



Development and Construction of a Personalized Pinnacle 21 Report Management Platform
Using R and Other Languages: Enhancing Data Compliance and Efficiency

林孟记,统计编程副总监
北京方胜远临医药科技有限公司



Meet the Speaker

林孟记

Title: 统计编程副总监

Organization: 北京方胜远临医药科技有限公司

超过10年的行业经验，在SAS编程方面有丰富经验，参与100多个临床试验项目的统计编程工作，其中多个产品成功获得上市批准；熟悉SAS/VBA/Python/R等多种编程语言，并领导实现了多种自动化技术产品

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.*
- *{Please disclose any financial relationship or conflict of interest relevant to this presentation here OR}*
- *The author(s) have no real or apparent conflicts of interest to report.*



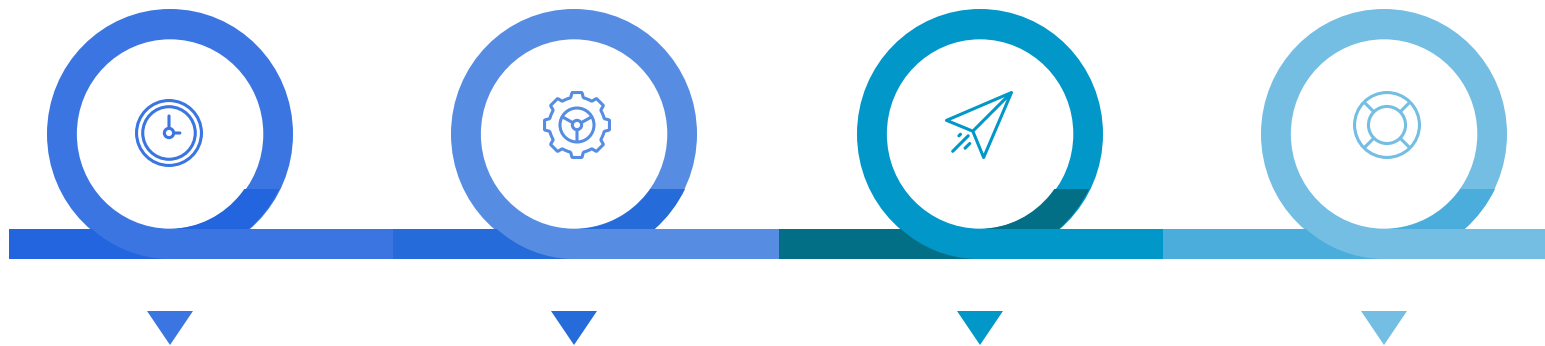
Agenda

1. 开发背景
2. Pinnacle 21报告管理
3. Specification验证
4. 数据转换
5. NMPA规则验证



开发背景

Why to Do?



Pinnacle 21 Report

版本管理?
前后版本比对?
Data Issue一键上报?

Specification验证

SDTM/ADaM Spec规范性?
Pinnacle 21规则提前验证?

数据转换

平台一键转换?
支持多种格式?
XPT? Excel? Json?

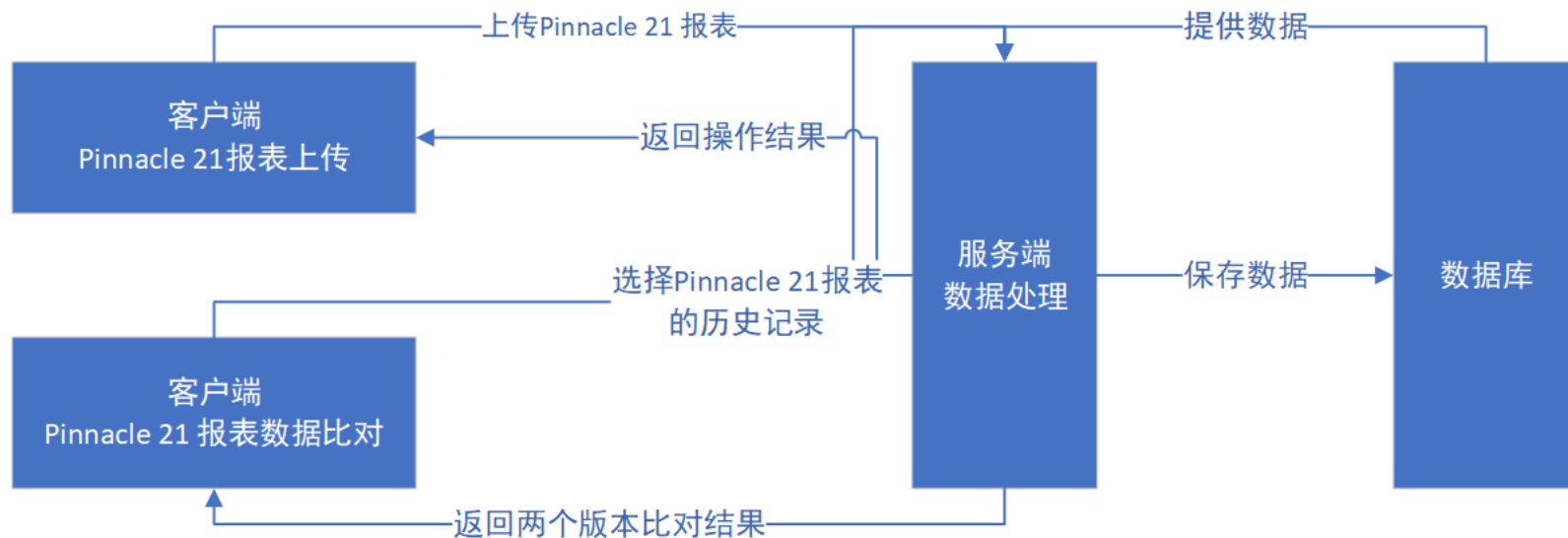
NMPA规则查验

CDE递交合规性?
项目一键验证?
保证递交项目一致性?

