

# Highlights of Define-XML 2.0 Implementation and FAQs

Chao Wang

April 12, 2019



# Disclaimer

- The purpose of this webinar is to provide examples of implementation and should not be considered official recommendations by the standards development teams or CDISC unless otherwise stated in the presentation.
- This webinar is not considered to be an authorized CDISC course, is not developed or delivered under CDISC procedures, and should not replace a published CDISC IG or UG. Please refer to the latest published standards documents for the most authoritative implementation information.

# Agenda

- Define-XML & Analysis Result Metadata
- How to create Define-XML
- Define-XML Inspection
- Implementation in China Studies
- FAQs & Panel Discussion

# Define-XML

- Transmits metadata that describes any tabular dataset structure.
- Define-XML is required by the FDA and PMDA for every study in each electronic submission to inform the regulators which datasets, variables, controlled terms, and other specified metadata were used.

## CDISC01

[Annotated Case Report](#)  
[Supplemental Document](#)  
[Datasets](#)  
[Controlled Terminology](#)  
[Methods](#)

Expand all VLM

Collapse all VLM

**Standard** SDTM-IG 3.1.2  
**Study Name** CDISC01  
**Study Description** CDISC Test Study  
**Protocol Name** CDISC01  
**Metadata Name** Study CDISC01, Data Definitions  
**Metadata Description** Study CDISC01, Data Definitions

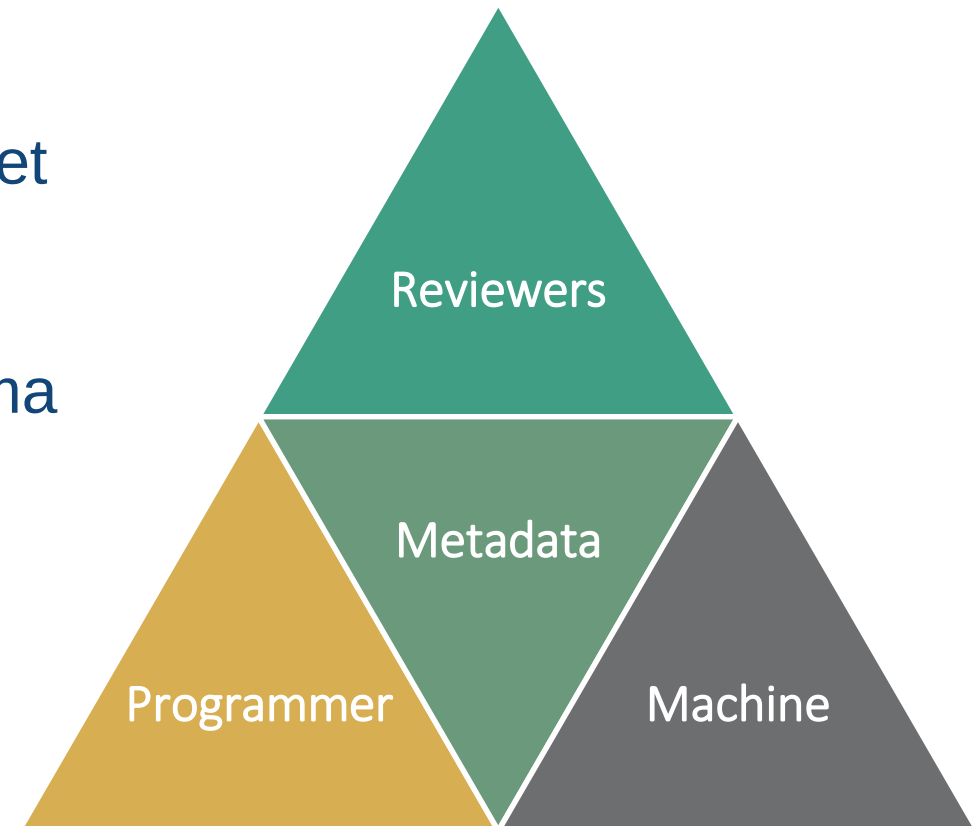
Date/Time of Define-XML document generation: 2013-03-03T17:04:44  
 Define-XML version: 2.0.0  
 Stylesheet version: 2018-11-21

