

Is your organization ready for CDISC 360 View and Challenges?

Sunil Gupta
Executive Director, Data Sciences
TalentMine

Disclaimer – information shared is work in progress, my views do not necessarily represent CDISC 360 views.



Is your organization ready for CDISC 360 View and Challenges?

Agenda

- ✓ Control Terminology Codelist: CDASH to SDTM Relationships
- ✓ Industry Best Practices
 - ✓ Balance between standards and customization
- ✓ CDISC Mission 360
 - ✓ Requirements
 - ✓ Scope
- ✓ Metadata Standards Process
- ✓ CDISC 360 Work Streams
 - ✓ Work Stream 4 (User Case 1 – End to Start Specifications)
 - ✓ Work Stream 5 (User Case 2 – Start to End Study Metadata)
 - ✓ Work Stream 6 (User Case 3 – Start to End Data Processing)
- ✓ Goal: Apply Metadata to create SDTMs
- ✓ Summary
 - ✓ Levels of Metadata Programming
 - ✓ Extract intelligence information from metadata and macro processing



Control Terminology Updates (Quarterly)

Benefits of staying informed

- ✓ Know about reasons for current CDISC releases
- ✓ Make better decisions on new studies
- ✓ Be in the know for SDTMs, ADaM and Define.xml industry directions

Package 39 Publication Release 2019-09

- ✓ Scope
 - ✓ Brief update based on public review
 - ✓ General, Devices, ECG, General, Labs, Units and Questionnaires
 - ✓ SDTM, SEND, Protocol, Microbiology and Selected Therapeutic Area
- ✓ Key Updates
 - ✓ Remove duplicate terms and old terms
 - ✓ Clean up old terms
 - ✓ Merge related terms
 - ✓ Add new terms that may have been synonym to existing terms before
 - ✓ Review of selected extensible terms
 - ✓ Codetable mapping file updates

Control Terminology Updates (20 Dec 2019)

- ✓ Rules
 - ✓ For all Codelists
 - ✓ For Labs and Units
- ✓ Control Terminology Updates
 - ✓ SDTM Variable to Codelist Map
 - ✓ Changes in Control Terminology
- ✓ CDASH to SDTM Relationships
 - ✓ Many to One Relationship
 - ✓ One to Many Relationship

CDISC Controlled Terminology is maintained and distributed as part of the **NCI Thesaurus** on an NCI File Transfer Protocol (FTP) site and is available for direct download in Excel, text, odm.xml, pdf, html and OWL/RDF formats from the **CDISC Controlled Terminology resources page** on the National Cancer Institute website.

Protocol	CDASH
Glossary	SDTM
SEND	ADaM
Define-XML	

<https://www.cdisc.org/standards/terminology>

Control Terminology Rules

For All Codelists

General Rules

- A CDISC terminology concept is defined as the c-code, synonym(s), and definition.
 - o CDISC terminology concept can be associated with multiple codelists.
 - o CDISC submission value can be the same or different across multiple codelists.
 - o CDISC terminology concept used across multiple codelists must have the same NCI C-code, CDISC synonym(s) and CDISC definition.
- The CDISC submission value, synonym(s), and definition should use the American English spelling.

For Labs and Units

Rules for Laboratory Test Code/Name Codelists

- General
 - o Lab tests where results can be expressed as qualitative, semi-quantitative or quantitative should all have the same test name and code.
 - o Lab terminology includes lab tests for drugs of abuse, e.g., Methamphetamine.
 - o Lab terminology does not include lab tests to perform therapeutic drug level monitoring, e.g., Digoxin.
 - o If an instrument or lab commonly does the calculation and reports the observed and calculated value then the calculated results should have lab terminology, e.g., clearance and rates
 - Conversely, derived values, (i.e., delta, such as change from baseline) that require two or more separate collections or time points will not be included in lab terminology.

Control Terminology Codelist Updates

SDTM Variable to Codelist Map

Variable Name	Variable Label	Type	Codelist	Codelist Long Name	Role	Description
SEX	Sex	Char	(SEX)	Sex	Qualifier Variable	Sex of the subject.
RACE	Race	Char	(RACE)	Race	Qualifier Variable	Race of the subject. Sponsors should refer to "Collection of Race and Ethnicity Data in Clinical Trials" (FDA, September 2005) for guidance regarding the collection of race (http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm126396.pdf)
ETHNIC	Ethnicity	Char	(ETHNIC)	Ethnic Group	Qualifier Variable	The ethnicity of the subject. Sponsors should refer to "Collection of Race and Ethnicity Data in Clinical Trials" (FDA, September 2005) for guidance regarding the collection of ethnicity (http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm126396.pdf)

SDTMIG3.2_Variable_to_Codelist_Codetable.xls

Changes in SDTM Control Terms

How		What		Codelist Name		Change	Original	→	New
Release Date	Request Code	Change Type	NCI Code	CDISC Term Type	CDISC Codelist (Short Name)	CDISC Codelist (Long Name)	Change Summary	Original	New
2019-12-20	CDISC-1780; CDISC-4043	Update	C161625	CDISC Submission Value	BPRS01TC	Brief Psychiatric Rating Scale 1988 Version Clinical Classification Test Code	Update CDISC Submission Value	BPR02TC	BPRS01TC
2019-12-20	CDISC-1780; CDISC-4043	Update	C161741	CDISC Submission Value	BPRS01TC	Brief Psychiatric Rating Scale 1988 Version Clinical Classification Test Code	Update CDISC Submission Value	BPR0201	BPRS0101
2019-12-20	CDISC-1780; CDISC-4043	Update	C161741	CDISC Synonym	BPRS01TC	Brief Psychiatric Rating Scale 1988 Version Clinical Classification Test Code	Add new CDISC Synonym	---	BPRS01-Somatic Concern

SDTM Terminology Changes.xls

CodeTable Mapping Files Define Relationships

Many to One Relationship

1. DM: Collected Race (ASIAN, CHINESE) to ASIAN

C-code (Concept Code)	Race As Collected (RACEC) (codelist code = C128689)		C-code (Concept Code)	Race (RACE) (codelist code = C74457)
C41260	ASIAN		C41260	ASIAN
C43391	CHINESE		C41260	ASIAN

Map Many Row Collected

To One Control Term Value

C-code (Concept Code)	Codelist Name	Race As Collected (RACEC) (codelist code = C128689)	Raw Value		C-code (Concept Code)	Codelist Value	Race (RACE) (codelist code = C74457)
C18237		ALASKA NATIVE			C41259		AMERICAN INDIAN OR ALASKA NATIVE
C43877		AMERICAN INDIAN			C41259		AMERICAN INDIAN OR ALASKA NATIVE
C41259		AMERICAN INDIAN OR ALASKA NATIVE			C41259		AMERICAN INDIAN OR ALASKA NATIVE
C77810		CARIBBEAN INDIAN			C41259		AMERICAN INDIAN OR ALASKA NATIVE
C44953		SOUTH AMERICAN INDIAN			C41259		AMERICAN INDIAN OR ALASKA NATIVE
C44270		YUPIK ESKIMO			C41259		AMERICAN INDIAN OR ALASKA NATIVE
C41260		ASIAN	Group Raw Values		C41260	One	ASIAN
C43671		BANGLADESHI		C41260	Control Term	ASIAN	
C43675		TAIWANESE		C41260	ASIAN		
C43400		THAI		C41260	ASIAN		
C43396		VIETNAMESE		C41260	ASIAN		
C42331		AFRICAN			C16352		BLACK OR AFRICAN AMERICAN
C128937		AFRICAN AMERICAN			C16352		BLACK OR AFRICAN AMERICAN
C41226		AFRICAN CARIBBEAN			C16352		BLACK OR AFRICAN AMERICAN
C43835		ZAIREAN			C16352		BLACK OR AFRICAN AMERICAN
C154874		KHOISAN			C16352		BLACK OR AFRICAN AMERICAN
C43409		MELANESIAN			C41219		NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER

CodeTable Mapping Files Define Relationships

Many to One Relationship

2. VS: SYSBP and DIABP to mmHg

Paired Codelists

C-code (Concept Code)	Vital Signs Test Code (VSTESTCD) (codelist code = C66741)	Vital Signs Test Name (VSTEST) (codelist code = C67153)	C-code (Concept Code)	Units for Vital Signs Results (VSRESU) (codelist code = C66770)	C-code (Concept Code)
C25299	DIABP	Diastolic Blood Pressure	C49670	mmHg	C90480
C25298	SYSBP	Systolic Blood Pressure	C49670	mmHg	C77532

Map Many Row Collected

To One Control Term Value

C-code (Concept Code)	Vital Signs Test Code (VSTESTCD) (codelist code = C66741)	Vital Signs Test Name (VSTEST) (codelist code = C67153)	C-code (Concept Code)	Units for Vital Signs Results (VSRESU) (codelist code = C66770)	C-code (Concept Code)	Position (POSITION) (codelist code = C71148)
C103346	ABSKNF	Abdominal Skinfold Thickness	C28251	mm		
C16358	BMI	Body Mass Index	C49671	kg/m2		
C126083	BMR	Basal Metabolic Rate	C139135	kcal/day		
C126083	BMR	Basal Metabolic Rate	C42549	Watt		
C81298	BODLNTH	Body Length	C49668	cm		
C81298	BODLNTH	Body Length	C48500	in		
C81298	BODLNTH	Body Length	C28251	mm		
C122232	BODYFATM	Body Fat Measurement	C25613	%		
C25157	BSA	Body Surface Area	C42569	m2		
C25299	DIABP	Diastolic Blood Pressure	C49670	mmHg	C77532	DECUBITUS
C25299	DIABP	Diastolic Blood Pressure	C49670	mmHg	C62173	FOWLERS
C25299	DIABP	Diastolic Blood Pressure	C49670	mmHg	C100758	LATERAL DECUBITUS
C25299	DIABP	Diastolic Blood Pressure	C49670	mmHg	C62172	LEFT LATERAL DECUBITUS
C25299	DIABP	Diastolic Blood Pressure	C49670	mmHg	C62165	PRONE

CodeTable Mapping Files Define Relationships

One to Many Relationship

EG: EGTESTCD='MI' to many EGSTRESC values

Paired Codelists

Codelist Name		Codelist Name		Codelist Name	
C-code (Concept Code)	ECG Test Code (EGTESTCD) (codelist code = C71153)	ECG Test Name (EGTEST) (codelist code = C71152)	C-code (Concept Code)	ECG Result (EGSTRESC) (codelist code = C71150)	C-code (Concept Code)
C111280	MI	Myocardial Infarction	C71065	ACUTE ANTERIOR WALL MYOCARDIAL INFARCTION	C71065
C111280	MI	Myocardial Infarction	C102591	ACUTE ANTEROLATERAL WALL MYOCARDIAL INFARCTION	C102591
C111280	MI	Myocardial Infarction	C102592	ACUTE ANTEROSEPTAL WALL MYOCARDIAL INFARCTION	C102592

Map One Row Collected

To Many Control Term Values

Best practices are to create raw and CDISC codelist lookup tables to be used to dynamically create format catalogs from Proc Format to apply PUT() functions to convert raw data values to SDTM control terminology instead of hard coding each raw value.

