How do companies use metadata driven programming for SDTM, ADaM, and Define XML Creation?

PhUSE Single Day Event - Bridgewater

Agenda

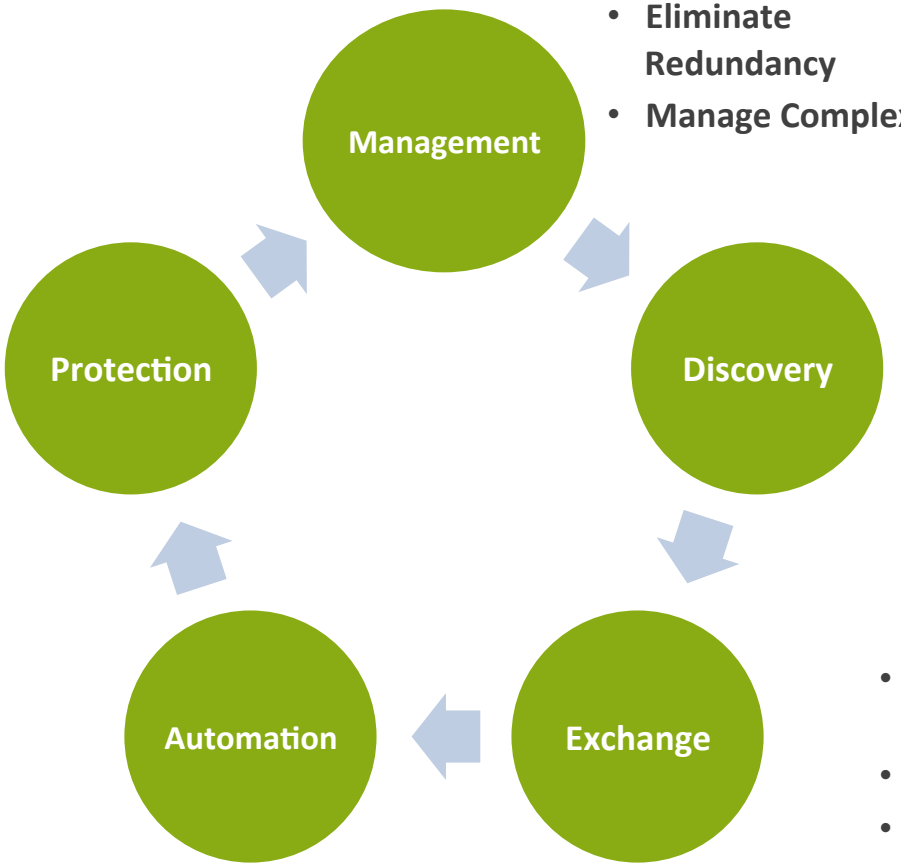
- **Traceability and Metadata**
- **Metadata Flow**
- **Metadata Development**
- **Metadata Enabled Data Processing**
  - Legacy Data Conversion
  - End-to-End Data Management
- **Metadata Management**
- **Implementation Example**
- **Discussion**

Traceability Ensures Confidence

- Traceability is a major concern for the FDA. Traceability means understanding the relationship between the results and the source. It facilitates transparency and is an essential component for building confidence in all stages of clinical information processing.
- Building traceability is establishing the path between an element and its immediate predecessor.
- Metadata Traceability: provides an explanation of how any down-stream information has been created from its predecessors by describing the algorithms and deviations used.
- Data Point Traceability: provides data lineage of how down-stream records are related to the predecessor records(s): are the records originating from source or are these derived records.

Importance of Metadata

- Improve Accuracy
- Improve Freshness
- Eliminate Redundancy
- Manage Complexity



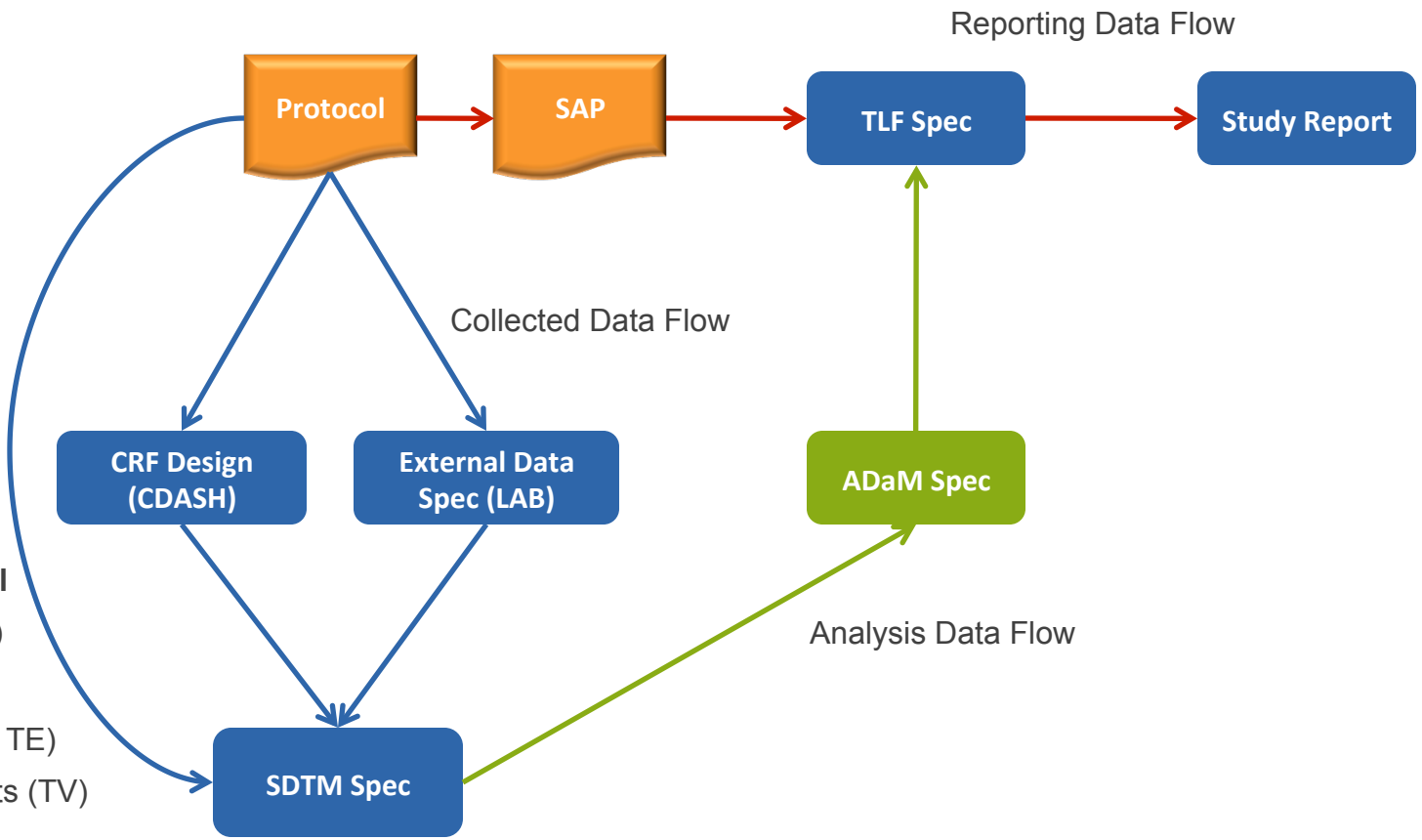
- Offer Traceability
- Facilitate Distribution
- Improve Searchability
- Encourage Usage

