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.

TALENTMINE
KNOW WHERE TO LOOK™

Is your organization ready for CDISC 360 View and Challenges?

Agenda

- ✓ Industry Best Practices
  - ✓ Balance between standards and customization
- ✓ CDISC Mission 360
  - ✓ Requirements
- ✓ 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)
- ✓ Analysis Results Metadata (ARM)
- ✓ Pinnacle 21 Define.XML specification template
- ✓ Goal: Apply Metadata to create SDTMs
- ✓ Summary
  - ✓ Levels of Metadata Programming
  - ✓ Extract intelligence information from metadata and macro processing

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.

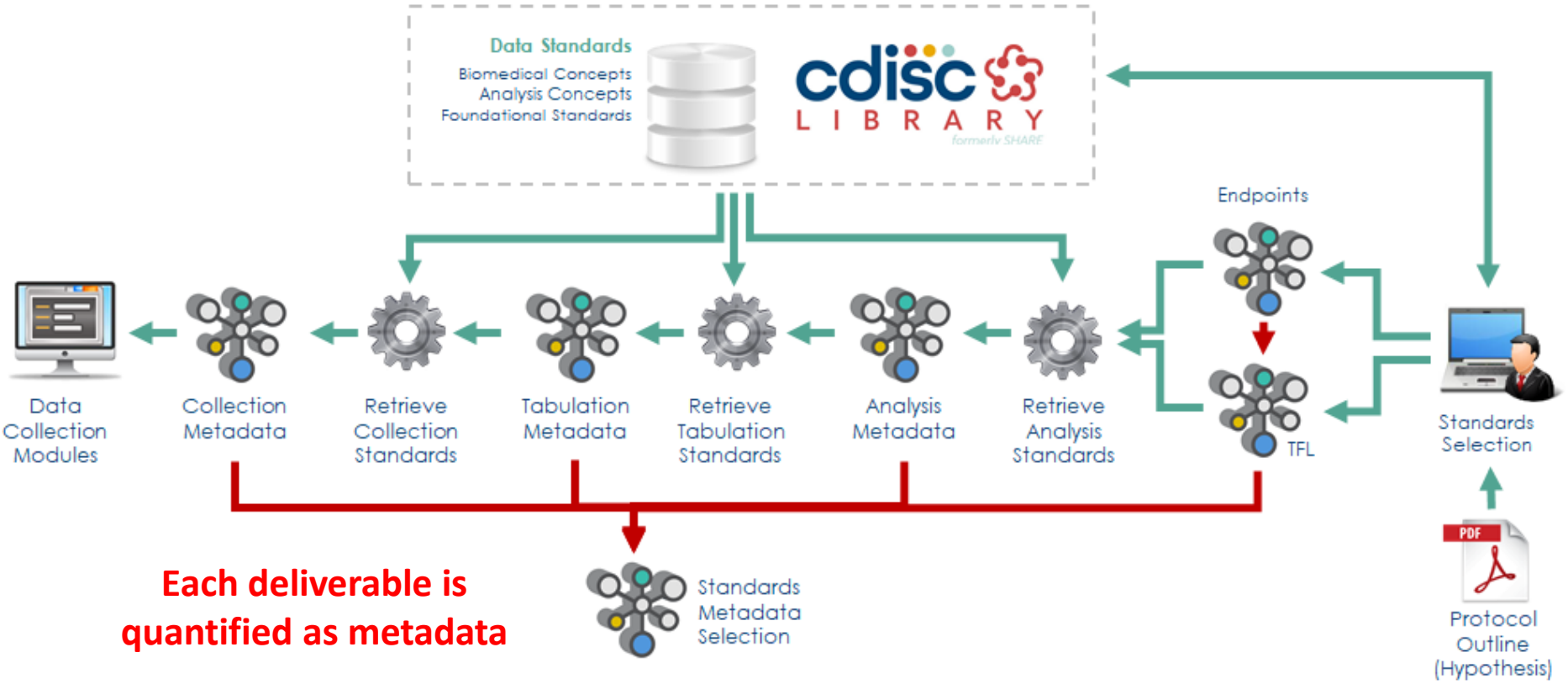
Evolving CDISC to the Next Decades: The CDISC Proof of Concept, Peter Van Reusel, CDISC, Sam Hume, CDISC

