



2023

CHINA

INTERCHANGE

BEIJING | 25-26 AUGUST



Preparing for Data Driven Submissions

李鹏 Peng LI, Statistical Programmer, B&P, Sanofi



Meet the Speaker

李鹏 Peng LI

Title: Preparing for Data Driven Submissions

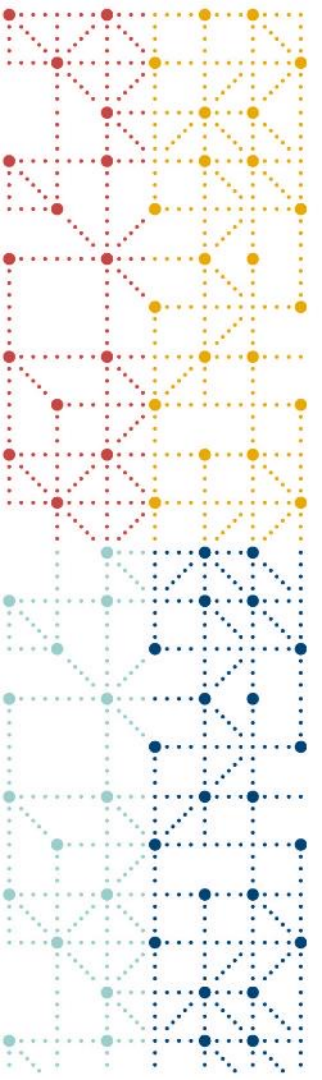
Organization: Sanofi

Graduated from Chongqing Medical University majoring in Applied Statistics, be with Sanofi since 2020 as Statistical Programmer in Statistical Programming department, be experienced in data sharing and innovation development in analytics and reporting tools for submission.



