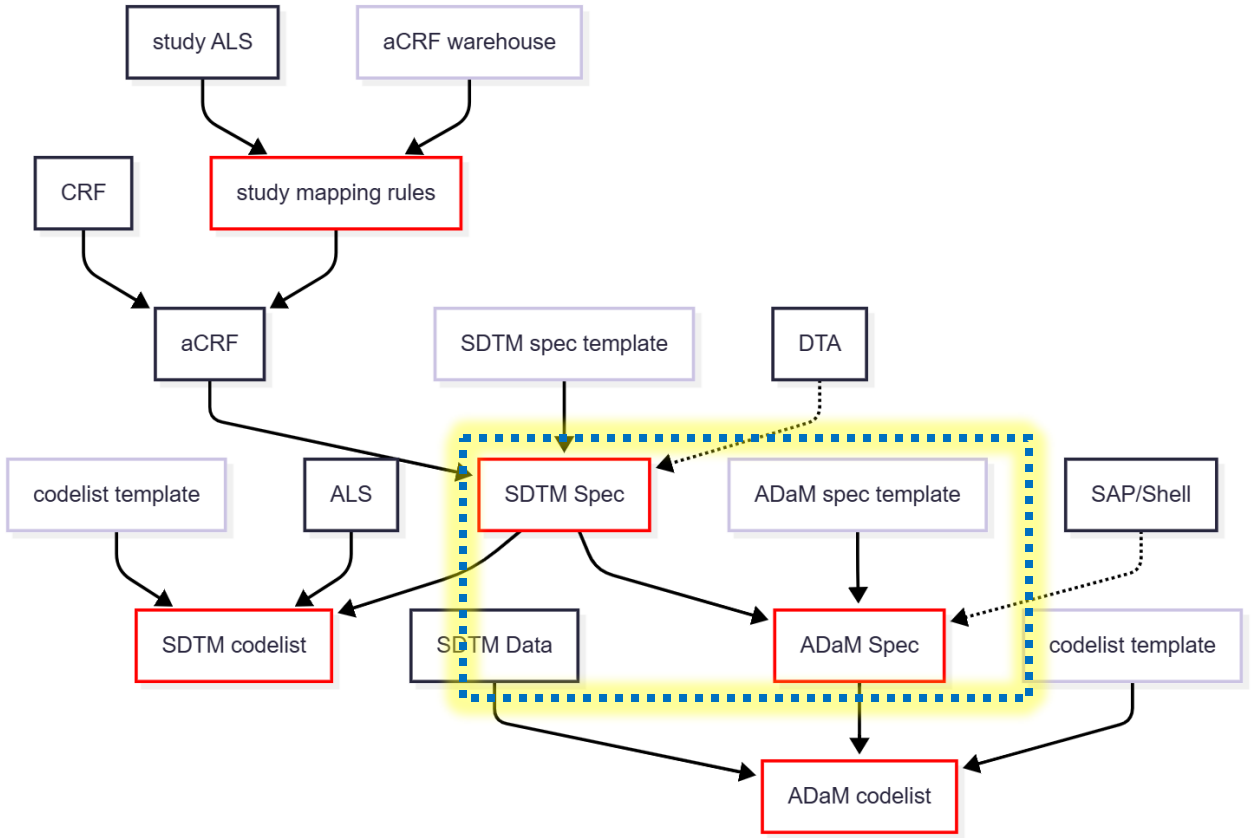
SDTM Codelist

从SDTM spec中抓取CT name；与CDISC controlled terminology匹配获得对应CDSIC CT；从mapping rules中抓取对应raw variable；产生SDTM codelist。



