

# PINNACLE<sup>21</sup>

## PREPARING CDISC SUBMISSIONS FOR CHINA'S NMPA

Max Kanevsky – Founder and CEO

September 29, 2020



P21



# MAX KANEVSKY

## CEO

- ▶ Founder of Pinnacle 21
- ▶ Validation solution architect at FDA and PMDA
- ▶ SME on FDA JumpStart and coreDF

# AGENDA

- ▶ New NMPA requirements
- ▶ Challenges in current data validation
- ▶ Pinnacle 21 implementation plan
- ▶ First NMPA engine
- ▶ Next steps



P<sup>21</sup>



**P**<sup>21</sup>

# NEW NMPA REQUIREMENTS

# ACCEPTING CDISC SUBMISSIONS

- ▶ Guidance published on July 20, 2020
  - ▶ <http://www.cde.org.cn/news.do?method=viewInfoCommon&id=7a43c3abfde95950>
- ▶ Effective October 1, 2020 for drugs and biologics
- ▶ TBD for Chinese medicine

# KEY REQUIREMENTS

- ▶ Data in Chinese language
- ▶ SAS XPORT v5 format
- ▶ SDTM and ADaM
- ▶ Define (XML or PDF)
- ▶ DM and ADSL are required, other datasets if applicable

## 外文数据库

- 中文为主，与《药品注册管理办法》一致
- 翻译内容的考虑：提高审评效率并考虑申办者工作量
- 翻译最低要求如下：
  - 数据库：数据集标签和变量标签；在临床总结报告等文件中出现的不良事件名称、合并用药名称、病史名称。
  - 数据说明文件：数据库中各数据集的描述/标签和说明；数据集中各变量的描述/标签和衍生过程；涉及疗效指标的取值或编码列表。
  - aCRF：为了收集数据所设计的问题描述；涉及疗效指标问题的取值或编码。
  - 数据审阅说明应为中文。
- **注意事项：申报资料不同文件之间的中文保持一致**

# MINIMUM REQUIREMENTS FOR TRANSLATION

- ▶ Dataset contents
  - ▶ Dataset and variable labels
  - ▶ Adverse events and medical history terms
  - ▶ Concomitant medications
  - ▶ Clinical endpoints
- ▶ Define.xml
- ▶ aCRF
- ▶ Data Reviewer's Guides
- ▶ Keep Chinese translations consistent between datasets and different documents (e.g. clinical summary reports)



## 数据集及变量相关要求

- 数据集名称的最大长度为8个字节
- 变量名称的最大长度为8个字节
- 数据集标签和变量标签应使用**中文**
- 拆分：存储大小，可仅递交拆分后的数据集

**注意事项：**在数据审阅说明中，应详细说明数据集的拆分规则及合并的详细步骤

# DATASETS REQUIREMENTS

- ▶ Dataset and variable names should not exceed 8 bytes
- ▶ Variable labels should not exceed 40 bytes
- ▶ Variable length should not exceed 200 bytes
- ▶ Dataset and variable labels should be in Chinese
- ▶ Depending on encoding a character may use 1-4 bytes
- ▶ When splitting datasets, make sure to provide rules for splitting and merging details in Data Reviewer's Guide



