2026 CDISC China Interchange

Multi-Regional Data Submission/AI Implementation
多地区数据递交/人工智能应用

Victor Wu吴崇胜
北京迪时咨询有限公司
C3C Honored Chair, Instructor
victor.wu@datascie.com

Victor Wu吴崇胜

- Ph.D
- C3C Honored Chair, CDISC Authorized Instructor (CASH, SDTM & ADaM)
- Associate Director, CMAC Data Management & Intelligentization Working Group
- 20 years' experience in data science

Victor Wu吴崇胜
EVP, Data Science
北京迪时咨询有限公司

Agenda

1. Regulator requirements of data submission
2. Validation Rules
3. AI implementation thoughts

1. Regulator requirements of data submission

CDISC 用于数据递交现状

- 中国NMPA：《药物临床试验数据递交指导原则（试行）》(2020)
  - *2020年10月1日生效*
  - *鼓励申办方参照CDISC标准递交临床试验数据及相关的申报资料*

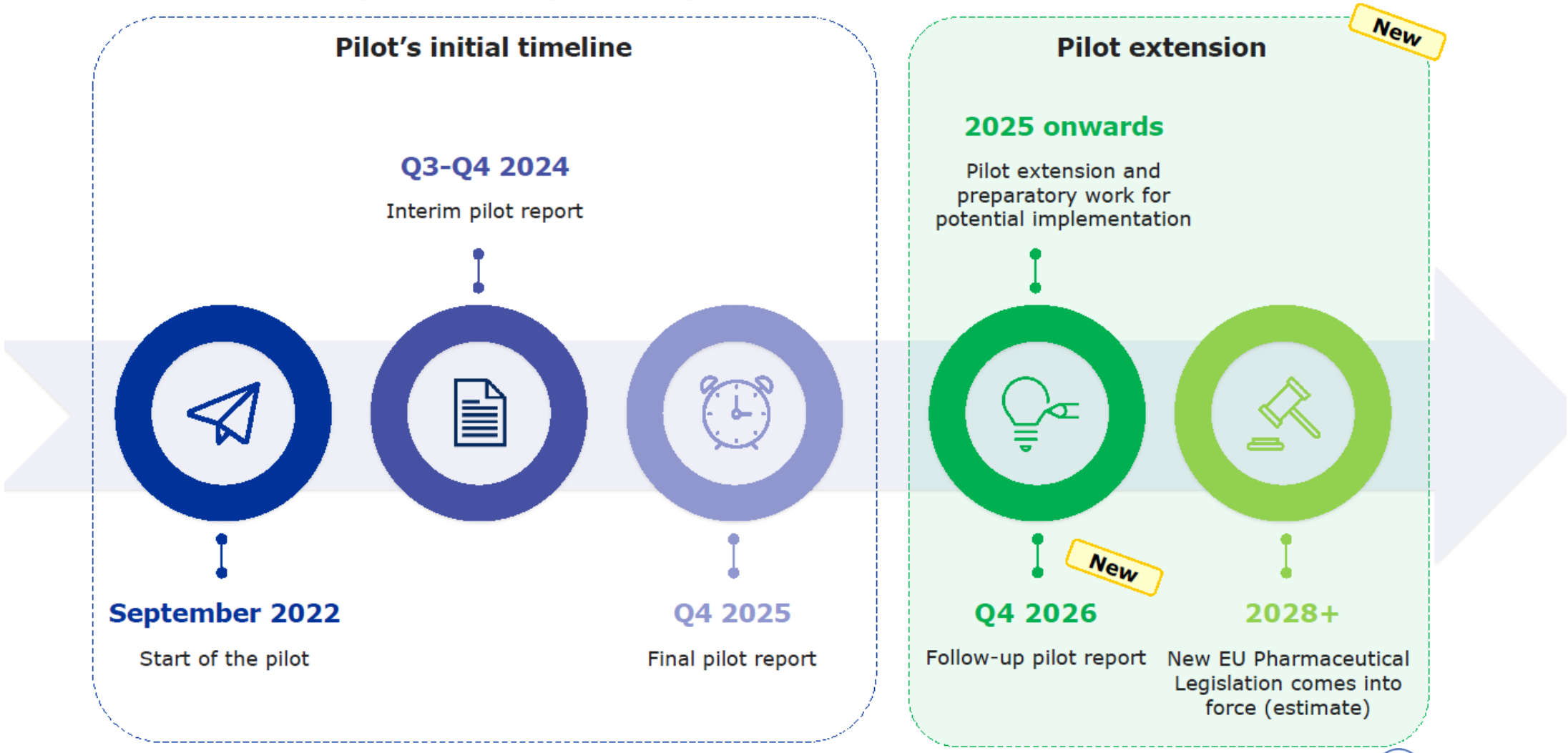
- 美国FDA:
  - 2016Dec 开始的临床试验强制要求遵循CDISC标准
- 日本PMDA:
  - 2020年4月起，强制要求遵循CDISC标准