ADaM Specification

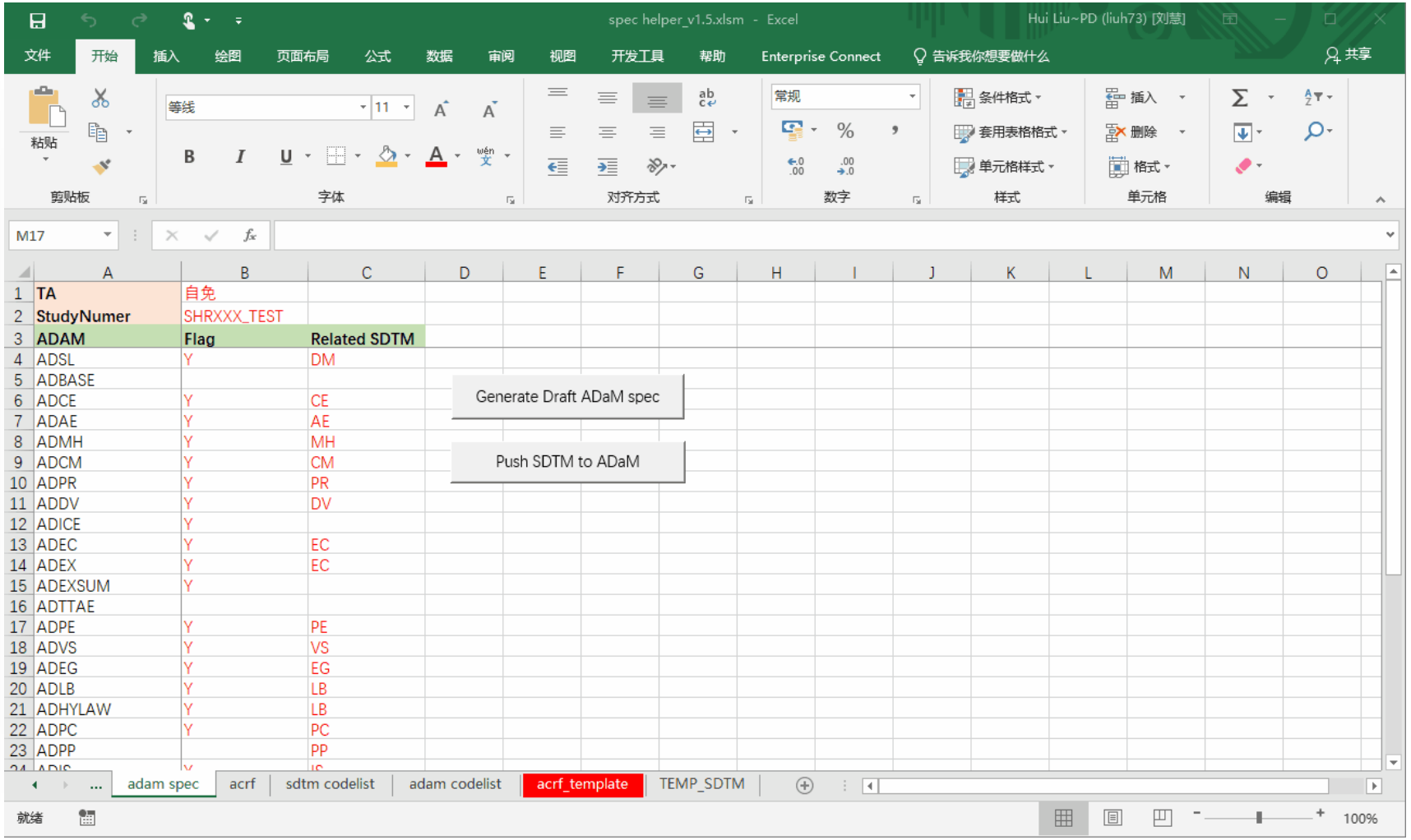
填写SDTM与ADaM对应关系，
自动产生项目ADaM spec



03 工具功能演示

ADaM Specification

填写SDTM与ADaM对应关系，
自动产生项目ADaM spec

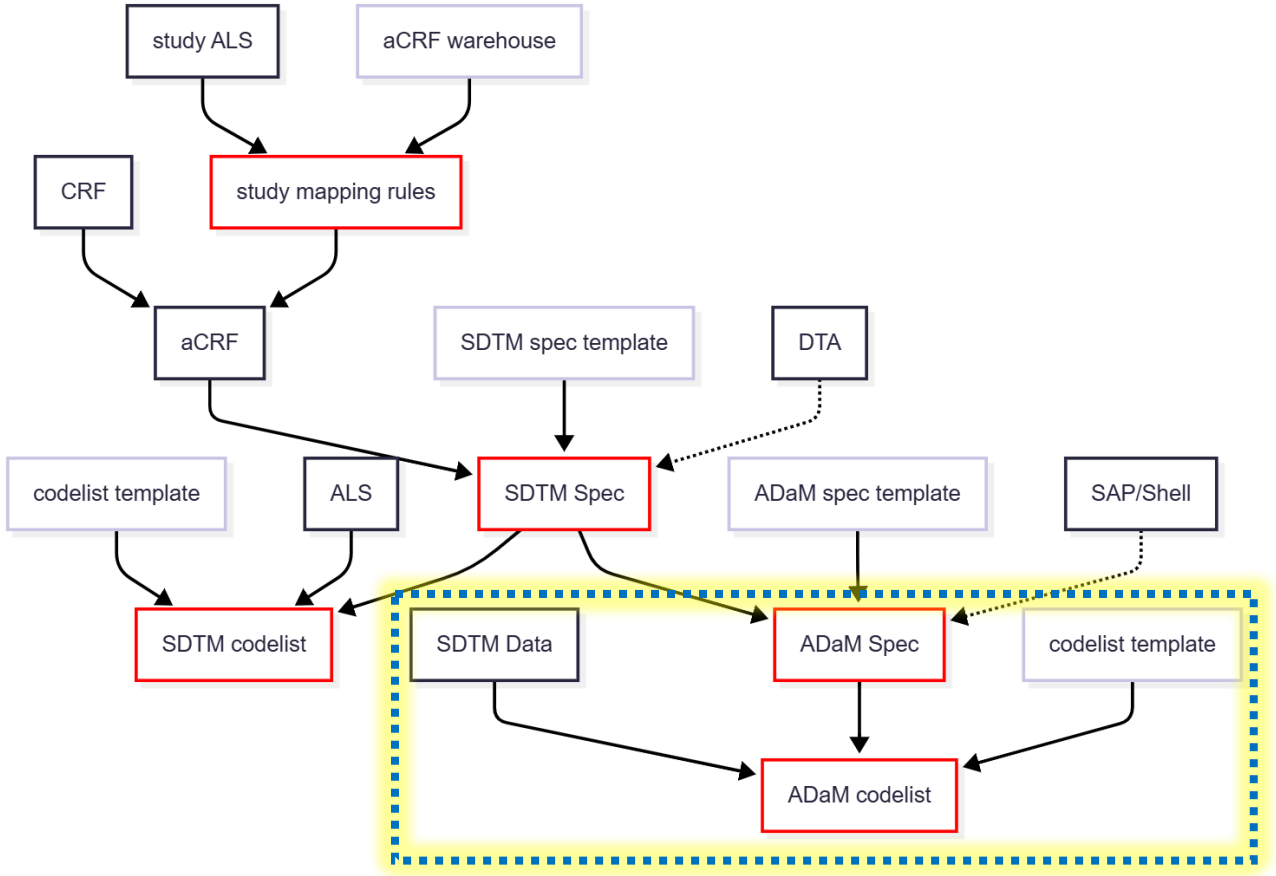


	A	B	C
1	TA	自免	
2	StudyNumer	SHRXXX_TEST	
3	ADAM	Flag	Related SDTM
4	ADSL	Y	DM
5	ADBASE		
6	ADCE	Y	CE
7	ADAE	Y	AE
8	ADMH	Y	MH
9	ADCM	Y	CM
10	ADPR	Y	PR
11	ADDV	Y	DV
12	ADICE	Y	
13	ADEC	Y	EC
14	ADEX	Y	EC
15	ADEXSUM	Y	
16	ADTTAE		
17	ADPE	Y	PE
18	ADVS	Y	VS
19	ADEG	Y	EG
20	ADLB	Y	LB
21	ADHYLAW	Y	LB
22	ADPC	Y	PC
23	ADPP		PP