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

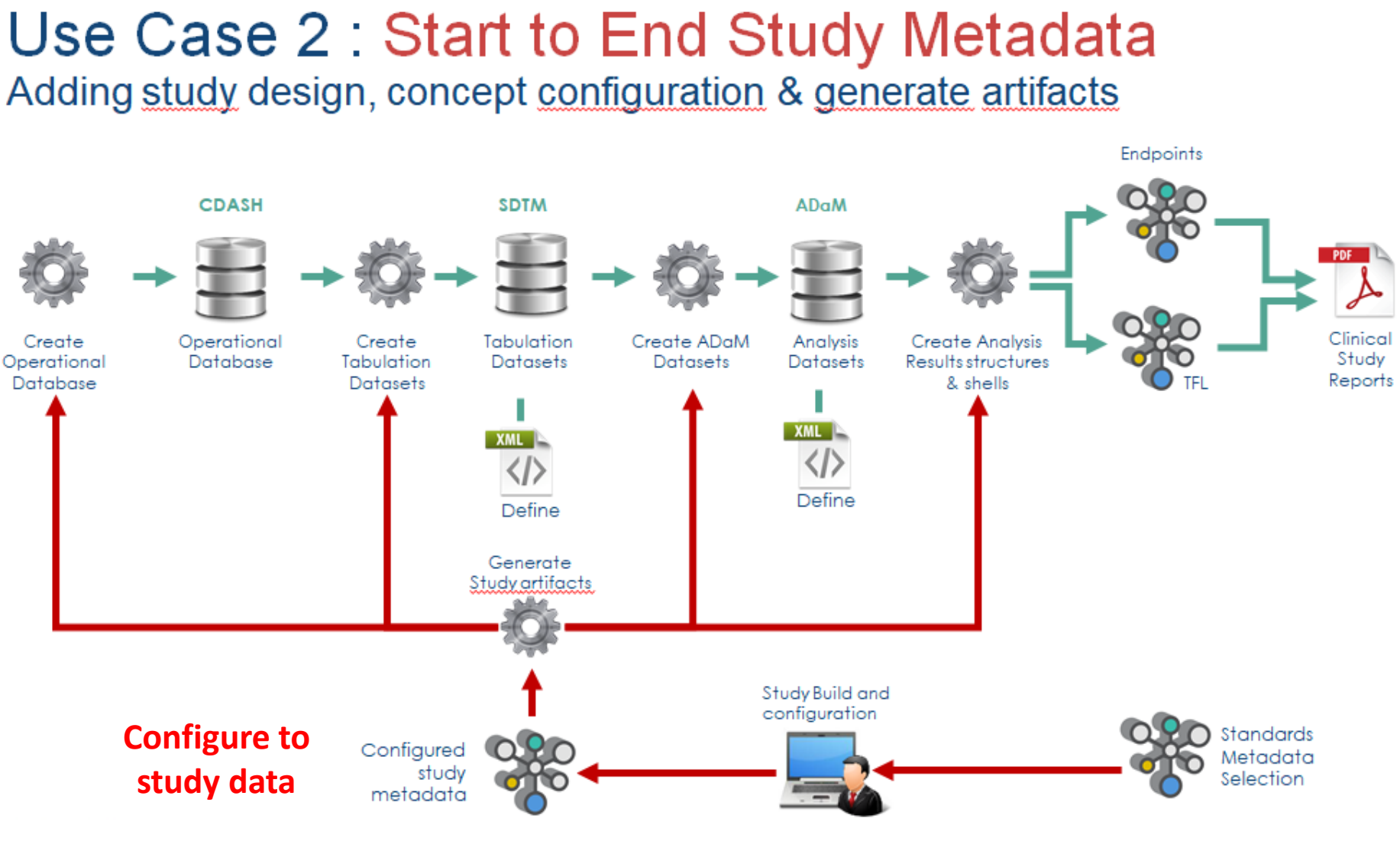


Each deliverable is quantified as metadata

CDISC 360 Work Stream 5 (User Case 2)

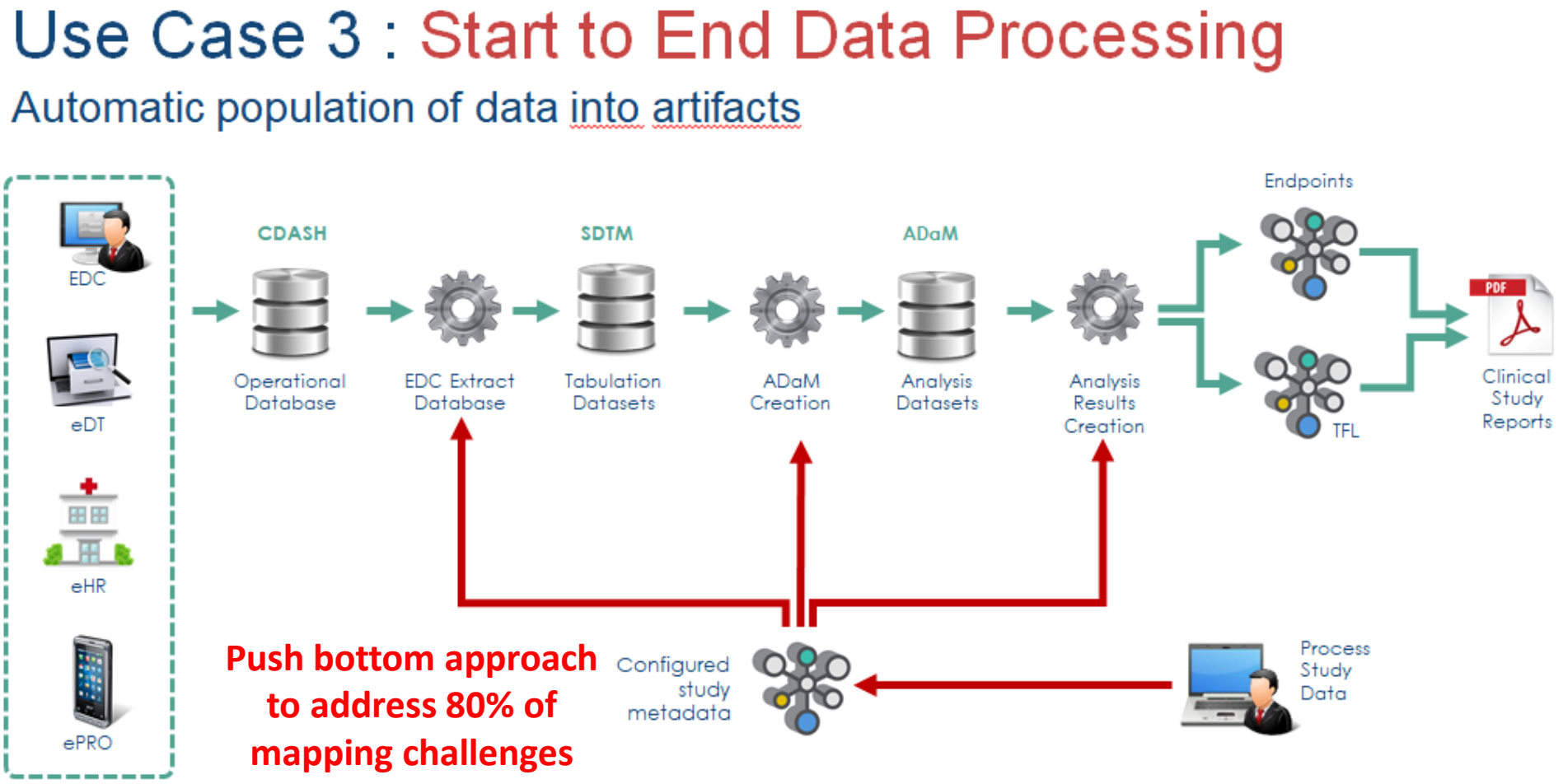
Industry configures SDTM and ADaM specification excel files to their studies.