- 欧盟EMA
  - raw data proof-of-concept pilot for industry
  - 预试验2022：Data standards: Data from clinical trials should comply with Clinical Data Interchange Standards Consortium (CDISC) …

EMA PoC clinical study data pilot

- The PoC pilot has been launched in September 2022, i.e. the date when clinical study data may be submitted for the first regulatory procedure included in the PoC pilot. The PoC pilot has initially been designed with an inclusion target of approximately ten procedures and an estimated duration of two to three years. Building on the insights the pilot has generated so far and captured in the pilot's interim report (link), the pilot is now extended until further notice. This will allow for inclusion of additional procedures, beyond the initially targeted number of ten applications. The extension aims to generate further learnings on the use of clinical study data in support of regulatory assessment, with a focus on addressing remaining knowledge gaps and further validation of the already identified benefits of clinical study data into regulatory decision-making.

Clinical study data pilot phases and timelines



EMA

cdisc

Eftychia-Eirini Psarelli, Data Analytics and Methods Task force, European; Medicines Agency

Information about the clinical study data proof-of-concept pilot for industry

- Data standards: Data from clinical trials should comply with Clinical Data Interchange Standards Consortium (CDISC) Analysis Data Model (ADaM) and Study Data Tabulation Model (SDTM) study data standards. Furthermore, the Define-XML and optionally also the Analysis Results Metadata (ARM for Define-XML) should be submitted. Where a similar submission is planned/submitted to the FDA or PMDA, the same data standards may be followed. There are no specific requirements for non-interventional study data.

Main deliverables and achievements of 2025

.....

Contribution of individual patient data to evidence generation

- New pilot procedures onboarded in 2025
- Follow-up surveys on pilot learnings run throughout 2025
- Network Community on Raw Data membership expanded to include contact points from more NCAs
- Network Advisory Group on Raw Data membership expanded to include CAT, PDCO and NDSG members
- Change management intensified (several presentations to EMA/EMRN and public fora, Industry Group focusing on Clinical Study Data and Network Advisory Group on Raw Data meetings held) over the course of 2025
- Collection of business requirements to support IT development ahead of implementation has been endorsed by EMRN Portofolio Board in July 2025 and is currently underway

CHMP work plan 2026

Danish Medicines Agency

- During 2023 and 2024 the Danish Medicines Agency conducted a pilot study in which we explored how bioequivalence data in CDISC format could be used to assess the bioequivalence studies included in abridged applications.

- **From September 1, 2025 the Danish Medicines Agency will request submission of bioequivalence data in CDISC format**

- The submitted CDISC-data must meet the following:
  - ✓ File format: SAS Transport Format (XPORT) Version 5, Adobe Portable Document Format (pdf), Extensible Markup Language (xml)
  - ✓ Data standards: CDISC SDTM 1.2 or higher is requested. ADaM + programming statements for creation of ADaM data sets and analysis output (optional) including SDTMIG & define.xml

https://laegemiddelstyrelsen.dk/en/licensing/licensing-of-medicines/marketing-authorisation/application-for-marketing-authorisation/

《药物临床试验数据递交指导原则（试行）》 (NMPA 2020)

当前位置：新闻中心>>工作动态>>通知公告>>新闻正文

国家药监局药审中心关于发布《药物临床试验数据递交指导原则（试行）》的通告（2020年第16号）

发布日期：20200720

为规范药品注册申请人递交药物临床试验数据及相关资料，配合新修订的药品注册申报资料要求，提高药品审评效率，药审中心组织制定了《药物临床试验数据递交指导原则（试行）》（见附件）。根据《国家药监局综合司关于印发药品技术指导原则发布程序的通知》（药监综药管〔2020〕9号）要求，经国家药品监督管理局审核同意，现予发布。