Industry Best Practices

- ✓ Utility Macros
 - ✓ Proc SQL Dictionary tables to access metadata (datasets, variables, etc.)
 - ✓ Create format catalog from codelist tables to map to SDTM control terms
 - ✓ Scan SAS logs for errors and warnings
 - ✓ Create SAS generated code to run independently
 - ✓ Defensive programming to display user messages
 - ✓ Program index of table, list and graph titles and footnotes to SAS programs
 - ✓ Analysis Results Metadata for 'one-proc away' in SAS programs
 - ✓ Cross-reference SAS source and qc program file date time stamps
 - ✓ Populate define.xml template excel file to create define.xml

- ✓ SDTM/ADaM Macros
 - ✓ Apply PUT() and format catalog to convert raw to SDTM control terms
 - ✓ Apply attributes (Name, Label, Type, Length)
 - ✓ Apply variable and record sort order
 - ✓ Create ISO Dates
 - ✓ Merge XX with SUPPXX



CDISC 360 Mission Requirements

1. Machine-readable standards.
2. Add more meaning to metadata with 'semantic relationships'.
3. Apply and customize all standards directly from metadata files to study specific metadata files.
4. Access and integrate the latest standard files including control terminology. (Application Program Interface, OpenSource account)
5. Metadata driven process for higher-level quality control and customization without manual efforts.
6. SAS generated code for independent creation and submission.
7. GUI interface based on metadata and standards when user input is required such as data mapping, domain shells, ADaM specs and table shells.
8. Proof of concept tests metadata standards with macro-level programming techniques to create deliverables with SAS generated code.
9. Test data is used for proof of concept and is independent between process components
10. Metadata complements each other - ADaM and ARM.

CDISC 360 Mission Scope

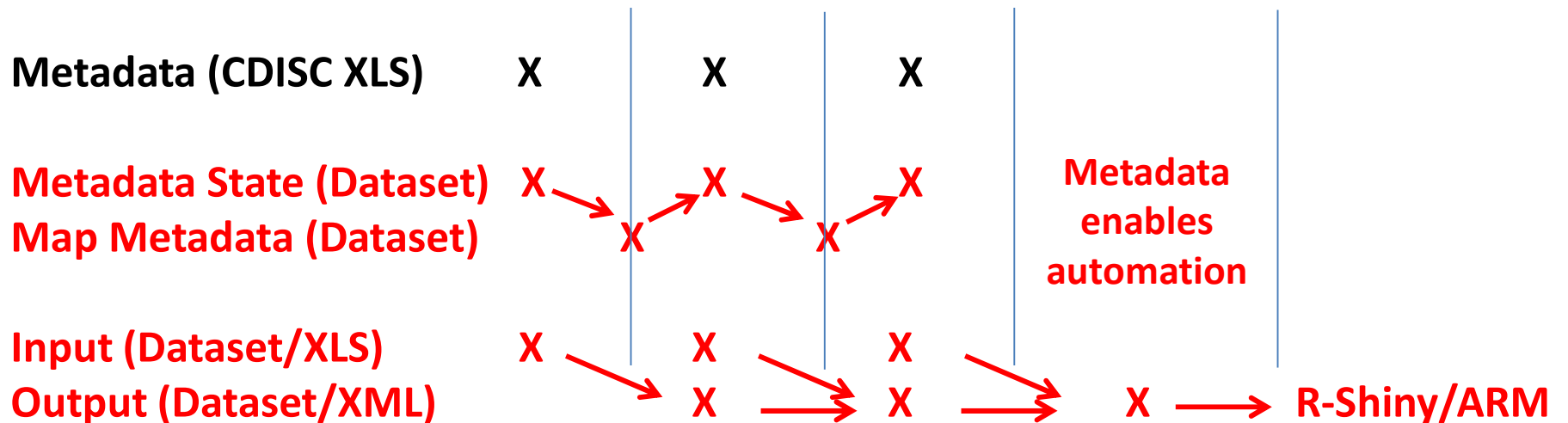
USING THE CDISC TYPE 1 DIABETES TAUG SELECT THE FOLLOWING:

- 1 or 2 statistical end points
 - Analysis of Glycated Hemoglobin
 - Summary of Hypoglycemic episodes
- ~3-4 ADaM datasets
 - ADSL(Subject-Level Analysis Data (ADSL))
 - Hemoglobin A1C Analysis Dataset (HbA1c Analysis Dataset)
 - Hypoglycemic Episodes Analysis Dataset (Hypoglycemic Episodes Analysis Dataset)
 - Hypoglycemic Episodes Summary Dataset (Hypoglycemic Episodes Summary Dataset)
- ~7-8 SDTM datasets
 - DM (Demographics, so support standard variables in ADSL)
 - VS (Vital Signs, for height and weight in ADSL)
 - CM (Concomitant Medications, to support stratification by background treatment, and for treatments of hypoglycemic events)
 - LB (for Hemoglobin A1C data)
 - CE and FACE (for data on hypoglycemic events)
 - EX, ML (for data about meals and study treatments relative to hypoglycemic events)
 - Trial Design datasets (for arms, visit schedule, definition of hypoglycemic events as disease milestones)
- ~15 CDASH CRFs
 - CDASH CRFs needed to support SDTM datasets above. One CRF will support collection of data about hypoglycemic events that will be mapped to multiple SDTM domains.

Evolving CDISC to the Next Decades: The CDISC Proof of Concept, Peter Van Reusel, CDISC, Sam Hume, CDISC

Metadata Standards Process

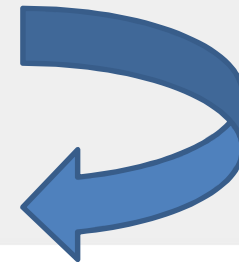
PRM > eCRF > Raw/CDASH > SDTMs > ADaMs > Define.xml > ARM > TLFs



Project Workstreams

Following is our best initial understanding of the project workstreams, subject to change as the project's agile sprints unfold.

- Workstream 1 - Analysis concept development - human-consumable
- Workstream 2 - Analysis concept development - machine-consumable
- Workstream 3 - Extended standard dataset definitions to include transformations
- Workstream 4 - End-to-start standards-based specification development (Use Case 1)
- Workstream 5 - Start-to-end study-specific metadata development (Use Case 2)
- Workstream 6 - Transform data start-to-end (Use Case 3)



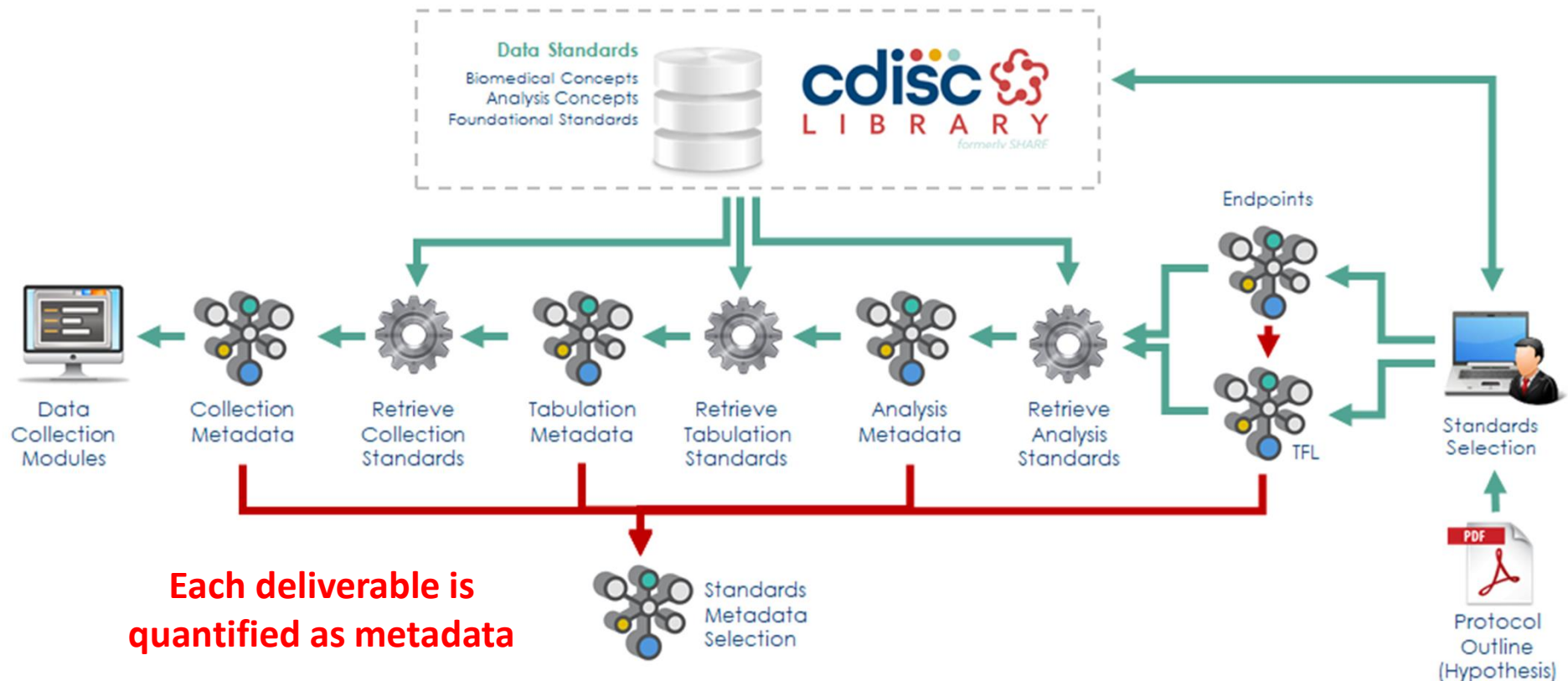
Protocol Representation Model (PRM), <https://www.cdisc.org/restricted/cdisc-360/moa>