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



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.



Analysis Results Metadata (ARM)

Metadata Field	Metadata
DISPLAY IDENTIFIER	Table 12.3.1.1
DISPLAY NAME	Mean NRS Pain Score Over the Last 5 Days for Overall Pain. Full Analys Set
RESULT IDENTIFIER	Treatment difference results (Mean, confidence interval, p-value)
PARAM	Overall Pain Score during the 5-day Period
PARAMCD	PLPNOV
ANALYSIS VARIABLE	CHG, BASE, TRT02AN, GEOREGN
REASON	Primary efficacy analysis as pre-specified in protocol
DATASET	ADQS
SELECTION CRITERIA DOCUMENTATION	fas1fl='Y', paramcd='PLPNOV', trt01pn~=., avisit='EoT' See Protocol Section XX for details. Program: program_ex1.sas. NRS sco 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 a covariate.

PROGRAMMING STATEMENTS	`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;`

TFL Metadata

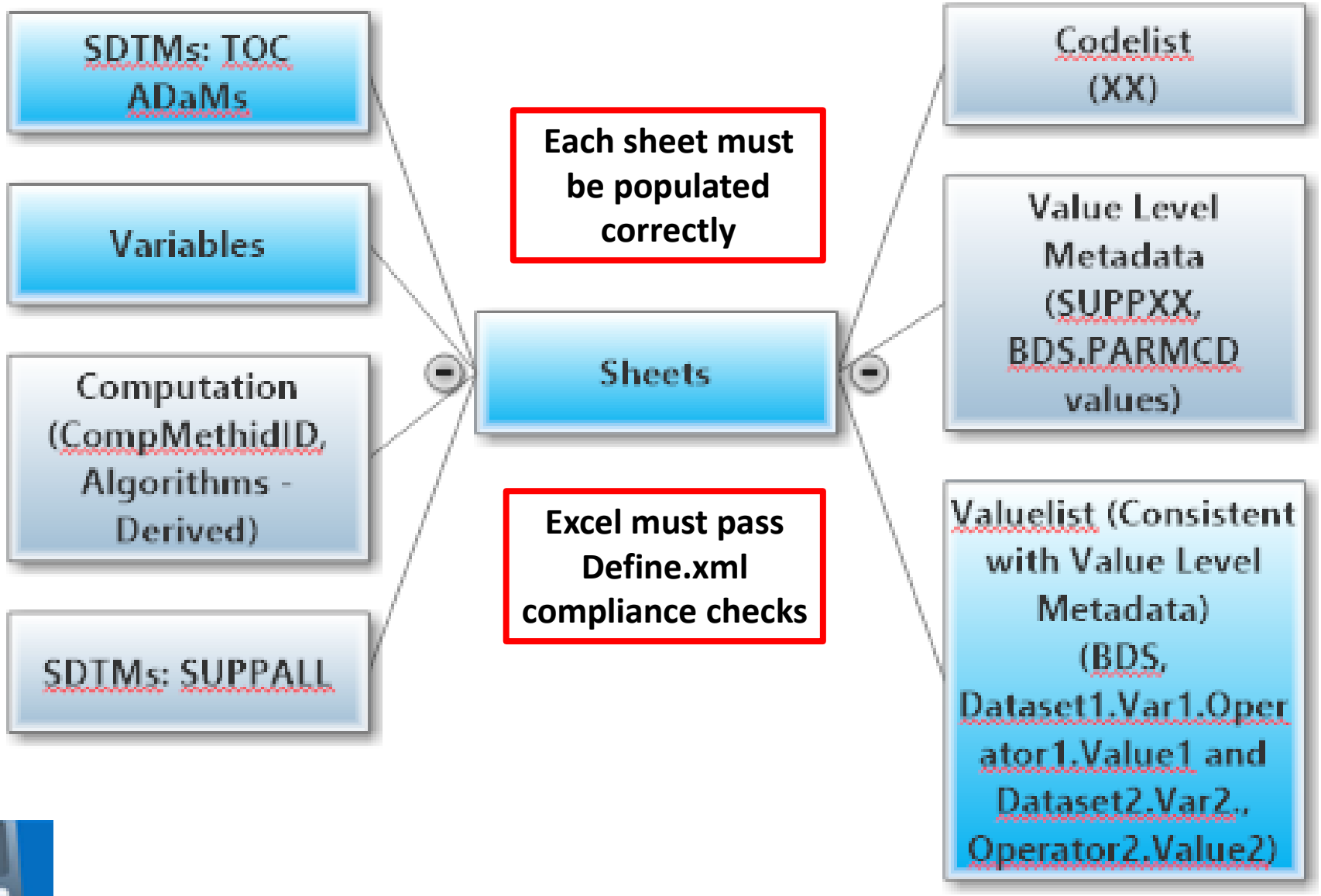
TFL Specifications

ADaM Metadata

Protocol /SAP

SAS Code

Pinnacle 21's Define.xml Specification Template



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	QEVAL	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