化学药品、生物制品自2020年10月1日起实施。中药实施日期按国家药监局发布中药注册分类及申报资料要求的通告中相关规定执行。

特此通告。

国家药品监督管理局药品审评中心

2020年7月20日

附件 1：	《药物临床试验数据递交指导原则（试行）》.pdf

http://www.cde.org.cn/news.do?method=viewInfoCommon&id=7a43c3abfde95950

NMPA递交准备Tips

- 相关文件齐全、格式正确
  - 数据说明
    - xml: define.xml(推荐)
    - pdf: define.pdf (不推荐)
  - aCRF: pdf
    - ⊗ 不是EDC数据集的aCRF
    - 而是递交的xpt原始数据集对应的aCRF
  - 数据审阅说明: pdf （原始数据库/分析数据库）

- 数据文件
  - 文件格式： XPT V5或更高版本
    - XPT v5: SAS (proc copy)
    - XPT v8 (推荐):
    - ⊗ http://support.sas.com/kb/46/944.html
    - *注意：SAS CPORT做的xpt是不符合要求的*
  - Encoding: 需要在《数据审阅说明》中说明
    - UTF8：推荐
    - EUC-CN
    - …

- 必备数据集
  - 原始数据库：人口学数据集 – dm.xpt
    - 命名为dm; 每人一条记录
    - 必须变量：STUDYID/USUBJID/SUBJID
  - 分析数据库：受试者水平分析数据集 – adsl.xpt
    - 命名为adsl; 每人一条记录
    - 包括：人口学、重要的基线特征/分层因素、治疗组、预后因素、重要日期、分析人群划分等信息
- 其他数据集：
  - 必须变量：STUDYID/USUBJID
  - 原始数据库
    - VISITNUM、VISIT一一对应
    - 数据集命名参照CDISC：ae, cm, ds, lb, vs, qs…
  - 分析数据库
    - 通常以“adxxxxxx”命名
    - 分析数据集的命名应尽量与原始数据集保持对应，如：adcm、adae、adlb等

申办方应【在数据审阅说明中】说明所用编码（Encoding, 如utf-8、euc-cn等），以避免所递交的数据集出现乱码的情形

外文数据库 Foreign language database

- What to translate 翻译内容

✓ 数据库：CSR报表中涉及的内容 Values referred in CSR:
  - AE/MH/CM 名称: *coded values*
  - 疗效指标
  - …

✓ 数据说明文件meta: 数据集标签、变量标签；衍生过程说明；涉及疗效指标的取值或编码列表

✓ 数据审阅说明

✓ aCRF：问题描述；涉及疗效指标问题的取值或编码

C3C解读: 相关文件、数据集/变量标签、CSR中图表中涉及的数据内容应为中文.

FDA Study Data Standards Resources

1. FDA Data Standards Catalog
2. FDA Guidances
3. Technical Guides
4. FDA Business and Validator Rules

https://www.fda.gov/industry/fda-data-standards-advisory-board/study-data-standards-resources

Providing Regulatory Submissions in Electronic Format - Standardized Study Data

- Sponsors or applicants are encouraged to use the latest version listed in the Catalog.
- However, when there are multiple versions of a standard listed, sponsors or applicants can select the version to use for their study.