## Datasets

| Dataset            | Description                        | Class        | Structure                              | Purpose    | Keys                    | Documentation | Location                                                                                                   |
|--------------------|------------------------------------|--------------|----------------------------------------|------------|-------------------------|---------------|------------------------------------------------------------------------------------------------------------|
| <a href="#">TA</a> | Trial Arms                         | TRIAL DESIGN | One record per planned Element per Arm | Tabulation | STUDYID, ARMCD, TAETORD |               | <a href="#">ta.xpt</a>  |
| <a href="#">IE</a> | Trial Elements                     | TRIAL DESIGN | One record per planned Element         | Tabulation | STUDYID, ETCOD          |               | <a href="#">ie.xpt</a>  |
| <a href="#">II</a> | Trial Inclusion/Exclusion Criteria | TRIAL DESIGN | One record per I/E criterion           | Tabulation | STUDYID, IETESTCD       |               | <a href="#">ii.xpt</a>  |

# What is Define-XML?

- Data Definition Tables, accompanies FDA/PMDA submissions
- An XML File, present data with stylesheet
- Extension to the ODM foundation schema
- Structured metadata on different levels



# Legacy Define Samples

define.pdf - Adobe Acrobat Pro

File Edit View Window Help

Create 1 / 101 75% Tools Comment Share

Bookmarks

- Datasets for Study XXX
- Code Names
- External Dictionaries
- Supplemental Documents

| Dataset | Description                             | Link to Data |
|---------|-----------------------------------------|--------------|
| AE      | Adverse Event                           | ae.xpt       |
| AEC     | Adverse Event - Coded                   | aec.xpt      |
| ABCM    | AE/Concomitant Medication Questionnaire | abcm.xpt     |
| BAC     | Breath Alcohol Concentration Test       | bac.xpt      |
| CM      | Concomitant Medication                  | cm.xpt       |
| CMC     | Concomitant Medication - Coded          | cmc.xpt      |
| CO      | CO Breath Test                          | co.xpt       |
| DA      | Study Drug Administration               | da.xpt       |
| DEEFILE | DEEFILE Dataset                         | deepline.xpt |
| DM      | Demographic/Informed Consent            | dm.xpt       |
| DSM     | DSM-IV/SCID-9/8                         | dsm.xpt      |
| DU      | Recreational Drug Use History           | du.xpt       |
| DUD     | Recreational Drug Use History Detail    | dud.xpt      |
| EG      | 12-Lead Electrocardiogram               | eg.xpt       |
| HTWT    | Height/Weight/BMI                       | htwt.xpt     |
| IA      | Informed Administration Record          | ia.xpt       |
| IE      | Inclusion/Exclusion                     | ie.xpt       |
| EST     | Dosing and Assessment Phase Inc./Exc.   | est.xpt      |
| LB      | Clinical Laboratory                     | lb.xpt       |
| LBD     | Clinical Laboratory Assay Data          | lbd.xpt      |

Page 1 of 101

Annotated Case Report Form  
 Note to File XXX Hardcoding  
 Note to File XXX Treatment  
 Received Hardcoding  
 Reviews Guide  
 Datasets  
 Value Level Metadata  
 Computational Algorithms  
 Controlled Terminology

