



Ideas and Systems for Implementing SDTM Datasets Review Report

Presented by Xuejia Xu, Senior SAS programmer II, Program Department, GCP ClinPlus Co., Ltd



Meet the Speakers

Xuejia Xu

Title: Senior SAS programmer II

Organization: GCP ClinPlus Co., Ltd

高级SAS程序师II，普瑞盛程序部各模块负责人，6年行业经验和丰富的CDISC经验，在独立Lead项目、作为PM审核项目的同时，也负责包括SDTM、ADaM、TFL各模块项目质量控制、流程完善及开发需求挖掘、辅助开发团队开发等团队建设管理事宜。



Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*
- *The author(s) have no real or apparent conflicts of interest to report.*



Agenda

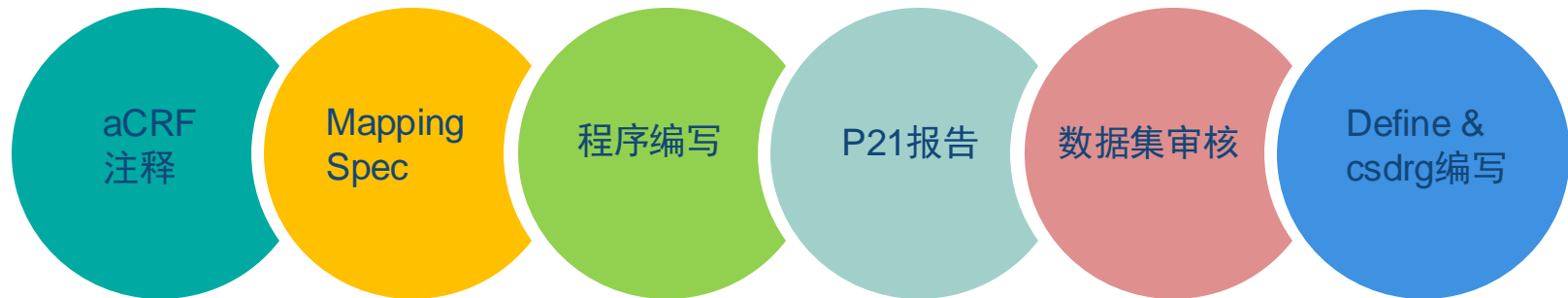
1. 背景
2. SDTM Datasets Review工具
3. SDTM全流程自动化简介
4. Summary



背景

背景

基于目前越来越多的临床研究项目采用CDISC标准，SDTM Datasets作为统计分析使用的基础数据，实现高效、高标准、高准确度的完成SDTM Datasets异常重要。保证aCRF注释、SDTM Mapping Spec、SDTM Datasets、SDTM的审核、SDTM Define & csdrg生成等各个环节的准确完成是基础，但往往需要投入大量的制作及审阅时间。所以如何能快速且准确的完成SDTM各环节以及保证各环节的一致性，提高递交质量是我们一直在思考的问题，自动化工具无疑是最佳答案。



背景

P21 community是Review SDTM Datasets的一个重要工具，它基本能核查SDTM Datasets中所有违背SDTM IG的情况，但是对于aCRF注释，Raw Data & SDTM Data之间的一致性无法核查。我们根据SDTM标准并结合我们以往项目经验，形成一系列Check List，并开发SDTM Datasets Review工具来弥补P21 community的不足，减轻人工审核负担，减少后期返工。