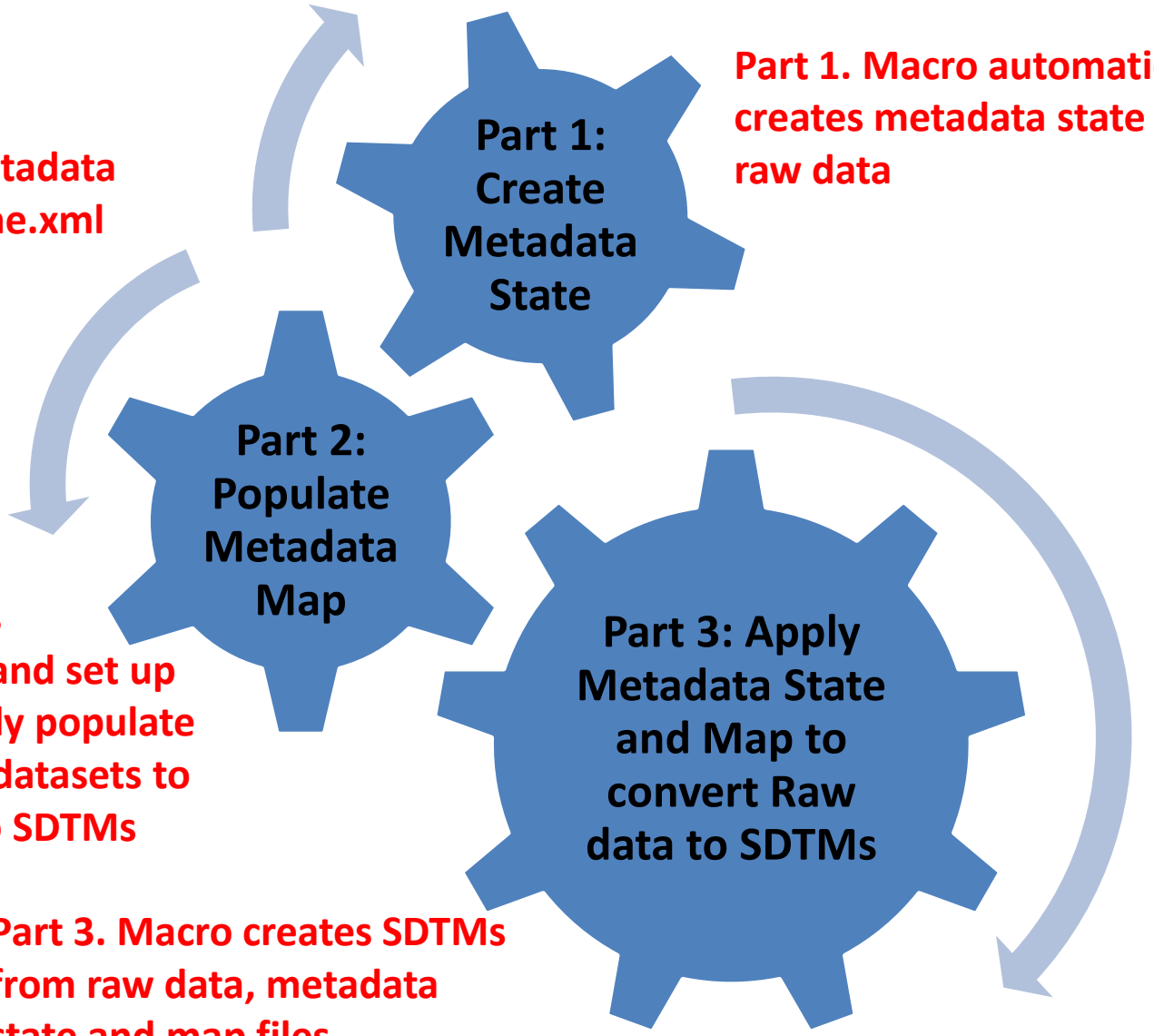
Creating define.xml is not trivial

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

All sheets are interconnected

Study | Datasets | **Variables** | ValueLevel | WhereClauses | Codelists | Dictionaries | Methods | Comments | Documents | SDTM Rules | ADaM Rules

Goal: Apply Metadata to create SDTMs



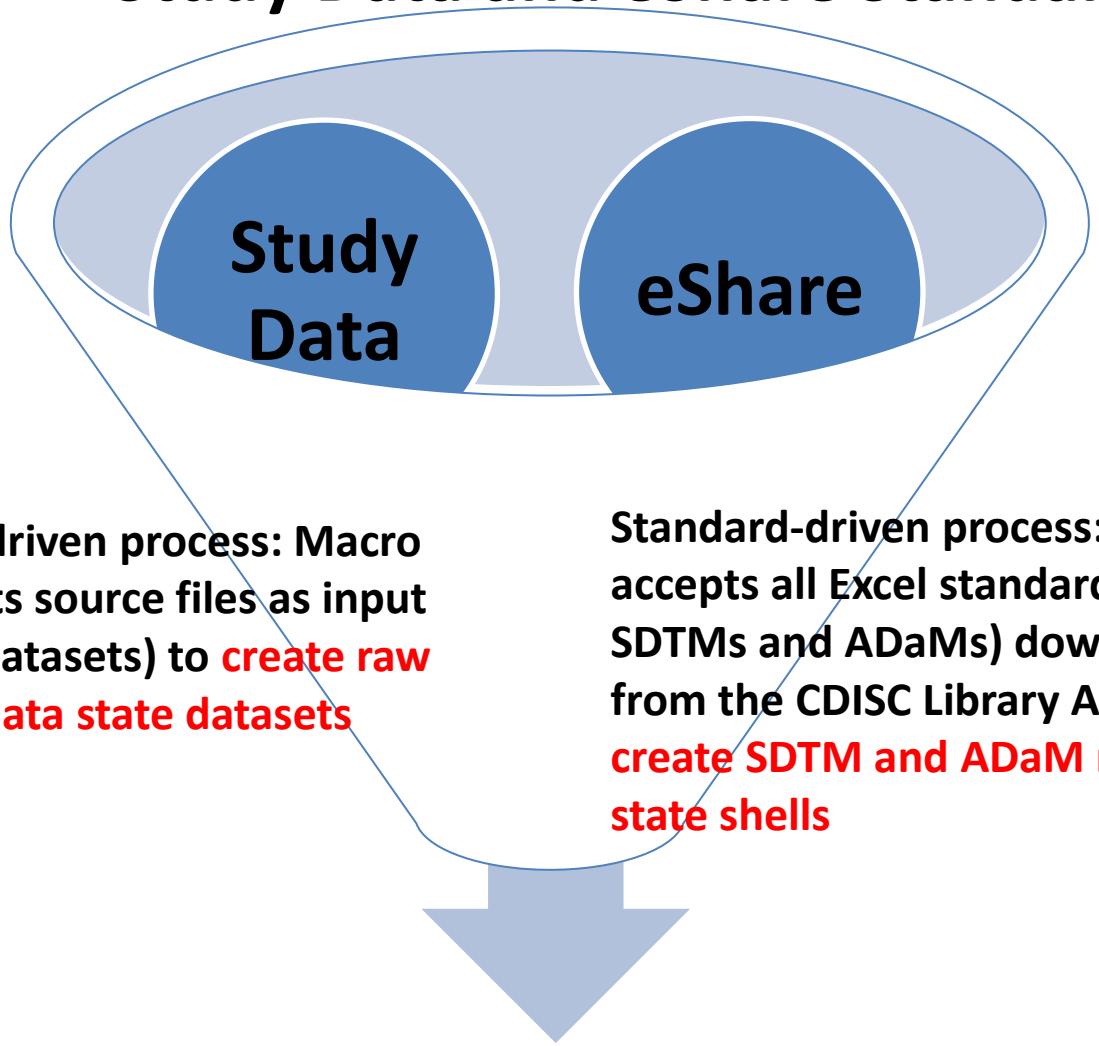
Part 1. Macro automatically creates metadata state from raw data