| Datasets for Study XXX |                                                    |                 |                                                                                        |            |                                              |                        |
|------------------------|----------------------------------------------------|-----------------|----------------------------------------------------------------------------------------|------------|----------------------------------------------|------------------------|
| Dataset                | Description                                        | Class           | Structure                                                                              | Purpose    | Keys                                         | Location               |
| TA                     | <a href="#">Trial Arms</a>                         | TRIAL DESIGN    | One record per planned Element per Arm                                                 | Tabulation | STUDYID, ARMCD, TAETORD                      | <a href="#">ta.xpt</a> |
| TE                     | <a href="#">Trial Elements</a>                     | TRIAL DESIGN    | One record per planned Element                                                         | Tabulation | STUDYID, ETCD                                | <a href="#">te.xpt</a> |
| TI                     | <a href="#">Trial Inclusion/Exclusion Criteria</a> | TRIAL DESIGN    | One record per I/E criterion                                                           | Tabulation | STUDYID, IETESTCD                            | <a href="#">ti.xpt</a> |
| TS                     | <a href="#">Trial Summary</a>                      | TRIAL DESIGN    | One record per trial summary parameter value                                           | Tabulation | STUDYID, TSPARMCD, TSSEQ                     | <a href="#">ts.xpt</a> |
| TV                     | <a href="#">Trial Visits</a>                       | TRIAL DESIGN    | One record per planned Visit per Arm                                                   | Tabulation | STUDYID, VISITNUM, ARMCD                     | <a href="#">tv.xpt</a> |
| DM                     | <a href="#">Demographics</a>                       | SPECIAL PURPOSE | One record per subject                                                                 | Tabulation | STUDYID, USUBJID                             | <a href="#">dm.xpt</a> |
| SE                     | <a href="#">Subject Elements</a>                   | SPECIAL PURPOSE | One record per actual Element per subject                                              | Tabulation | STUDYID, USUBJID, ETCD                       | <a href="#">se.xpt</a> |
| SV                     | <a href="#">Subject Visits</a>                     | SPECIAL PURPOSE | One record per actual visit per subject                                                | Tabulation | STUDYID, USUBJID, VISITNUM                   | <a href="#">sv.xpt</a> |
| CM                     | <a href="#">Concomitant Medications</a>            | INTERVENTIONS   | One record per recorded medication occurrence or constant-dosing interval per subject. | Tabulation | STUDYID, USUBJID, CMTRT, CMSTDTC             | <a href="#">cm.xpt</a> |
| EX                     | <a href="#">Exposure</a>                           | INTERVENTIONS   | One record per constant dosing interval per subject                                    | Tabulation | STUDYID, USUBJID, EXSTDTC                    | <a href="#">ex.xpt</a> |
| AE                     | <a href="#">Adverse Events</a>                     | EVENTS          | One record per adverse event per time point per subject                                | Tabulation | STUDYID, USUBJID, AEDECOD, AESTDTC           | <a href="#">ae.xpt</a> |
| DS                     | <a href="#">Disposition</a>                        | EVENTS          | One record per disposition event per subject.                                          | Tabulation | STUDYID, USUBJID, DSTERM                     | <a href="#">ds.xpt</a> |
| DV                     | <a href="#">Protocol Deviations</a>                | EVENTS          | One record per protocol deviation per time point per subject                           | Tabulation | STUDYID, USUBJID, DVTERM, DVSTDTC            | <a href="#">dv.xpt</a> |
| MH                     | <a href="#">Medical History</a>                    | EVENTS          | One record per medical history event per subject                                       | Tabulation | STUDYID, USUBJID, MHDECOD                    | <a href="#">mh.xpt</a> |
| DA                     | <a href="#">Drug Accountability</a>                | FINDINGS        | One record per drug accountability finding per time point per subject                  | Tabulation | STUDYID, USUBJID, DATESTCD, DADTC            | <a href="#">da.xpt</a> |
| EG                     | <a href="#">ECG Test Results</a>                   | FINDINGS        | One record per ECG observation per time point per visit per subject                    | Tabulation | STUDYID, USUBJID, EGTESTCD, VISITNUM, EGPNUM | <a href="#">eg.xpt</a> |

# Define-XML - Elements

- Study Metadata
- Dataset
- Variable
- Value/Parameter Level Metadata
- Controlled Terminology
- Methods (Computational Algorithms/Comments/Documentations/Derivations)
- External Supporting Documents (csdrg, adrg, acrf, hard coding memo and etc.)
- Analysis Results Metadata

# Define-XML Elements by Examples

- SDTM Define-XML

- ADaM Define-XML

Demonstration by Files from CDISC

- ADaM Define-XML with Analysis Result Metadata

# Define-XML 2.0 Style Sheet

- Human readable when viewed with a stylesheet.
- Original: 2013-04-24
- Revised version with ARM 1.0: 2015-01-16
- **Newest Version** (published by PhUSE): 2018-11-21
- Download the newest stylesheet through the link below,  
<https://wiki.cdisc.org/display/PUB/Stylesheet+Library>