SDTM Datasets Review工具

SDTM Datasets Review工具

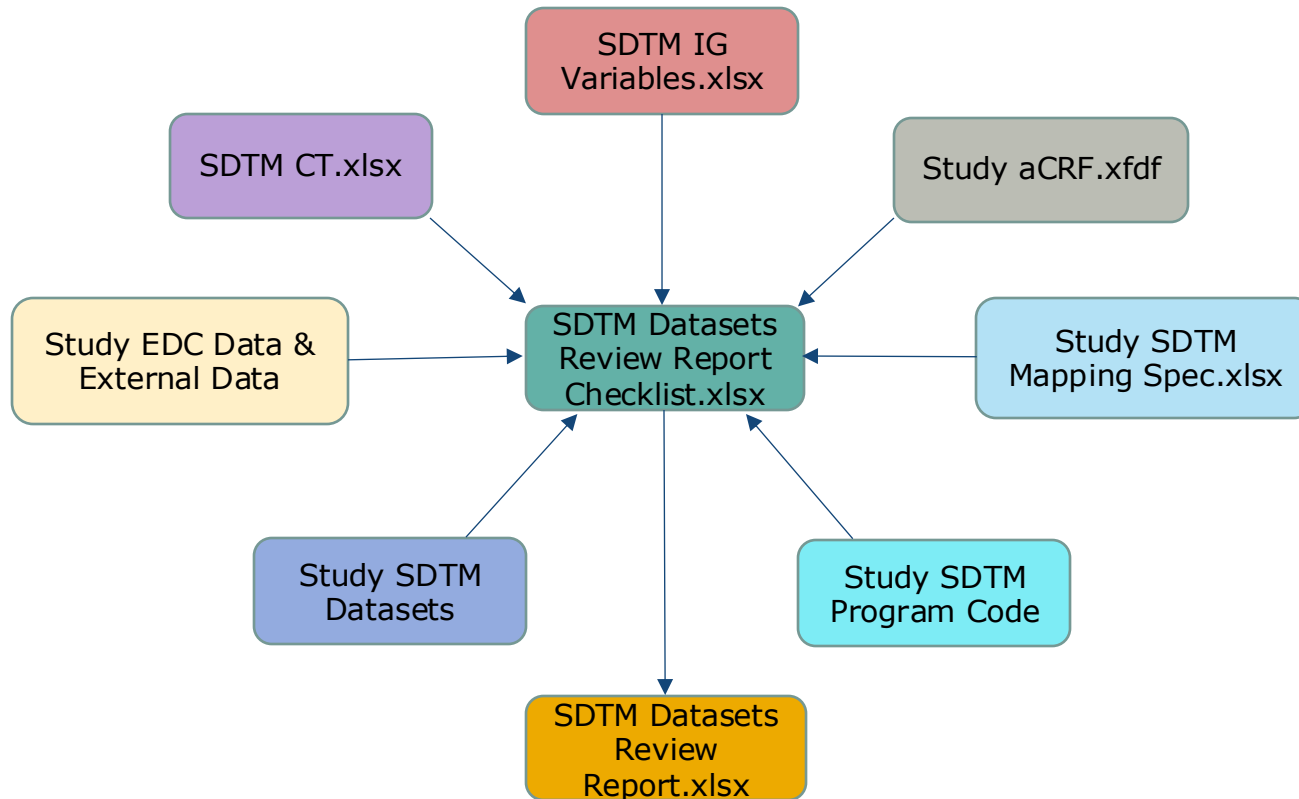
1. 你是否也常因为遗漏加工那些未在crf上记录但在edc数据中存在的非系统变量而苦恼？你是否也常因为项目进行过程中改库，新增页面或增加变量，而SDTM程序忘记更新而苦恼？
2. 你是否也常因为aCRF遗漏注释、SDTM Datasets遗漏变量加工、FDA递交约定中需提交的许可变量未添加而烦恼？
3. 你是否也常因为文件间版本号不一致问题而烦恼？
4. 你是否也常因为制作Define时大量修改Mapping Spec及CRF Page No.而痛苦？
5. 你是否也常因为QNAM长度大于8、QLABEL的长度大于40、SUPP中遗漏实际值长度超过200的变量内容、相同变量在不同数据集里信息不一致等这些细节而担忧？
6. 你是否也常因为锁库后才发现转换因子或数据录入有误造成返工而烦恼？

SDTM Datasets Review工具

核查要点:

- ※ EDC Data & External Data和Study SDTM Program Code的一致性核查
- ※ Study aCRF.xfdf和Study SDTM Datasets的一致性核查
- ※ Study aCRF.xfdf和Study SDTM Program Code的一致性核查
- ※ Study aCRF.xfdf和Study SDTM Mapping Spec的一致性核查
- ※ 文件间版本号的一致性核查
- ※ Study SDTM Mapping Spec各信息间的一致性核查
- ※ 转换因子和数据正确性的核查

SDTM Datasets Review工具



SDTM Datasets Review工具

1. EDC Data & External Data和Study SDTM Program Code的一致性核查:

Raw_data	Problem_description
aeI code	1.01: This EDC dataset exists but is not used in any of the SDTM programs, please check!
cm	1.01: This EDC dataset exists but is not used in any of the SDTM programs, please check!
sv_uns	1.01: This EDC dataset exists but is not used in any of the SDTM programs, please check!

Doamin_all	Library Member Name	Variable Name	Variable Label	Problem_description
"ae"	AEI	AEYN	是否发生不良事件?	1.03: Variable in this dataset but not used in this SDTM programs, please check!
"ae"	AEI	MDRALANG	编码语言	1.03: Variable in this dataset but not used in this SDTM programs, please check!
"ae"	AEI	MDRAVER	字典版本	1.03: Variable in this dataset but not used in this SDTM programs, please check!
"cm"	CMCODE	CMINDCO	其他, 请描述	1.03: Variable in this dataset but not used in this SDTM programs, please check!
"cm"	CMCODE	CMYN	是否存在既往及合并用药?	1.03: Variable in this dataset but not used in this SDTM programs, please check!
"dc" "dm"	SUBJ	SUBJID2	受试者筛选号	1.03: Variable in this dataset but not used in this SDTM programs, please check!
"dc" "dm" "ds"	DSIC	DSVER1	更新后的方案版本号	1.03: Variable in this dataset but not used in this SDTM programs, please check!
"dc" "dm" "ec"	EC1	ECDOSTOS_UNIT	计划给药量_UNIT	1.03: Variable in this dataset but not used in this SDTM programs, please check!