# CHALLENGES IN CURRENT DATA VALIDATION

**P**<sup>21</sup>

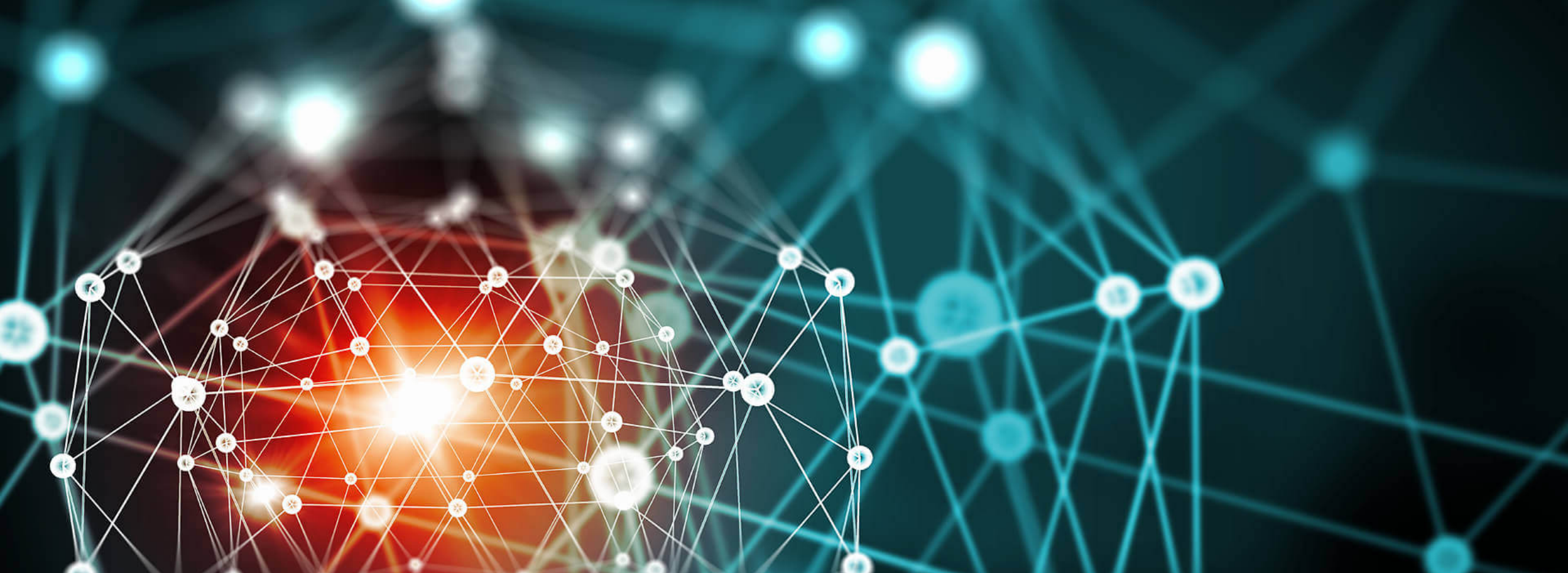


# NEW REQUIREMENTS FOR VALIDATION

- ▶ Support for Chinese character encodings
  - Originally: LATIN1
  - Needed: UTF-8, EUC-CN, GBK, ...
- ▶ Messages and Descriptions in Chinese
- ▶ Subset of rules for NMPA submissions

# CHALLENGES

- ▶ CDISC is still working on Chinese translation of standards and terminology
  - ▶ SDTM-IG 3.2 is available, SDTM-IG 3.3 is **in progress**
  - ▶ ADaM-IG 1.1 is **in progress**
  - ▶ CDISC CT translation for common codelists is **in progress**
- ▶ Translation of other documents also needed
  - ▶ Define.xml stylesheet
  - ▶ PhUSE Data Reviewer's Guide templates
- ▶ NMPA business rules are not finalized