03 工具功能演示

ADaM Codelist

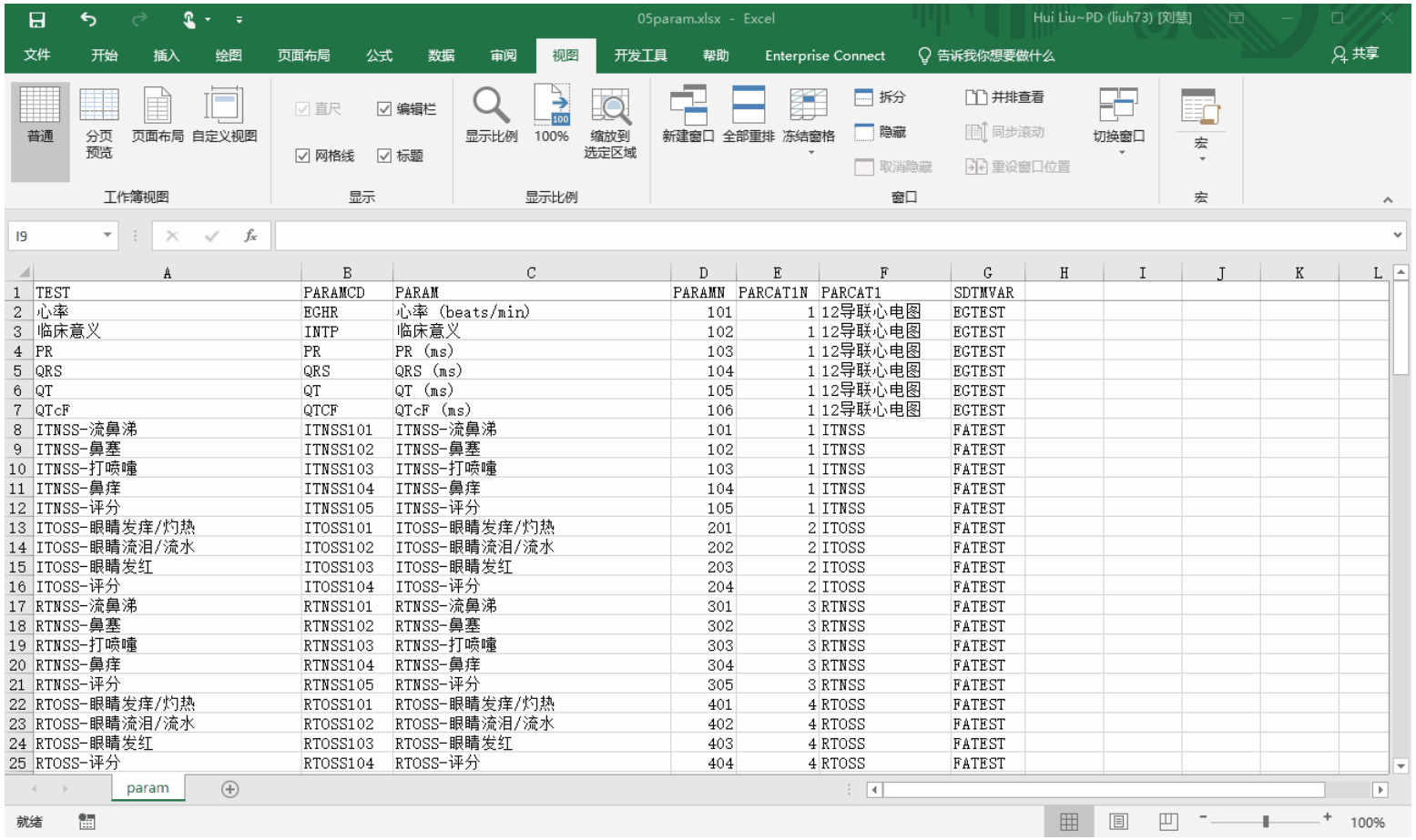
从ADaM spec中抓取CT name；与CDISC controlled terminology匹配获得对应CDSIC CT；标记SDTM CT；标记对应SDTM变量；产生ADaM codelist。