- Support Platform Independency
- Facilitate Collaboration
- Enhance Information Sharing
- Avoid Vendor Lock-in

- Improve Efficiency
- Reduce Chance of Error
- Encourage Code Reuse

- Enhance Entitlement Management
- Enhance Information Security

Metadata Flow in Clinical Trial Data Processing



Trial Design Model

- Registration (TS)
- In/Exclusion (TI)
- Trial Design (TA, TE)
- Scheduled Events (TV)

Metadata Development Framework

Standards

- Establish Conceptual Models
- Establish CRF, SDTM, and ADaM Standards
- Store in MDR
- Governed and Maintained

Library

- Establish Libraries at Different Levels in Different Areas
- Library would grow to incorporate new scenarios
- CRF Libraries, SDTM Templates, ADaM Templates, TLF Templates

Design

- Extract Metadata from Protocol and SAP
- Study Activities drive CRF Design
- Study Endpoints drive Variable Derivation
- Study Analysis drives TLF Design, ADaM Structure

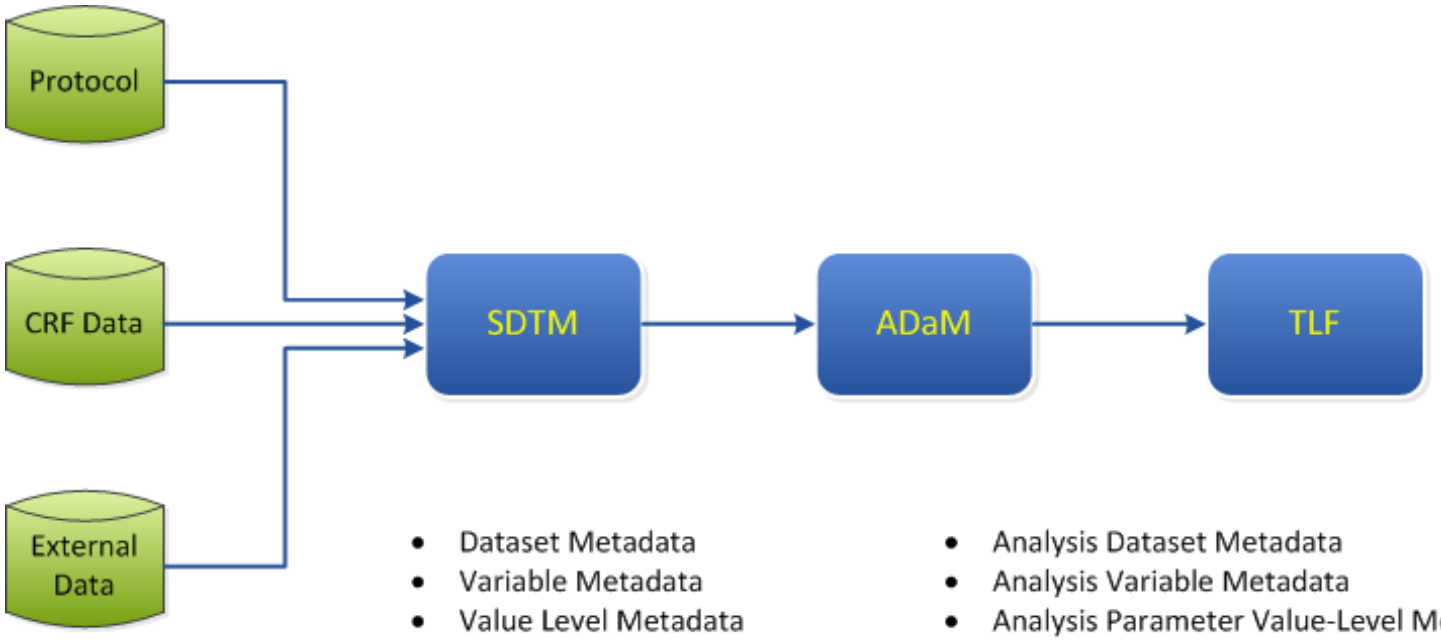
Customize

- Study Specific Configuration
- Conformance and Deviation can be monitored
- Mechanism exists to publish to Library

Metadata Development

- **Legacy Data Conversion**
  - Convert existing studies to follow CDSIC standards
  - Lessor initial efforts, lessor disruption of existing processes
  - Lessor dependency on CDISC standards during study startup
  - Lessor control over metadata lineage
  - More manual and repeated work to develop metadata
- **End-to-End Metadata Management**
  - Ideal situation, but requires significant initial efforts
  - Depends on CDISC standard development
  - Study Design -> Study Conduct -> Study Analysis -> Study Submission
  - Best utilization of metadata
    - Consistency
    - Accuracy
    - Efficiency

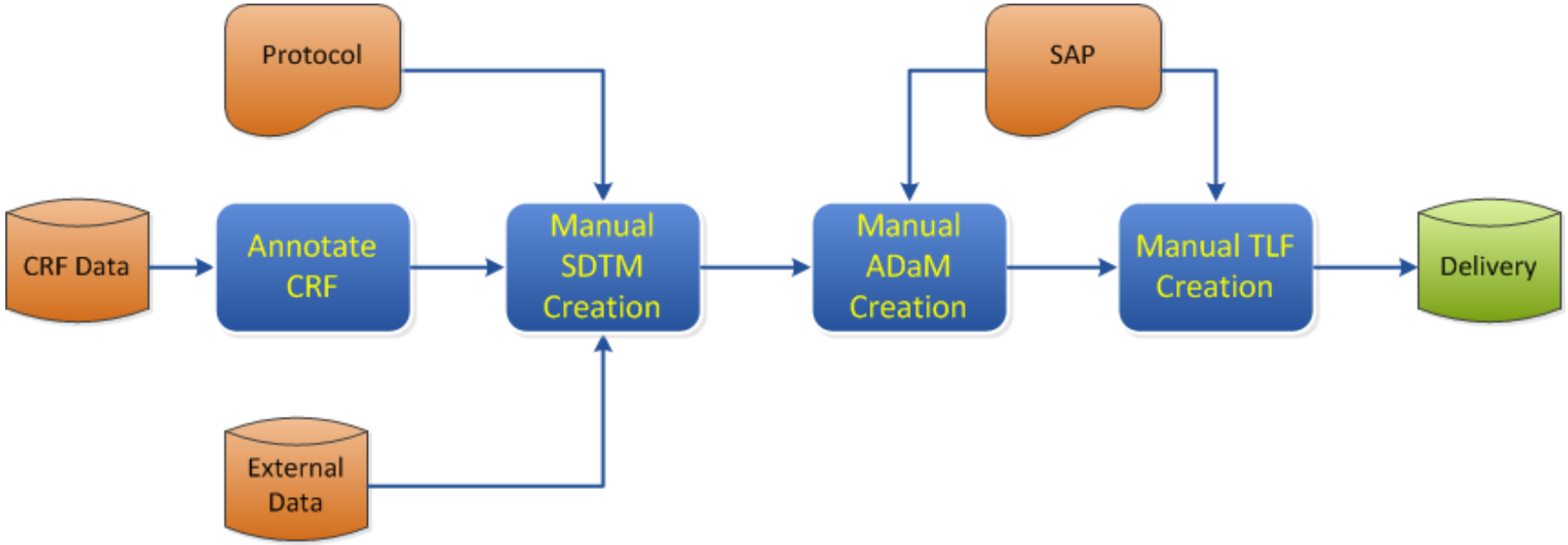
Required Metadata – Legacy Data Conversion