CDISC 360 Work Stream 4 (User Case 1)

Industry downloads SDTM and ADaM specification excel files. SAS programs read and convert to variable attributes. **New SDTM and ADaM Metadata specification datasets will be introduced for industry to download, understand and populate.**

Use Case 1 : End to Start specification

Selecting standards concepts and linked metadata needed for a study



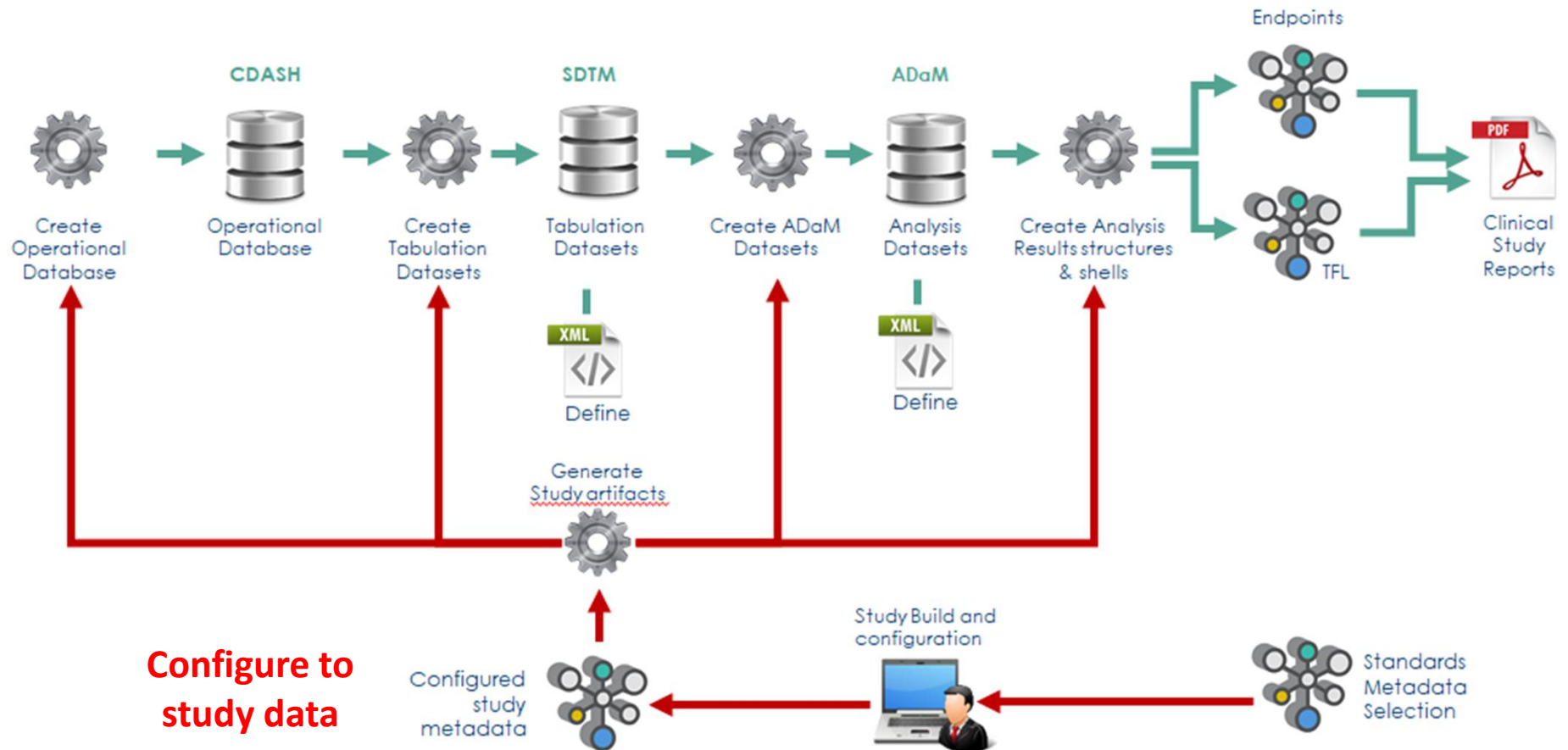
CDISC 360 Work Stream 5 (User Case 2)

Industry configures SDTM and ADaM specification excel files to their studies.

Industry needs to configure SDTM and ADaM Metadata State and Mapping datasets to their studies. Mapping datasets require ADaM most of the work.

Use Case 2 : Start to End Study Metadata

Adding study design, concept configuration & generate artifacts

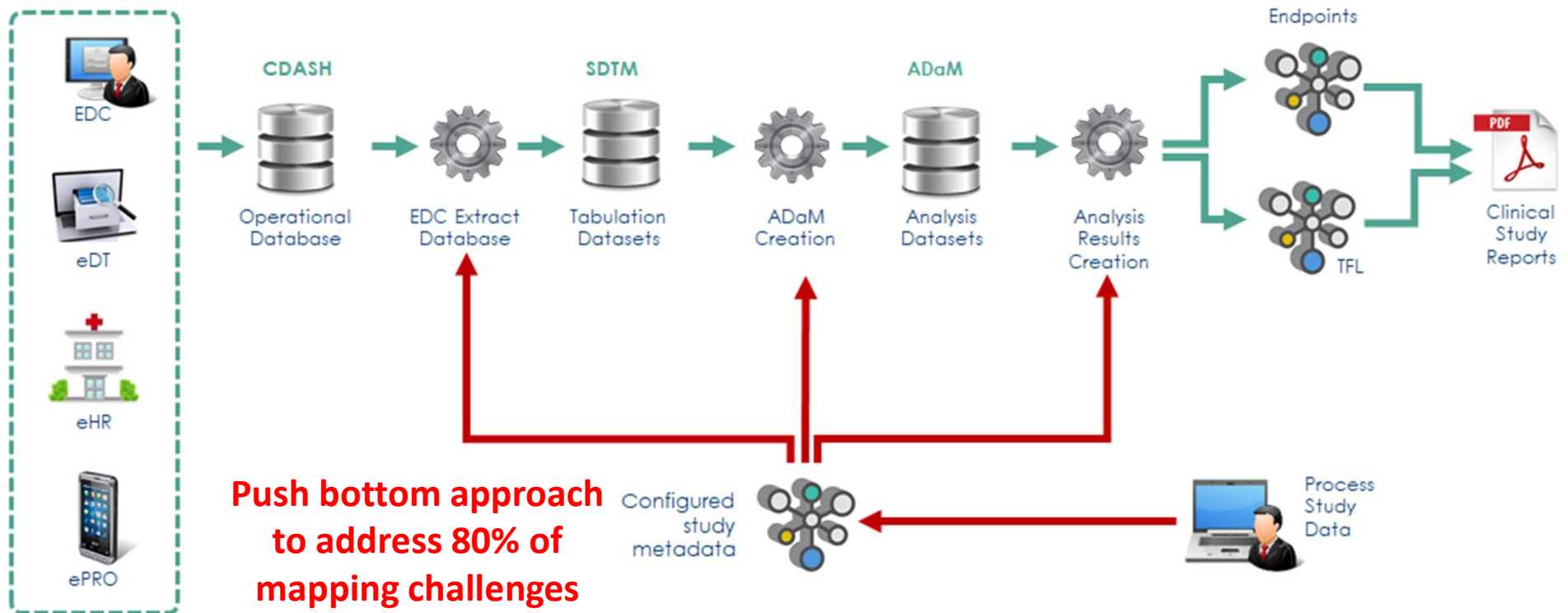


CDISC 360 Work Stream 6 (User Case 3)

Industry runs macros to automate processing SDTM and ADaM specification excel files, Raw Metadata State and Mapping and Data to create SDTMs, ADaMs and Define.xml files. Metadata design has options for dataset transpose, record and variable derivations.