- 同时支持的多个版本，申办方可以自选；推荐用最新的

FDA Data Standards Catalog

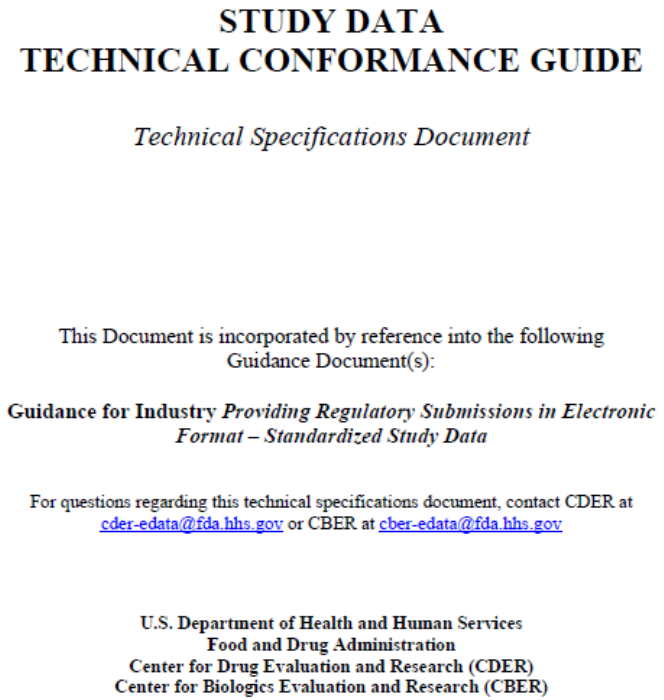
SDO	Property	Related Properties	Date Support Begins [8]	Date Support Ends	Date Requirement Begins [10] [11]	Date Requirement Ends
CDISC	SDTMIGv3.2		08/18/2015	12/13/2023 [12]	03/15/2018 [1] 03/15/2019 [2]	12/13/2023 [12]
CDISC	SDTMIGv3.3		07/07/2020		03/15/2022 [1] 03/15/2023 [2]	
CDISC	SENDIG-ARv1.0		03/26/2024		03/15/2027 [1], [2]	
CDISC	SENDIG-ARv1.0		03/11/2020		03/15/2022 [1] 03/15/2023 [2]	
CDISC	SDTMIGv3.4		12/13/2023 [12]		03/15/2025 [12]	

Property	Related Properties	Date Support Begins [8]	Date Support Ends	Date Requirement Begins [10] [11]	Date Requirement Ends
ADaMv2.1		**Ongoing**		**12/17/2016 [1] 12/17/2017 [2]**	
ADaMIGv1.0		Ongoing	03/15/2019 [1] [12] 03/15/2020 [2] [12]	12/17/2016 [1] 12/17/2017 [2]	03/15/2019 [1] [12] 03/15/2020 [2] [12]
ADaMIGv1.1		10/02/2017		03/15/2019 [1] 03/15/2020 [2]	
ADaMIGv1.3		07/18/2022		03/15/2024	

- https://www.fda.gov/media/185927/download

Technical Guide

- This Study Data Technical Conformance Guide (Guide) provides specifications, recommendations, and general considerations on how to submit standardized study data using FDA-supported data standards located in the **FDA Data Standards Catalog** (Catalog).



STUDY DATA
TECHNICAL CONFORMANCE GUIDE

Technical Specifications Document

This Document is incorporated by reference into the following Guidance Document(s):

Guidance for Industry ***Providing Regulatory Submissions in Electronic Format – Standardized Study Data***

For questions regarding this technical specifications document, contact CDER at cder-edata@fda.hhs.gov or CBER at cber-edata@fda.hhs.gov

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

https://www.fda.gov/media/153632/download

Demographics for Multiple Participations (DC)

FDA Tech conformance guide:

- For subjects with multiple enrollments within a single study, the primary enrollment should be submitted in DM. Additional enrollments should be included in a custom domain with a similar structure to DM. Clarifying statements in the RG would be helpful.
- For subjects with multiple screenings and no subsequent enrollment, include the primary screening in DM with additional screenings in a custom domain with a structure similar to DM.
- For subjects with multiple screenings and subsequent enrollment, include the enrollment in DM with screenings in a custom domain with a structure similar to DM.