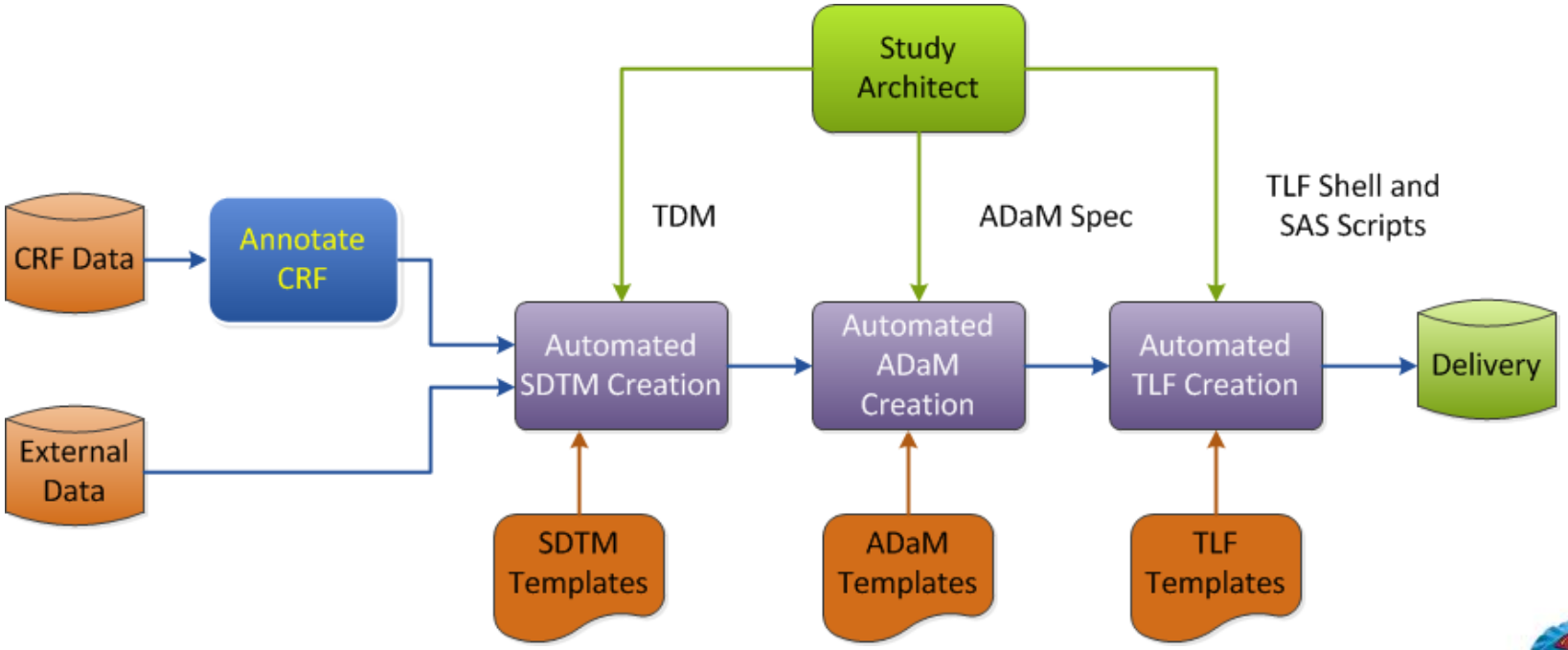
- Dataset Metadata
- Variable Metadata
- Value Level Metadata
- Controlled Terms
- Computational Algorithms
- External Dictionaries

- Analysis Dataset Metadata
- Analysis Variable Metadata
- Analysis Parameter Value-Level Metadata
- Analysis Results Metadata