Use Case 3 : Start to End Data Processing

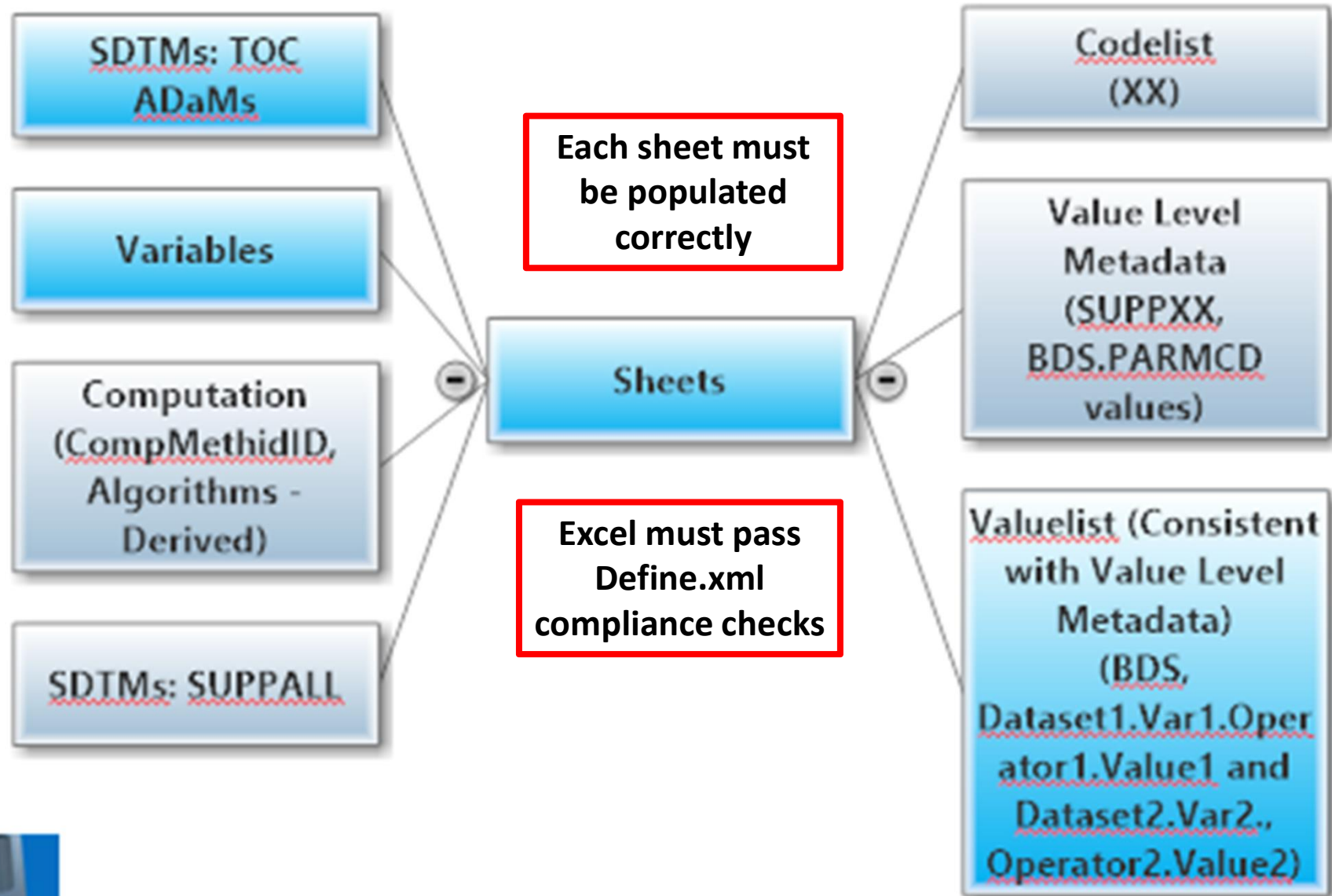
Automatic population of data into artifacts



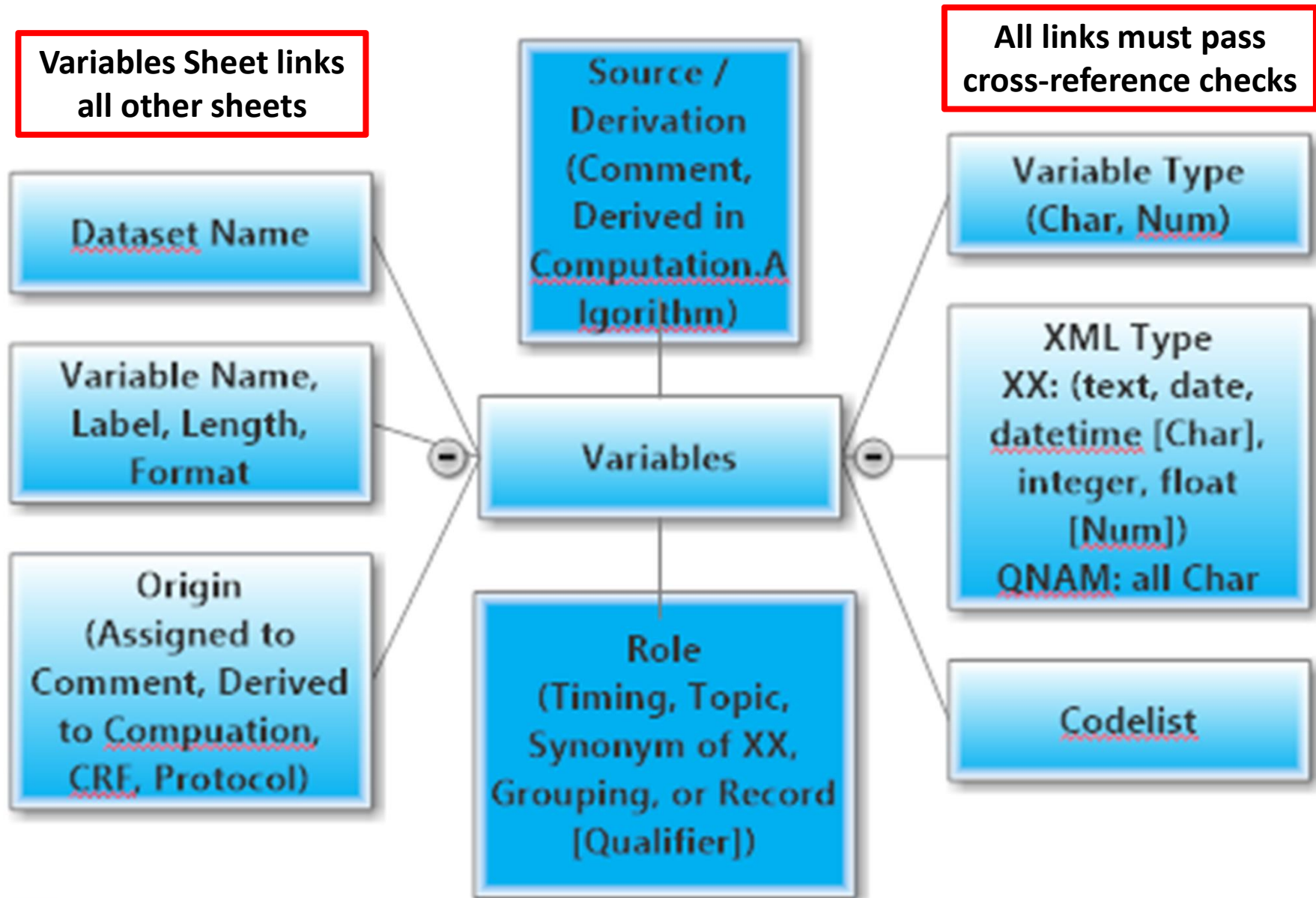
Analysis Results Metadata

Metadata Field	Metadata	
DISPLAY IDENTIFIER	Table 12.3.1.1	TFL
DISPLAY NAME	Mean NRS Pain Score Over the Last 5 Days for Overall Pain. Full Analysis Set	Metadata
RESULT IDENTIFIER	Treatment difference results (Mean, confidence interval, p-value)	TFL
PARAM	Overall Pain Score during the 5-day Period	Specifications
PARAMCD	PLPNOV	
ANALYSIS VARIABLE	CHG, BASE, TRT02AM, GEOREGN	
REASON	Primary efficacy analysis as pre-specified in protocol	ADaM
DATASET	ADQS	Metadata
SELECTION CRITERIA DOCUMENTATION	fas1fl='Y', paramcd='PLPNOV', trt01pn~=., avisit='EoT' See Protocol Section XX for details. Program: program_ex1.sas. NRS scores were analysed using an ANCOVA model which included dose group and region (REG1 and REG2) as fixed factors and baseline NRS pain score of overall pain as a covariate.	Protocol /SAP
PROGRAMMING STATEMENTS	<pre> data pain; set adam.adqs; where fas1fl="Y" and paramcd="PLPNOV" and avisit="EoT"; run; proc mixed data=pain; class &trt georegn; model chg=base &trt georegn; lsmeans &trt/cl adjust=dunnett; estimate 'Linear trend' &trt -2 -1 0 1 2; ods output type3=pvalue; ods output lsmeans=lsm; ods output diffs=dif; ods output estimates=trend; run; </pre>	SAS Code