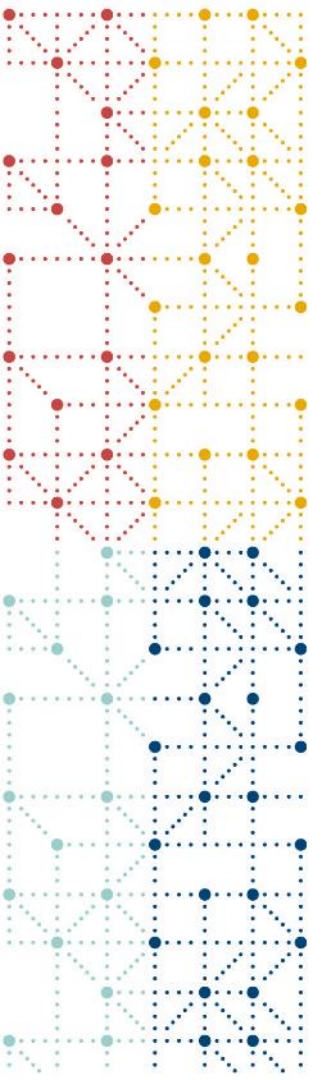
Disclaimer and Disclosures

- *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.*



Agenda

1. Overview
2. Data-Driven Metadata
3. Data-Driven Controlled Terminology
4. Data-Driven SAS Transport File Format



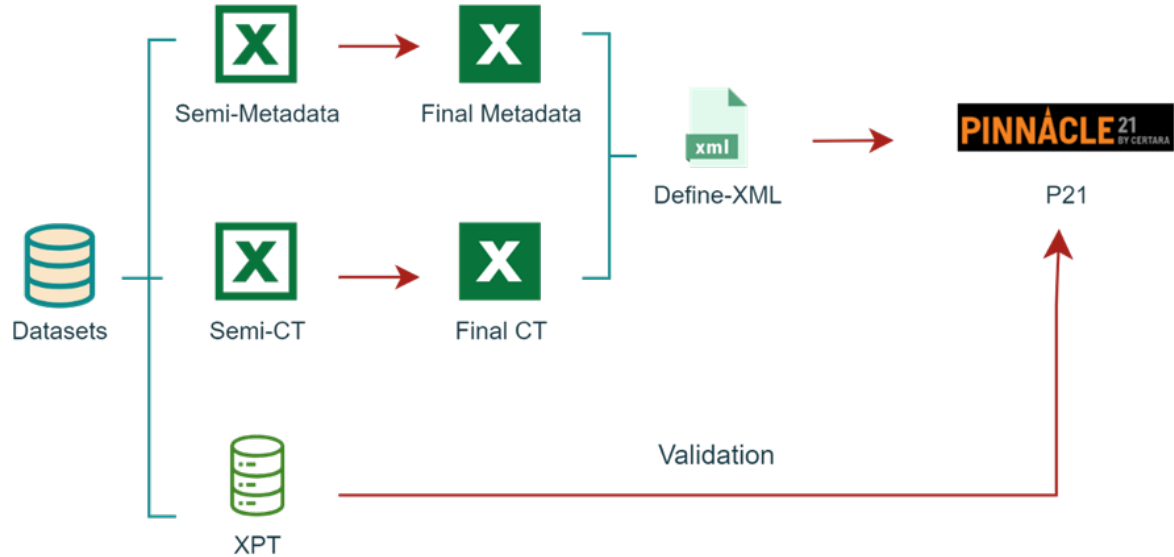
OVERVIEW

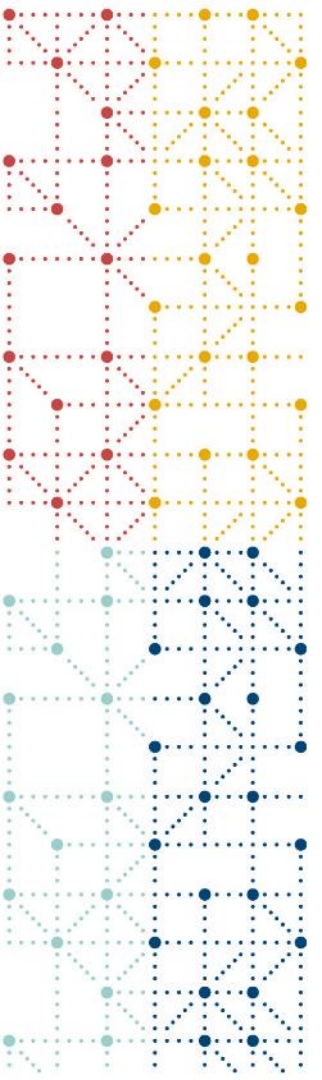


Background

- **HEALTH AUTHORITIES (HA):** Data standards (eCTD, CDISC, etc.) enable Health Authorities to streamline the review process, and improve efficiency and consistency of review decisions.
- **SPONSOR:** Data-driven automated standard submissions also benefit sponsors in terms of efficiency and consistency.

Process of data-driven automated Define-XML





DATA-DRIVEN METADATA



Metadata Structure

- Taking Define-XML 2.0 as an example, we divide metadata into the following 7 dimensions: STUDY INFORMATION, DATASET LEVEL, VARIABLE LEVEL, VALUE LEVEL, METHODS, EXTERNAL DICTIONARY, ADDITIONAL DOCUMENTATION.

Attribute Derivation Rule - Study Information

Attribute	Derivation Rule
Standard	The "SDTIGVER" parameter of the TS domain for SDTM
Study Name	From the STUDYID column of the dataset
Study Description	The "TITLE" parameter of the TS domain
Metadata Name	Derived from study name
Metadata Description	Derived from study name

Attribute Derivation Rule - Dataset Level

Attribute	Derivation Rule
Dataset	Dataset property
Description	Dataset property
Class	Refer to the SDTMIG/ADaMIG
Structure	Refer to the SDTMIG/ADaMIG
Purpose	Refer to the SDTMIG/ADaMIG
Keys	Dataset contents

Attribute Derivation Rule - Variable Level

Attribute	Derivation Rule
Variable	Variable property
Label / Description	Variable property
Type	Calculate variable type from variable actual value
Length or Display Format	Calculate variable length from variable actual value

Attribute Derivation Rule - Value Level

Attribute	Derivation Rule
Variable	Variables that satisfy the preset where Clause rule
Where Condition	Conditions that satisfy the preset where Clause rule
Label / Description	Label variables that satisfy the preset where Clause rule
Type	The type of the actual value under the where Clause
Length or Display Format	The length of the actual value under the where Clause