03 工具功能演示

ADaM Codelist

从ADaM spec中抓取CT name；与CDISC controlled terminology匹配获得对应CDSIC CT；标记SDTM CT；标记对应SDTM变量；产生ADaM codelist。



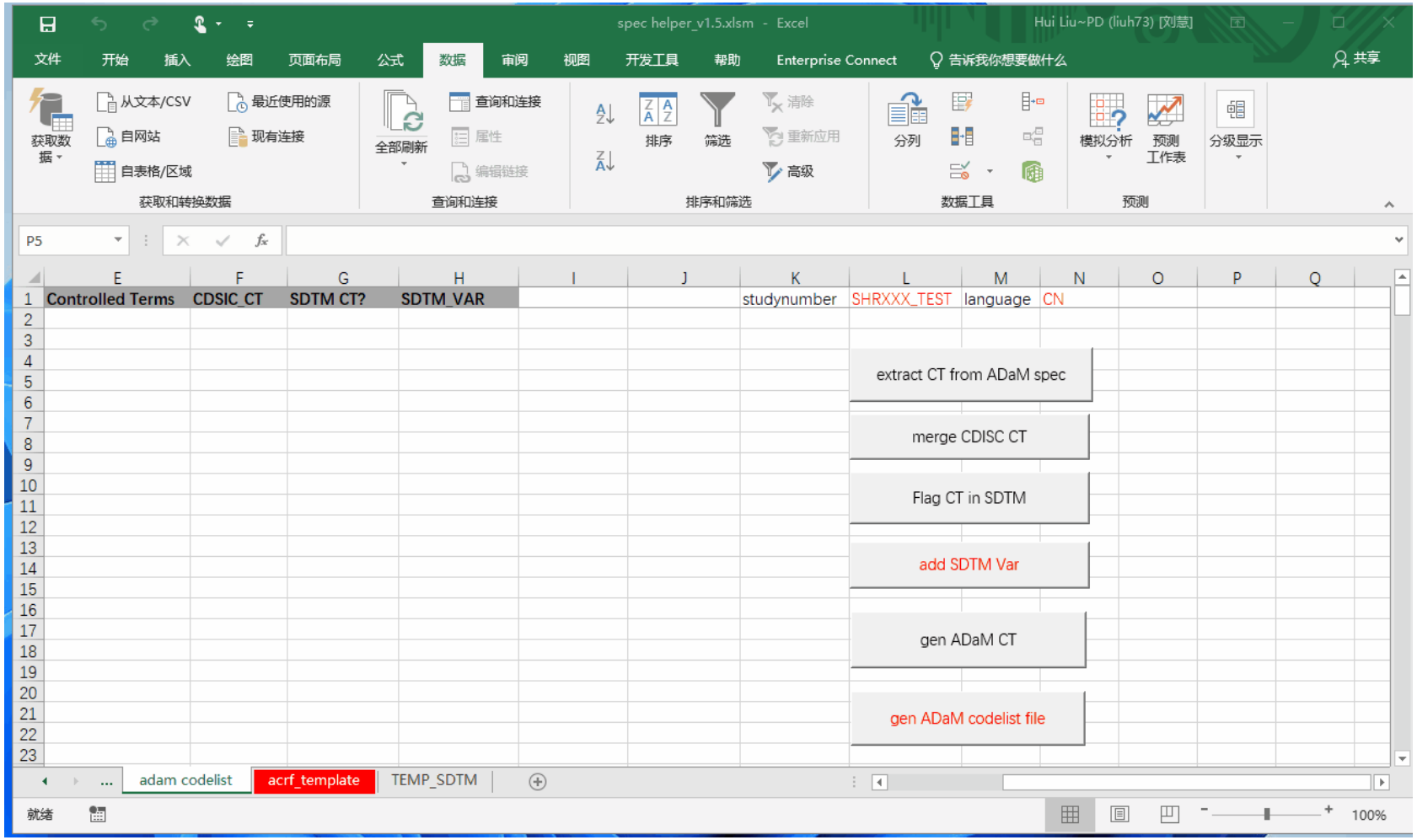
05param.xlsx - Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1	TEST	PARAMCD	PARAM	PARAMN	PARCAT1N	PARCAT1	SDTMVAR					
2	心率	EGHR	心率 (beats/min)	101	1	12导联心电图	EGTEST					
3	临床意义	INTP	临床意义	102	1	12导联心电图	EGTEST					
4	PR	PR	PR (ms)	103	1	12导联心电图	EGTEST					
5	QRS	QRS	QRS (ms)	104	1	12导联心电图	EGTEST					
6	QT	QT	QT (ms)	105	1	12导联心电图	EGTEST					
7	QTcF	QTCF	QTcF (ms)	106	1	12导联心电图	EGTEST					
8	ITNSS-流鼻涕	ITNSS101	ITNSS-流鼻涕	101	1	ITNSS	FATEST					
9	ITNSS-鼻塞	ITNSS102	ITNSS-鼻塞	102	1	ITNSS	FATEST					
10	ITNSS-打喷嚏	ITNSS103	ITNSS-打喷嚏	103	1	ITNSS	FATEST					
11	ITNSS-鼻痒	ITNSS104	ITNSS-鼻痒	104	1	ITNSS	FATEST					
12	ITNSS-评分	ITNSS105	ITNSS-评分	105	1	ITNSS	FATEST					