Pinnacle 21's Define.xml Specification Template



Define.xml Specification: Variable Sheet



Pinnacle 21's P21_MappingSpec_Template_V3.xls

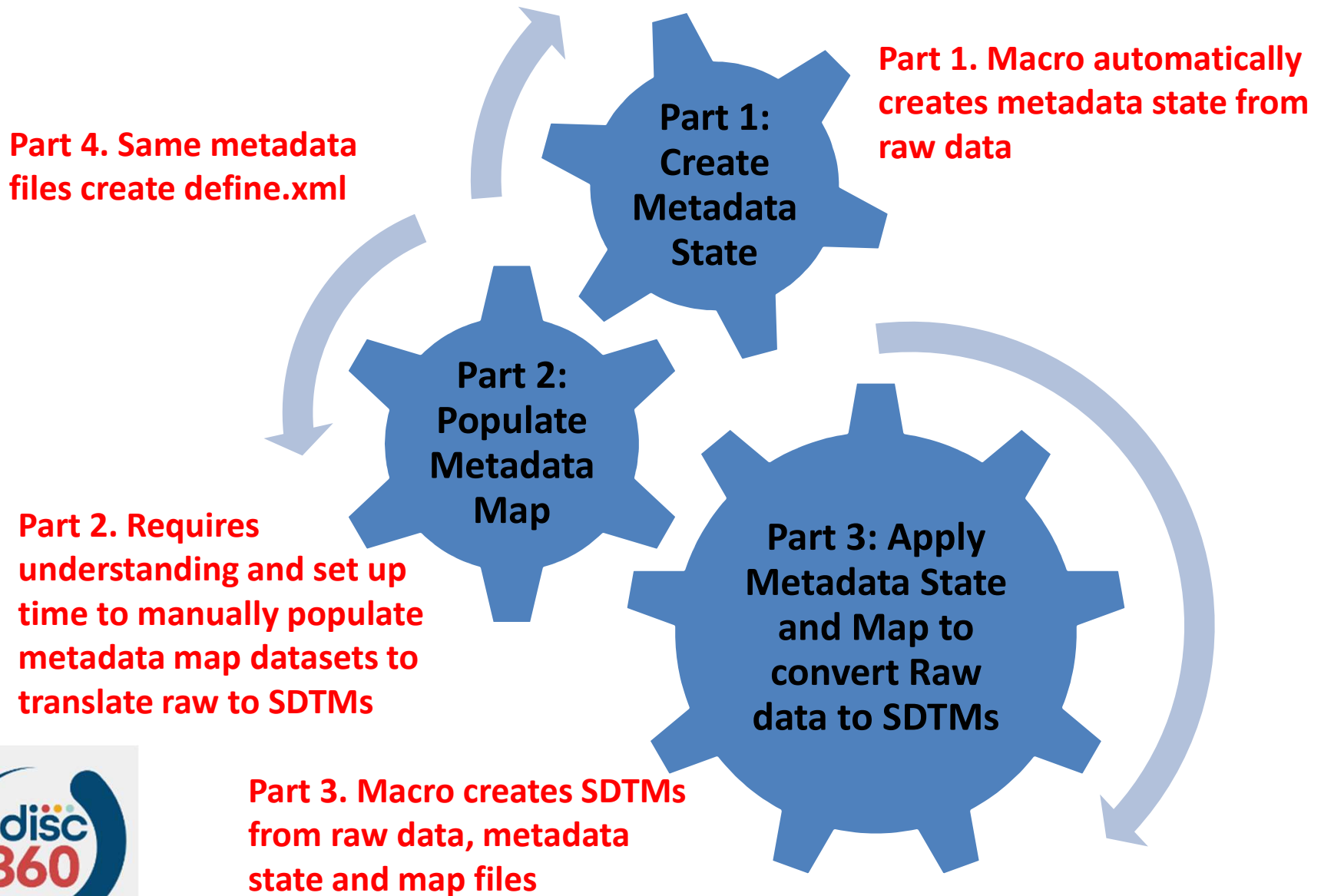
	A	B	C	D	E	F	G	H	I	J	K
1	Order	Dataset	Variable	Label	Data Type	Length	Significant Dig	Format	Mandatory	Codelist	Origin
2	1	AE	STUDYID	Study Identifier	text	11			Yes		Assigned
3	2	AE	DOMAIN	Domain Abbreviation	text	2			Yes	(DOMAIN)	Assigned
4	3	AE	USUBJID	Unique Subject Identifier	text	19			No		Derived
5	4	AE	AESEQ	Sequence Number	integer	8			Yes		Derived
6	5	AE	AETERM	Reported Term for the Adv	text	104			Yes		CRF
7	8	AE	AEDECOD	Dictionary-Derived Term	text	44			Yes	MedDRA	Assigned
8	14	AE	AECAT	Category for Adverse Even	text	23			No	(AECAT)	Assigned
9	15	AE	AESCAT	Subcategory for Adverse E	text	20			No	(AESCAT)	Assigned
10	16	AE	AEBODSYS	Body System or Organ Clas	text	67			No	MedDRA	Assigned
11	20	AE	AESER	Serious Event	text	1			No	(NY)	CRF
12											
13	1	SUPPAE	STUDYID	Study Identifier	text	11			Yes		Assigned
14	2	SUPPAE	RDOMAIN	Related Domain Abbreviat	text	2			Yes	(DOMAIN)	Assigned
15	3	SUPPAE	USUBJID	Unique Subject Identifier	text	19			No		Derived
16	4	SUPPAE	IDVAR	Identifying Variable	text	5			No		Assigned
17	5	SUPPAE	IDVARVAL	Identifying Variable Value	text	3			No		Derived
18	6	SUPPAE	QNAM	Qualifier Variable Name	text	8			Yes		Assigned
19	7	SUPPAE	QLABEL	Qualifier Variable Label	text	38			Yes		Assigned
20	8	SUPPAE	QVAL	Data Value	text	164			Yes		CRF
21	9	SUPPAE	QORIG	Origin	text	3			Yes		Assigned
22	10	SUPPAE	QVAL	Evaluator	text	1			No		Assigned
23											
24	6	CM	CMTRT	Reported Name of Drug, M	text	191			Yes		CRF
25	22	CM	CMSTDTC	Start Date/Time of Medica	text	10			No		CRF
26	23	CM	CMENDTC	End Date/Time of Medicat	text	10			No		CRF
27	24	CM	CMSTDY	Study Day of Start of Medi	integer	8			No		Derived
28	25	CM	CMENDY	Study Day of End of Medic	integer	8			No		Derived
29											
30	5	VS	VSTESTCD	Vital Signs Test Short Nam	text	8			Yes	(VSTESTCD)	Assigned
31	6	VS	VSTEST	Vital Signs Test Name	text	24			Yes	(VSTEST)	CRF

Creating define.xml is not trivial

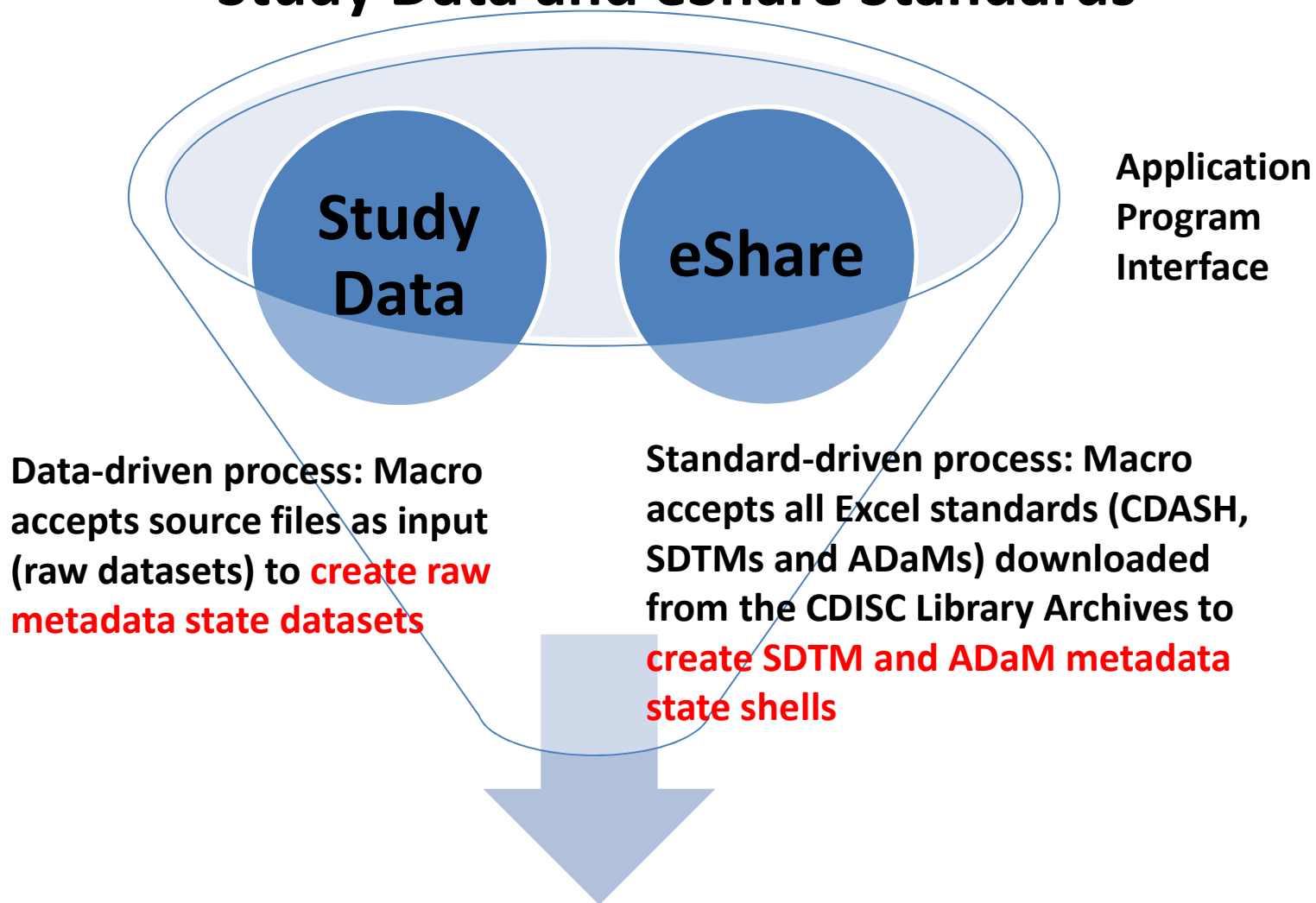
Pinnacle 21's template forced industry to be structured and organized for traceability to collect all SDTM components required for creating and meeting define.xml specifications