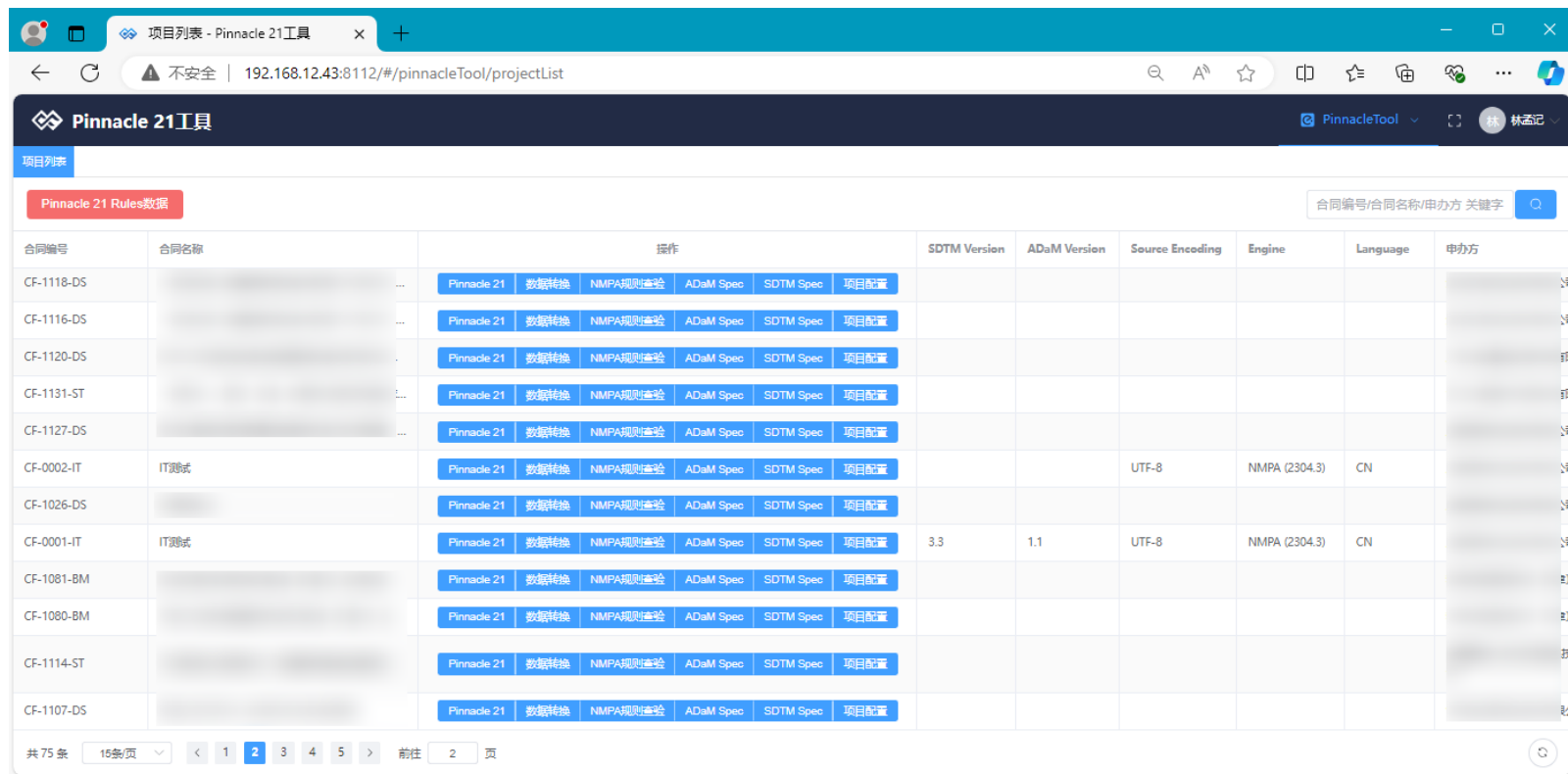


Pinnacle 21 报告管理

Pinnacle 21 报告管理



Pinnacle 21 报告管理



The screenshot displays the Pinnacle 21 web application interface. At the top, there is a browser window with the address bar showing "192.168.12.43:8112/#/pinnacleTool/projectList". The application header includes the "Pinnacle 21 工具" logo and a user profile for "林孟记". Below the header, there is a search bar with the text "Pinnacle 21 Rules数据" and a search icon. The main content area is a table with the following columns: "合同编号", "合同名称", "操作", "SDTM Version", "ADaM Version", "Source Encoding", "Engine", "Language", and "申办方". The table lists several projects, each with a set of action buttons: "Pinnacle 21", "数据转换", "NMPA规则验证", "ADaM Spec", "SDTM Spec", and "项目配置". The "CF-0001-IT" and "CF-1107-DS" rows have additional data in the "SDTM Version", "ADaM Version", "Source Encoding", and "Language" columns. At the bottom of the table, there is a pagination control showing "共 75 条" and "15 条/页", with page numbers 1 through 5 and a "前往" button.

合同编号	合同名称	操作	SDTM Version	ADaM Version	Source Encoding	Engine	Language	申办方
CF-1118-DS		Pinnacle 21 数据转换 NMPA规则验证 ADaM Spec SDTM Spec 项目配置						
CF-1116-DS		Pinnacle 21 数据转换 NMPA规则验证 ADaM Spec SDTM Spec 项目配置						
CF-1120-DS		Pinnacle 21 数据转换 NMPA规则验证 ADaM Spec SDTM Spec 项目配置						
CF-1131-ST		Pinnacle 21 数据转换 NMPA规则验证 ADaM Spec SDTM Spec 项目配置						
CF-1127-DS		Pinnacle 21 数据转换 NMPA规则验证 ADaM Spec SDTM Spec 项目配置						
CF-0002-IT	IT测试	Pinnacle 21 数据转换 NMPA规则验证 ADaM Spec SDTM Spec 项目配置			UTF-8	NMPA (2304.3)	CN	
CF-1026-DS		Pinnacle 21 数据转换 NMPA规则验证 ADaM Spec SDTM Spec 项目配置						
CF-0001-IT	IT测试	Pinnacle 21 数据转换 NMPA规则验证 ADaM Spec SDTM Spec 项目配置	3.3	1.1	UTF-8	NMPA (2304.3)	CN	
CF-1081-BM		Pinnacle 21 数据转换 NMPA规则验证 ADaM Spec SDTM Spec 项目配置						
CF-1080-BM		Pinnacle 21 数据转换 NMPA规则验证 ADaM Spec SDTM Spec 项目配置						
CF-1114-ST		Pinnacle 21 数据转换 NMPA规则验证 ADaM Spec SDTM Spec 项目配置						
CF-1107-DS		Pinnacle 21 数据转换 NMPA规则验证 ADaM Spec SDTM Spec 项目配置						

Pinnacle 21 报告管理

The screenshot displays the Pinnacle 21 web application interface. The browser address bar shows the URL `192.168.12.43:8112/#/pinnacleTool/projectList`. The page title is "Pinnacle 21 工具".

At the top, there is a search bar with the text "cf-00" and a search icon. Below this is a table with the following columns: "合同编号", "合同名称", "操作", "SDTM Version", "ADaM Version", "Source Encoding", "Engine", "Language", and "申办方".

合同编号	合同名称	操作	SDTM Version	ADaM Version	Source Encoding	Engine	Language	申办方
CF-0002-IT	IT测试	Pinnacle 21 数据转换 NMPA规则验证 ADaM Spec SDTM Spec 项目配置			UTF-8	NMPA (2304.3)	CN	成都倍特药业股份有限公司
CF-0001-IT	IT测试	Pinnacle 21 数据转换 NMPA规则验证 ADaM Spec SDTM Spec 项目配置	3.3	1.1	UTF-8	NMPA (2304.3)	CN	成都倍特药业股份有限公司

An orange circle highlights the "项目配置" link in the second row of the table, with an arrow pointing to the configuration panel below. The configuration panel is titled "[CF-0001-IT] 项目配置" and contains the following fields:

- SDTM IG Version: 3.3
- ADaM IG Version: 1.1
- Source Encoding: UTF-8
- Engine: NMPA (2304.3)
- Language: 中文

At the bottom of the configuration panel, there are "取消" (Cancel) and "确定" (Confirm) buttons. The footer of the page shows "共 2 条" (Total 2 items), "15 条/页" (15 items/page), and "1" page.

Pinnacle 21 报告管理

The screenshot displays the Pinnacle 21 web application interface. The browser address bar shows the URL: 192.168.12.43:8112/#/pinnacleTool/pinnacleValidatorReport?projectId=55&contractNum=CF-0002-IT. The application header includes the Pinnacle 21 logo and the text "Pinnacle 21工具". Below the header, there is a navigation bar with "项目列表" and "Pinnacle21". A search bar contains "CF-0002-IT" and a red button labeled "导入Pinnacle 21报告". Below the search bar, there are tabs for "SDTM", "SDTM_Define", "ADaM", and "ADaM_Define", with "ADaM" currently selected. A search filter is set to "source/Pinnacle 21 ID/Message 关键字". There are buttons for "查询", "导出到Excel", and "发送Issue". Below the search filters is a table with the following columns: Source, Pinnacle 21 ID, Message, Found, Change, Status, Explanation, Comments, Data Issue, and 操作. The table is currently empty, displaying a message "暂无数据" (No data available) with a folder icon.