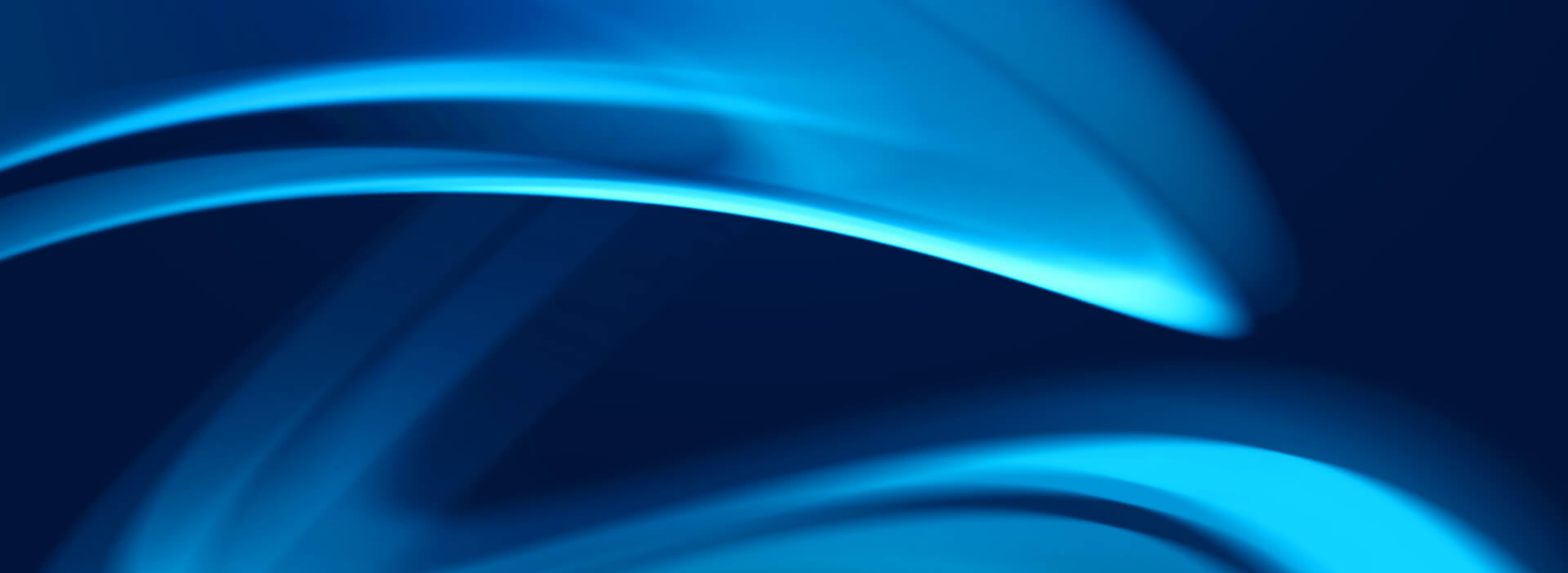


**P**<sup>21</sup>

# PINNACLE 21 IMPLEMENTATION PLAN

# SPEEDING UP IMPLEMENTATION

- ▶ Quick initial release
  - ▶ Basic functionality
- ▶ Monthly updates
  - ▶ New functionality
  - ▶ Adding support for new standards and versions
  - ▶ Tuning rules
- ▶ Close collaboration with all stakeholders
  - ▶ NMPA
  - ▶ CDISC
  - ▶ Users



# FIRST NMPA ENGINE



# AUGUST

- ▶ Engine NMPA 1907.2 released on August 24, 2020
  - ▶ Community: NMPA 1907.3
- ▶ SDTM-IG 3.2 and 3.1.3
  - ▶ 288 rules
  - ▶ Chinese language MedDRA
- ▶ UTF-8 encoding
- ▶ Messages and Descriptions in simplified Chinese
- ▶ Pinnacle 21 Chinese-speaking forum

# SELECTING THE NEW NMPA ENGINE

## Community

## Enterprise

Validator | *check compliance with SDTM, SEND, ADaM, and Define.xml*

✓ Validate Data

Engine	NMPA (1907.3) ▲	Standard
Source Format	NMPA (1907.3)	Configuration
Source Data	FDA (1907.2)	
	FDA Legacy (1903.1)	
	PMDA (1810.3)	
	PMDA Legacy (1511.6)	

Drop your files here to select the engine or

Home > Configure > Studies > My First NMPA submission > New Data Package

### New Data Package

Details

Compatibility

**Validation Engine**

- P21 1907.2
- FDA 1907.2
- FDA 1903.1 (Legacy)
- PMDA 1810.3
- PMDA 1511.6 (Legacy)
- NMPA 1907.2