SDTM Datasets Review工具

2. Study aCRF.xfdf和Study SDTM Datasets的一致性核查:

Anno_text	Varname	Page	Problem_description
EGCAT = INTERVAL	EGCAT	37	2.04: The variable in the annotation CRF but not in the SDTM data sets, please check!
EGCAT = MEASUREMENT	EGCAT	37	2.04: The variable in the annotation CRF but not in the SDTM data sets, please check!
SVPRESP = Y When Visit is Scheduled	SVPRESP	11	2.04: The variable in the annotation CRF but not in the SDTM data sets, please check!
SVREASOC	SVREASOC	11	2.04: The variable in the annotation CRF but not in the SDTM data sets, please check!
QSDESC	QSDESC	40	2.04: The variable in the annotation CRF but not in the SDTM data sets, please check!
CMDOSEU	CMDOSEU	113	2.04: The variable in the annotation CRF but not in the SDTM data sets, please check!

SDTM Datasets Review工具

3. Study aCRF.xfdf和Study SDTM Program Code的一致性核查:

Check Item	Item Value	域名缩写	Page	Problem_description
SDTM PROGRAM	if ^missing(DTHDTC) or ^missing(DDREAS) then DTHFL="&_yes.";	dc		3.19: Pages for AE or DD are not annotated with DTHFL or subjects with DTHFL='Y' are inconsistent between datasets DM, AE or DD, please check!
SDTM PROGRAM	if ^missing(DTHDTC) or ^missing(DDREAS) then DTHFL="&_yes.";	dm		3.19: Pages for AE or DD are not annotated with DTHFL or subjects with DTHFL='Y' are inconsistent between datasets DM, AE or DD, please check!

SDTM Datasets Review工具

4. Study aCRF.xfdf和Study SDTM Mapping Spec的一致性核查:

Domain	Variable_Name	Origin	Problem_description
DA	DACAT	CRF	6.03: Origin is "CRF", there is no corresponding annotation in aCRF, please check mapping spec!
DV	DVCAT	CRF	6.03: Origin is "CRF", there is no corresponding annotation in aCRF, please check mapping spec!
DV	DVSTDTC	CRF	6.03: Origin is "CRF", there is no corresponding annotation in aCRF, please check mapping spec!
EC	ECDOSFRM	CRF	6.03: Origin is "CRF", there is no corresponding annotation in aCRF, please check mapping spec!
EC	ECROUTE	CRF	6.03: Origin is "CRF", there is no corresponding annotation in aCRF, please check mapping spec!
EX	EXDOSFRM	CRF	6.03: Origin is "CRF", there is no corresponding annotation in aCRF, please check mapping spec!
EX	EXROUTE	CRF	6.03: Origin is "CRF", there is no corresponding annotation in aCRF, please check mapping spec!
IE	IECAT	CRF	6.03: Origin is "CRF", there is no corresponding annotation in aCRF, please check mapping spec!

Domain	Variable Name	Variable Label	Origin	CRF_Page_No	Supp Domain	CRF_Page_No_xfdf	DIFF	Problem_description
MH	MHSTDTC	开始日期/时间	CRF /eDT	12		11	Y	6.13: Page numbers in the Mapping spec is inconsistent with CRF, please check mapping spec!
DM	RFICDTC	知情同意日期/时间	CRF			9	Y	6.13: Page numbers in the Mapping spec is inconsistent with CRF, please check mapping spec!
DS	DSTERM	受试者分布事件报告名称	CRF			9, 42, 43, 45	Y	6.13: Page numbers in the Mapping spec is inconsistent with CRF, please check mapping spec!

SDTM Datasets Review工具

5. 文件间版本号的一致性核查:

Domain Abbreviation	Protocol Criteria Versions	nTIVERS	Problem_description
TI	1	4	3.14: Only one PROTOCOL version or PROTOCOL version discontinuity in SDTM.TI, please check that the incl/excl criteria of all PROTOCOL version have been included!
TI	1.1	4	3.14: Only one PROTOCOL version or PROTOCOL version discontinuity in SDTM.TI, please check that the incl/excl criteria of all PROTOCOL version have been included!
TI	1.2	4	3.14: Only one PROTOCOL version or PROTOCOL version discontinuity in SDTM.TI, please check that the incl/excl criteria of all PROTOCOL version have been included!
TI	2	4	3.14: Only one PROTOCOL version or PROTOCOL version discontinuity in SDTM.TI, please check that the incl/excl criteria of all PROTOCOL version have been included!