Source	Pinnacle 21 ID	Message	Found	Change	Status	Explanation	Comments	Data Issue	操作
暂无数据									

Pinnacle 21 报告管理

The screenshot displays the Pinnacle 21 web application interface. The browser address bar shows the URL: `192.168.12.43:8112/#/pinnacleTool/pinnacleValidatorReport?projectId=53&contractNum=CF-0001-IT`. The application title is "Pinnacle 21工具".

The main content area is titled "导入Pinnacle 21 报告" (Import Pinnacle 21 Report). It includes a "Validation Summary" section with a "Details" link. Below this, it states "共 2733 条数据" (Total 2733 records).

A table displays the validation data with the following columns: Domain, Record, Count, Variables, Values, Pinnacle 21 ID, and Message. The table contains 13 rows of data.

Domain	Record	Count	Variables	Values	Pinnacle 21 ID	Message
GLOBAL			DATASET	SUPPSV	SD0061	defi
AE			Variable, Length	AEACNOTH, 0	SD1078	缺失
AE			Variable, Length	AESCONG, 0	SD1078	缺失
AE			Variable, Length	AESDISAB, 0	SD1078	缺失
AE			Variable, Length	AESDTH, 0	SD1078	缺失
AE			Variable, Length	AESHOSP, 0	SD1078	缺失
AE			Variable, Length	AESLIFE, 0	SD1078	缺失
AE			Variable, Length	AESMIE, 0	SD1078	缺失
DM			Variable, Length	ACTARMUD, 0	SD1149	预测
DM			Variable, Length	DTHDTC, 0	SD1149	预测
DM			Variable, Length	DTHFL, 0	SD1149	预测
DM	1		USUBIID	BT-INT056-BE-01-01-S001	SD1240	D5i

At the bottom of the report window, there are two buttons: "请选择要导入的文件" (Please select the file to import) and "确定导入[SDTM]Pinnacle21报告数据" (Confirm import of [SDTM]Pinnacle21 report data).

Pinnacle 21 报告管理

Tab 操作菜单 全 | 192.168.12.43:8112/#/pinnacleTool/pinnacleValidatorReport?projectId=53&contractNum=CF-0001-IT

Pinnacle 21工具 PinnacleTool 林 林志记

项目列表 Pinnacle21 x

CF-0001-IT **首次导入报告结果** 导入Pinnacle 21报告

SDTM SDTM_Define ADaM ADaM_Define

最新: pinnacle21-report-2024-08-10T17: 对比 对比版本 source/Pinnacle 21 ID/Message 关键字 查询 导出到Excel 发送Issue

Source	Pinnacle 21 ID	Message	Found	Change	Status	Explanation	Comments	Data Issue	操作
GLOBAL	SD0061	define.xml中有参照的域, 但数据集缺失	1	NEW					
AE	SD1078	缺失所有记录中的许可变量值	7	NEW					
DM	SD1149	预期变量缺失所有记录的值	3	NEW					
	SD1240	DS域内没有取得受访者知情同意的记录	153	NEW					
DS	SD0022	开始时间点缺失	153	NEW			原始数据未收集相关时间变量		
	SD1118	DSSTDTC值、DSDTC值、DSSTDY值同时缺失	153	NEW			原始数据未收集相关时间变量		
EC	SD1076	模型许可变量被添加到标准域中	2	NEW					
EG	SD1023	VISIT/VISITNUM的域域T域的数据不一致	18	NEW					
	SD1078	缺失所有记录中的许可变量值	3	NEW					
EX	SD1076	模型许可变量被添加到标准域中	2	NEW					
	SD1023	VISIT/VISITNUM的域域T域的数据不一致	144	NEW			存在计划外访视		

Pinnacle 21 报告管理

The screenshot displays the Pinnacle 21 web application interface. The browser address bar shows the URL: `192.168.12.43:8112/#/pinnacleTool/pinnacleValidatorReport?projectId=53&contractNum=CF-0001-IT`. The application header includes the Pinnacle 21 logo and the text "Pinnacle 21工具".

The main content area shows a table of data issues. A modal window titled "DS [SD1118] 修改" (Edit DS [SD1118]) is open, allowing for the modification of a specific data issue. The modal contains the following fields:

- Status: Closed (dropdown menu)
- Explanation: 筛选结论页面未收集日期 (text input)
- Comments: 原始数据未收集相关时间变量 (text input)
- Datassue: 请选择 (dropdown menu)

Buttons for "取消" (Cancel) and "确定" (Confirm) are located at the bottom of the modal.

The background table lists various data issues with columns for Source, Pinnacle 21 ID, Message, and a "NEW" status indicator. The issue SD1118 is highlighted, corresponding to the modal.

Source	Pinnacle 21 ID	Message	NEW
GLOBAL	SD0061	define.xml中有参数的域, 但数据集缺失	
AE	SD1078	缺失所有记录中的许可变量值	
DM	SD1149	预期变量缺失所有记录的值	
	SD1240	DS域内没有取得受访者知情同意的记录	
DS	SD0022	开始时间点缺失	
	SD1118	DSSTDTc值、DSDTC值、DSSTDY值同时缺失	NEW
EC	SD1076	模型许可变量被添加到源域中	NEW
EG	SD1023	VISIT/VISITNUM的域域TV域的数据不一致	NEW
	SD1078	缺失所有记录中的许可变量值	NEW
EX	SD1076	模型许可变量被添加到源域中	NEW
	SD1023	VISIT/VISITNUM的域域TV域的数据不一致	NEW

Pinnacle 21 报告管理

项目列表 Pinnacle21

CF-0001-IT 导入Pinnacle 21报告

SDTM SDTM_Define ADaM ADaM_Define

最新: pinnacle21-report-2024-08-10T17: 对比 对比版本 source/Pinnacle 21 ID/Message 关键字 查询 导出到Excel **发送Issue**

Source	Pinnacle 21 ID	Message	Found	Change	Status	Explanation	Comments	Data Issue	操作
DM	SD1149	预期变量缺失所有记录的值	3	NEW	Closed	筛选失败受试者未分组			
	SD1240	DS域内没有取得受试者知情同意的记录	153	NEW	Closed	DS域内有受试者的知情同意记录			
DS	SD0022	开始时间点缺失	153	NEW	Closed	筛选讨论页面未收集日期	原始数据未收集相关时间变量		
	SD1118	DSSTDTC值、DSDTC值、DSSTDY值同时缺失	153	NEW	Closed	筛选讨论页面未收集日期	原始数据未收集相关时间变量		
EC	SD1076	模型许可变量被添加到排除域中	2	NEW	Closed	原始数据收集了访视信息			
EG	SD1023	VISIT/VISITNUM的域TV域的数据不一致	18	NEW	Closed	计划外访视			
	SD1078	缺失所有记录中的许可变量值	3	NEW	Closed	原始数据集为空			
EX	SD1076	模型许可变量被添加到排除域中	2	NEW	Closed	原始数据收集了访视信息			
LB	SD1023	VISIT/VISITNUM的域TV域的数据不一致	144	NEW	Closed	存在计划外访视	存在计划外访视		
	SD1078	缺失所有记录中的许可变量值	1	NEW	Open	原始数据收集了未查原因的变...		Yes	

Pinnacle 21 报告管理

项目列表 Pinnacle21 x

CF-0001-IT 导入Pinnacle 21报告

SDTM SDTM_Define ADaM ADaM_Define

最新: pinnacle21-report-2024-08-11T14:~ 对比: pinnacle21-report-2024-08-10T17:~ source/Pinnacle 21 ID/Message 关键字 查询 导出到Excel 发送Issue