13	ITOSS-眼睛发痒/灼热	ITOSS101	ITOSS-眼睛发痒/灼热	201	2	ITOSS	FATEST					
14	ITOSS-眼睛流泪/流水	ITOSS102	ITOSS-眼睛流泪/流水	202	2	ITOSS	FATEST					
15	ITOSS-眼睛发红	ITOSS103	ITOSS-眼睛发红	203	2	ITOSS	FATEST					
16	ITOSS-评分	ITOSS104	ITOSS-评分	204	2	ITOSS	FATEST					
17	RTNSS-流鼻涕	RTNSS101	RTNSS-流鼻涕	301	3	RTNSS	FATEST					
18	RTNSS-鼻塞	RTNSS102	RTNSS-鼻塞	302	3	RTNSS	FATEST					
19	RTNSS-打喷嚏	RTNSS103	RTNSS-打喷嚏	303	3	RTNSS	FATEST					
20	RTNSS-鼻痒	RTNSS104	RTNSS-鼻痒	304	3	RTNSS	FATEST					
21	RTNSS-评分	RTNSS105	RTNSS-评分	305	3	RTNSS	FATEST					
22	RTOSS-眼睛发痒/灼热	RTOSS101	RTOSS-眼睛发痒/灼热	401	4	RTOSS	FATEST					
23	RTOSS-眼睛流泪/流水	RTOSS102	RTOSS-眼睛流泪/流水	402	4	RTOSS	FATEST					
24	RTOSS-眼睛发红	RTOSS103	RTOSS-眼睛发红	403	4	RTOSS	FATEST					
25	RTOSS-评分	RTOSS104	RTOSS-评分	404	4	RTOSS	FATEST					

03 工具功能演示

ADaM Codelist

从ADaM spec中抓取CT name；与CDISC controlled terminology匹配获得对应CDSIC CT；标记SDTM CT；标记对应SDTM变量；产生ADaM codelist。