Revision_History_latest	Revision_MappingspecName	Problem_description
V0.1	V0.2	6.14: The version in sheet [Revision History] is inconsistent with the latest version of mapping spec,please check mapping spec!
V0.8	V1.0	6.14: The version in sheet [Revision History] is inconsistent with the latest version of mapping spec,please check mapping spec!

SDTM Datasets Review工具

6. Study SDTM Mapping Spec各信息间的一致性核查:

Check Item	Variable Name	Domain all	Check Value	Problem_description
Mapping Rule	USUBJID	AE, CM, DA, DD, DM, DS, EC, EG, EX, FA, IE, LB, MH, ML, PC, PE, PR, QS, RS, SE, SU, SUPQUAL, SV, VS	对于随机受试者, 将研究标识符、中心号、受试者编号用"-"连接起来。	6.17: The Origin or Mapping rule of variable STUDYID or USUBJID are inconsistent, please check mapping spec!
Mapping Rule	USUBJID	PP	set SDTM.DM.USUBJID	6.17: The Origin or Mapping rule of variable STUDYID or USUBJID are inconsistent, please check mapping spec!

Domain	Variable_Order	QNAM	Origin	Mapping Rule	Problem_description
FA	8	FATESTCD	Assigned		6.02: Origin is "Derived" or "Assigned", mapping rule is null, please check mapping spec!
MB	24	MBSTNRLO	Derived		6.02: Origin is "Derived" or "Assigned", mapping rule is null, please check mapping spec!
MB	25	MBSTNRHI	Derived		6.02: Origin is "Derived" or "Assigned", mapping rule is null, please check mapping spec!
QS	20	QSBLFL	Derived		6.02: Origin is "Derived" or "Assigned", mapping rule is null, please check mapping spec!

SDTM Datasets Review工具

7. 转换因子的核查:

Category for Lab Test	Specimen Type	Method of Test or Examination	Lab Test or Examination Name	Original Units	Standard Units	Domain Abbreviation	factor	Problem_description
CHEMISTRY			Alanine Aminotransferase	IU/L	U/L	LB	1	4.01: The unit has been converted, please confirm the conversion factor!
CHEMISTRY			Alkaline Phosphatase	IU/L	U/L	LB	1	4.01: The unit has been converted, please confirm the conversion factor!
CHEMISTRY			Aspartate Aminotransferase	IU/L	U/L	LB	1	4.01: The unit has been converted, please confirm the conversion factor!
CHEMISTRY			Bilirubin	g/L	umol/L	LB	1710	4.01: The unit has been converted, please confirm the conversion factor!
CHEMISTRY			Gamma Glutamyl Transferase	IU/L	U/L	LB	1	4.01: The unit has been converted, please confirm the conversion factor!
CHEMISTRY			Urea	umol/L	mmol/L	LB	0.001	4.01: The unit has been converted, please confirm the conversion factor!
HEMATOLOGY			Hematocrit	%	RATIO	LB	0.001	4.01: The unit has been converted, please confirm the conversion factor!
HEMATOLOGY			Hematocrit	L/L	RATIO	LB	0.01	4.01: The unit has been converted, please confirm the conversion factor!

SDTM Datasets Review工具

7. 转换因子的核查:

Category for Lab Test	Specimen Type	Method of Test or Examination	Lab Test or Examination Name	Original Units	Standard Units	Domain Abbreviation	STRESN_MIN	STRESN_MAX	STRESN_MEAN	STRESN_STD	factor	Problem_description
CHEMISTRY			Alanine Aminotransferase	IU/L	U/L	LB	9	42.3	16.997297297	6.869760834	1	4.02: There are differences between the mean and standard deviation of each original unit and standard unit, please check!
CHEMISTRY			Alanine Aminotransferase	U/L	U/L	LB	4	152	17.090571737	11.049458007		4.02: There are differences between the mean and standard deviation of each original unit and standard unit, please check!
HEMATOLOGY			Erythrocytes	10 ¹² /L	10 ¹² /L	LB	2.9	5.95	4.2812469734	0.4495670855		4.02: There are differences between the mean and standard deviation of each original unit and standard unit, please check!
HEMATOLOGY			Erythrocytes	10 ⁹ /L	10 ¹² /L	LB	0.00518	0.00542	0.005303333	0.0001201388	0.001	4.02: There are differences between the mean and standard deviation of each original unit and standard unit, please check!
HEMATOLOGY			Hematocrit	%	RATIO	LB	0.0016	0.489	0.377047891	0.0644553882	0.01	4.02: There are differences between the mean and standard deviation of each original unit and standard unit, please check!
HEMATOLOGY			Aspartate Aminotransferase	RATIO	RATIO	LB	0.295	0.393	0.32496	0.0245992547		4.02: There are differences between the mean and standard deviation of each original unit and standard unit, please check!