Source	Pinnacle 21 ID	Message			Comments	Data Issue	操作
AE	SD1078	缺失所有记录中的					
	SD1443	AETERM中的AEBODSYS值不一致	1	NEW			
	SD1444	AETERM中的 AESOC值不一致	1	NEW			
DM	SD1149	预期变量缺失所有记录的	3	0	Closed	筛选失败受试者未分组	
	SD1240	DS域内没有取得受试者知情同意的记录	149	4-	Closed	DS域内有受试者的知情同意记录	
DS	SD0022	开始时间点缺失	145	8-	Closed	筛选结论页面未收集日期	原始数据未收集相关时间变量
	SD1118	DSSTDTC值、DSSTDY值同时缺失	148	5-	Closed	筛选结论页面未收集日期	原始数据未收集相关时间变量
EC	SD1076	模型许可变量被添加到筛选域中	2	0	Closed	原始数据收集了访视信息	
EG	SD1023	VISIT/VISITNUM的值域TV域的数据不一致	18	0	Closed	计划外访视	
	SD1078	缺失所有记录中的许可变量值	3	0	Closed	原始数据集为空	
FX	SD1076	模型许可变量被添加到筛选域中	2	0	Closed	原始数据收集了访视信息	

Pinnacle 21 报告管理

最新: pinnacle21-report-2024-08-11T14: 对比: pinnacle21-report-2024-08-10T17: source/Pinnacle 21 ID/Message 关键字 查询 导出到Excel 发送Issue

Source	Pinnacle 21 ID	Message	Found	Change	Status	Explanation	Comments	Data Issue	操作
AE	SD1078	缺失所有记录中的许可变量值	8	1 +	Closed	原始数据收集为空			
	SD1443	AETERM中的AEBODSYS值不一致	1	NEW					
	SD1444	AETERM中的 AESOC值不一致	1	NEW					
DM	SD1149	预期变量缺失所有记录的值得	3	0	Closed	筛选失败受试者未分组			
	SD1240	DS域内没有取得受试者知情同意的记录	149	4 -	Closed	DS域内有受试者的知情同意记录			
DS	SD0022	开始时间点缺失	145	8 -	Closed	筛选讨论页面未收集日期	原始数据未收集相关时间变量		
	SD1118	DSSTDTC值、DSDTC值、DSDTDV值同时缺失	148	5 -	Closed	筛选讨论页面未收集日期	原始数据未收集相关时间变量		
EC	SD1076	模型许可变量被添加到新增域中	2	0	Closed	原始数据收集了访视信息			
EG	SD1023	VISIT/VISITNUM的值域TV域的数据不一致	18	0	Closed	计划外访视			
	SD1078	缺失所有记录中的许可变量值	3	0	Closed	原始数据收集为空			
FX	SD1076	模型许可变量被添加到新增域中	2	0	Closed	原始数据收集了访视信息			

Pinnacle 21 报告管理

下载excel格式结果

	A	B	C	D	E	F	G	H
	Source	Pinnacle 21 ID	Message	Found	Change	Status	Explanation	Comments
1								
2	AE	SD1078	缺失所有记录中的许可变量值	8	1 +	Closed	原始数据收集为空	
3		SD1443	AETERM中的AEBODSYS值不一致	1	NEW			
4		SD1444	AETERM中的 AESOC值不一致	1	NEW			
5	DM	SD1149	预期变量缺失所有记录的值	3	0	Closed	筛选失败受试者未分组	
6		SD1240	DS域内没有取得受试者知情同意的记录	149	4 -	Closed	DS域内有受试者的知情同意记录	
7	DS	SD0022	开始时间点缺失	145	8 -	Closed	筛选结论页面未收集日期	原始数据未收集相关时间变量
8		SD1118	DSSTDTC值、DSDTC值、DSSTDY值	148	5 -	Closed	筛选结论页面未收集日期	原始数据未收集相关时间变量
9	EC	SD1076	模型许可变量被添加到标准域中	2	0	Closed	原始数据收集了访视信息	
10	EG	SD1023	VISIT/VISITNUM的值域TV域的数据不一致	18	0	Closed	计划外访视	
11		SD1078	缺失所有记录中的许可变量值	3	0	Closed	原始数据收集为空	



Specification验证

Specification验证

TYPE	No.	RULE			
SDTM	1.0	instruction 这个 sheet 只是模板说明文件, 最后应该删掉。			
SDTM	2.0	所有的 Sheet History, SU	TYPE	No.	Rule
SDTM	3.1	对于 Content	ADaM	1.0	instruction 这个 sheet 只是模板说明文件, 最后应该删掉。
SDTM	3.2	对于 Content Purpose, p	ADaM	2.0	所有的 Sheet 在 Content 中的“Dataset”都有对应的数据集, Cover Page, Instruction, CONTENT, Revision History以及带有“metadata”、“detail”、“codelist”字符 (不区分大小写) 的 Sheet 除外。
SDTM	3.3	对于 Content	ADaM	3.1	对于 Content, 当“Dataset”不为空时: “Active”列的值应为“Y”或空
SDTM	3.4	对于 Content 的变量	ADaM	3.2	对于 Content, 当“Dataset”不为空时: 若“Active”列的值为“Y”, 则“Order, Class, Description, Structure, Purpose, program name, Key Variables, Sort Order”不能为空
SDTM	3.5	对于 Content	ADaM	3.3	对于 Content, 当“Dataset”不为空时: “Dataset”列的数据集应存在对应的 Sheet
SDTM	3.6	对于 Content	ADaM	3.4	对于 Content, 当“Dataset”不为空时: “Key Variables”列的变量必须来自于数据集中标 mapping action = x 的变量
SDTM	3.7	对于 Content	ADaM	3.4	对于 Content, 当“Dataset”不为空时: “Key Variables”列的变量必须来自于数据集中标 mapping action = x 的变量
SDTM	4.1	对于 Content Type, Role	ADaM	3.5	对于 Content, 当“Dataset”不为空时: “Dataset”列的数据集名长度必须小于等于 8
SDTM	4.2	对于 Content	ADaM	3.6	对于 Content, 当“Dataset”不为空时: “Description”列的值长度必须小于等于 40
SDTM	4.3	对于 Content	ADaM	3.7	对于 Content, 当“Dataset”不为空时: “Description”列的值不能有单独不对的 < > ' () {} [] 符号
SDTM	4.4	对于 Content “() {} [] 符号”	ADaM	4.1	对于 Content 中的数据集, 当“Variable Name”不为空时: “Dataset, Variable Order, Variable Label, Type, Origin, Mapping Action”不为空
SDTM	4.5	对于 Content	ADaM	4.2	对于 Content 中的数据集, 当“Variable Name”不为空时: “Variable Name”列值的长度必须小于等于 8
SDTM	4.6	对于 Content Identifier/Tc	ADaM	4.3	对于 Content 中的数据集, 当“Variable Name”不为空时: “Variable Label”列的值长度必须小于等于 40
SDTM	4.7	对于 Content Qualifiers/V	ADaM	4.4	对于 Content 中的数据集, 当“Variable Name”不为空时: “Variable Label”列的值不能有单独不对的 < > ' () {} [] 符号
SDTM	4.8	对于 Content (Expected)	ADaM	4.5	对于 Content 中的数据集, 当“Variable Name”不为空时: “Type”列的值只能为: Num/Char
SDTM	4.8	对于 Content “只能为“x”	ADaM	4.5	对于 Content 中的数据集, 当“Variable Name”不为空时: “Type”列的值只能为: Num/Char
			ADaM	4.6	对于 Content 中的数据集, 当“Variable Name”不为空时: “Origin”列的值只能为: CRF/Derived/Assigned/Protocol/eDT 以及以上值的组合 (英文逗号分隔)