**Version: 1907.2**  
Chinese-language validation engine. Use to prepare for NMPA submission.



# ISSUE DETAILS WITH CHINESE DATA

Domain	Record	Count	Variables	Values	Pinnacle 21 ID	Publisher ID	Message
AE		3	AEOUT	痊愈/缓解	<a href="#">SD0037</a>		未能在(Outcomes of Events for AE)用户定义的代码列表中找到AEOUT变量的值
AE		2	AEPRESP	是	<a href="#">SD0037</a>		未能在(No Yes Response (Yes only))用户定义的代码列表中找到AEPRESP变量的值
AE		1	AEREL	可能有关	<a href="#">SD0037</a>		未能在(Causality)用户定义的代码列表中找到AEREL变量的值
AE		2	AEREL	有关	<a href="#">SD0037</a>		未能在(Causality)用户定义的代码列表中找到AEREL变量的值
AE		1	AESEV	严重	<a href="#">SD0037</a>		未能在(Severity/Intensity Scale for AE)用户定义的代码列表中找到AESEV变量的值
AE		1	AESEV	中度	<a href="#">SD0037</a>		未能在(Severity/Intensity Scale for AE)用户定义的代码列表中找到AESEV变量的值
AE		1	AESEV	轻度	<a href="#">SD0037</a>		未能在(Severity/Intensity Scale for AE)用户定义的代码列表中找到AESEV变量的值
DM		1	ETHNIC	西班牙或拉丁裔	<a href="#">SD0037</a>		未能在(Ethnic Group)用户定义的代码列表中找到ETHNIC变量的值
DM		5	ETHNIC	非西班牙或拉丁裔	<a href="#">SD0037</a>		未能在(Ethnic Group)用户定义的代码列表中找到ETHNIC变量的值
DM		1	RACE	夏威夷原住民或其他太平洋岛国人	<a href="#">SD0037</a>		未能在(Race)用户定义的代码列表中找到RACE变量的值
DM		2	RACE	白人	<a href="#">SD0037</a>		未能在(Race)用户定义的代码列表中找到RACE变量的值
DM		1	RACE	美国印第安人或阿拉斯加原住民	<a href="#">SD0037</a>		未能在(Race)用户定义的代码列表中找到RACE变量的值
DM		1	RACE	阿拉斯加加人	<a href="#">SD0037</a>		未能在(Race)用户定义的代码列表中找到RACE变量的值
DM		1	RACE	黑人或非洲裔美国人	<a href="#">SD0037</a>		未能在(Race)用户定义的代码列表中找到RACE变量的值
DM			VARIABLE	DTHDTC	<a href="#">SD0057</a>	CG0016	未能找到SDTM规定的期望变量DTHDTC
DM			VARIABLE	DTHFL	<a href="#">SD0057</a>	CG0016	未能找到SDTM规定的期望变量DTHFL
DM			VARIABLE	RFICDTC	<a href="#">SD0057</a>	CG0016	未能找到SDTM规定的期望变量RFICDTC
DM			VARIABLE	RFICTTC	<a href="#">SD0058</a>	CG0013, CG0542, CG0564, CG0565, CG0566, CG0567	变量在数据集中出现, 但是不在SDTM模型中