Part 2. Requires understanding and set up time to manually populate metadata map datasets to translate raw to SDTMs

Part 3. Macro creates SDTMs from raw data, metadata state and map files

Part 4. Same metadata files create define.xml

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



Application Program Interface

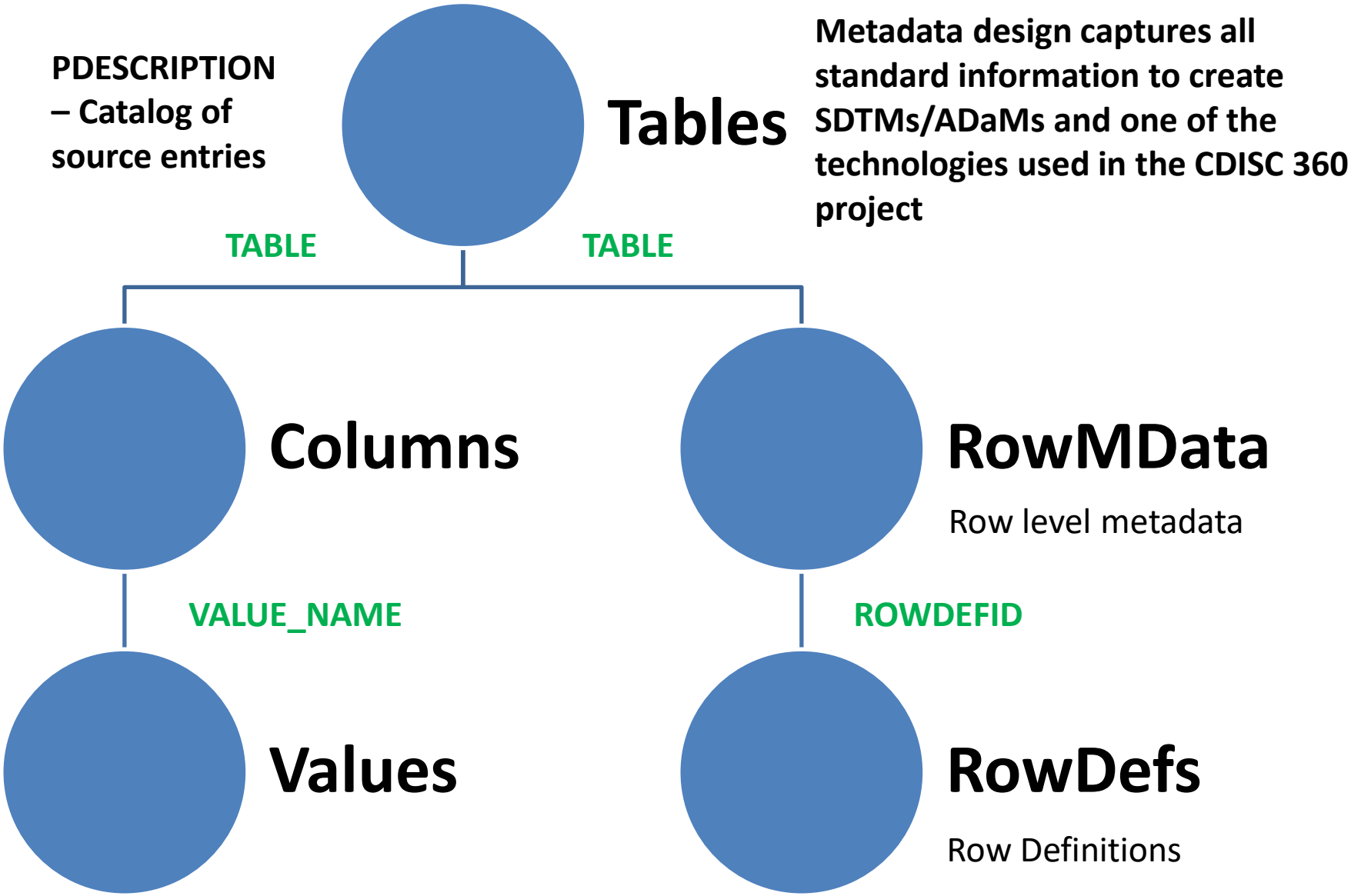
Data-driven process: Macro accepts source files as input (raw datasets) to create raw metadata state datasets