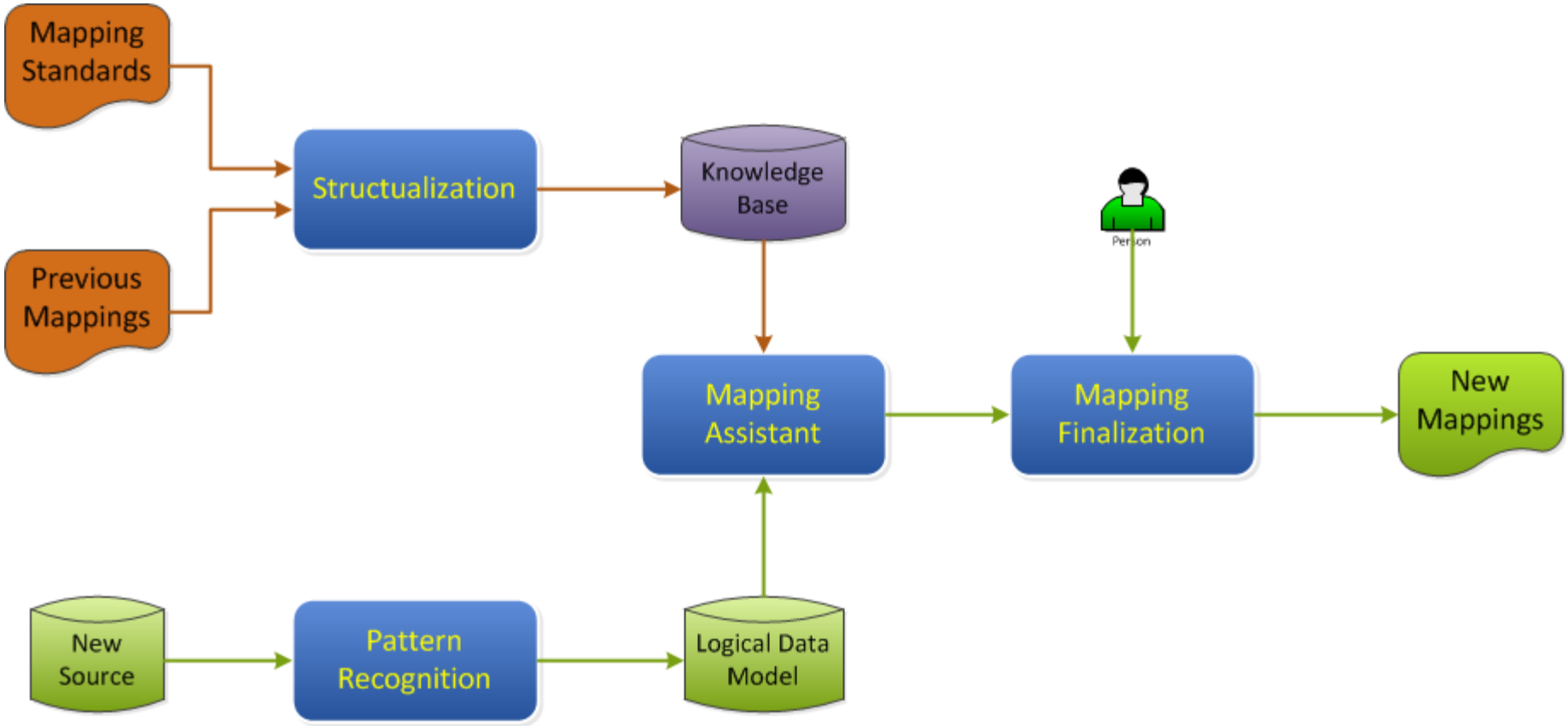
Typically It's a Manual Process



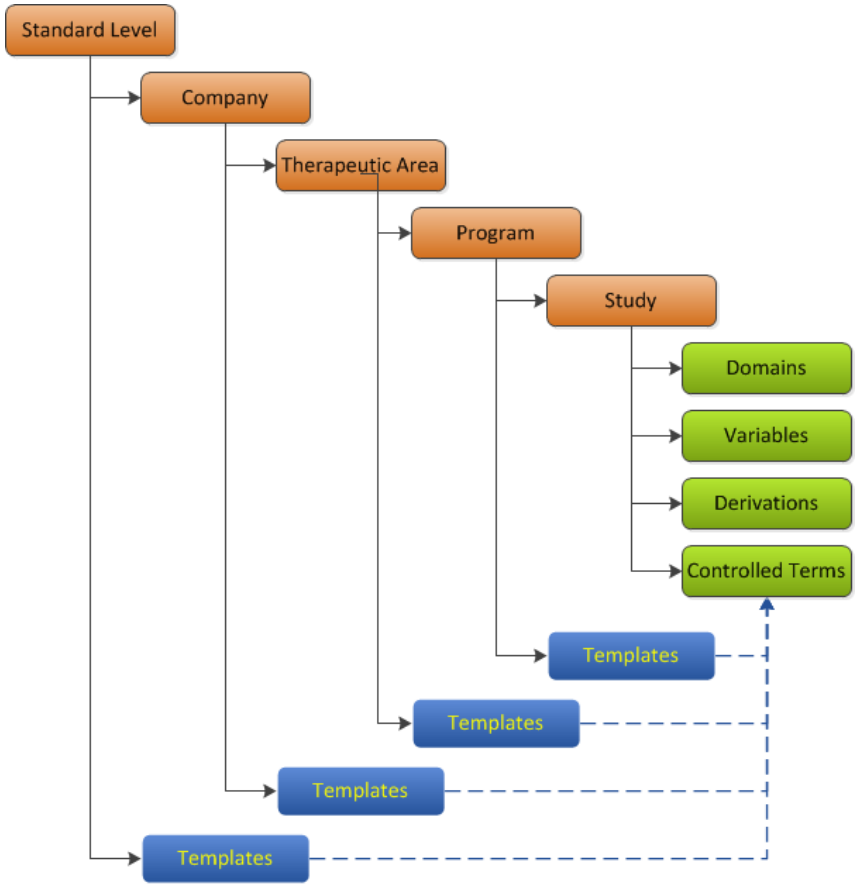
Metadata Empowered Legacy Data Conversion