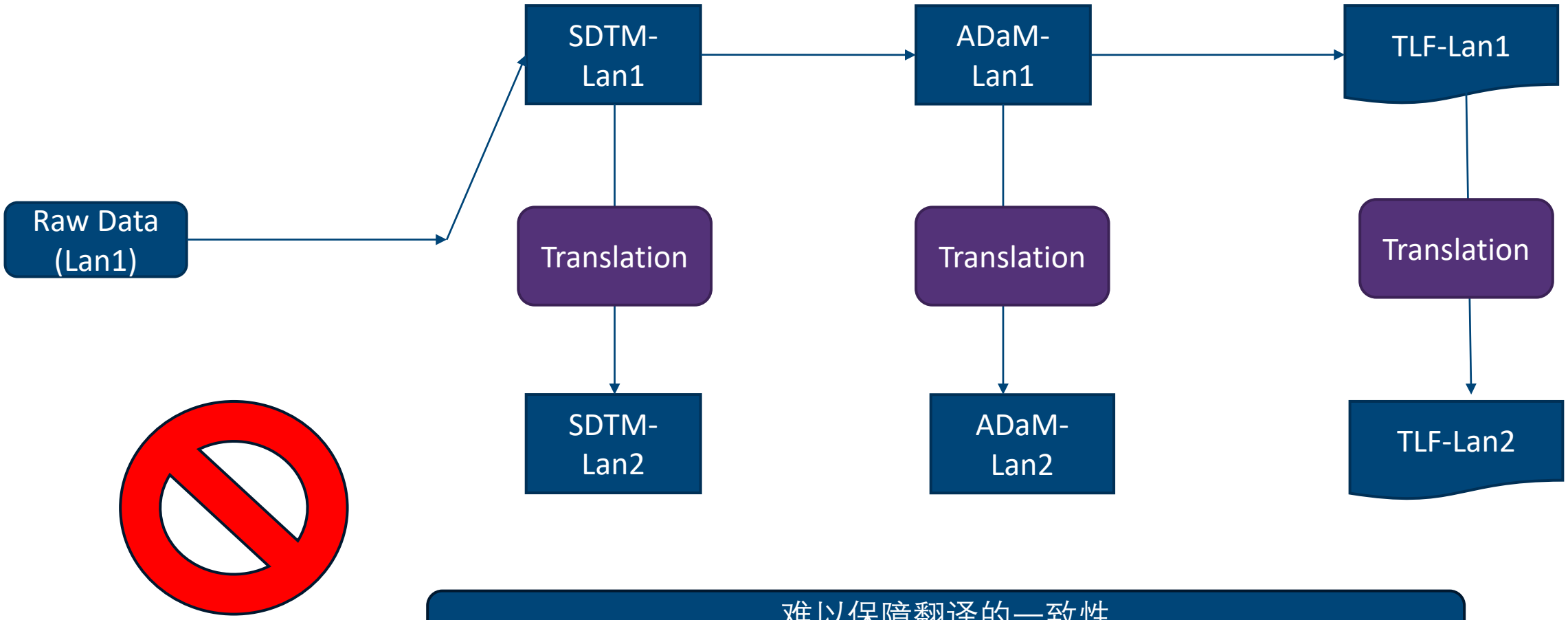
CDISC: SDTM IG V4.0 (not yet released)

- The DC domain is created only if a study includes subjects with multiple participations.
- The DC domain includes every participation, rather than only those considered non-primary, which means the sponsor does not have to decide which participation is "primary".
- Participations are distinguished by values in the SUBJID variable.

LB and LC Domain (Laboratory and Conventional Laboratory)

- The LB domain should contain SI units in LBSTRESU for the SI results in the LBSTRESC and LBSTRESN fields.
- An additional custom domain called LC structured identically to LB should contain conventional units in –STRESU for the results in conventional units in the –STRESC and –STRESN variables.
- If there is no conversion factor then the same value would appear in both LB and LC.