Specification验证

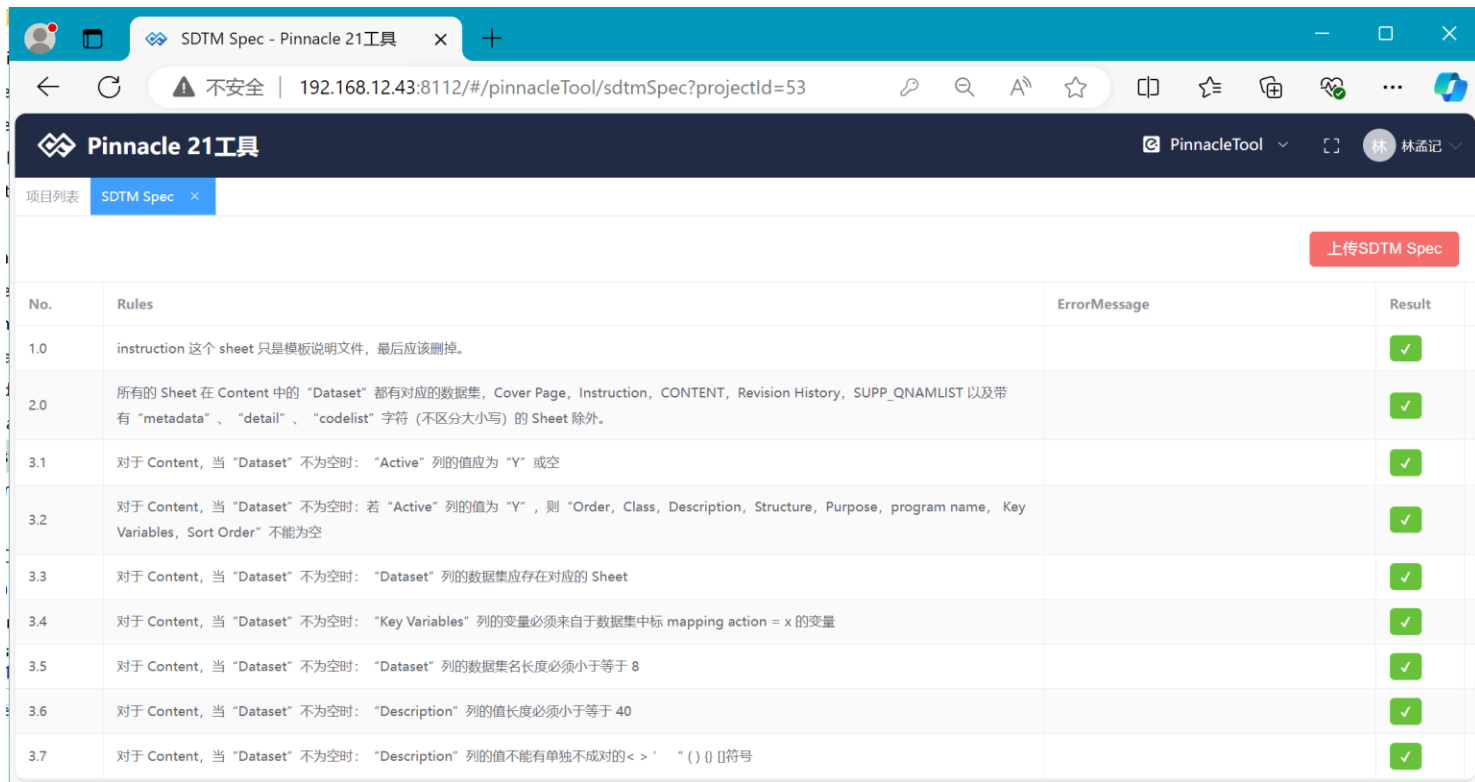
The screenshot shows the Pinnacle 21 tool interface with a file selection dialog box open. The dialog box is titled "打开" (Open) and shows the file "adam-mapping-spec-xxxxxx-cdisc.xls" selected. The file name is entered in the "文件名(N):" field. The file type is set to "自定义文件 (*.xls;*.xlsx)". The "打开(O)" button is highlighted. A yellow circle highlights a button labeled "请选择要导入的文件" (Please select the file to import) at the bottom of the dialog box.

No.	Rules
1.0	instruction 这个 sheet 只是模板说明文件, 最后应该...
2.0	所有的 Sheet 在 Content 中的 "Dataset" 都有对应的...
3.1	对于 Content, 当 "Dataset" 不为空时: "Active" 列...
3.2	对于 Content, 当 "Dataset" 不为空时: 若 "Active"...
3.3	对于 Content, 当 "Dataset" 不为空时: "Dataset"...
3.4	对于 Content, 当 "Dataset" 不为空时: "Key Variab..."
3.5	对于 Content, 当 "Dataset" 不为空时: "Dataset"...
3.6	对于 Content, 当 "Dataset" 不为空时: "Descripti..."
3.7	对于 Content, 当 "Dataset" 不为空时: "Descripti..."
4.1	对于 Content 中的数据集, 当 "Variable Name" 不为...
4.2	对于 Content 中的数据集, 当 "Variable Name" 不为空时: "Variable Name" 列值的长度必须小于等于 8
4.3	对于 Content 中的数据集, 当 "Variable Name" 不为空时: "Variable Label" 列值的长度必须小于等于 40
4.4	对于 Content 中的数据集, 当 "Variable Name" 不为空时: "Variable Label" 列的值不能有单独不成功的 < >
4.5	对于 Content 中的数据集, 当 "Variable Name" 不为空时: "Type" 列的值只能为: Num/Char

Specification验证

No.	Rules	ErrorMessage	Result
1.0	instruction 这个 sheet 只是模板说明文件, 最后应该删掉。		✓
2.0	所有的 Sheet 在 Content 中的 "Dataset" 都有对应的数据集, Cover Page, Instruction, CONTENT, Revision History以及带有 "metadata"、"detail"、"codelist" 字符 (不区分大小写) 的 Sheet 除外。		✓
3.1	对于 Content, 当 "Dataset" 不为空时: "Active" 列的值应为 "Y" 或空		✓
3.2	对于 Content, 当 "Dataset" 不为空时: 若 "Active" 列的值为 "Y", 则 "Order, Class, Description, Structure, Purpose, program name, Key Variables, Sort Order" 不能为空		✓
3.3	对于 Content, 当 "Dataset" 不为空时: "Dataset" 列的数据集应存在对应的 Sheet		✓
3.4	对于 Content, 当 "Dataset" 不为空时: "Key Variables" 列的变量必须来自于数据集中标 mapping action = x 的变量		✓
3.5	对于 Content, 当 "Dataset" 不为空时: "Dataset" 列的数据集名长度必须小于等于 8		✓
3.6	对于 Content, 当 "Dataset" 不为空时: "Description" 列的值长度必须小于等于 40		✓
3.7	对于 Content, 当 "Dataset" 不为空时: "Description" 列的值不能有单独不成功的 < > ' " () [] 符号		✓

Specification验证



项目列表 SDTM Spec ×

上传SDTM Spec

No.	Rules	ErrorMessage	Result
1.0	instruction 这个 sheet 只是模板说明文件, 最后应该删掉。		✓
2.0	所有的 Sheet 在 Content 中的 "Dataset" 都有对应的数据集, Cover Page, Instruction, CONTENT, Revision History, SUPP_QNAMLIST 以及带有 "metadata"、"detail"、"codelist" 字符 (不区分大小写) 的 Sheet 除外。		✓
3.1	对于 Content, 当 "Dataset" 不为空时: "Active" 列的值应为 "Y" 或空		✓
3.2	对于 Content, 当 "Dataset" 不为空时: 若 "Active" 列的值为 "Y", 则 "Order, Class, Description, Structure, Purpose, program name, Key Variables, Sort Order" 不能为空		✓
3.3	对于 Content, 当 "Dataset" 不为空时: "Dataset" 列的数据集应存在对应的 Sheet		✓
3.4	对于 Content, 当 "Dataset" 不为空时: "Key Variables" 列的变量必须来自于数据集中标 mapping action = x 的变量		✓
3.5	对于 Content, 当 "Dataset" 不为空时: "Dataset" 列的数据集名长度必须小于等于 8		✓
3.6	对于 Content, 当 "Dataset" 不为空时: "Description" 列的值长度必须小于等于 40		✓
3.7	对于 Content, 当 "Dataset" 不为空时: "Description" 列的值不能有单独不对的 < > ' " () [] 符号		✓