Standard-driven process: Macro accepts all Excel standards (CDASH, SDTMs and ADaMs) downloaded from the CDISC Library Archives to create SDTM and ADaM metadata state shells

Six Metadata State Datasets

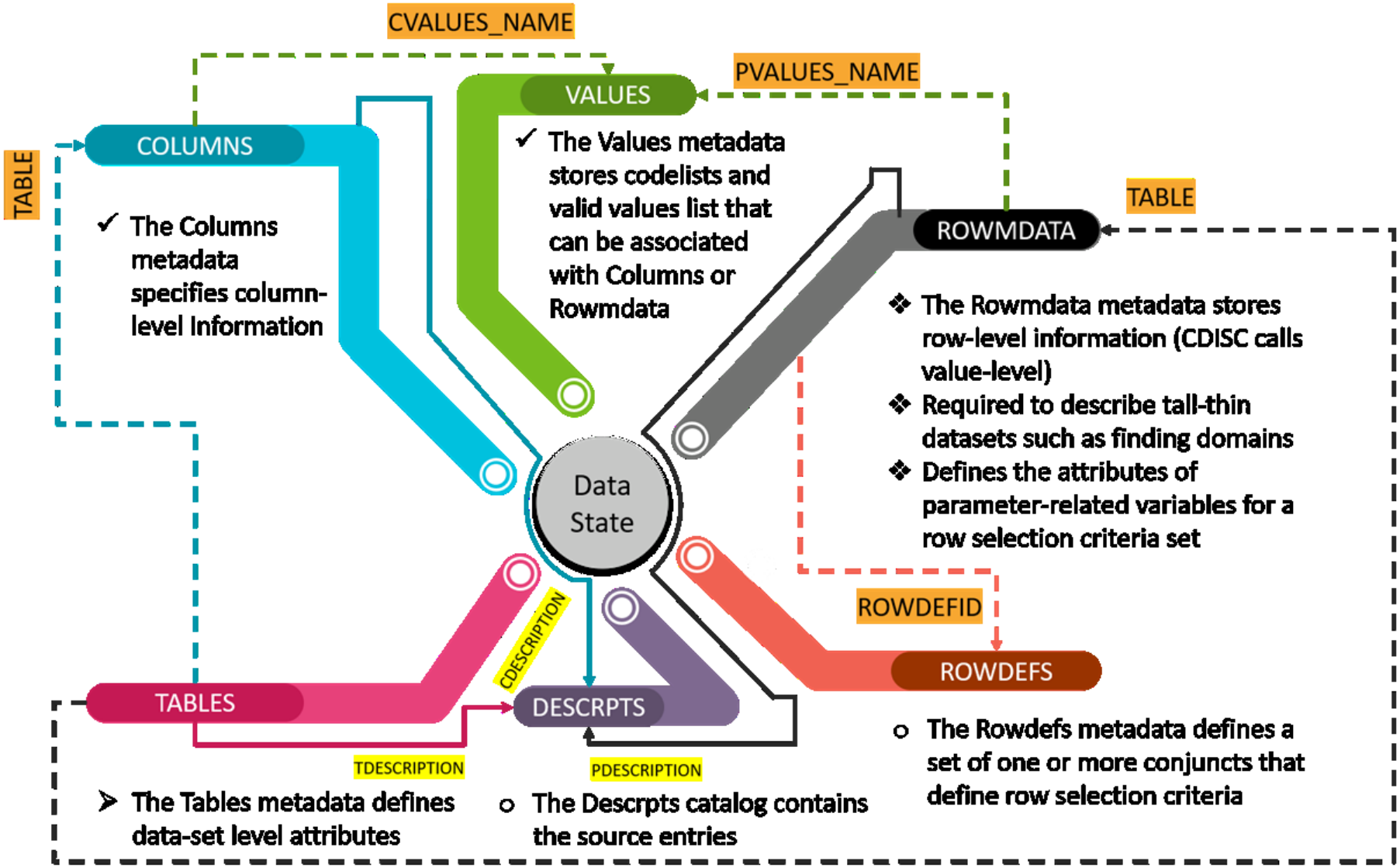
CDISC plans to publish metadata datasets once evaluation is completed

Part 1: Create Six Metadata State Datasets



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.



Metadata State Content – Six Datasets

- ✓ **All SDTM/ADaM variable attributes**
  - ✓ Name, Label, Type, Length, etc.
- ✓ **Meet's Pinnacle 21's SDTM Compliance Test**
  - ✓ Control Terminology Dictionary
  - ✓ Basic Data Structure (BDS) – Variable-Level Metadata
  - ✓ Variable Order
- ✓ **TABLES links to COLUMNS and ROWMDATA**
  - ✓ TABLES.TABLE = COLUMNS.TABLE = ROWMDATA.TABLE = 'VS'
- ✓ **COLUMNS links to ROWMDATA**
  - ✓ COLUMNS.TABLE = ROWMDATA.TABLE = 'VS'
- ✓ **COLUMNS and ROWMDATA links to VALUES**
  - ✓ COLUMNS.CVALUES_NAME = VALUES.VALUES_NAME = 'V1_'
- ✓ **ROWMDATA links to ROWDEFS**
  - ✓ ROWMDATA.ROWDEFID = ROWDEFS.ROWDEFID = 'ROWDEFID1'