# TRANSLATED RULE MESSAGES AND DESCRIPTIONS

Pinnacle 21 ID	Publisher ID	Message	Description	Category
DD0101		define.xml文件缺失	监管机构提供的指南表明，递交eCTD时，4号模组（非临床）与5号模组（临床）中每个研究都必须有一个define.xml文件。备注：验收标准及其版本请参考相关监管机构的数据标准目录。	Presence
SD0001	CG0408	数据源中没有记录	域名表中应至少有一个记录。	Presence
SD0002	CG0014, CG0208	必需Variable变量的值为空值（NULL）	任何记录中必需变量（"核心"属性为'Req'的变量）都不能为空（NULL）。	Presence
SD0003	CG0238	ISO 8601 Variable变量值无效	日期/时间变量值（*DTC）必须符合ISO 8601国际标准。	Format
SD0004	CG0413	DOMAIN值不一致	域名缩写（DOMAIN）变量应与数据集的名称一致。	Consistency
SD0005	CG0028	-SEQ变量值重复	在一个域中，每条记录的顺序变量（-SEQ）的数值都必须是唯一的。该域中存在的每个唯一受试者标识符（USUBJID），池标识符（POOLID），或申办者设备标识符（SPDEVID）变量值记录的顺序变量（-SEQ）也必须是唯一的。	Consistency
SD0007		标准单位值不一致	测量、测试、或检查简称（-TESTCD）、类别（-CAT）、样本类型（-SPEC）、以及测试或检查方法（-METHOD）相同的记录，其标准化单位（-STRESU）必须保持一致。	Consistency
SD0008	CG0379	未能在MedDRA词典中找到-DECOD的值	词典衍生术语（-DECOD）变量的值必须用define.xml指定的MedDRA词典中的词典首选术语（Preferred Term）填写（不分大小写）。	Terminology
SD0008C		-DECOD值大小写错误	词典衍生术语（-DECOD）变量必须首字母大写，并使用define.xml指定版本的MedDRA词典中的词典首选术语（Preferred Term）（区分大小写）。	Terminology
SD0012		-STDY在-ENDY之后	事件、接触、或观察的开始日期（-STDY）必须小于或等于事件、接触、或观察的结束日期（-ENDY）。	Limit
SD0013		-STDTC在-ENDTC之后	事件、接触、或观察的开始日期/时间（-STDTC）必须小于或等于事件、接触、或观察的结束日期/时间（-ENDTC）	Limit
SD0014		-DOSE为负值	剂量（-DOSE）值没有缺失时必须大于等于0。	Limit
SD0015		-DUR为负值	事件、接触、或观察的持续时间（-DUR）值没有缺失时必须大于等于0。	Limit
SD0021		结束时间点缺失	一次事件或干预发生时，应填入一个结束时间点变量（比如：除非是否发生的变量（-OCCUR）='N'，或完成状态（-STAT）!='，或持续时间（-DUR）!='，否则事件或干预的结束日期/时间（-ENDTC）、或相对于参照期间的结束（-ENRF）、或相对于参照期间的结束（-ENRTPPT）不应缺失）。	Consistency



**P21**

# NEXT STEPS

# SEPTEMBER

- ▶ SDTM
  - ▶ Additional rules with support for Chinese language CT
- ▶ Define-XML
  - ▶ Translation of Messages and Descriptions
  - ▶ Chinese language stylesheet
- ▶ Additional Encodings
  - ▶ UTF-8
  - ▶ EUC-CN, GB2312
  - ▶ MS936, CP936, GBK
  - ▶ GB18030
  - ▶ Big5

# SELECTING DATASET ENCODING

## Community

Validator | *check compliance with SDTM, SEND, ADaM, and Define.xml*

✓ Validate Data

Engine: NMPA (1907.3) | Standard: SDTM

Source Format: SAS® Transport (XPORT) | Configuration: SDTM-IG 3.2 (NMPA)

Source Encoding: EUC-CN, GB2312 (selected)

Source Data: [Empty field]

Drop your files here to select them  
or  
Browse...

Source Encoding dropdown options: UTF-8, EUC-CN, GB2312, MS936, CP936, GBK, GB18030, Big5

## Enterprise

Home > Validate Data

### Validate Data

Validation Setup

Project: NMPA submission

Study: My First NMPA submission

Data Package: SDTM

Data Encoding: UTF-8 (selected)

Comment: [Empty field] (Optional)

Data Encoding dropdown options: UTF-8, EUC-CN, GB2312, MS936, CP936, GBK, GB18030, Big5