数据转换

数据转换

The screenshot shows the Pinnacle 21 web interface. The browser address bar displays the URL: 192.168.12.43:8112/#/pinnacleTool/projectList. The page title is 'Pinnacle 21工具'. The main content area shows a table of projects with the following data:

合同编号	合同名称	操作	SDTM Version	ADaM Version	Source Encoding	Engine	Language	申办方
CF-0002-IT	IT测试	Pinnacle 21 数据转换 NMPA规则查验 ADaM Spec SDTM Spec 项目配置			UTF-8	NMPA (2304.3)	CN	成都普特药业股份有限公司
CF-0001-IT	IT测试	Pinnacle 21 数据转换 NMPA规则查验 ADaM Spec SDTM Spec 项目配置	3.3	1.1	UTF-8	NMPA (2304.3)	CN	成都普特药业股份有限公司

The '数据转换' (Data Conversion) button in the second row is highlighted with a red circle, and an orange arrow points to it. The page footer shows '共 2 条' (Total 2 items) and '15 条/页' (15 items/page).

数据转换

项目列表 - Pinnacle 21工具

192.168.12.43:8112/#/pinnacleTool/projectList

Pinnacle 21工具

项目列表

Pinnacle 21 Rules数据

合同编号	合同名称
CF-0002-IT	IT测试
CF-0001-IT	IT测试

[CF-0001-IT] 数据转换

To数据类型 请选择要转换的数据类型

输入路径

必填, 路径为: \\192.168.12.40\s\xx\xxx

结果路径

选填, 默认在输入路径下新建文件夹如sas/xpt/json

取消 确定

[CF-0001-IT] 数据转换

To数据类型 请选择要转换的数据类型

输入路径 请选择要转换的数据类型

结果路径

取消 确定转换

共 2 条 15条/页 1 前往 1 页

TM Version	ADaM Version	Source Encoding	Eng
	1.1	UTF-8	NM
	1.1	UTF-8	NM

数据转换

The screenshot displays the Pinnacle 21 web application interface. A modal dialog box is open, titled "[CF-0001-IT] 数据转换". The dialog contains the following information:

- To数据类型:** sas2json
- 输入路径:** \\192.168.12.40\I\studies\cdf\test001\11x-statpr\testdir\dataset\s
- Message:** 数据转换操作
sas2json 操作成功, 成功数量:[42]!
- Buttons:** 确定 (blue), 取消 (grey), 确定转换 (blue)

In the background, a table of project data is visible. The table has columns for "合同编号" (Contract ID), "合同名称" (Contract Name), "TM Version", "ADaM Version", "Source Encoding", and "Eng". The visible rows are:

合同编号	合同名称	TM Version	ADaM Version	Source Encoding	Eng
CF-0002-IT	IT测试		1.1	UTF-8	NM
CF-0001-IT	IT测试		1.1	UTF-8	NM

数据转换

The image shows a Windows File Explorer window with the path `11x-statpr > testdir > dataset > sdtm`. The file list includes `ae.sas7bdat`, `cm.sas7bdat`, `co.sas7bdat`, `dm.sas7bdat`, `ds.sas7bdat`, `dv.sas7bdat`, `ec.sas7bdat`, `eg.sas7bdat`, `ex.sas7bdat`, `ie.sas7bdat`, `lb.sas7bdat`, `mh.sas7bdat`, `pc.sas7bdat`, `pe.sas7bdat`, `pp.sas7bdat`, `pr.sas7bdat`, `qs.sas7bdat`, and `relrec.sas7bdat`. A sub-window shows the `json` folder containing files like `ae.json`, `cm.json`, `co.json`, etc. A Notepad window titled `ae.json - 记事本` displays the following JSON content:

```
{
  "creationDateTime": "2024-08-14T22:42:36",
  "datasetJSONVersion": "1.0.0",
  "clinicalData": {
    "itemGroupData": {
      "IG.AE": {
        "records": 463,
        "name": "AE",
        "label": "Adverse Event",
        "items": [
          {
            "OID": "ITEMGROUPDATASEQ",
            "name": "ITEMGROUPDATASEQ",
            "label": "Record Identifier",
            "type": "integer"
          },
          {
            "OID": "IT.STUDYID",
            "name": "STUDYID",
            "label": "Study Identifier",
            "type": "String"
          }
        ]
      }
    }
  }
}
```



NMPA规则验证

NMPA规则验证

RuleID	Category	Rules	Severity	Instruction
NMPA001	文件格式	数据集必须是要求的文件格式，即SAS Xport V5或更高版本	拒收	
NMPA002	必备文件	必须有数据说明文件	拒收	
NMPA003AD	必备文件	递交中必须包括受试者水平分析数据集(adsl)。	拒收	仅适用于分析数据库
NMPA003RD	必备文件	递交中必须包括人口学(dm)数据集。	拒收	仅适用于原始数据库
NMPA004RD	必备文件	必须有注释病例报告表。	拒收	仅适用于原始数据库
NMPA005	必备文件	应当有数据审阅说明，并在其中说明数据集所用编码(Encoding)。	错误	
NMPA006	文件格式	数据说明文件最好为xml格式，符合define.xml标准(CDISC)。	警示	
NMPA007	文件格式	如递交define.xml格式数据说明文件，对应的可扩展样式表语言(XSL)文件也应一并递交。	警示	
NMPA008	文件大小	数据集文件大小应满足申报资料相关要求。	错误	
NMPA009	必须变量	数据集中必须包含[必须变量]且类型正确。	错误	
NMPA010	必须变量	[必须变量]应当有值。	警示	
NMPA011	数据集记录	数据集应有数据记录。	错误	
NMPA012	数据集/变量名称	数据集名称/变量名称应只包含字母和数字，并且必须以字母开头；数据集名称中的英文字母必须为小写。	错误	
NMPA013RD	数据集名称	原始数据集名称应为二个小写英文字母。	错误	
NMPA013AD	数据集名称	分析数据集名称应以“ad”开头。	错误	
NMPA014	数据集/变量名称	数据集名称/变量名称的最大长度为8个字节。	错误	
NMPA015	数据集/变量标签	数据集/变量应有标签。	警示	
NMPA016	数据集/变量标签	数据集/变量标签应使用中文。	警示	
NMPA017	数据集/变量标签	数据集/变量标签长度应小于或等于40字节。	警示	