KNOWLEDGE-BASE ASSISTED DATA STANDARDIZATION



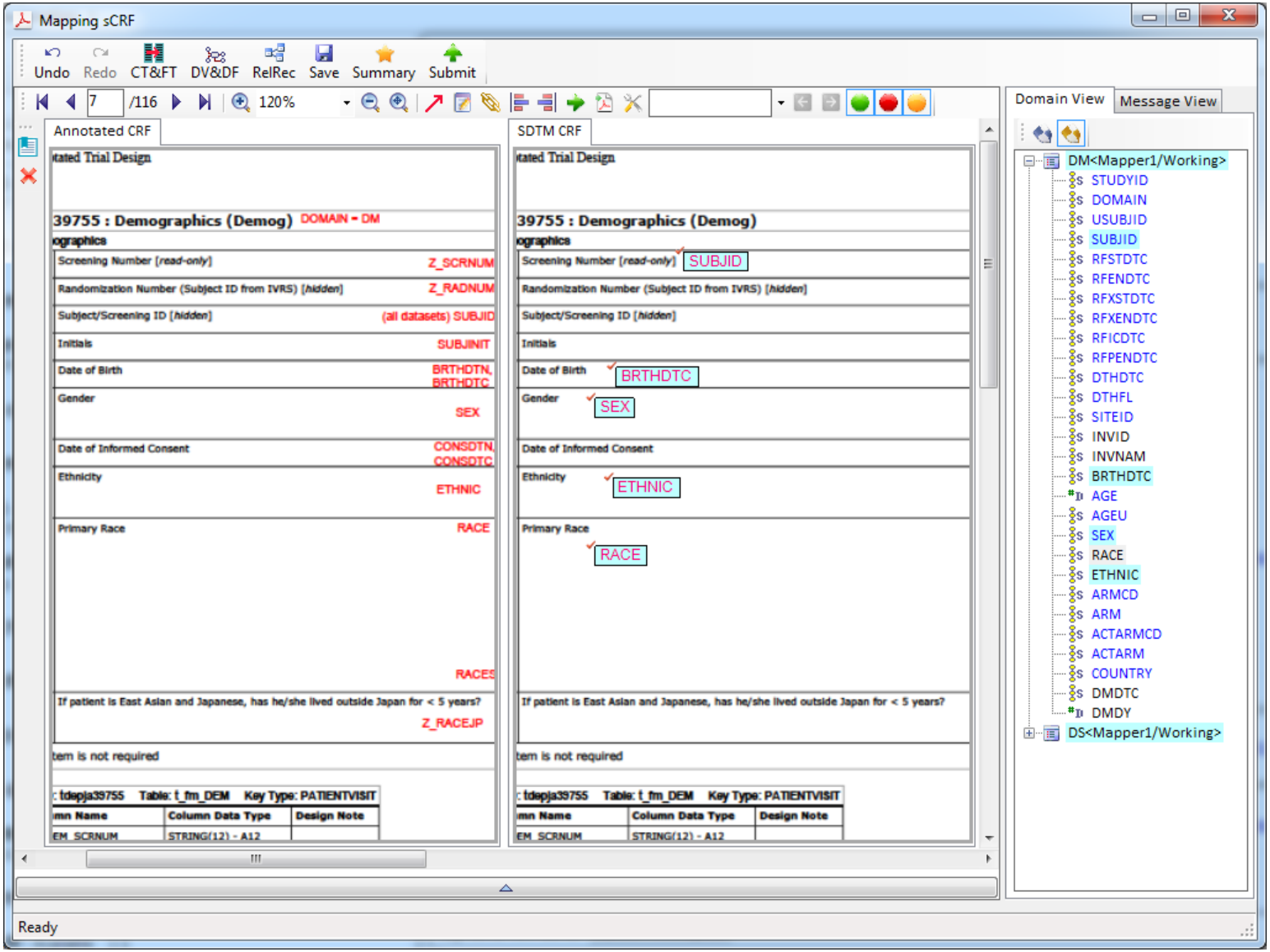
METADATA MANAGEMENT



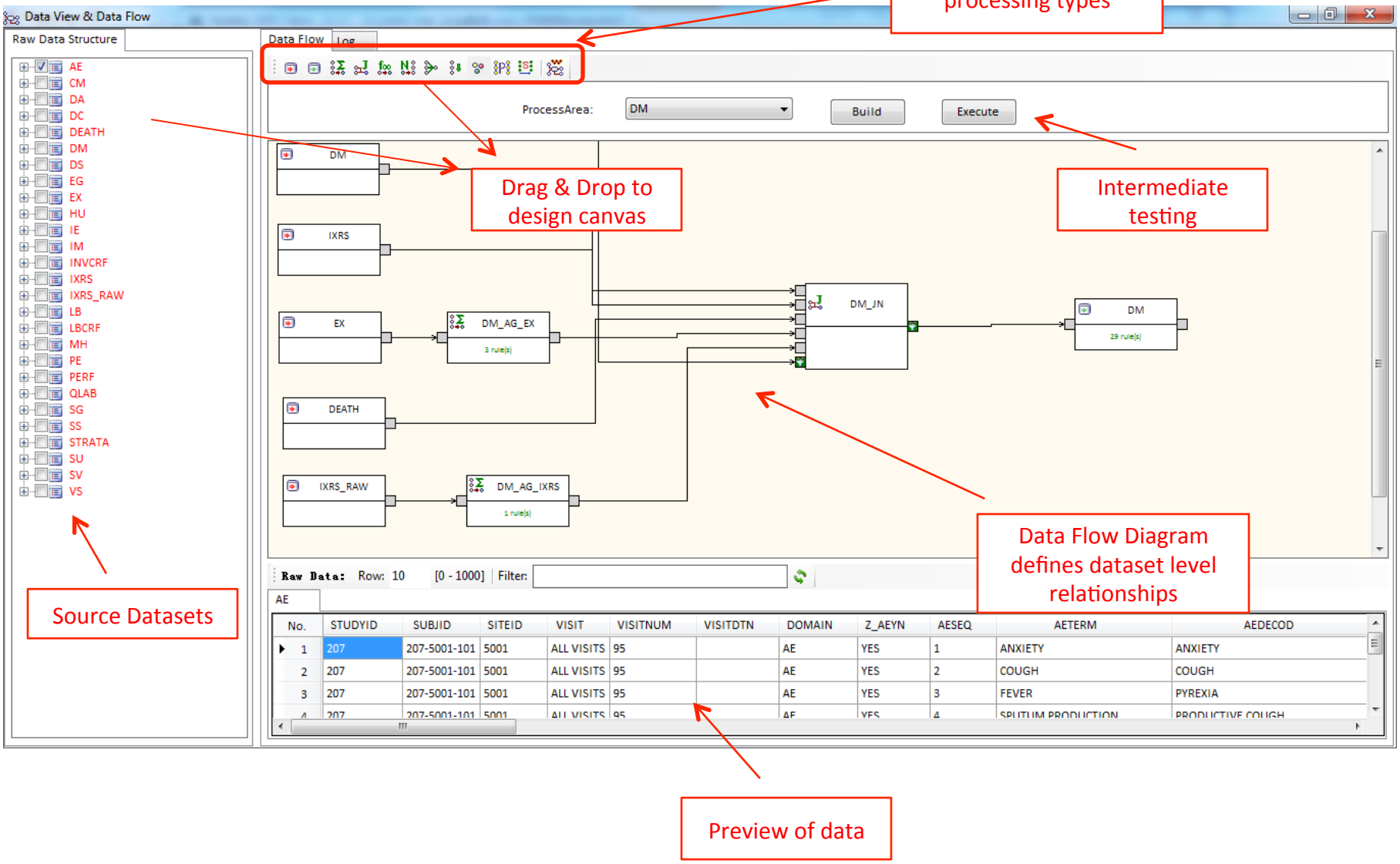
Metadata Repository Structure

- Templates can be used for mapping reuse
- Hierarchical structure help to provide granular standardization while maintain consistency
- New derivations can be generated based on auto-matching of standards/ templates at different levels
- Metadata need to be managed by dedicated team and governing processes