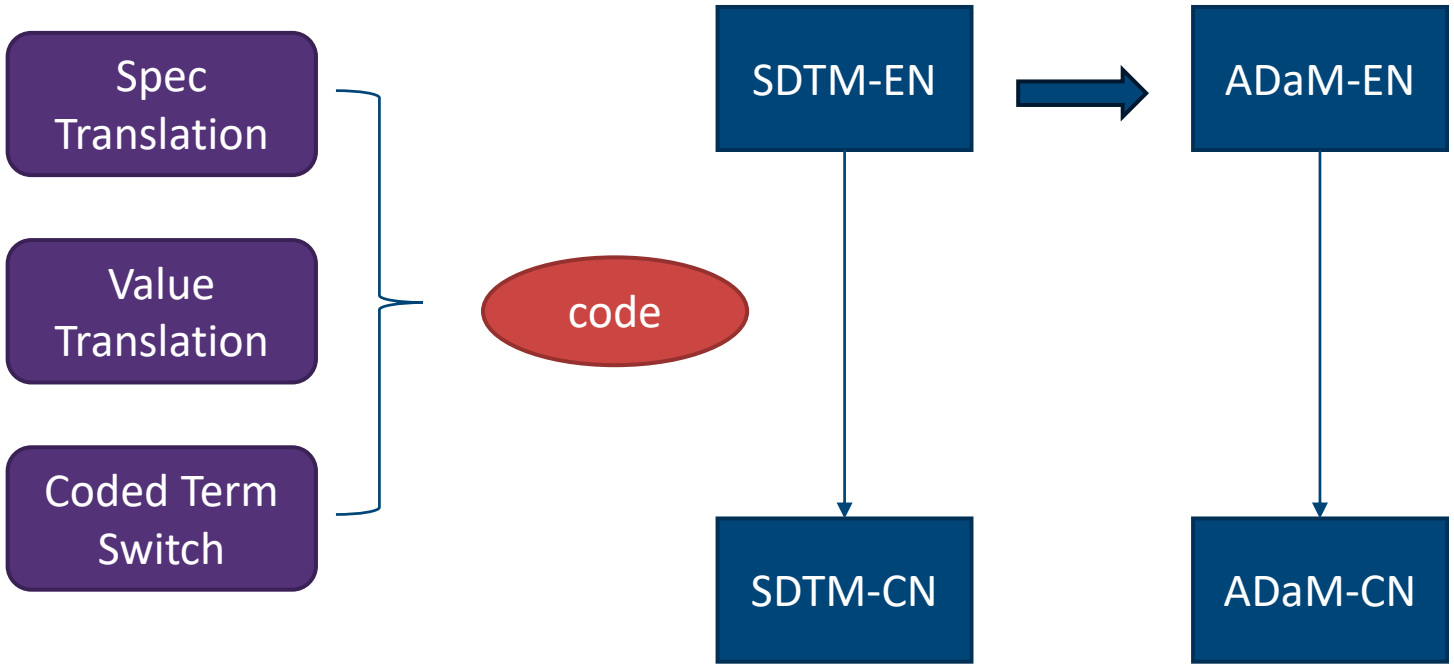
中英双语 – Option 1 分别翻译



难以保障翻译的一致性
标准化部分，只能切换，不适合翻译

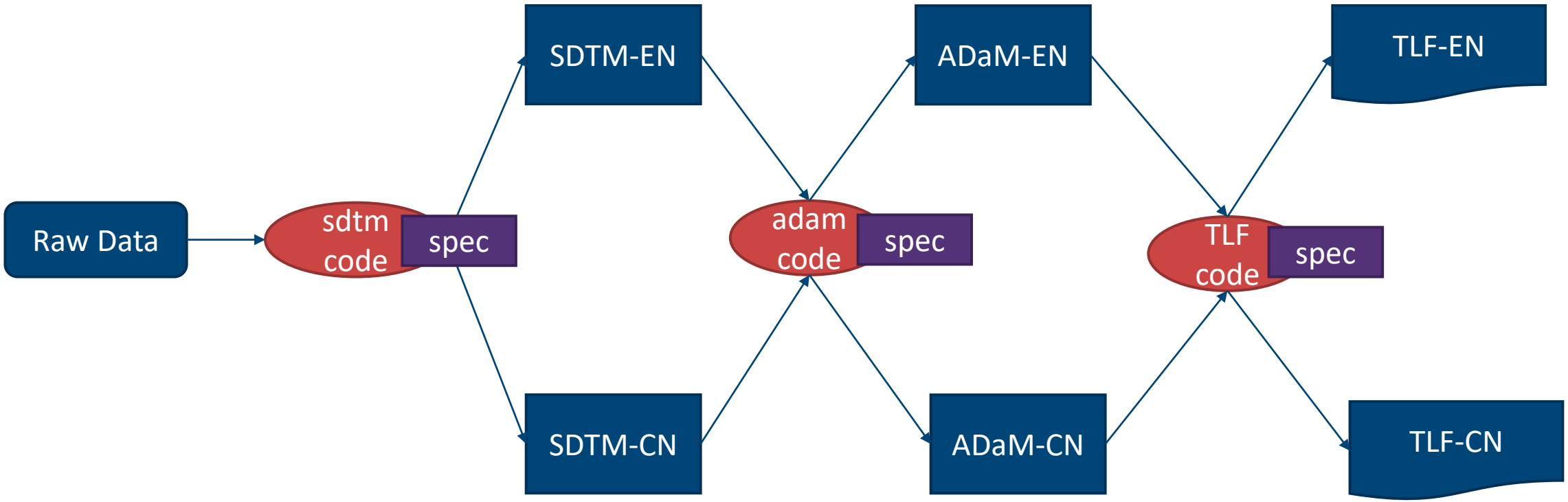
中英双语 – Option 2 统一翻译

Senario: 已有英文递交包



要点：统一翻译，然后再替换到数据集中

中英双语 – Option 3 统一生成



Key point 要点

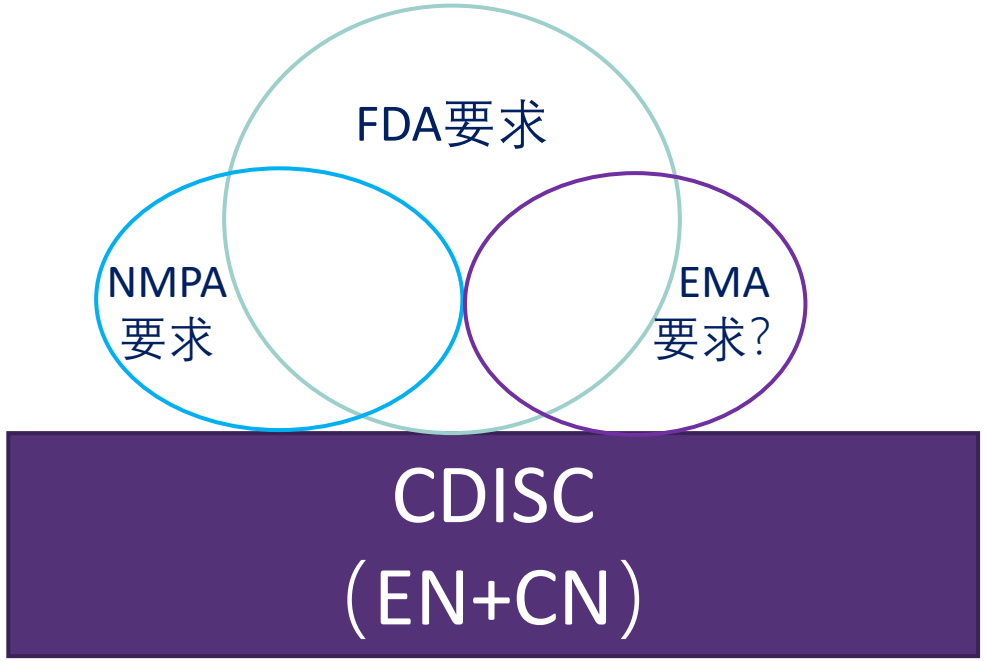
- 提前规划，不同语言的数据，不仅仅是翻译
- 方案、CRF等，推荐先制作英文版本
- 如果只有中国研究中心，主版本建议为中文（数据质量更重要）
- 需要由专业人员来负责不同语言版本，不能靠翻译公司

2. Validation rules

Types of Study Data Validation Rules – FDA example

- Standards Development Organizations (e.g., CDISC) provide rules that assess conformance to its published standards (See www.CDISC.org).
- FDA eCTD Technical Rejection Criteria for Study Data that assess conformance to the standards listed in the Catalog (See section 7.1, section 8.2.2, and Appendix F).
- FDA Business and Validator rules to assess that the data support regulatory review and analysis.

CDISC标准+不同监管机构要求



CDISC 发布的符合性规则

- SDTM and SDTMIG Conformance Rules v2.0
- ADaM Conformance Rules v5.0
- Conformance Rules for Define-XML v2.1

监管机构

- FDA
  - eCTD Technical Rejection
  - FDA Business and Validator rules
  - *Study Data Technical Conformance Guide*
- PMDA Validation Rules
  - SDTM
  - ADaM
  - Define