□□□□□□
□□□AI□□
□□□□


95%□□□□
AI□□


AI


AI
AI

精准提问与需求架构能力

AI需要精准的指令。你必须能将模糊的临床研究问题，转化为清晰、逻辑严密、可被AI理解的任务描述。这要求深厚的领域知识与逻辑分解能力。

批判性评估与迭代优化能力

AI的产出并非总是正确或最优。你必须具备“专家眼光”，能审慎评估其代码的逻辑、效率和统计正确性，并通过多轮对话引导AI优化。

复杂系统整合与工程化能力

AI擅长写片段，但将多个模块安全、可靠地整合为可维护、可复用的自动化工具或流程，仍需人类的系统工程思维。

持续学习与工具驾驭能力

AI工具迭代飞速，保持好奇心，如思考如何将AI与专业工具（SAS/R/Python）、自动化脚本和临床数据库组合起来，形成一个端到端分析流水线。

04 AI的应用与思考

举个例子：

C	D	E	F	G	H	I
Variable Label	**Variable LabelCN**	**Controlled Terms**	**CDSIC_CT**	**RAW_DS**	**RAW_VAR**	**RAW_CT**
Domain Abbreviation	域名缩写	DOMAIN	DOMAIN			
Pre-specified	预设	NY_Y	NY			
Occurrence	发生	NY	NY	SV	SVYN	YN
Epoch	时期	EPOCH	EPOCH			
Domain Abbreviation	域名缩写	DOMAIN	DOMAIN			
Category for Medication	药物的类别	CMCAT				
Dose Units	剂量单位	CM_UNIT	UNIT	CM	CMDOSU	CMDOSU
Dose Form	剂型	CM_FRM	FRM	CM	CMDOSFRM	CMDOSFRM
Dosing Frequency per Interval	服药频率	CM_FREQ	FREQ	CM	CMDOSFRQ	CMDOSFRQ
Route of Administration	给药途径	CM_ROUTE	ROUTE	CM	CMROUTE	CMROUTE
Epoch	时期	EPOCH	EPOCH			
End Relative to Reference Time Point	结束时间与参照时点的关系	STENRF	STENRF	CM	CMONGO	YN
Reason of Treatment	用药原因	CMREAS		CM	CMREAS	CMREAS

编码名ID	编码名称	编码值	编码描述
AEACN	对试验用药	INTERRUP	暂停用药
AEACN	对试验用药	NA	不适用
AEACN	对试验用药	NCHANGE	剂量不变
AEACN	对试验用药	WITHDRA\	永久停药
AEOUT	AE转归	DEAD	死亡
AEOUT	AE转归	NREC	未恢复/未
AEOUT	AE转归	REC	恢复/解决
AEOUT	AE转归	RECING	恢复中
AEOUT	AE转归	RECSEQ	恢复/解决
AEOUT	AE转归	UNK	未知
REL	与试验用药	NRELATED	无关
REL	与试验用药	POSSIBLY	可能有关
REL	与试验用药	PROBABLY	很可能有关
REL	与试验用药	RELATED	肯定有关
REL	与试验用药	UNLIKELY	可能无关

	A	C	D	E
1	**Variable Label**	**Controlled Terms**	**RAW_CT**	**编码描述**
28	Sex	SEX	SEX	女
29	Sex	SEX	SEX	男
35	Reason Arm and/or Actual	ARMNULRS	RANDTRIG	否
36	Reason Arm and/or Actual	ARMNULRS	RANDTRIG	是
42	Occurrence	NY	YN	否
43	Occurrence	NY	YN	是
45	CM-Dose Units	CM_UNIT	CMDOSU	胶囊
46	CM-Dose Units	CM_UNIT	CMDOSU	滴
47	CM-Dose Units	CM_UNIT	CMDOSU	g
48	CM-Dose Units	CM_UNIT	CMDOSU	喷/吸
49	CM-Dose Units	CM_UNIT	CMDOSU	IU