SDTM全流程自动化简介

SDTM全流程自动化简介



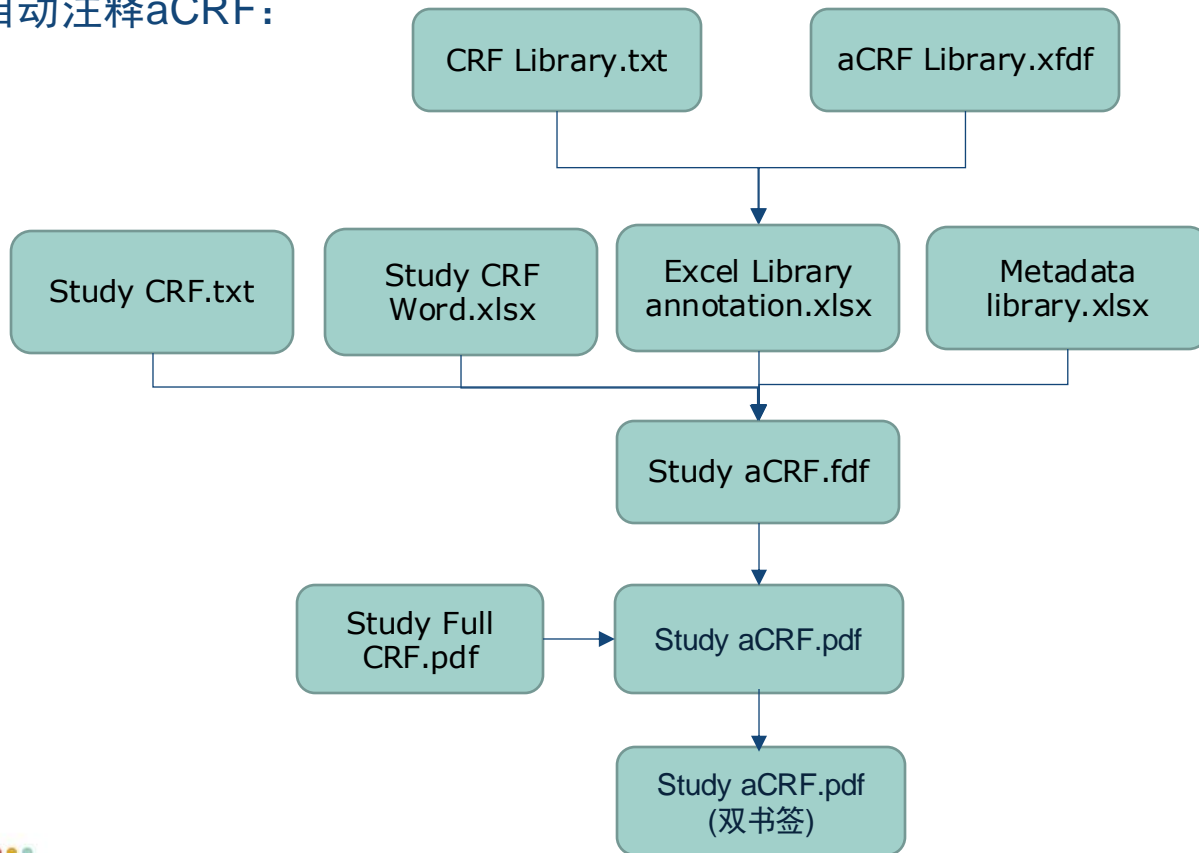
SDTM全流程自动化简介



自动化工具	使用工具前所需时长	使用工具后所需时长
自动注释aCRF	1-3 Working Days	<=8 Hours
自动化添加aCRF双书签	>=1 Working Days	<=0.5 Hours
自动化创建Mapping Spec	3-5 Working Days	1-2 Working Days
自动化创建SDTM Datasets	2-3 Weeks	About 1 Week
自动化审核SDTM	2 Working Days	4-6 Hours
自动化生成Define	3-5 Working Days	<=1 Working Days
自动生成cSDRG	About 2 Working Days	About 4 Hours

SDTM全流程自动化简介

1. 自动注释aCRF:



SDTM全流程自动化简介

1. 自动注释aCRF:

LB (实验室检查)

凝血功能 **LBCAT = 凝血功能**

是否进行了凝血功能检查? 是: **[NOT SUBMITTED]**
 否: **LBSTAT = 未查 and LBTESTCD = LBALL when 否 is checked**