Sheets

Goal: Apply Metadata to create SDTMs



Part 1: Create Metadata State from Study Data and eShare Standards

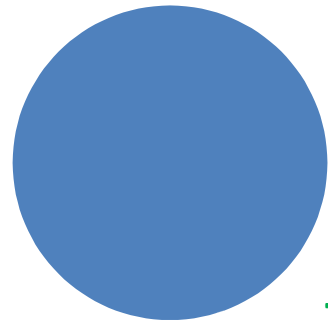


Six Metadata State Datasets

CDISC plans to publish metadata datasets once evaluation is completed

Part 1: Create Six Metadata State Datasets

PDESCRIPTION
– Catalog of
source entries

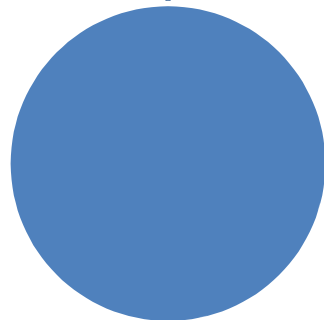


Tables

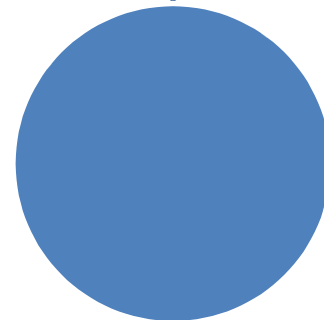
Metadata design captures all standard information to create SDTMs/ADaMs and one of the technologies used in the CDISC 360 project

TABLE

TABLE



Columns

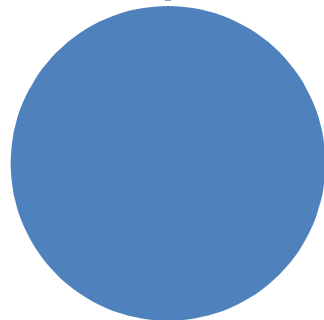


RowMData

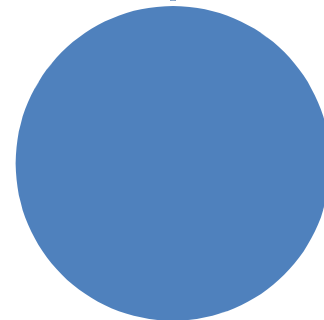
Row level metadata

VALUE_NAME

ROWDEFID



Values

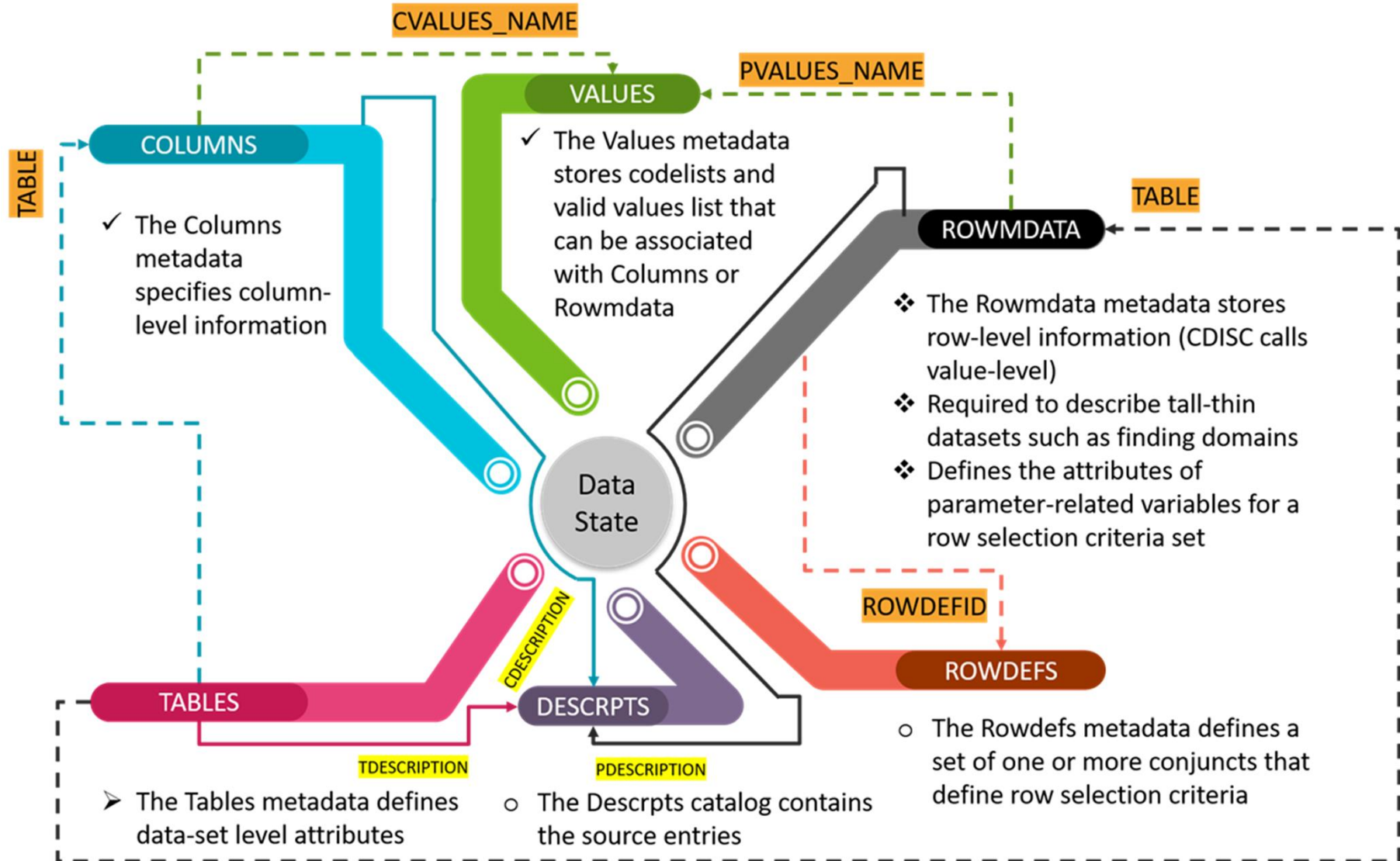


RowDefs

Row Definitions

Metadata State Content – Six Datasets

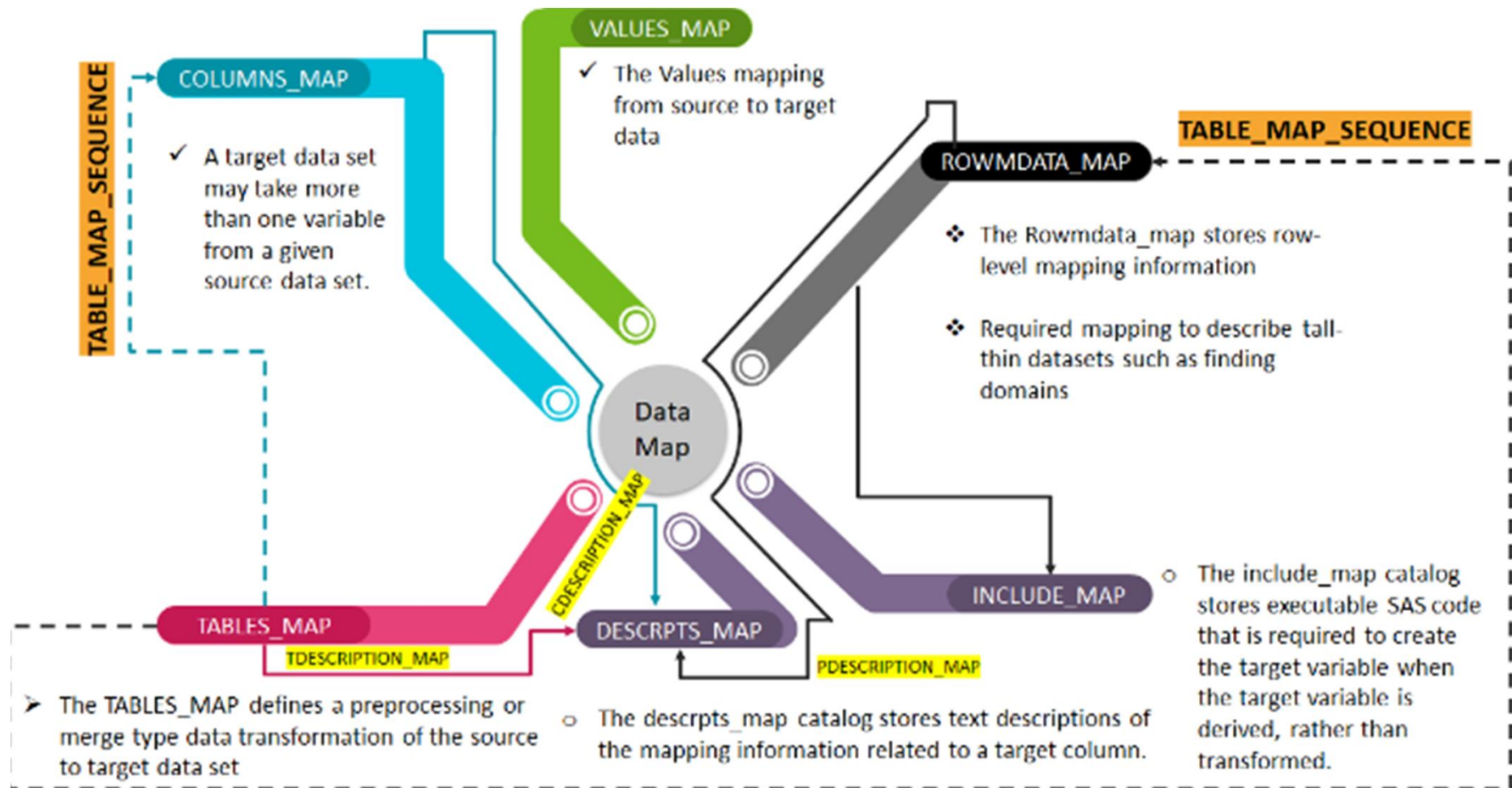
All datasets are integrated with key variables. Datasets house structure and variables for ODM requirements such as specifications, crf and xpts to create define.xml. Variables can be populated from excel file and CDISC metadata as needed.