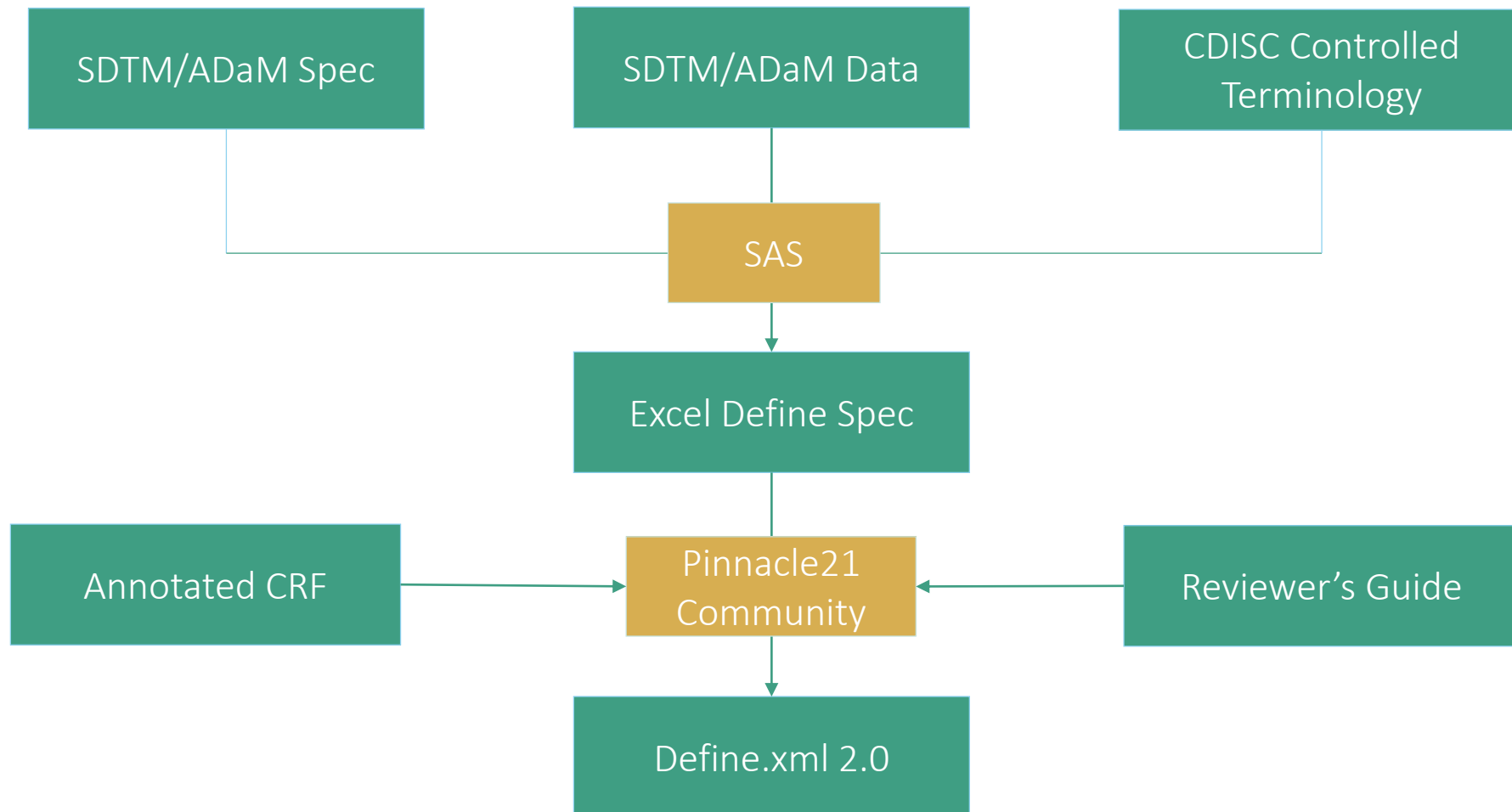
# Key Components

- Class:
  - SDTM and SEND: SPECIAL PURPOSE, FINDINGS, EVENTS, INTERVENTIONS, TRIAL DESIGN, RELATIONSHIP
  - ADaM: SUBJECT LEVEL ANALYSIS DATASET, BASIC DATA STRUCTURE, ADAM OTHER
- Keys
- Length: optimization?
- Data Type: integer, float, text, datetime, date, durationDatetime, parialDatetime
- Controlled Terms/ISO Format
  - NCI Code/Extension
  - External Dictionary
- Origin: CRF, Derived, Assigned, Protocol, eDT, Predecessor

# How to Create Define-XML

- Programming from Scratch
- Tools:
  - SAS Clinical Toolbox
  - Pinnacle 21 Community/Enterprise
- XML Editor Tools:
  - WYSIWYG Editors (by Dmitry Kolosov and Sergey Krivtcov)
  - <http://defineeditor.com/>

# Define-XML Creation Process



# Input Document

sdtrn.xlsx - Excel Chao Wang

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

Clipboard Font Alignment Number Styles Cells

Calibri 11 A A Wrap Text Merge & Center General Normal Bad Good Neutral Calculation Check Cell Explanatory ... Input