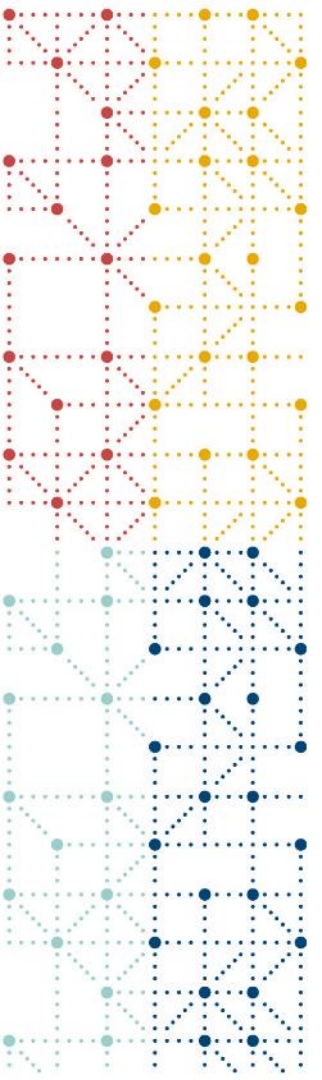
Attribute Derivation Rule - External Dictionary

Attribute	Derivation Rule
Reference Name	The dataset contains variables or parameters corresponding to the dictionary in SDTMIG/ADaMIG
External Dictionary	The dataset contains variables or parameters corresponding to the dictionary in SDTMIG/ADaMIG
Dictionary Version	The version variable corresponding to the dictionary in the dataset



Attribute Derivation Rule – Methods and Additional Documentation

- Methods and Additional Documentation can not be obtained from the dataset at all, and this part of information needs to be manually added to the metadata.