否, 请说明原因 **LBREASND**

样本采集日期 **LBDC**

样本采集时间

年龄 (只读) **LBAGE in SUPPLB** **LBAGEU in SUPPLB** 岁

检查项 **LBTEST** 凝血酶原时间

检查结果 **LBORRES when LBTESTCD = PT**

临床意义判断 正常 **LBCLSIG1 in SUPPLB**
 异常无临床意义 **LBCLSIG in SUPPLB**
 异常有临床意义
 未查 **LBSTAT = 未查**

相关病史 **LBMHNO1 in SUPPLB**

相关不良事件 **LBAENO1 in SUPPLB**

单位 **LBORRESU**

aCRF注释工具:

- 1) 能精准根据CRF字段自动注释;
- 2) --TESTCD可根据检查项自动注释。

SDTM全流程自动化简介

2. 自动化aCRF双书签示例图：

The image displays two screenshots of a CDISC SDTM data dictionary interface, illustrating automated aCRF double bookmarking. The left screenshot shows a tree view where the '筛选期' (Selection Period) node is expanded, and its child nodes are bookmarked. The right screenshot shows a tree view where the '表单' (Form) node is expanded, and its child nodes are bookmarked.

Left Screenshot (Selection Period):

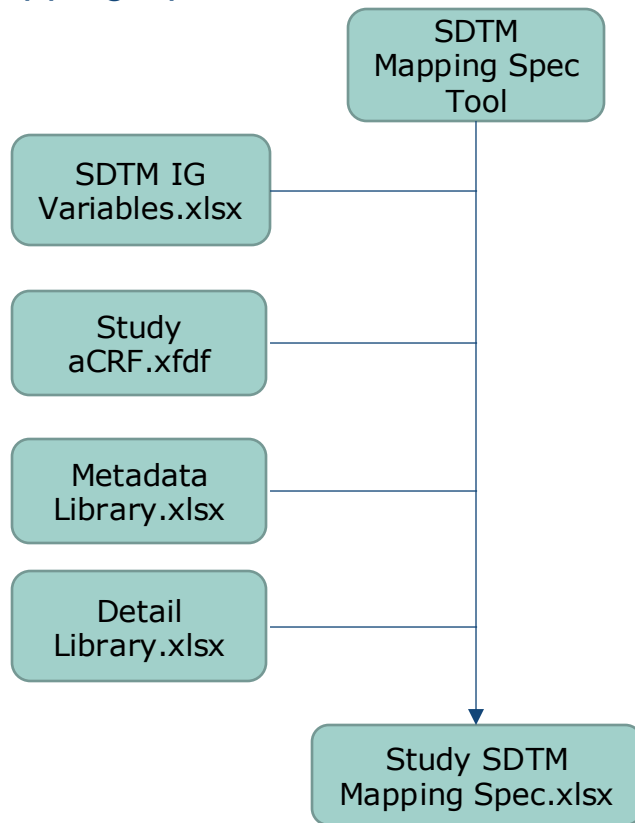
- √ 访问
 - √ 筛选期
 - 访问日期
 - 知情同意
 - 入选/排除标准
 - 人口统计学资料
 - 吸烟史
 - > C 1 D1
 - > C 1 D15
 - > C 2 D1
 - > C 2 D15
 - > C 3 D1

Right Screenshot (Form):

- √ 表单
 - > 新增受试者
 - √ 访问日期
 - 筛选期
 - C 1 D1
 - C 1 D15
 - C 2 D1
 - C 2 D15
 - C 3 D1
 - > 知情同意
 - > 入选/排除标准
 - > 人口统计学资料
 - > 吸烟史

SDTM全流程自动化简介

3. 自动生成SDTM Mapping Spec:



SDTM全流程自动化简介

3. 自动化生成SDTM Mapping Spec:

①根据acrf.xfdf文件获取的CRF变量以及SDTM IG中的必须变量和期望变量的Mapping Action均会自动给Y；且来源于CRF的变量会自动填充CRF Page No。