NMPA规则验证

读取项目路径下的数据集文件信息

- 数据集信息

- 名称
- 标签

- 变量信息

- 名称
- 标签
- 类型
- 长度
- 值

```
function(CheckPath) {  
  # 数据集文件夹路径  
  anad_path <- paste0(CheckPath, "\\analysis\\adam\\datasets")  
  anlg_path <- paste0(CheckPath, "\\analysis\\legacy\\datasets")  
  tasd_path <- paste0(CheckPath, "\\tabulations\\sdtm")  
  talg_path <- paste0(CheckPath, "\\tabulations\\legacy")  
  
  # 获取文件夹下所有文件  
  anad_files <- list.files(anad_path, recursive = TRUE, full.names = TRUE)  
  anlg_files <- list.files(anlg_path, recursive = TRUE, full.names = TRUE)  
  tasd_files <- list.files(tasd_path, recursive = TRUE, full.names = TRUE)  
  talg_files <- list.files(talg_path, recursive = TRUE, full.names = TRUE)  
  
  ...  
  
  # 读取 SAS 文件  
  sas_data <- read_xpt(path)  
  metadata <- attributes(sas_data)  
  datatable <- metadata$label  
  
  # 获取xpt数据集 - 变量信息 (名称 标签 类型 长度)  
  variable_names <- list()  
  variable_labels <- list()  
  variable_formats <- list()  
  variable_types <- list()  
  variable_lengths <- list()  
  
  dataname <- gsub("/", "\\", path)  
  sasFileName <- basename(path)  
  upperName <- toupper(gsub(".xpt", "", sasFileName))  
  datasize <- (file_info$size / 1024)  
  variable_names <- append(variable_names, names(sas_data))  
  variable_labels <- append(variable_labels, sapply(sas_data, attr, "label"))  
  variable_formats <- append(variable_formats, sapply(sas_data, attr, "format"))  
}
```

NMPA规则验证

读取项目路径下的Define文件信息

- 数据集信息

- 名称
- 标签

- 变量信息

- 名称
- 标签
- 类型
- 长度

- Codelist

```
# 读取XML文件
adam_xml <- paste0(anad_path, "\\define.xml")
sdtm_xml <- paste0(tasd_path, "\\define.xml")
if (file.exists(adam_xml)) {
  # 读取adam -define.xml
  adamxml <- read_xml(adam_xml)
  # 去除命名空间
  xml_ns_strip(adamxml)
  # 查找节点 Item
  adam_group <- xml_find_all(adamxml, "//ItemGroupDef")
  # 提取 数据集 Name
  adamName <- xml_attr(adam_group, "Name")
  # 检查 xml name 是否全部存在
  adam_elements <- setdiff(adamName, adam_xpt_files)
  definename_result <- append(definename_result, adam_elements)
  # 提取 数据集 标签
  adam_desp <- xml_find_first(adam_group, "./Description")
  adam_desptxt <- xml_text(xml_find_first(adam_desp, "./TranslatedText"))
  adamlable_elements <- setdiff(adam_desptxt, adam_lable)
  definelable_result <- append(definelable_result, adamlable_elements)
  # 提取 数据集 变量
  adam_ref <- xml_find_all(adam_group, "./ItemRef")
  adam_ref_oid <- xml_attr(adam_ref, "ItemOID")
  xml_variable <- append(xml_variable, adam_ref_oid)
} else {
  definename_result <- append(definename_result, "adam-define.xml文件不存在")
  definelable_result <- append(definelable_result, "adam-define.xml文件不存在")
  definevariable_result <- append(definevariable_result, "adam-define.xml文件不存在")
}
```


NMPA规则验证

读取项目路径下的RG文件信息

获取数据所用编码

```
adrpdf <- paste0(anad_path, "\\adrg.pdf")
# 判断文件是否存在
if (file.exists(adrpdf)) {
  adrgtext <- pdf_text(adrpdf)
  # 将多行文本合并为一个字符串
  adrgtext <- paste(adrgtext, collapse = "\n")
  # 判断是否存在字符串 "数据所用编码"
  adrg_encoding <- str_detect(adrgtext, "数据所用编码")
  if (adrg_encoding == FALSE) encoding_result <- append(encoding_
} else {
  encoding_result <- append(encoding_result, "adrg.pdf文件不存在")
}
# 读取PDF文件内容
csdrpdf <- paste0(tasd_path, "\\csdrg.pdf")
# 判断文件是否存在
if (file.exists(csdrpdf)) {
  csdrtext <- pdf_text(csdrpdf)
  # 将多行文本合并为一个字符串
  csdrtext <- paste(csdrtext, collapse = "\n")
  # 判断是否存在字符串 "数据所用编码"
  csdr_encoding <- str_detect(csdrtext, "数据所用编码")
  if (csdr_encoding == FALSE) encoding_result <- append(encoding_
} else {
  encoding_result <- append(encoding_result, "csdrg.pdf文件不存在")
}
```

NMPA规则验证

依次根据规则对项目文件夹内文件进行验证，返回验证结果

```
# 分析数据集 判断变量AVISITN、AVISIT是否同时存在
if (grepl("analysis", path)) {
  if ("AVISITN" %in% variable_names || "AVISIT" %in% variable_names) {
    if ("AVISIT" %in% variable_names) {
    } else {
      message <- paste0(sasFileName, "缺少AVISIT")
      avisitn_noexist <- append(avisitn_noexist, message)
    }
    if ("AVISITN" %in% variable_names) {
    } else {
      message <- paste0(sasFileName, "缺少AVISITN")
      avisitn_noexist <- append(avisitn_noexist, message)
    }
  }
}

# AVISITN、AVISIT的值为1:1对应关系。
if ("AVISITN" %in% variable_names && "AVISIT" %in% variable_names) {
  avisitn_data <- sas_data[c("AVISITN", "AVISIT")]
  avisitn_data <- unique(avisitn_data)
  # 使用dplyr进行分组和筛选
  avisitnum_group <- avisitn_data %>%
    group_by(AVISITN) %>%
    filter(n() > 2) %>%
    nrow() > 0

  avisit_group <- avisitn_data %>%
    group_by(AVISIT) %>%
    filter(n() > 2) %>%
    nrow() > 0
  if (avisitnum_group == FALSE && avisit_group == FALSE) {
  } else {
    message <- paste0(sasFileName, "不符")
  }
}
```

NMPA规则验证

The screenshot shows the Pinnacle 21 tool interface. At the top, there is a search bar with the text "Pinnacle 21 Rules数据" and a search input field containing "CF-000". Below the search bar is a table with columns for "合同编号", "合同名称", "操作", "SDTM Version", "ADaM Version", "Source Encoding", and "Eng". The table contains two rows of data. The first row is for contract "CF-0002-IT" and the second row is for "CF-0001-IT". Both rows have a "操作" column with several buttons: "Pinnacle 21", "数据转换", "NMPA规则验证", "ADaM Spec", "SDTM Spec", and "项目配置". The "NMPA规则验证" button in the second row is circled in red. At the bottom of the page, there is a pagination bar showing "共 2 条", "15条/页", and "1" page.

合同编号	合同名称	操作	SDTM Version	ADaM Version	Source Encoding	Eng
CF-0002-IT	IT测试	Pinnacle 21 数据转换 NMPA规则验证 ADaM Spec SDTM Spec 项目配置	3.2	1.1	UTF-8	NM
CF-0001-IT	IT测试	Pinnacle 21 数据转换 NMPA规则验证 ADaM Spec SDTM Spec 项目配置	3.3	1.1	UTF-8	NM