- **NMPA**
  - 未发布专门的关于符合性规则的指南
  - 符合性规则-从《药物临床试验数据递交指导原则（试行）》中提取

NMPA Rules

- CMAC ICH指导原则落地研究课题

NMPA 数据递交核验规则

RuleID	Category	Rules	Severity	适用于原始数据库	适用于分析数据库	Instruction/Comments
NMPA001	文件格式	数据集文件格式必须符合要求，即 SAS Xport V5 或更高版本(V8)。	重大	Y	Y	
NMPA002	必备文件	必须有数据说明文件(define.xml 或 define.pdf)	重大			
NMPA003AD	必备文件	递交中必须包括必备数据集：adsl(受试者水平分析数据集)。	重大			
NMPA003RD	必备文件	递交中必须包括必备数据集：dm(人口学)。	重大			
NMPA004RD	必备文件	必须有注释病例报告表(acrf.pdf)。	重大			
NMPA005RD	必备文件	应当有数据审阅说明：原始数据库审阅说明(csdrg.pdf)。	错误			
NMPA005AD	必备文件	应当有数据审阅说明：分析数据库审阅说明(adrg.pdf)。	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			

NMPA 数据递交人工核查规则

类别	规则描述	Severity	适用于原始数据库	适用于分析数据库	来源	Instruction/Comments
文件格式	每个 XPT 文件应对应一个数据集，且 XPT 文件名与数据集名称一致。	重大	Y	Y	原文第三(四)节	
语言及编码	数据审阅说明应为中文。	错误	Y	Y	原文第四(三)节	
语言及编码	数据审阅说明应说明数据集所用编码(Encoding)。	错误	Y	Y	原文第二(四)节	
语言及编码	数据说明文件中涉及疗效指标的取值或编码列表应为中文。（数据集中相应内容与数据说明文件一致）	错误	Y	Y	原文第四(三)节	
语言及编码	注释病例报告表中涉及疗效指标问题的取值或编码应为中文。（且与数据说明文件、数据集一致）	错误	Y		原文第四(三)节	
基本规则	数据集不应有重复记录。（根据数据内容及结构进行判断）	错误	Y	Y		
基本规则	原始数据库中的缺失数据不应进行填补。	错误	Y		原文第二(一)节	

3. AI implementation thoughts

从业者心态

AI 与标准化

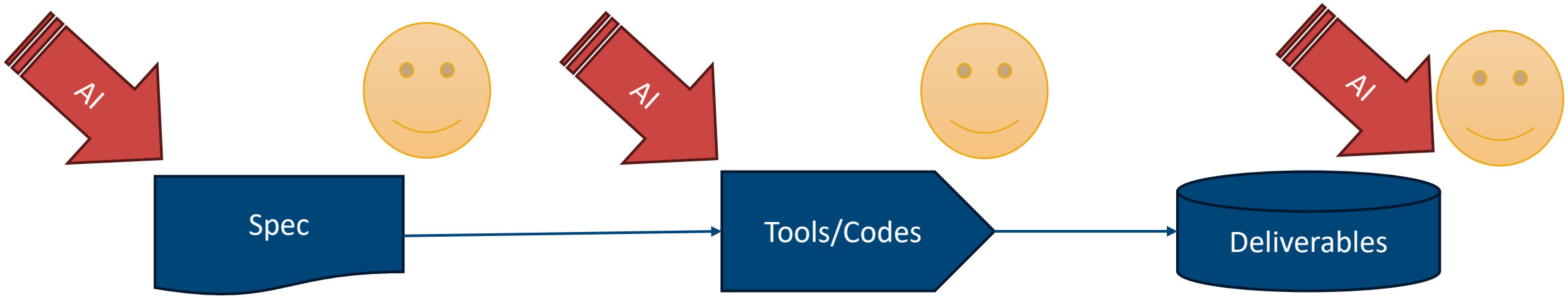
- 为AI提供高质量、可理解的“训练语料”
  - ✓ AI模型，尤其是大语言模型，其训练效果高度依赖于数据的质量和结构。标准化的数据（如CDISC标准）将原始、混乱的临床数据转化为结构统一、定义清晰的格式，这直接提升了AI的训练效率和准确性。

- AI加速标准的建立与执行
  - ✓ 自动化数据映射
  - ✓ 生成标准数据集
  - ✓ 自动生成元数据

职业规划

- 专业领域知识的不可替代性
- 学习一门编程语言
  - 不懂代码，并不可能开发出合格的工具
- AI提高了我们的学习效率，而不是替我们懂了
- AI与普通工具并没有太大的区别，最终负责任，决定结果好坏的，还是使用的人

重要的是如何落地，仅有想法不够



Human in the loop, human takes responsibility

迷幻剂 vs 御剑术

Questions?