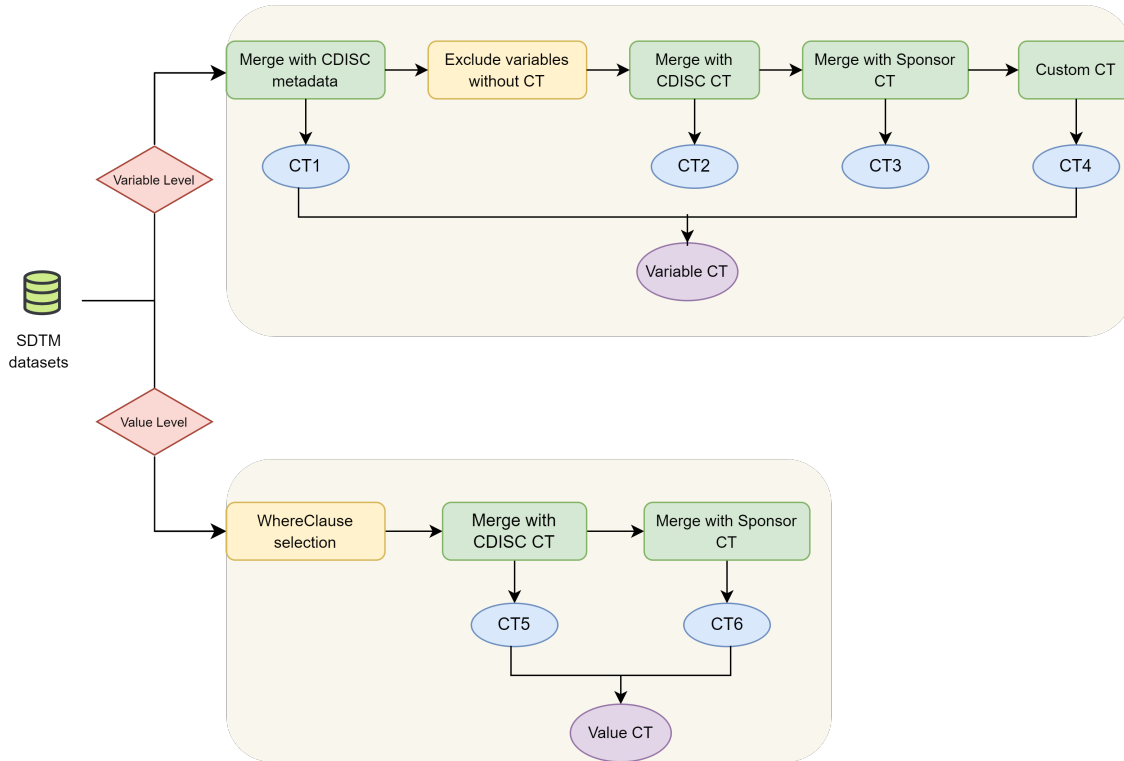
DATA-DRIVEN CONTROLLED TERMINOLOGY



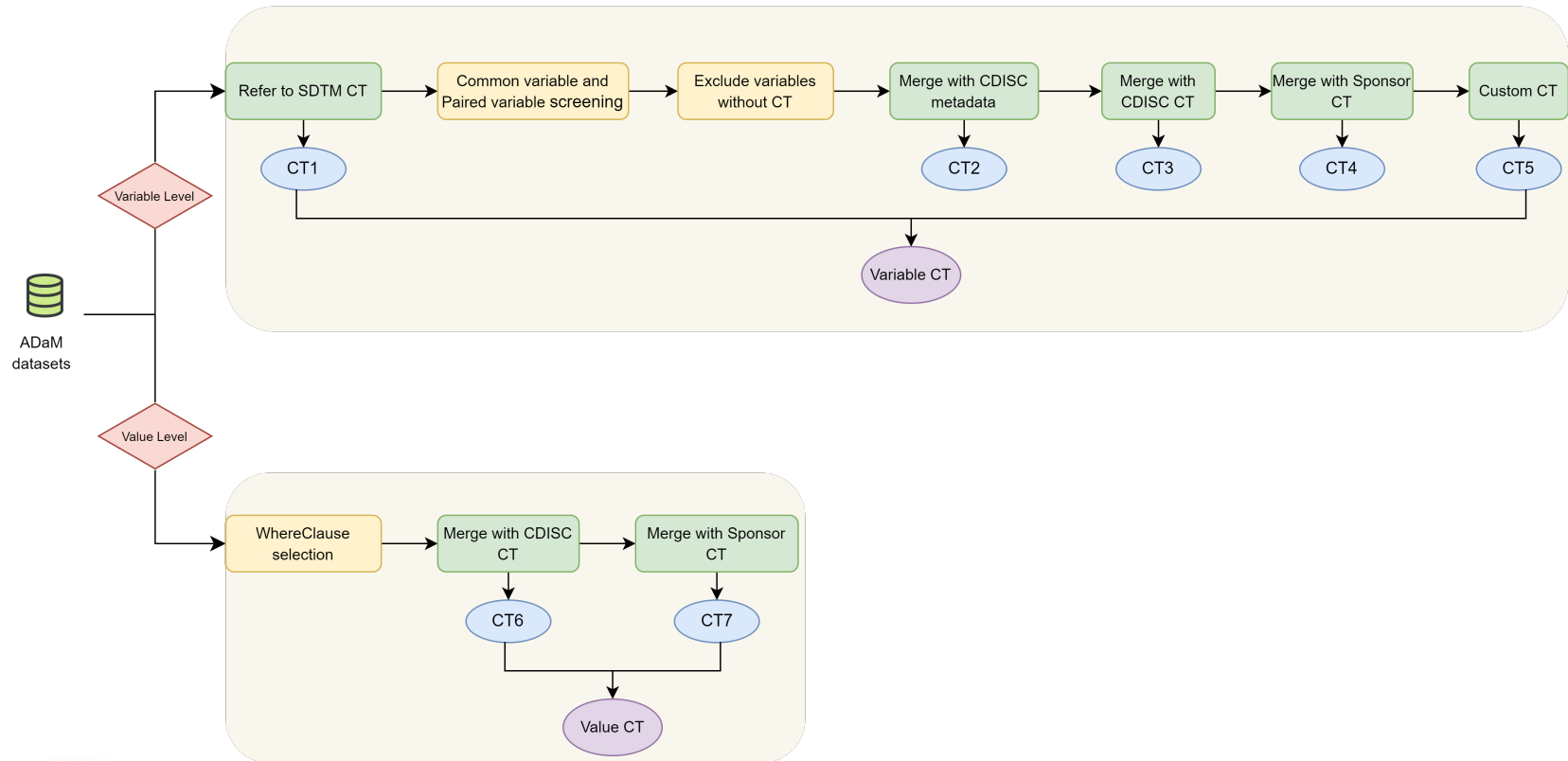
Controlled Terminology Library

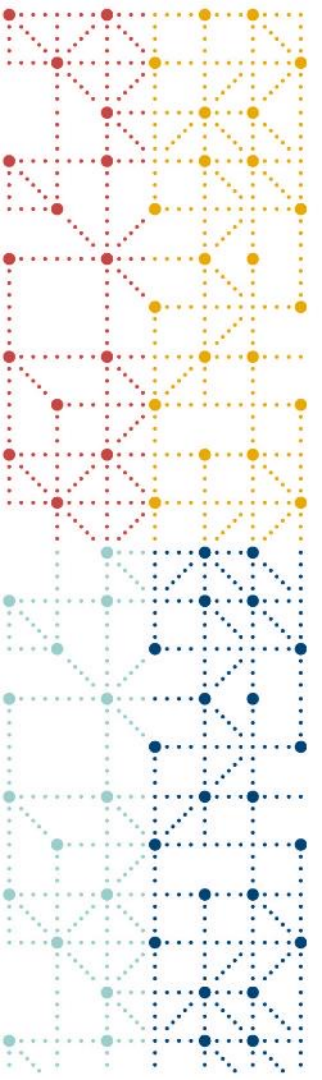
- **CDISC METADATA:** Show the relationships between published Controlled Terminology codelists/terminology subsets and those CDISC variables, Tests, Parameters, and NSVs
- **CDISC CT:** Standard CT for each version of CDISC
- **SPONSOR CT:** Sponsor-defined CT

SDTM Controlled Terminology Workflow



ADaM Controlled Terminology Workflow





DATA-DRIVEN SAS TRANSPORT FILE FORMAT

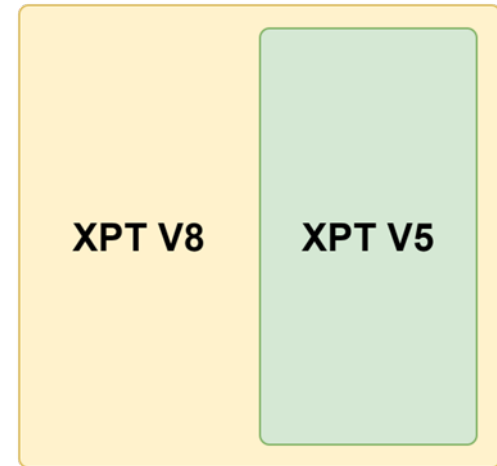


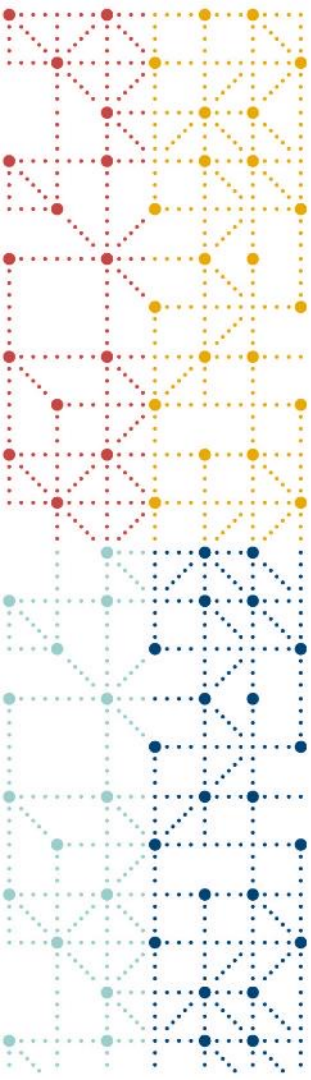
Dataset Resize

- **NMPA:** In latest electronic data declaration draft, it is stipulated that a single XPT file should not exceed 4 GB.
- **FDA:** The Study Data Technical Conformance Guide (TCG) stipulates that if the datasets greater than 5 GB in size, it should be split into smaller datasets no larger than 5 GB.

SAS Transport File Format Version

- **XPT V5:** Variable names may not be longer than 8 characters and are **case insensitive**; Variable labels may not be longer than 40 characters.
- **XPT V8:** Variable names can be up to 32 characters and are **case sensitive**; Variable labels can be up to 256 characters.





Thank You!

For further questions, please contact marco.li@sanofi.com