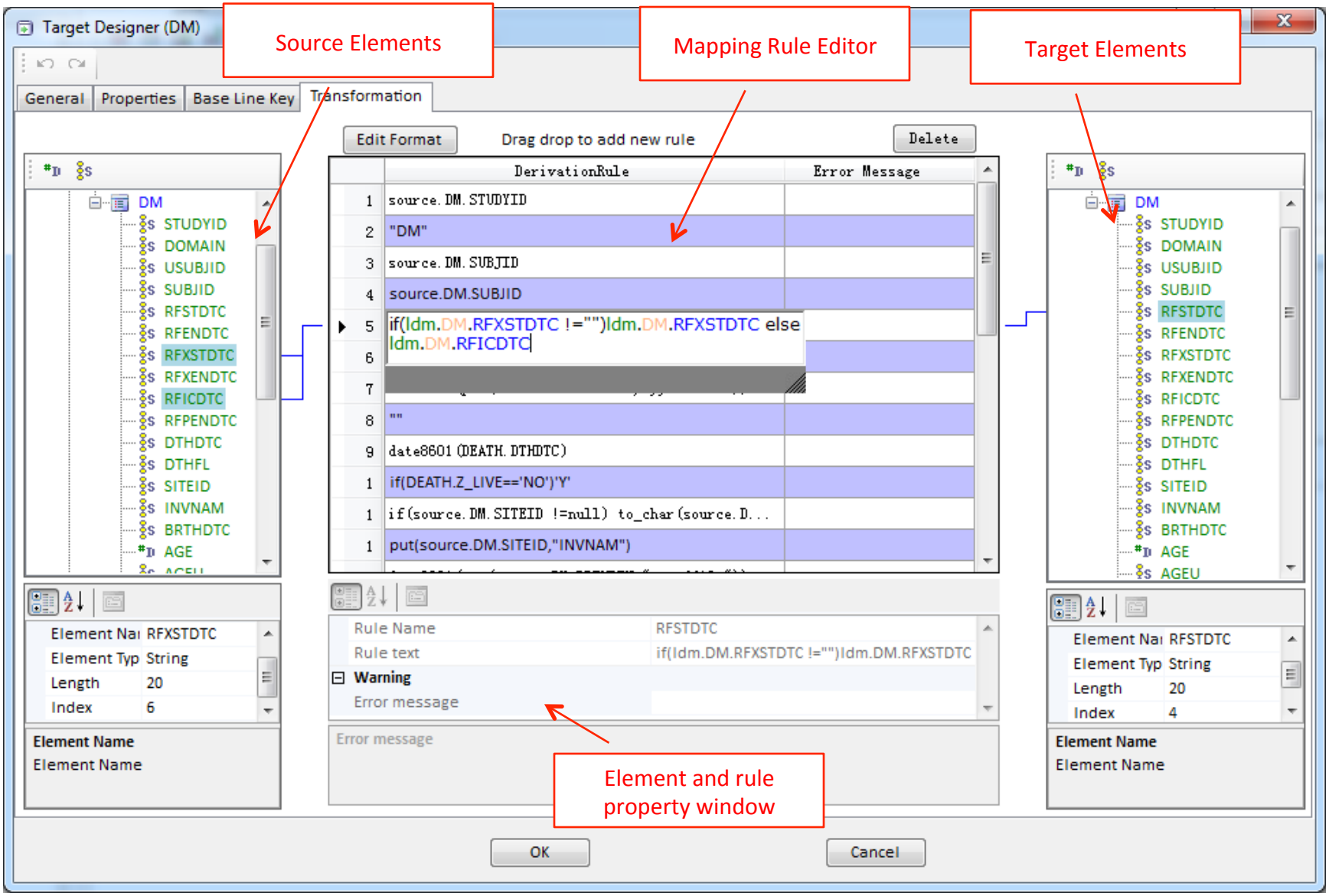
SDTM ANNOTATED CRFS



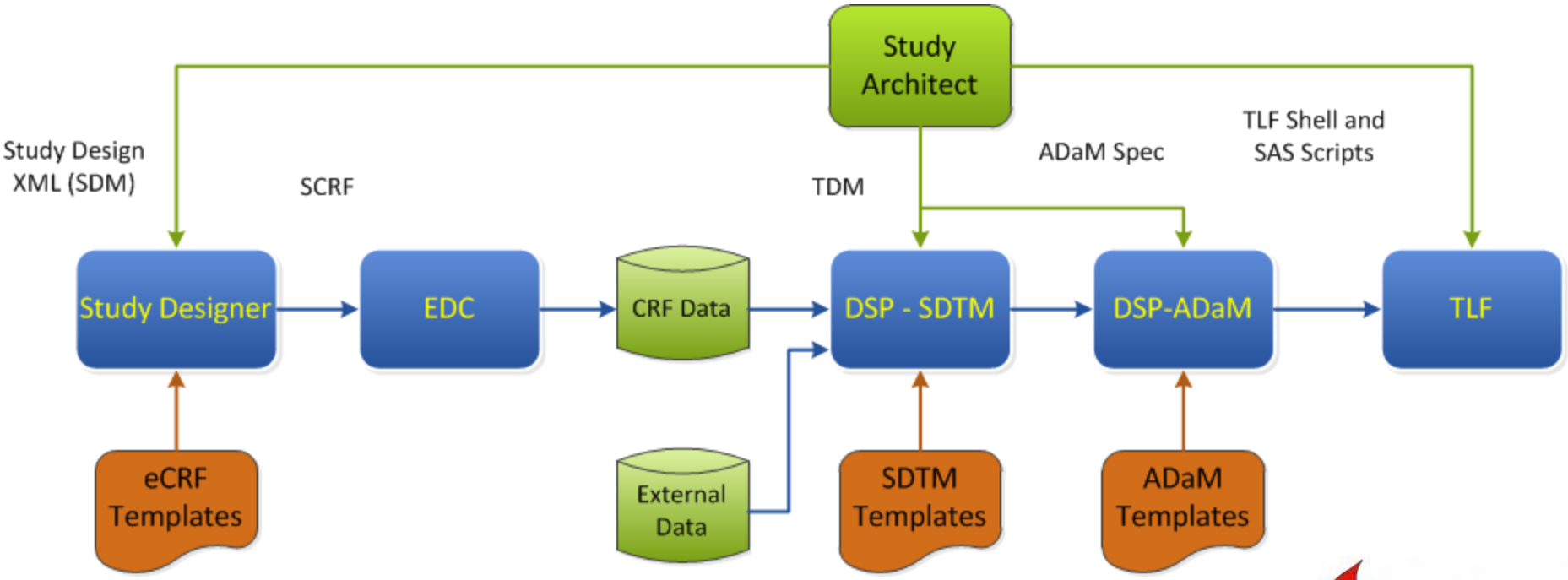
Data Flow Construction



VARIABLE LEVEL TRANSFORMATION DEFINITION

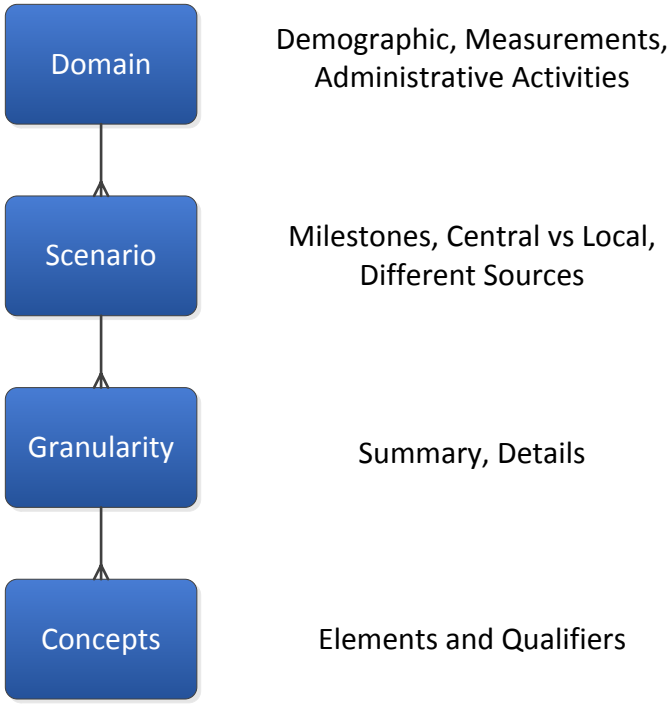


Metadata Empowered End-to-End Metadata Flow

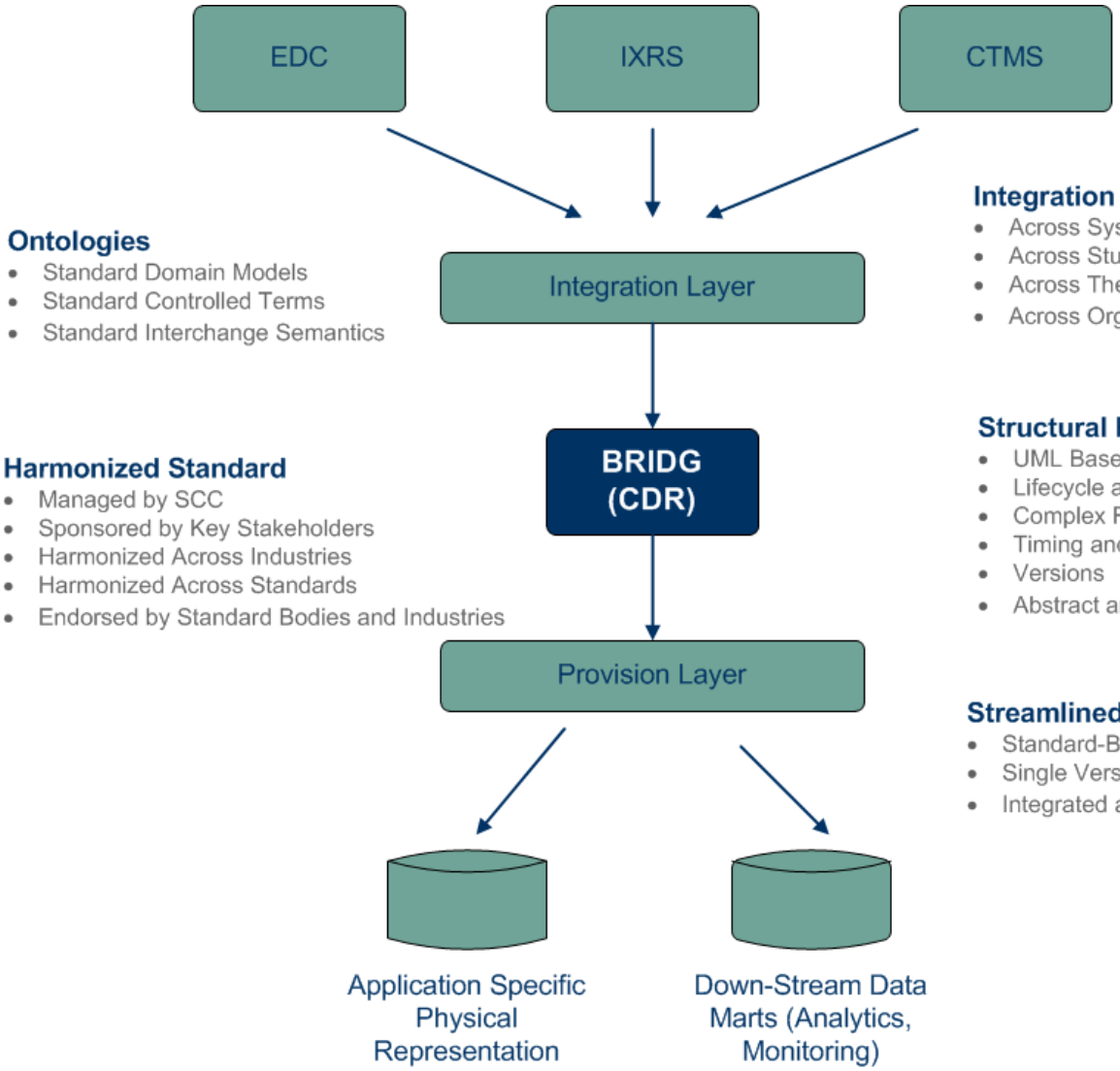


Step 1 – Establish Standards

- **Define Conceptual Model**
  - Leverage BRIDG concept
  - Cover all data origins
    - Trial Design
    - Protocol
    - Collected Subject Data
    - Derived Subject Data
  - Cover all trial stages
    - Study Design (TDM -> SDTM)
    - Study Conduct (CDASH -> SDTM)
    - Study Analysis (ADaM)
    - Study Reporting (TLF)
  - Use Controlled Term to add semantics
- **Define CRF, SDTM, and ADaM Standards**



WHY BRIDG?



Ontologies
- Standard Domain Models
- Standard Controlled Terms
- Standard Interchange Semantics