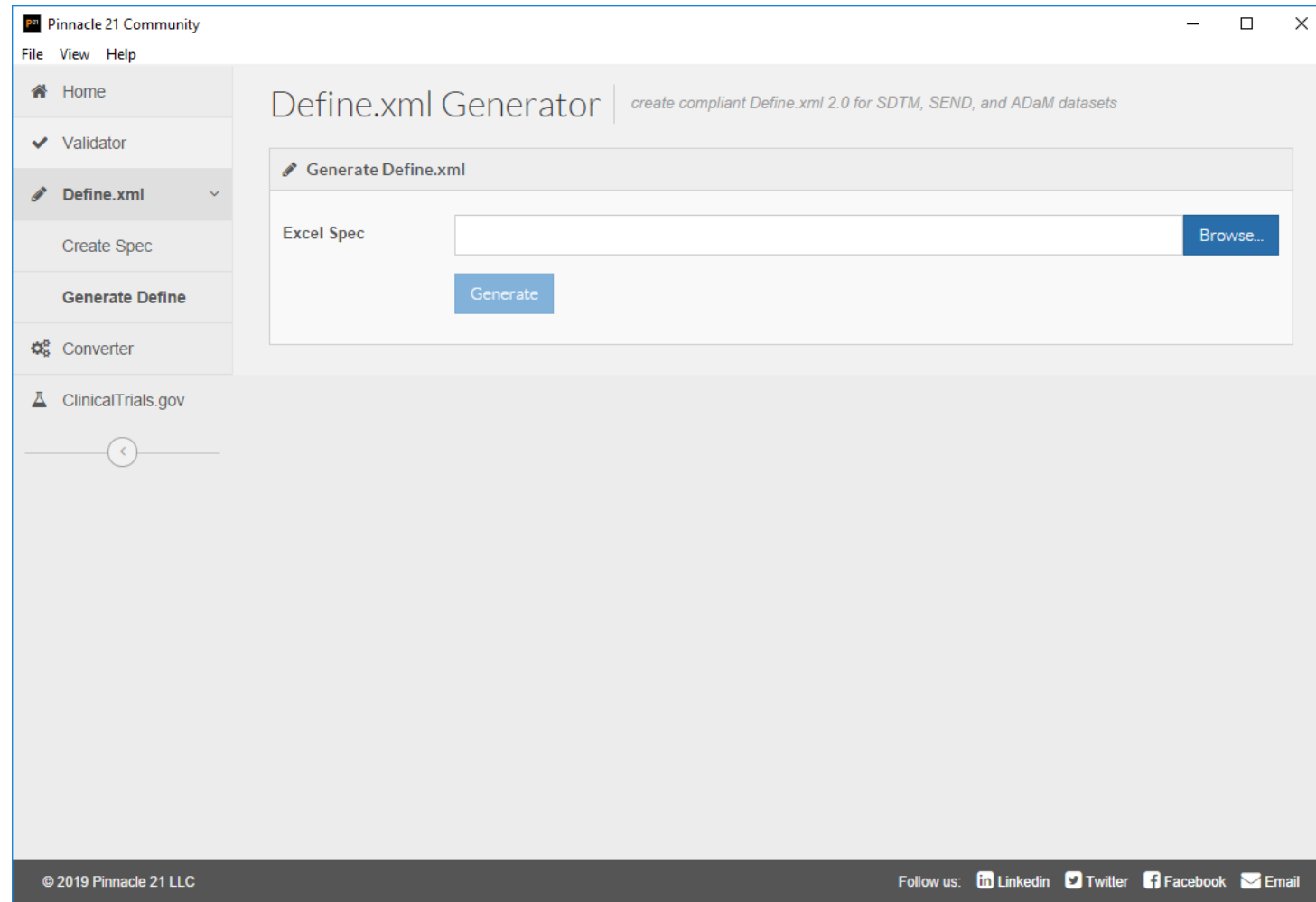
AutoSum Fill Clear

Dataset

|    | A       | B                            | C               | D                                                | E          | F                               | G         | H              | I         |
|----|---------|------------------------------|-----------------|--------------------------------------------------|------------|---------------------------------|-----------|----------------|-----------|
| 1  | Dataset | Description                  | Class           | Structure                                        | Purpose    | Key Variables                   | Repeating | Reference Data | Comment   |
| 2  | AE      | Adverse Events               | EVENTS          | One record per adverse event per subject         | Tabulation | STUDYID,USUBJID,AEDECOD,AESTDTC | Yes       | No             |           |
| 3  | CM      | Concomitant Medications      | INTERVENTIONS   | One record per recorded medication or procedure  | Tabulation | STUDYID,USUBJID,CMSTDTC,CMENDTC | Yes       | No             |           |
| 4  | DA      | Drug Accountability          | FINDINGS        | One record per drug accountability finding       | Tabulation | STUDYID,USUBJID,DATESTCD,DADTCD | Yes       | No             |           |
| 5  | DM      | Demographics                 | SPECIAL PURPOSE | One record per subject                           | Tabulation | STUDYID,USUBJID                 | No        | No             | DOMAIN.DM |
| 6  | DS      | Disposition                  | EVENTS          | One record per disposition status or procedure   | Tabulation | STUDYID,USUBJID,DSSTDY,DSSTDTC  | Yes       | No             |           |
| 7  | EG      | ECG Test Results             | FINDINGS        | One record per ECG observation per visit         | Tabulation | STUDYID,USUBJID,EGTESTCD,EGDTCD | Yes       | No             |           |
| 8  | EX      | Exposure                     | INTERVENTIONS   | One record per constant dosing interval          | Tabulation | STUDYID,USUBJID,EXSTDTC,EXENDTC | Yes       | No             |           |