Part 2: Map Metadata Content

Requires understanding and set up time (Active participation by TalentMine)

From: Raw or SDTMs
TABLES_MAP, COLUMNS_MAP, ROWMDATA_MAP, ROWDEFS_MAP, VALUES_MAP
To: SDTMs or ADaMs



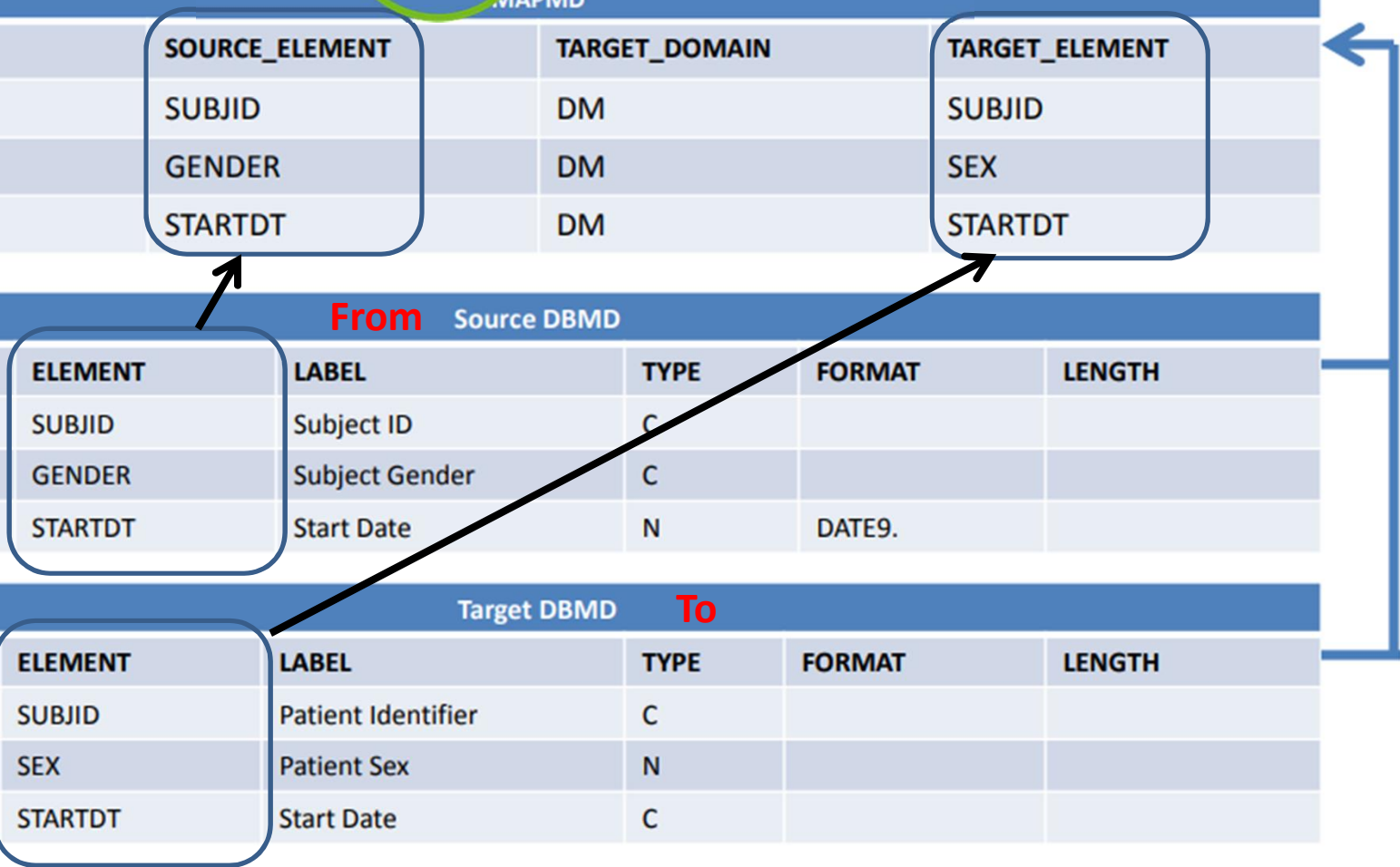
Source to Target Variables – DM Example



MAPMD			
SOURCE_DOMAIN	SOURCE_ELEMENT	TARGET_DOMAIN	TARGET_ELEMENT
DM	SUBJID	DM	SUBJID
DM	GENDER	DM	SEX
DM	STARTDT	DM	STARTDT

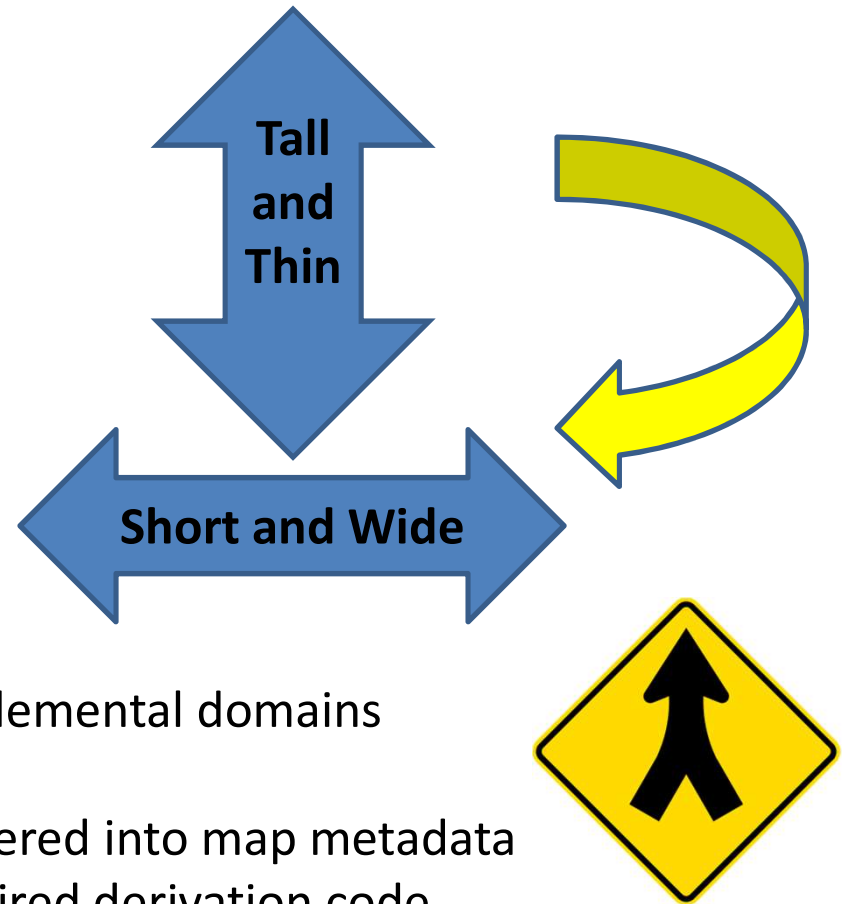
From Source DBMD					
DOMAIN	ELEMENT	LABEL	TYPE	FORMAT	LENGTH
DM	SUBJID	Subject ID	C		
DM	GENDER	Subject Gender	C		
DM	STARTDT	Start Date	N	DATE9.	

Target DBMD To					
DOMAIN	ELEMENT	LABEL	TYPE	FORMAT	LENGTH
DM	SUBJID	Patient Identifier	C		
DM	SEX	Patient Sex	N		
DM	STARTDT	Start Date	C		



Data Transfer Engine (DTE) Design

- ✓ Standard Process
 - ✓ Rename variables
 - ✓ Standardize control terms
 - ✓ Keep or Drop variables
- ✓ Variable/Record Derivations
 - ✓ Formulas
 - ✓ SAS Snippet Code Include
 - ✓ Transpose variable structure
 - ✓ Tall and Thin
 - ✓ Short and Wide
- ✓ Level 1
 - ✓ Applies attributes, creates supplemental domains
- ✓ Level 2
 - ✓ Adds derivation logic that is entered into map metadata
 - ✓ Assumes all variables have required derivation code
- ✓ Level 3
 - ✓ Adds merging of source data sets to gather the variables required by derivations and transformations



Part 3: Apply Metadata State and Map to convert Raw data to SDTMs

Process all metadata information and raw data to create SDTMs. CDISC 360 team is currently evaluating this metadata design for industry standard.