04 AI的应用与思考

success

举个例子：

帮我写个VBA：1.抓取当前ACTIVE workbook中的G不空时的C列和G列两列的值，并做一个去重处理；2.用户打开一个新的文件，读“一般编码”和“内嵌表编码”这两个sheet，抓取“编码名ID”和“编码描述”这两列的信息，抓到之后将两个SHEET中内容纵向合并放在一起；3.G列和“编码名ID”进行一个匹配，匹配上将“编码描述”的值MAP过来，一个G的值可能对应多个“编码描述”，这时候C G要复制多次；4.将C G “编码描述”这列的信息放到当前工作簿的一个叫做TEMP的sheet中；5.“编码名ID” 这列可能在部分情况不存在，这时我希望抓取“编码名”这列

If Trim(cell.value) = columnName Then 报运行错误13，类型不匹配

您遇到的错误是因为单元格内容可能不是文本类型。我来修改代码，增加类型检查和处理：

image.png
PNG 5.94KB

报错

这个错误通常发生在 `Columns.AutoFit` 方法上，可能是因为工作表被保护或者没有足够的权限。我来修复这个问题：

新建的TEMP工作表没看到

问题可能出现在创建TEMP工作表的过程中。我来修改代码，增加更详细的调试和确认步骤：

TEMP我希望放在又C G这两列这个ACTIVE WORKBOOK中

我明白了，您希望TEMP工作表创建在包含C列和G列的当前工作簿中。让我修改代码：

希望产生的TEMP三列的列名直接是抓过来的，如C列希望是原始C列的列名，没必要加一行列名“C列值”，另外抓取的时候不限定G列不空了，C G列全抓然后做去重处理

明白了！我来修改代码，直接使用原始列名，并且抓取所有C列和G列数据做去重处理：

image.png
PNG 9.14KB

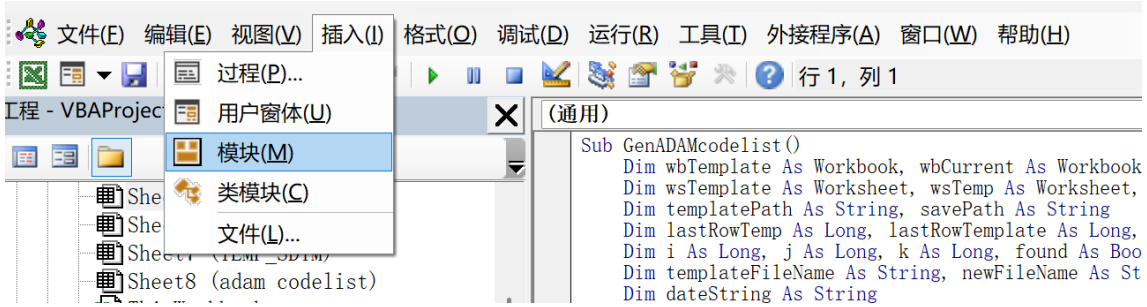
For Each key In dictSourceData.Keys

这个错误是因为没有声明 `dictSourceData` 变量。我来修复这个问题：

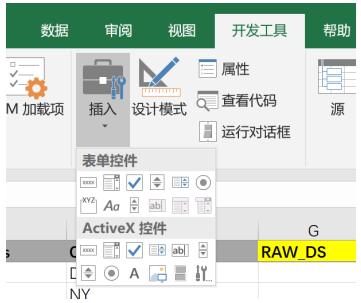
怎么在excel中配置宏?

在Excel中按Alt+F11打开VBA编辑器

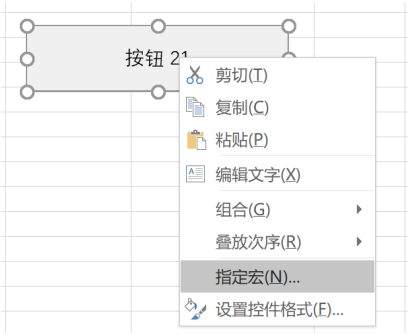
插入新模块，将AI生成的代码粘贴进去



开发工具-插入控件



修改控件名，指定宏




君子生非异也，
善假于物也
豆包AI生成

谢谢观看！

江苏恒瑞医药股份有限公司

JIANGSU HENGRUI PHARMACEUTICALS CO., LTD.