| 9  | IE      | Inclusion/Exclusion Criteria | FINDINGS        | One record per inclusion/exclusion criteria      | Tabulation | STUDYID,USUBJID,IETESTCD        | Yes       | No             |           |
| 10 | LB      | Laboratory Tests Results     | FINDINGS        | One record per analyte per visit per subject     | Tabulation | STUDYID,USUBJID,LBCAT,LBMETHCD  | Yes       | No             |           |
| 11 | MH      | Medical History              | EVENTS          | One record per medical history event per subject | Tabulation | STUDYID,USUBJID,MHCAT,MHTERM    | Yes       | No             |           |
| 12 | PE      | Physical Examination         | FINDINGS        | One record per body system or abnormal finding   | Tabulation | STUDYID,USUBJID,PETESTCD,PEDTCD | Yes       | No             |           |
| 13 | QSCG    | Questionnaire-QSCG           | FINDINGS        | One record per questionnaire per question        | Tabulation | STUDYID,USUBJID,QSCAT,QSTESTCD  | Yes       | No             | DOMAIN.QS |
| 14 | QSCS    | Questionnaire-QSCS           | FINDINGS        | One record per questionnaire per question        | Tabulation | STUDYID,USUBJID,QSCAT,QSTESTCD  | Yes       | No             | DOMAIN.QS |
| 15 | QSM     | Questionnaire-QSM            | FINDINGS        | One record per questionnaire per question        | Tabulation | STUDYID,USUBJID,QSCAT,QSTESTCD  | Yes       | No             | DOMAIN.QS |
| 16 | RELREC  | Related Records              | RELATIONSHIP    | One record per related record, group of records  | Tabulation | STUDYID,RDOMAIN,USUBJID,IDVATCD | Yes       | No             |           |
| 17 | SC      | Subject Characteristics      | FINDINGS        | One record per characteristic per subject        | Tabulation | STUDYID,USUBJID,SCTESTCD        | No        | No             |           |
| 18 | SE      | Subject Elements             | SPECIAL PURPOSE | One record per actual Element per subject        | Tabulation | STUDYID,USUBJID,SESTDTC,SEENDTC | Yes       | No             |           |

Study Datasets Variables ValueLevel WhereClauses CodeLists Dictionaries Methods Comments Documents