# CHINESE LANGUAGE DEFINE.XML

## 心病心绞痛治

Define-XML文档的创建日期/时间: 2020-09-11T21:11:07-04:00

Define-XML版本: 2.0.0

样式表版本: 2018-11-21

### Annotated Case Report For

▶ 补充文件

▼ 数据集

AE (不良事件)

DM (人口学)

TA (试验组)

▼ 受控术语

▼ 代码清单

实际分组描述

实际分组编码

对研究治疗采取的措施

转归

相关性

严重程度

年龄单位

计划分组描述

计划分组编码

国家

元素描述

时期

元素编码

族群

No Yes Response (No o

种族

性别

No Yes Response (Yes c

▶ 外部词典

标准	SDTM-IG 3.2
研究名称	心病心绞痛治
研究概要	冠心病心绞痛治疗领域
协议名称	002-心心绞
元数据名称	Study 心病心绞痛治 Data Definitions

### 数据集

数据集	描述	类	结构	目的	按键	文献资料	位置
<a href="#">AE</a>	不良事件	事件类	每个受试者每个不良事件一条记录	Tabulation	STUDYID, USUBJID, AEDECOD, AESTDTC		<a href="#">ae.xpt</a>
<a href="#">DM</a>	人口学	特殊目的	每个受试者一条记录	Tabulation	STUDYID, USUBJID		<a href="#">dm.xpt</a>
<a href="#">TA</a>	试验组	试验设计	每个试验组的每个元素对应一条记录	Tabulation	STUDYID, ARMCD, TAETORD		<a href="#">ta.xpt</a>

回到Define-XML文档的[顶部](#)

### TA (试验组) - 试验设计

位置: [ta.xpt](#)

变量	标签/描述	类型	角色	长度或显示格式	受控术语或ISO格式	来源/源/方法/注释
STUDYID	研究标识符	text	标识符	6		Protocol
DOMAIN	域名缩写	text	标识符	2		Assigned
ARMCD	计划分组 编码	text	主题	1	<a href="#">计划分组编码</a> <ul style="list-style-type: none"> <li>"A" = "药 A"</li> <li>"B" = "药 B"</li> <li>"P" = "安慰剂 ≤"</li> <li>"SCRNFAIL" = "筛查失败"</li> </ul>	Assigned
ARM	计划分组 描述	text	同义词修 饰语	13	<a href="#">计划分组描述</a> [6 术语]	Protocol



# OCTOBER

- ▶ ADaM
  - ▶ Initial support for ADaM-IG 1.1
  - ▶ Translated Messages and Descriptions
- ▶ SDTM
  - ▶ Continue to add rules for translated CDISC CT and standard metadata
  - ▶ Add any NMPA specific rules

# FUTURE

- Software localization
- Translate user interface and reports into Chinese and Japanese

# COLLABORATION

- ▶ NMPA
  - ▶ Clarifying business rules
- ▶ CDISC
  - ▶ Translation of standards
- ▶ PhUSE
  - ▶ Translation of Data Reviewer's Guide templates
- ▶ Users
  - ▶ Testing
  - ▶ Request for new functionality
  - ▶ Education

A blurred photograph of a modern office hallway with glass walls and a polished floor. Several people are walking through the hallway, their figures out of focus to convey a sense of movement and activity. The lighting is soft and even.

# QUESTIONS?

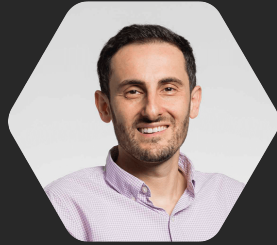
A cluster of decorative hexagonal shapes in the bottom-left corner. Some are solid orange, some are black with orange outlines, and one is a solid black hexagon containing the text 'P21'.

**P**21



THANK YOU ;)

# KEEP IN TOUCH!



**MAX KANEVSKY**

MAX@PINNACLE21.COM

You will receive an email when the slides are posted on our blog:

[www.pinnacle21.com/blog](http://www.pinnacle21.com/blog)

**P21**