Domain	Variable	Variable Name	Variable Label	Type	Length	Role	Core	Controlled T	Origin	CRF Page No	Mapping Action	Raw Variable	Mapping Rule
AE	1	STUDYID	研究标识符	Char		Identifier	Req		Protocol		x		
AE	2	DOMAIN	域名缩写	Char		Identifier	Req	DOMAIN	Assigned		x		AE
AE	3	USUBJID	受试者唯一标识符	Char		Identifier	Exp		Derived		x		将研究标识符、中心号、受试者编号用“-”连接起来。对于重复筛选受试者，将研究标识符、中心号、最后一次筛选的受试者编号用“-”连接起来，且不同的受试者编号都使用最后一次筛选的USUBJID。
AE	4	SUBJID	受试者标识符	Char		Identifier	Perm		Assigned		x		DC.SUBJID
AE	5	AESSEQ	序号	Num		Identifier	Req		Derived		x		按关键字量排序；为每个USUBJID分配从1开始的序列
AE	6	AEGRPID	组号ID	Char		Identifier	Perm		CRF	55	x		
AE	7	AEREFID	参考ID	Char		Identifier	Perm		CRF		NA		
AE	8	AESPID	自定义标识符	Char		Identifier	Perm		CRF		NA		
AE	9	AETERM	不良事件报告名称	Char		Topic	Req		CRF	55	x		
AE	10	AEMODIFY	修正词	Char		Synonym Qualifier	Perm		Assigned		NA		
AE	11	AELLT	低位语	Char		Variable Qualifier	Exp		Assigned		x		未源编码数据
AE	12	AELLTCD	低位语编码	Num		Variable Qualifier	Exp		Assigned		x		未源编码数据

SDTM全流程自动化简介

3. 自动化生成SDTM Mapping Spec:

②根据acrf.xfdf文件自动生成SUPP QNAMLIST, 且来源于CRF的变量会自动填充CRF Page No.

文件 开始 插入 页面布局 公式 数据 审阅 视图 帮助 Acrobat **SDTM Mapping Status**

显示高级权限按钮 通用规则检查项 中文字符检查项 读取aCRF创建 MappingSpec 获取Domain 和Variable Label 获取中/英调整格式

当前模块版本编号: v2.0
版本更新日期: 2023-12-09
作者: 崔允辉










权限相关 SDTM Mapping 检查项 Sheet 相关 版本相关



















N16

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	Domain	Varis	QNAM	QLABEL	RDO	IDVAR	QEVAL	QORIG	Controlled T	Origin	CRF Page No	Mappin	Raw Variable	Mapping Rule
1	SUPPAE	1	AEDIS	是否因此事件退出试验	AE	AESEQ		CRF		CRF	56	x		
2	SUPPAE	2	AESACOTH	对受试者采取的其他措施	AE	AESEQ		CRF		CRF	55	x		
4	SUPPAE	3	DICTNAVS	字典名称及版本	AE	AESEQ	CLINICAL STUDY SPONSOR	ASSIGNED	MedDRA	Assigned		x		来源于编码数据的词典版本
5	SUPPCM	1	ATC1	ATC Text 1	CM	CMSEQ	CLINICAL STUDY SPONSOR	ASSIGNED	WHODRUG	Assigned		x		源于编码数据
6	SUPPCM	2	ATC1CODE	ATC Code 1	CM	CMSEQ	CLINICAL STUDY SPONSOR	ASSIGNED	WHODRUG	Assigned		x		源于编码数据
7	SUPPCM	3	ATC2	ATC Text 2	CM	CMSEQ	CLINICAL STUDY SPONSOR	ASSIGNED	WHODRUG	Assigned		x		源于编码数据
8	SUPPCM	4	ATC2CODE	ATC Code 2	CM	CMSEQ	CLINICAL STUDY SPONSOR	ASSIGNED	WHODRUG	Assigned		x		源于编码数据
9	SUPPCM	5	ATC3	ATC Text 3	CM	CMSEQ	CLINICAL STUDY SPONSOR	ASSIGNED	WHODRUG	Assigned		x		源于编码数据
10	SUPPCM	6	ATC3CODE	ATC Code 3	CM	CMSEQ	CLINICAL STUDY SPONSOR	ASSIGNED	WHODRUG	Assigned		x		源于编码数据
11	SUPPCM	7	ATC4	ATC Text 4	CM	CMSEQ	CLINICAL STUDY SPONSOR	ASSIGNED	WHODRUG	Assigned		x		源于编码数据
12	SUPPCM	8	ATC4CODE	ATC Code 4	CM	CMSEQ	CLINICAL STUDY SPONSOR	ASSIGNED	WHODRUG	Assigned		x		源于编码数据
13	SUPPCM	9	CMAENO	相关不良事件	CM	CMSEQ		CRF		CRF	57	x		
14	SUPPCM	10	CMDOSFRO	其他给药频率	CM	CMSEQ		CRF		CRF	58	x		
15	SUPPCM	11	CMDOSUOT	其他单位	CM	CMSEQ		CRF		CRF	57	x		