# Demonstration



# How to validate Define-XML?

- ODM Schema
- CDISC Standards
- P21

```

24 <xs:attributeGroup name="MetaDataVersionAttributeExtension">
25   <xs:attributeGroup ref="MetaDataVersionAttributeExtension"/>
26   <xs:attribute ref="def:DefineVersion" use="required" fixed="2.0.0"/>
27   <xs:attribute ref="def:StandardName" use="required"/>
28   <xs:attribute ref="def:StandardVersion" use="required"/>
29 </xs:attributeGroup>
30
31 <xs:group name="MetaDataVersionPreIncludeElementExtension">
32   <xs:sequence>
33     <xs:group ref="MetaDataVersionPreIncludeElementExtension"/>
34     <xs:element ref="def:AnnotatedCRF" minOccurs="0" maxOccurs="1"/>
35     <xs:element ref="def:SupplementalDoc" minOccurs="0" maxOccurs="1"/>
36     <!-- moved to end of MetaDataVersion -->
37     <!-- <xs:element ref="def:leaf" minOccurs="0" maxOccurs="unbounded"/> -->
38     <!-- deprecated: replaced by odm:MethodDef -->
39     <!-- <xs:element ref="def:ComputationMethod" minOccurs="0" maxOccurs="unbounded"/> -->
40     <xs:element ref="def:ValueListDef" minOccurs="0" maxOccurs="unbounded"/>
41     <xs:element ref="def:WhereClauseDef" minOccurs="0" maxOccurs="unbounded"/>
42   </xs:sequence>
43 </xs:group>
44
45 <xs:group name="MetaDataVersionElementExtension">
46   <xs:sequence>
47     <xs:group ref="MetaDataVersionElementExtension"/>
48     <xs:element ref="def:CommentDef" minOccurs="0" maxOccurs="unbounded"/>
49     <xs:element ref="def:leaf" minOccurs="0" maxOccurs="unbounded"/>
50   </xs:sequence>
51 </xs:group>
52

```

# Checking (1)

Date/Time of Define-XML document generation: 2013-03-03T17:04:44

Define-XML version: 2.0.0

Stylesheet version: 2018-11-21

**Standard** SDTM-IG 3.1.2

**Study Name** CDISC01

**Study Description** CDISC Test Study

**Protocol Name** CDISC01

**Metadata Name** Study CDISC01, Data Definitions

**Metadata Description** Study CDISC01, Data Definitions

| Domain Abbreviation | Trial Summary Parameter Short Name | Trial Summary Parameter            | Parameter Value |
|---------------------|------------------------------------|------------------------------------|-----------------|
| TS                  | INDIC                              | Trial Disease/Condition Indication | XXXXXXXX        |
| TS                  | SPONSOR                            | Clinical Study Sponsor             | XXXXXXXX        |
| TS                  | SSTDTC                             | Study Start Date                   | XXXX-XX-XX      |
| TS                  | TITLE                              | Trial Title                        | XXXXXXXX        |

| Data Exchange Standard | Exchange Format | Standards Development Organization (SDO) | Supported Version | Supported Implementation Guide Version | FDA Center(s) | Date Support Begins (MM/DD/YYYY) | Date Support Ends (MM/DD/YYYY)   | Date Requirement Begins (MM/DD/YYYY) | Date Requirement Ends |
|------------------------|-----------------|------------------------------------------|-------------------|----------------------------------------|---------------|----------------------------------|----------------------------------|--------------------------------------|-----------------------|
| SDTM                   | XPT             | CDISC                                    | 1.2               | 3.1.2                                  | CDER, CBER    | 10/30/2009                       | 03/15/2019 [1]<br>03/15/2020 [2] | 12/17/2016 [1]<br>12/17/2017 [2]     |                       |

## Checking (2)

- Dataset order should in logical grouping order or alphabetic order
- All datasets should have observations
- All datasets should have correct description
- Check all the internal and external links, to ensure the target locations/files are correct
- Check the allowance value for Origin column. If multiple origins, should broke out in Value Level Metadata
- All the variables mentioned in the method column should have traceability

# Cross Checking

- aCRF Page Number
- Bookmark section has links to external documents
- All the metadata information (MedDRA, WhoDD and etc.) should be consistent across all the Define-XML, Data Reviewer's Guide and other study documents

# Define-XML to Define-PDF

- Converted by Tools
- Output by SAS ODS

# How to Update Define-XML

There are multiple ways:

- Input file – Metadata in Excel Format
- Define.xml directly
- XML Editor

Demonstration

# NMPA Requirement

- 2016年7月27日, 药物临床试验数据管理与统计分析的计划和报告指导原则 (P16)

- 以下报告附件作为关键性文件，应视为统计分析报告不可缺少的内容。
  - (1) 原始数据库、分析数据库及相应的变量说明文件（数据库应为SAS XPORT 传输格式，xpt格式）
  - (2) 受试者分布流程图
  - (3) 随机化方案（含随机分配表）
  - (4) 盲态审核决议
  - (5) 补充正文的统计附图和附表
  - (6) SAS分析代码（必要时）
  - (7) 统计方法的发表文献（必要时）