**Metadata
State: SDTM
Variable
Attributes**

**Metadata
Map, Raw
Data**

**SDTM SAS
Generated
Program**

Sample SDTMs Created

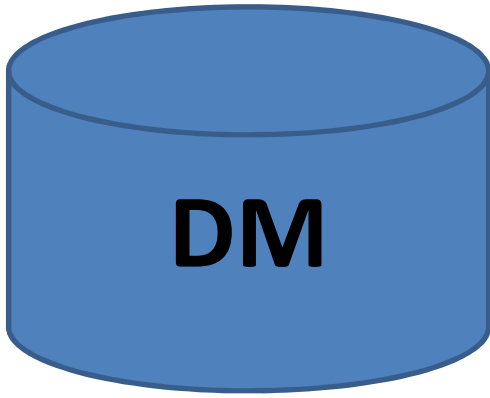


Table: WORK.DM | View:

Columns	Attributes
<input type="checkbox"/>	Select all
<input checked="" type="checkbox"/>	STUDYID
<input checked="" type="checkbox"/>	DOMAIN
<input checked="" type="checkbox"/>	USUBJID
<input checked="" type="checkbox"/>	POOLID
<input checked="" type="checkbox"/>	SUBJID
<input checked="" type="checkbox"/>	RFSTDTC
<input checked="" type="checkbox"/>	SPDEVID
<input type="checkbox"/>	DMSEQ
<input type="checkbox"/>	RFENDTC

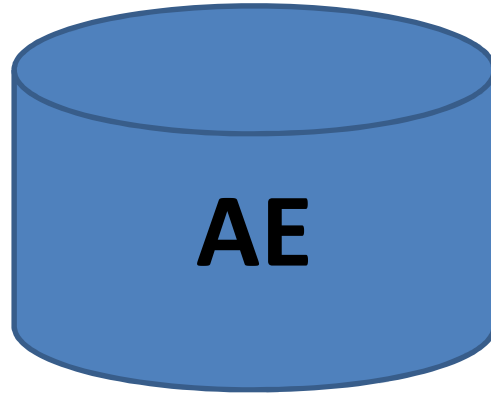


Table: WORK.AE | View:

Columns	Attributes
<input checked="" type="checkbox"/>	Select all
<input checked="" type="checkbox"/>	STUDYID
<input checked="" type="checkbox"/>	DOMAIN
<input checked="" type="checkbox"/>	USUBJID
<input checked="" type="checkbox"/>	AESEQ
<input checked="" type="checkbox"/>	POOLID
<input checked="" type="checkbox"/>	AEGRPID
<input checked="" type="checkbox"/>	SPDEVID
<input checked="" type="checkbox"/>	AEREFID
<input checked="" type="checkbox"/>	AESPID

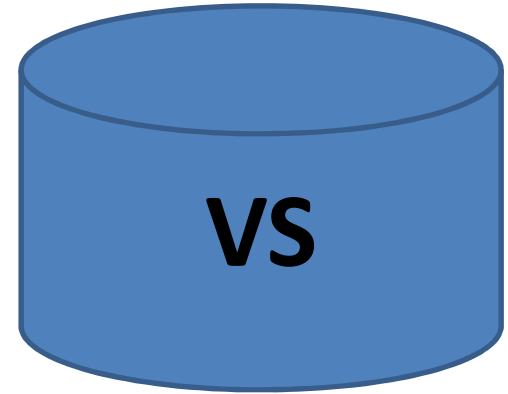


Table: WORK.VS | View:

Columns	Attributes
<input type="checkbox"/>	Select all
<input checked="" type="checkbox"/>	STUDYID
<input checked="" type="checkbox"/>	DOMAIN
<input checked="" type="checkbox"/>	USUBJID
<input checked="" type="checkbox"/>	POOLID
<input checked="" type="checkbox"/>	VSSEQ
<input checked="" type="checkbox"/>	SPDEVID
<input checked="" type="checkbox"/>	VSGRPID
<input checked="" type="checkbox"/>	VSSPID
<input checked="" type="checkbox"/>	VSTESTCD

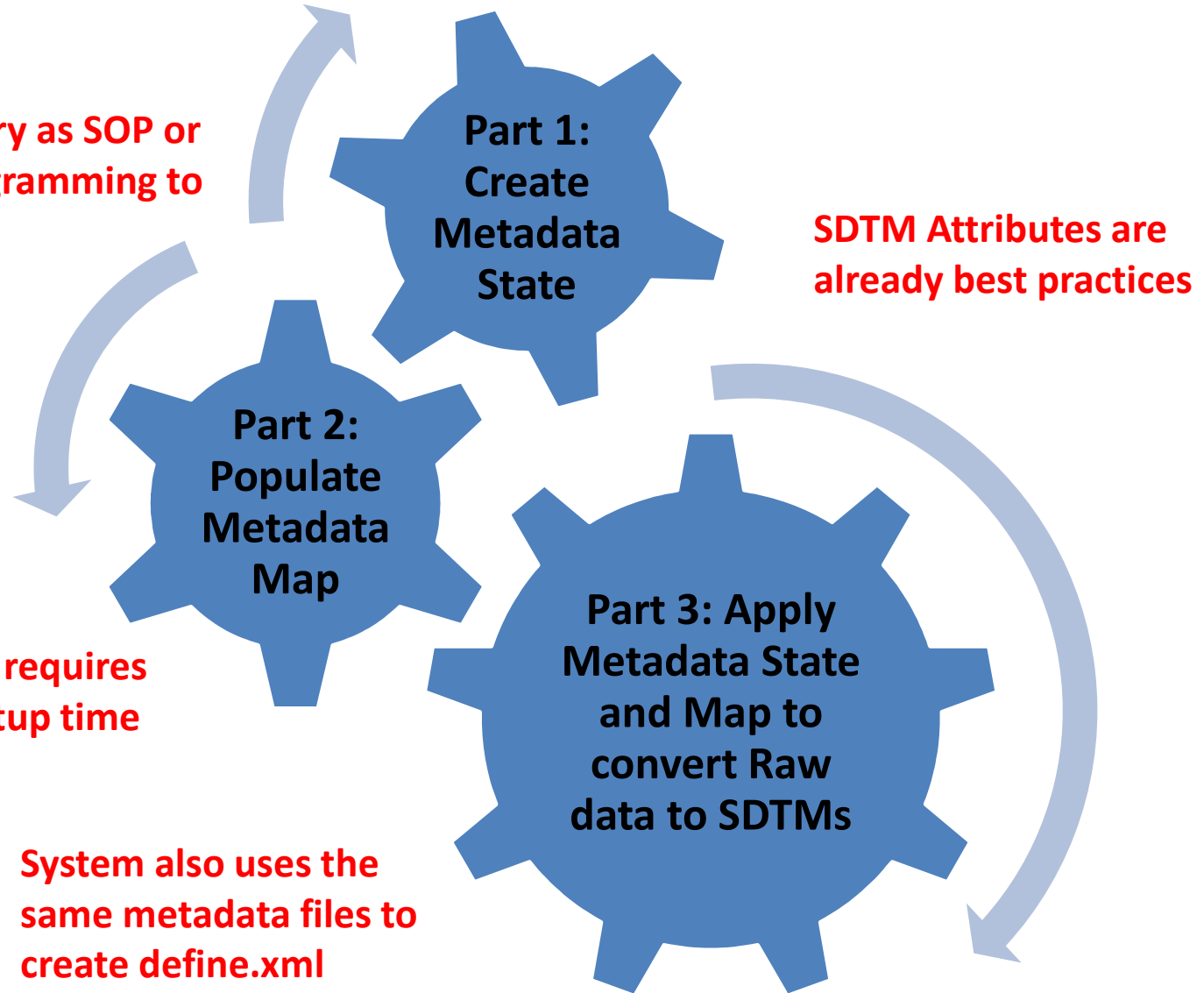
AE SAS Generated Program

Can customize independent SAS program to include raw data and create SDTM/ADaM.

```
*-----;
* Create the AE data set defined in the metadata;
*-----;
data work.AE ;SDTM/ADaM Attributes are already industry best practices
* .....;
* Define the length of each column;                Variable Length
* .....;
length STUDYID $ 200 DOMAIN $ 200 USUBJID $ 200 AESEQ 8 POOLID $ 200 AEGRPID
* .....;
* Define the label and format of each column;
* .....;
label STUDYID = "Study Identifier";
label DOMAIN = "Domain Abbreviation";
label USUBJID = "Unique Subject Identifier";        Variable Label
label AESEQ = "Sequence Number";
label POOLID = "Pool Identifier";
label AEGRPID = "Group ID";                        PUT() with format
label SPDEVID = "Sponsor Device Identifier";        catalog to convert raw
label AEREFID = "Reference ID";                    to SDTM control terms
label AESPID = "Sponsor-Defined Identifier";
```

Goal: SDTMs Automation Process

Ideal for primary as SOP or secondary programming to save time



Summary: Levels of Metadata Programming

What is your organization's metadata programming level expertise?

Evolution of Standardization and Automation

Intro
Level



1. Basic method to **standardize datasets** by applying dataset and variable attributes and applying program index table metadata in tables, lists and graph programs.

(Many organizations are already doing this as best practices)

2. Basic method to **automate** by processing a list of datasets or files to create inventory lists for example.

(Macro programming, Proc SQL and Dictionary table are utilized)

Advanced
Level



3. Advanced method to confirm **data loading compliance with specifications.**

(Create custom metadata and cross-reference with new data)

4. Advanced method to **standardize by creating codelists.**

(Automate creation of SDTMs and ADaMs codelists for cross-referencing)

5. One to one dataset mapping to apply derived logic.

(Build design and foundation to join datasets and apply variable level logic)

Expert
Level

6. Advanced method to handle standard and custom programming by transforming datasets based on **source to target variable mapping.**

(Automate to build standard process with custom options)

SAS: Extract intelligence information from metadata and macro processing

Data-Driven process is automatic, quality controlled, transparent and saves time!

INPUT:

Directory of Files (Excel files, Datasets, SAS Programs, Log and Proc Compare.lst)

SAS Tools

- Libnames
- Data Step
- SAS Macro Programming
- SAS & Dataset Functions
- Proc SQL
- Proc Compare
- Proc Means

OUTPUT: Metadata Attributes

- File pathname and names
- # of Files, Datetime stamps
- # and Type of variables
- Macro loop through all files
- Required datasets, variables, etc.
- Codelist dictionary, SDTM/ADaM attributes
- Maximum variable lengths
- Data cleaning & monitoring of valid variables and special characters
- Compare and contrast previous file
- Descriptive Statistics on categorical and continuous variables
- Search for ERRORS, WARNINGS or Notes in SAS Logs
- Search for QC differences in # of VARs, OBS, attributes and dups

Is your organization ready for CDISC 360 View and Challenges?

“Apply the 80/20 rule to ensure the Project automates 80% of the end-to-end metadata and data processing needed to generate study artifacts suitable for a regulatory submission.”