SDTM全流程自动化简介

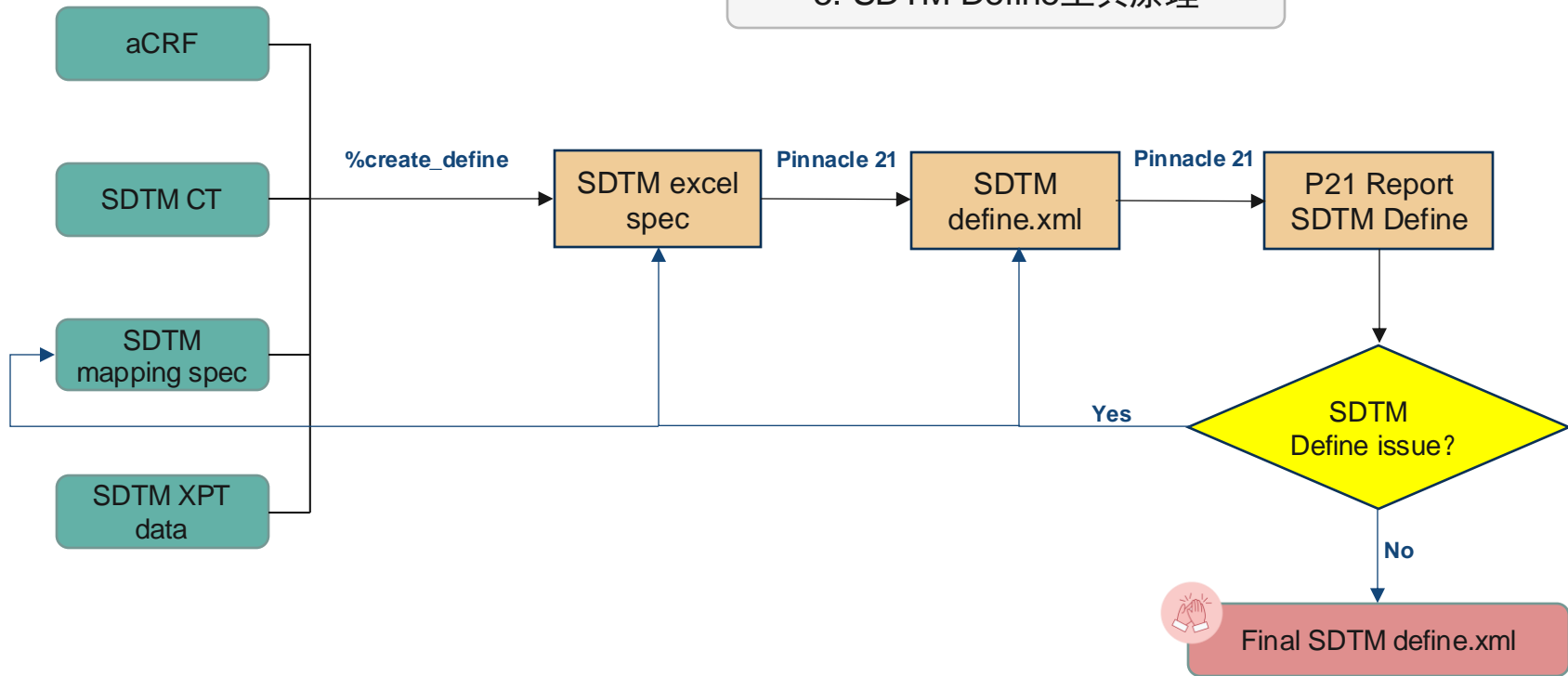
4. SDTM数据集模板程序部分程序示例：

-  ae
-  cm
-  dm
-  ds
-  ec
-  eg
-  ex from ec
-  finding
-  finding-nometadata

-  m-check-cdisciso8601-v1-2
-  m-check-varlenge200-v1-0
-  m-combine-supp-v1-1
-  m-convertvars-num2char-v1.0
-  m-create-cdisciso8601-v1-3
-  m-create-epoch-v1-2
-  m-create-maxdat-v1-2
-  m-create-orresn-v1-1
-  m-create-relrec-v1-0
-  m-create-teae-v1-0
-  m-findds2var-v1-0
-  m-get-dslist-v1-0
-  m-get-epochcode-v1-3
-  m-get-labelname-v1-0
-  m-merge-dtsbyedc-v1-0
-  m-remove-specialchar-v1-1
-  m-sdtm-blfl-v1-2
-  m-sdtm-seq-v1-0
- m-sdtm-supp-v1-1
- m-std-rawname-v1-1

SDTM全流程自动化简介

5. SDTM Define工具原理





Summary

Summary

普瑞盛数统团队以专研精神为基石，团队成员间紧密协作，利用各自的专业优势，迅速定位问题根源，并以创造性的解决方案迅速应对，将复杂任务化繁为简，让工作流程更加顺畅无阻。

如果您在此次分享中找到了共鸣，或是对更多行业前沿资讯、独家活动信息、实用技巧等内容抱有浓厚兴趣，我们诚挚地邀请您关注我们公司的官方公众号。在这里，您将享受到一个充满价值的知识宝库。让我们一起，在探索与成长的道路上携手前行！





Thank You!

cdisc