# China Study Implementation

## SDTM-IG 3.1.2

临床病例报告表-标注版

评审者指南

复杂逻辑

► 域模型数据集

► 变量值元数据

► 受控术语

► 计算方法

► 备注说明

文档生成日期: 2013-03-03T17:04:44

样式表版本: 2013-04-24

### 域模型数据集 CDISC01 (SDTM-IG 3.1.2)

| 数据集    | 描述                        | 分类      | 结构                                                          | 目的  | 关键变量                                                                                 | 地址                         | 备注说明                                                    |
|--------|---------------------------|---------|-------------------------------------------------------------|-----|--------------------------------------------------------------------------------------|----------------------------|---------------------------------------------------------|
| TA     | <a href="#">试验分组数据集</a>   | 试验设计数据集 | 每个planned Element和Arm一条记录                                   | 原始集 | STUDYID, ARMCD, TAETORD                                                              | <a href="#">ta.xpt</a>     |                                                         |
| DM     | <a href="#">人口学数据</a>     | 特殊目的域模型 | 每个受试者一条记录                                                   | 原始集 | STUDYID, USUBJID                                                                     | <a href="#">dm.xpt</a>     | 参考评审者指南<br>Section 2.1 人口学数据部分<br><a href="#">评审者指南</a> |
| CM     | <a href="#">既往和伴随用药</a>   | 干预类     | 每个人每次用药发生或连续服药一条记录                                          | 原始集 | STUDYID, USUBJID, CMSTDTC, CMENDTC, CMCAT, CMTRT, CMDOSTXT, CMDOSU, CMINDC, CMDOSFRQ | <a href="#">cm.xpt</a>     |                                                         |
| AE     | <a href="#">不良事件</a>      | 事件类     | 每个人每次不良反应一条记录                                               | 原始集 | STUDYID, USUBJID, AEDECOD, AESTDTC                                                   | <a href="#">ae.xpt</a>     |                                                         |
| VS     | <a href="#">生命体征</a>      | 发现类     | One record per vital sign measurement per visit per subject | 原始集 | STUDYID, USUBJID, VSTESTCD, VSDTC, VISITNUM, VSPOS                                   | <a href="#">vs.xpt</a>     |                                                         |
| SUPPCM | <a href="#">既往和伴随用药补充</a> | 关系类     | 按照IDVAR, IDVARVAL, QNAM, 每个人一条记录                            | 原始集 | STUDYID, , USUBJID, IDVAR, IDVARVAL, QNAM                                            | <a href="#">suppcm.xpt</a> |                                                         |

[返回define.xml的顶部](#)

# Preparation for China Study

- Chinese Version Stylesheet
- Validation Rule
- Balance Between English vs. Chinese
- Compatibility with NMPA eCTD System
- XPT Datasets with UTF-8 encoding

# Open Discussion

- When to create Define-XML?
- XPT V5, V8 or Dataset-XML?

# Discussion

- Panelist
  - John Wang, AD, dMed
  - Victor Wu, VP, Data Science Express
  - Ruiling Peng, CEO, Improve Quality



# CDISC Public Training in Shanghai, China



| Course                                      | Date      | Time        | Instructor                   | Registration Link        |
|---------------------------------------------|-----------|-------------|------------------------------|--------------------------|
| <a href="#">CDASH Implementation</a>        | 22 Apr    | 09:00-17:00 | <a href="#">Lily Zhao</a>    | <a href="#">Register</a> |
| <a href="#">SDTM Theory and Application</a> | 23-24 Apr | 09:00-17:00 | <a href="#">Zhijun Wei</a>   | <a href="#">Register</a> |
| <a href="#">ODM Implementation</a>          | 25 Apr    | 09:00-17:00 | <a href="#">Ruiling Peng</a> | <a href="#">Register</a> |
| <a href="#">Define-XML</a>                  | 26 Apr    | 09:00-17:00 | <a href="#">Ruiling Peng</a> | <a href="#">Register</a> |
| <a href="#">ADaM Primer</a>                 | 25 Apr    | 13:00-17:00 | <a href="#">Yanli Chang</a>  | <a href="#">Register</a> |
| <a href="#">ADaM Theory and Application</a> | 26 Apr    | 09:00-17:00 | <a href="#">Yanli Chang</a>  | <a href="#">Register</a> |

For questions, please contact [training@cdisc.org](mailto:training@cdisc.org).