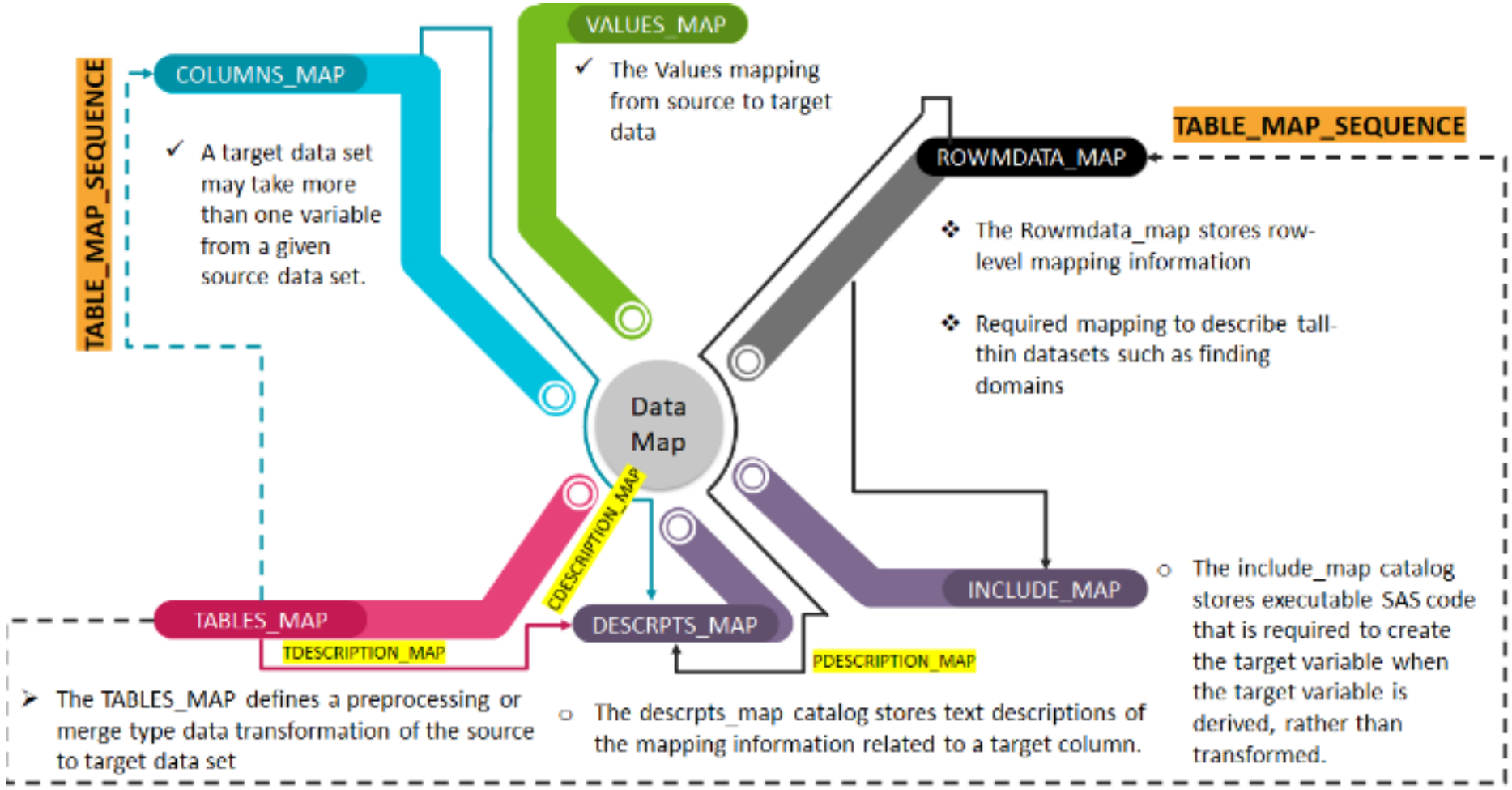
Part 2: Map Metadata Content

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

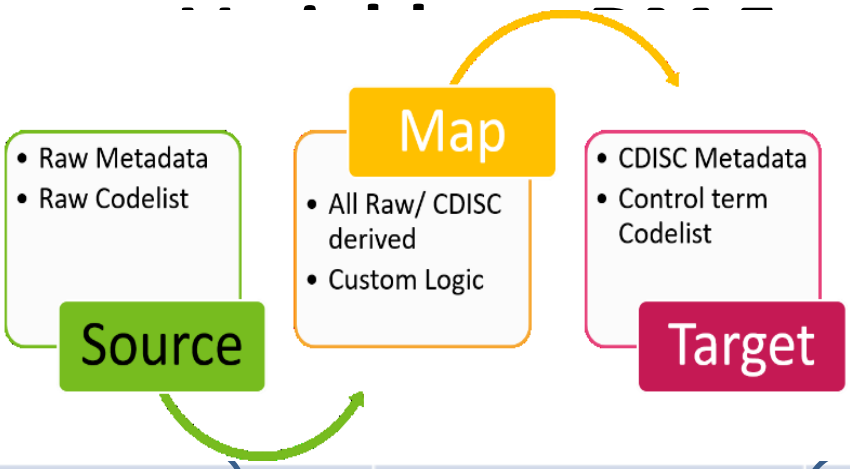
From: Raw or SDTMs

TABLES_MAP, COLUMNS_MAP, ROWMDATA_MAP, ROWDEFS_MAP , VALUES_MAP

To: SDTMs or ADaMs



Source to Ta mple



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

From Source DBMD

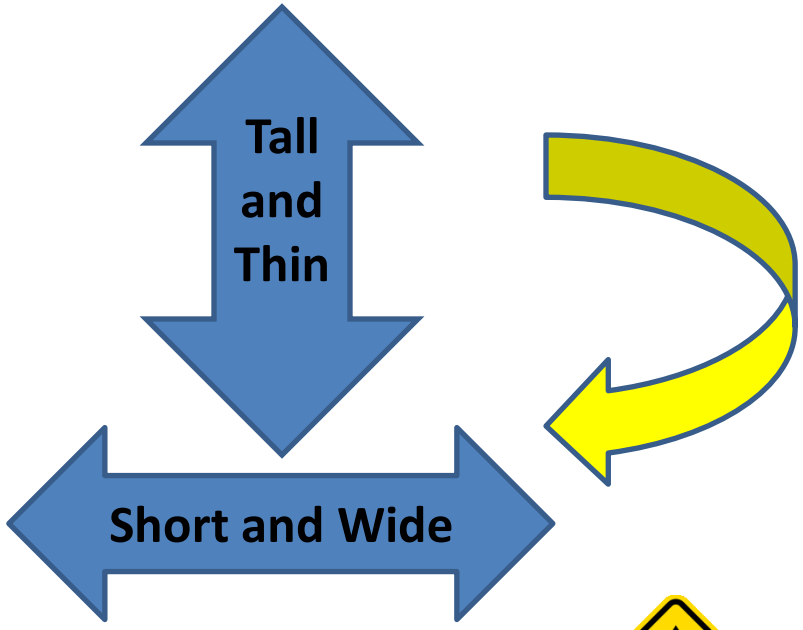
DOMAIN	ELEMENT	LABEL	TYPE	FORMAT	LENGTH
DM	SUBJID	Subject ID	C		
DM	GENDER	Subject Gender	C		
DM	STARTDT	Start Date	N	DATE9.	

Target DBMD **To**

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

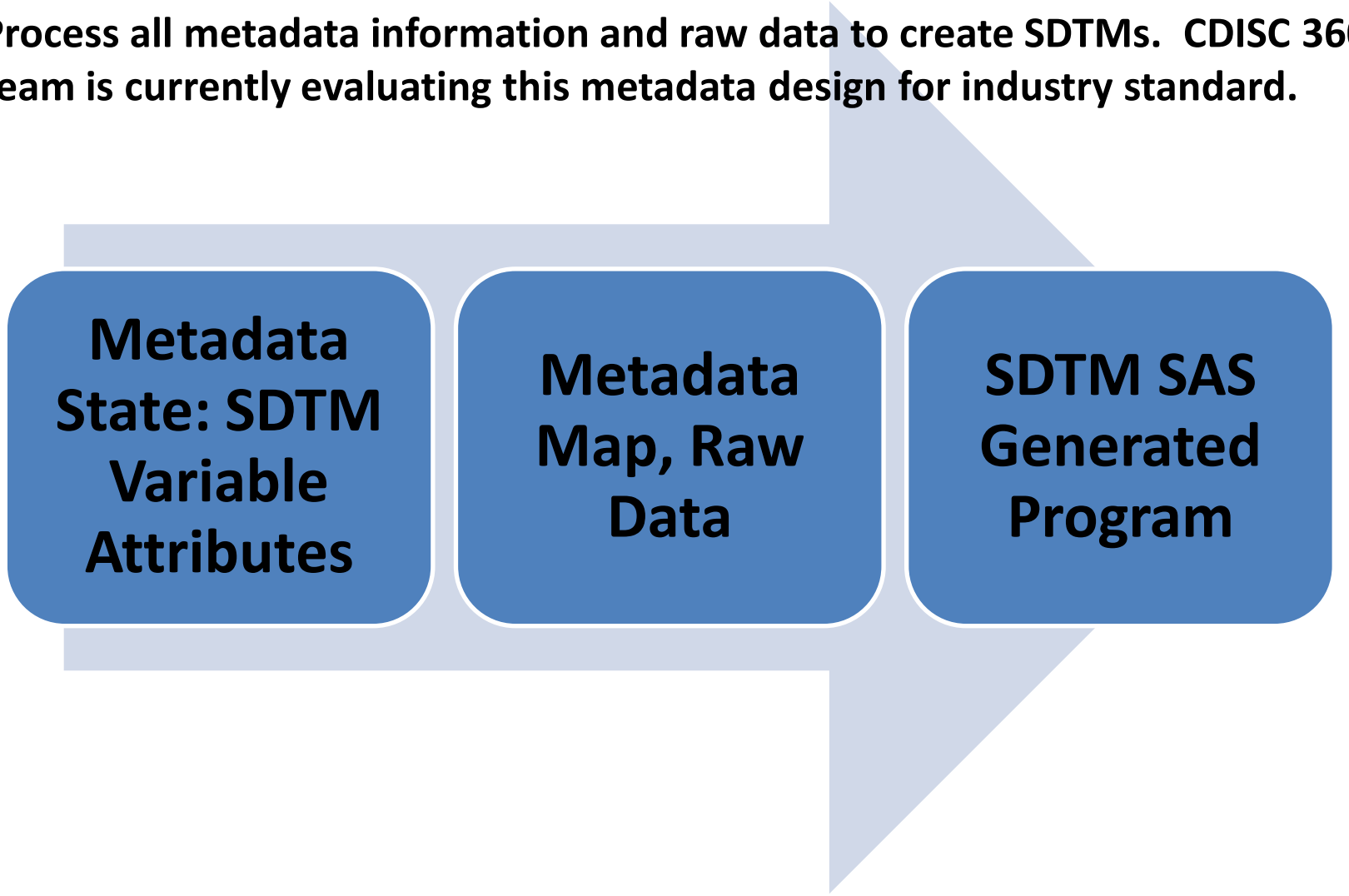
Data Transfer Engine (DTE) Design

- ✓ 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.



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 ;
*.....................................................................;
* Define the length of each column;
*.....................................................................;
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";
label AESEQ = "Sequence Number";
label POOLID = "Pool Identifier";
label AEGRPID = "Group ID";
label SPDEVID = "Sponsor Device Identifier";
label AEREFID = "Reference ID";
label AESPID = "Sponsor-Defined Identifier";
```

SDTM/ADaM Attributes are already industry best practices

Variable Length

Variable Label

PUT() with format catalog to convert raw to SDTM control terms

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."

TALENT MINE
KNOW WHERE TO LOOK™