NMPA规则验证

项目列表 NMPA规则查验 ×

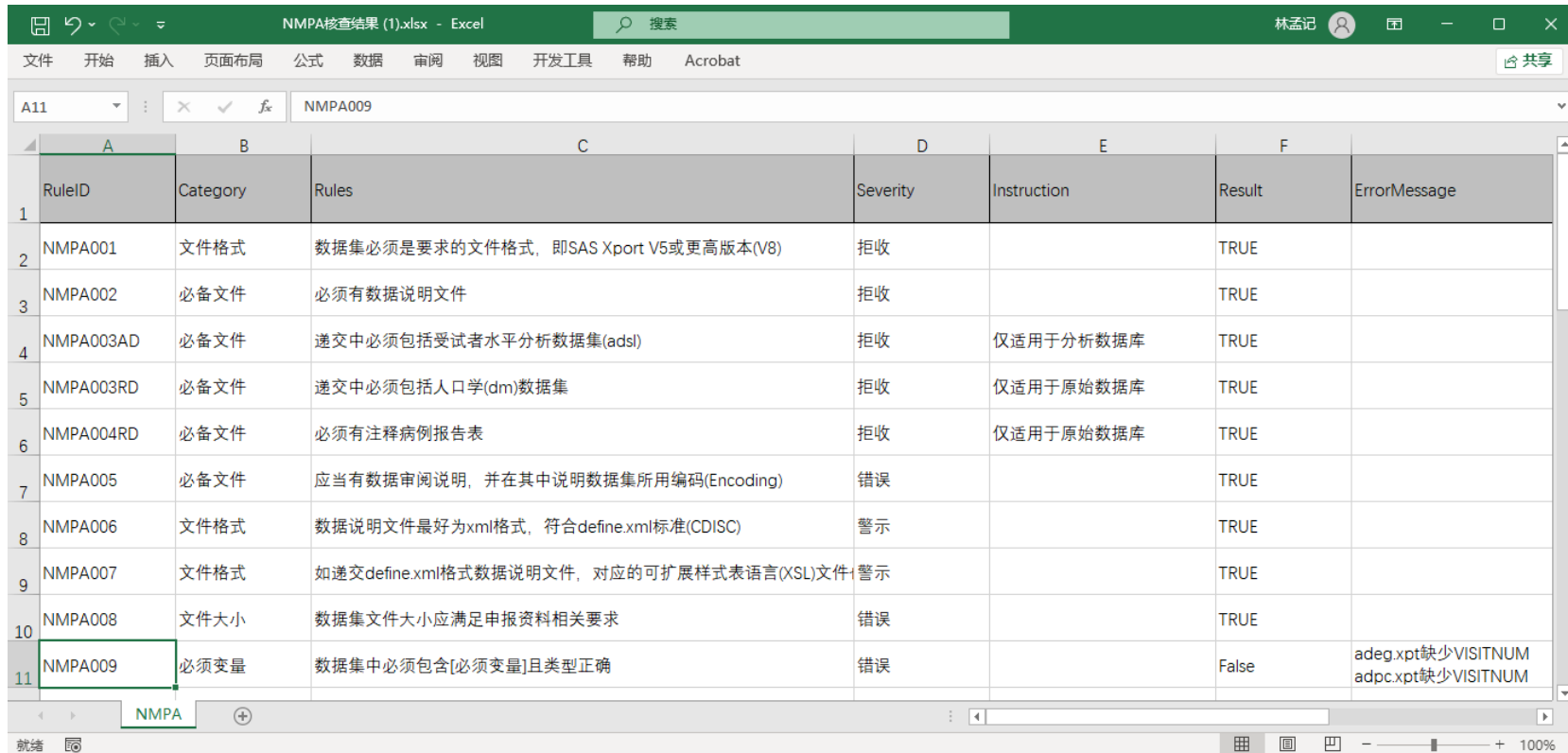
S:\Utility\Tools-dev\p21tools\m5\datasets\teststudy\ 运行 导出到Excel

RuleID	Category	Rules	Severity	Instruction	Describe	Result
暂无数据						

NMPA规则验证

RuleID	Category	Rules	Severity	Instruction	Result	ErrorMessage
NMPA001	文件格式	数据集必须是要求的文件格式, 即SAS Xport V5或更高版本(V8)	拒收		✓	
NMPA002	必备文件	必须有数据说明文件	拒收		✓	
NMPA003AD	必备文件	递交中必须包括受试者水平分析数据集(adsl)	拒收	仅适用于分析数据库	✓	
NMPA003RD	必备文件	递交中必须包括人口学(dm)数据集	拒收	仅适用于原始数据库	✓	
NMPA004RD	必备文件	必须有注释病例报告表	拒收	仅适用于原始数据库	✓	
NMPA005	必备文件	应当有数据审阅说明, 并在其中说明数据集所用编码(Encoding)	错误		✓	
NMPA006	文件格式	数据说明文件最好为xml格式, 符合define.xml标准(CDISC)	警告		✓	
NMPA007	文件格式	如递交define.xml格式数据说明文件, 对应的可扩展样式表语言(XSL)文件也应一并递交	警告		✓	
NMPA008	文件大小	数据集文件大小应满足申报资料相关要求	错误		✓	
NMPA009	必须变量	数据集中必须包含[必须变量]且类型正确	错误		✗	adae.xpt-VISIT-character... 查看更多
NMPA010	必须变量	[必须变量]应当有值	警告		✓	
NMPA011	数据集记录	数据集应有数据记录	错误		✓	
NMPA012	数据集/变量名称	数据集名称/变量名称应只包含字母和数字, 并且必须以字母开头。数据集名称中的英文字母必须为小写	错误		✓	

NMPA规则验证



RuleID	Category	Rules	Severity	Instruction	Result	ErrorMessage
NMPA001	文件格式	数据集必须是要求的文件格式，即SAS Xport V5或更高版本(V8)	拒收		TRUE	
NMPA002	必备文件	必须有数据说明文件	拒收		TRUE	
NMPA003AD	必备文件	递交中必须包括受试者水平分析数据集(adsl)	拒收	仅适用于分析数据库	TRUE	
NMPA003RD	必备文件	递交中必须包括人口学(dm)数据集	拒收	仅适用于原始数据库	TRUE	
NMPA004RD	必备文件	必须有注释病例报告表	拒收	仅适用于原始数据库	TRUE	
NMPA005	必备文件	应当有数据审阅说明，并在其中说明数据集所用编码(Encoding)	错误		TRUE	
NMPA006	文件格式	数据说明文件最好为xml格式，符合define.xml标准(CDISC)	警示		TRUE	
NMPA007	文件格式	如递交define.xml格式数据说明文件，对应的可扩展样式表语言(XSL)文件	警示		TRUE	
NMPA008	文件大小	数据集文件大小应满足申报资料相关要求	错误		TRUE	
NMPA009	必须变量	数据集中必须包含[必须变量]且类型正确	错误		False	adeg.xpt缺少VISITNUM adpc.xpt缺少VISITNUM

How to do in a single study?



导入平台验证, 核查文件结果



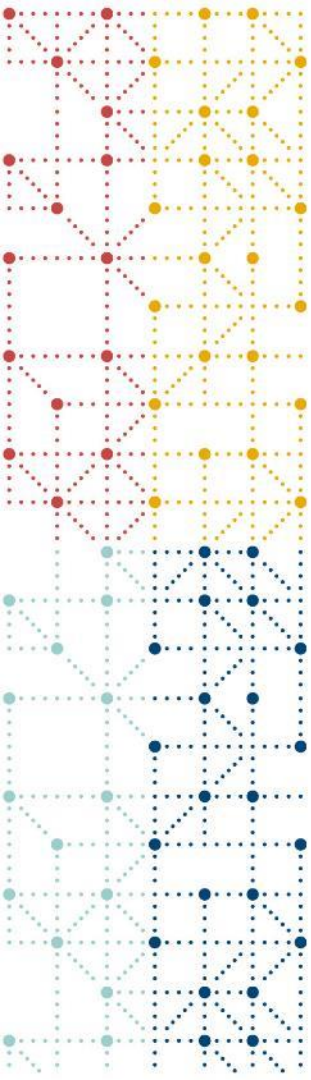
根据CDE数据递交标准, 平台一键完成



上传xpt运行验证报告, 一键导入平台, 版本比对管理,
Issue comments匹配



上传xpt运行验证报告, 核查验证结果



Thank You!

cdisc