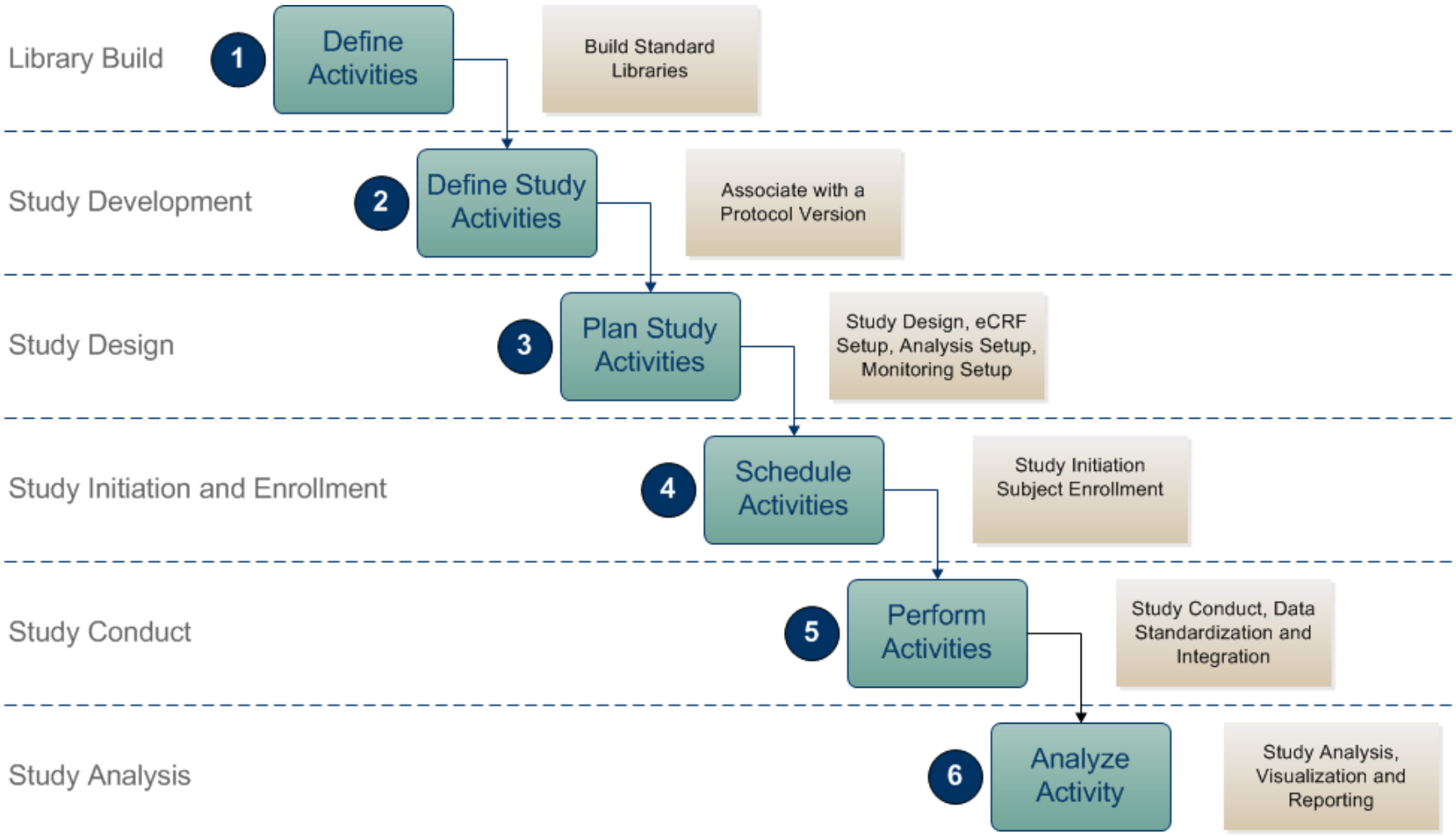
Integration
- Across Systems
- Across Study Stages
- Across Therapeutic Areas
- Across Organizations

Harmonized Standard
- Managed by SCC
- Sponsored by Key Stakeholders
- Harmonized Across Industries
- Harmonized Across Standards
- Endorsed by Standard Bodies and Industries

Structural Rigor
- UML Based Modeling
- Lifecycle and Context
- Complex Relationship
- Timing and Workflow
- Versions
- Abstract and Complex Data Types

Streamlined Provisioning
- Standard-Based Metadata-Driven
- Single Version of Truth
- Integrated and Streamlined

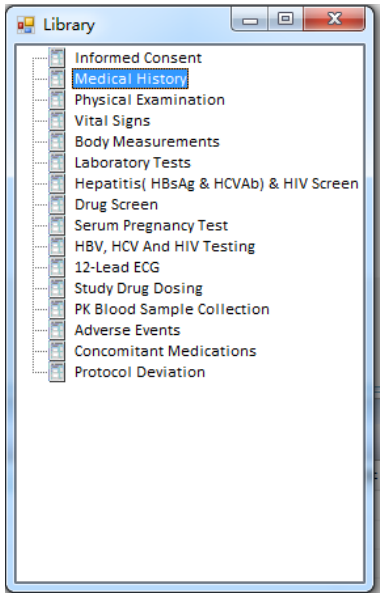
MANAGING STUDY USING BRIDG



Step 2 – Develop Libraries

- **Libraries at Different Levels**
  - Company
  - Therapeutic Area
  - Program
- **CRF Library**
  - Standard CRFs
  - Cover different scenarios
- **SDTM Template**
  - Standard Structure
  - Standard raw -> SDTM mapping
- **ADaM Template**
  - Standard Structure
  - Standard Derivations
- **TLF Template**
  - Break-down to Standard Units

STEP 2 : CREATE COMMON LIBRARIES – eCRF & TLF



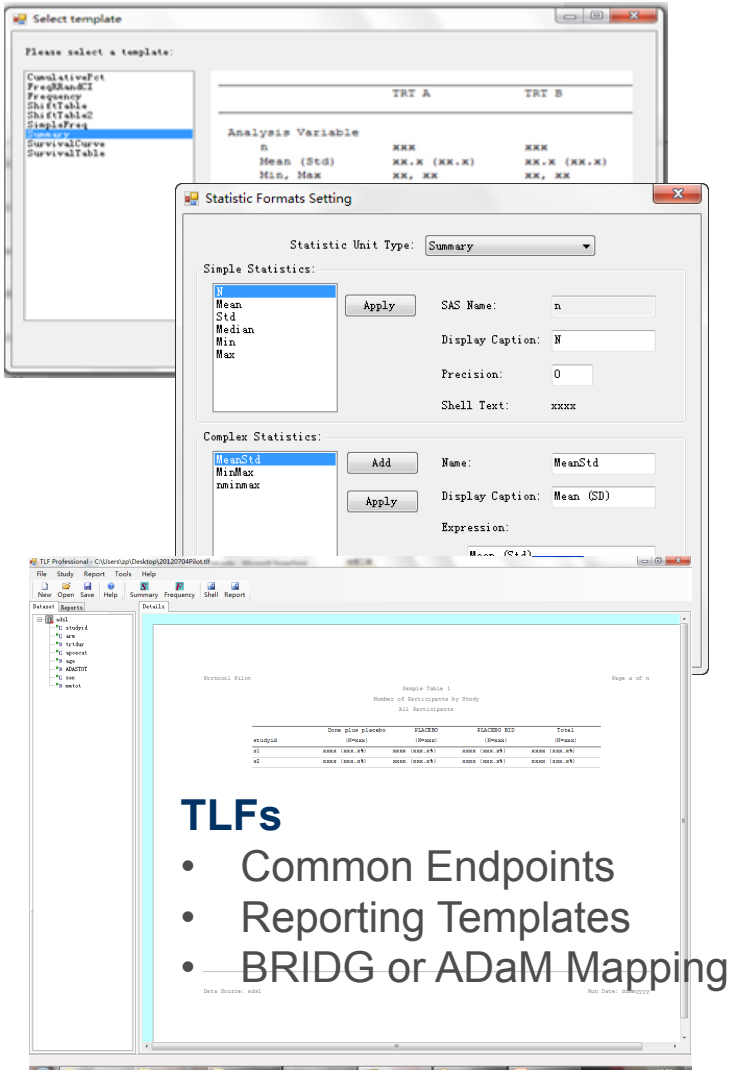
eCRF:

- Common Activities
- Associated Forms
- BRIDG or CDASH Mappings

MEDICAL HISTORY

Does the subject have any past clinically significant medical conditions? ◯ Yes ◯ No

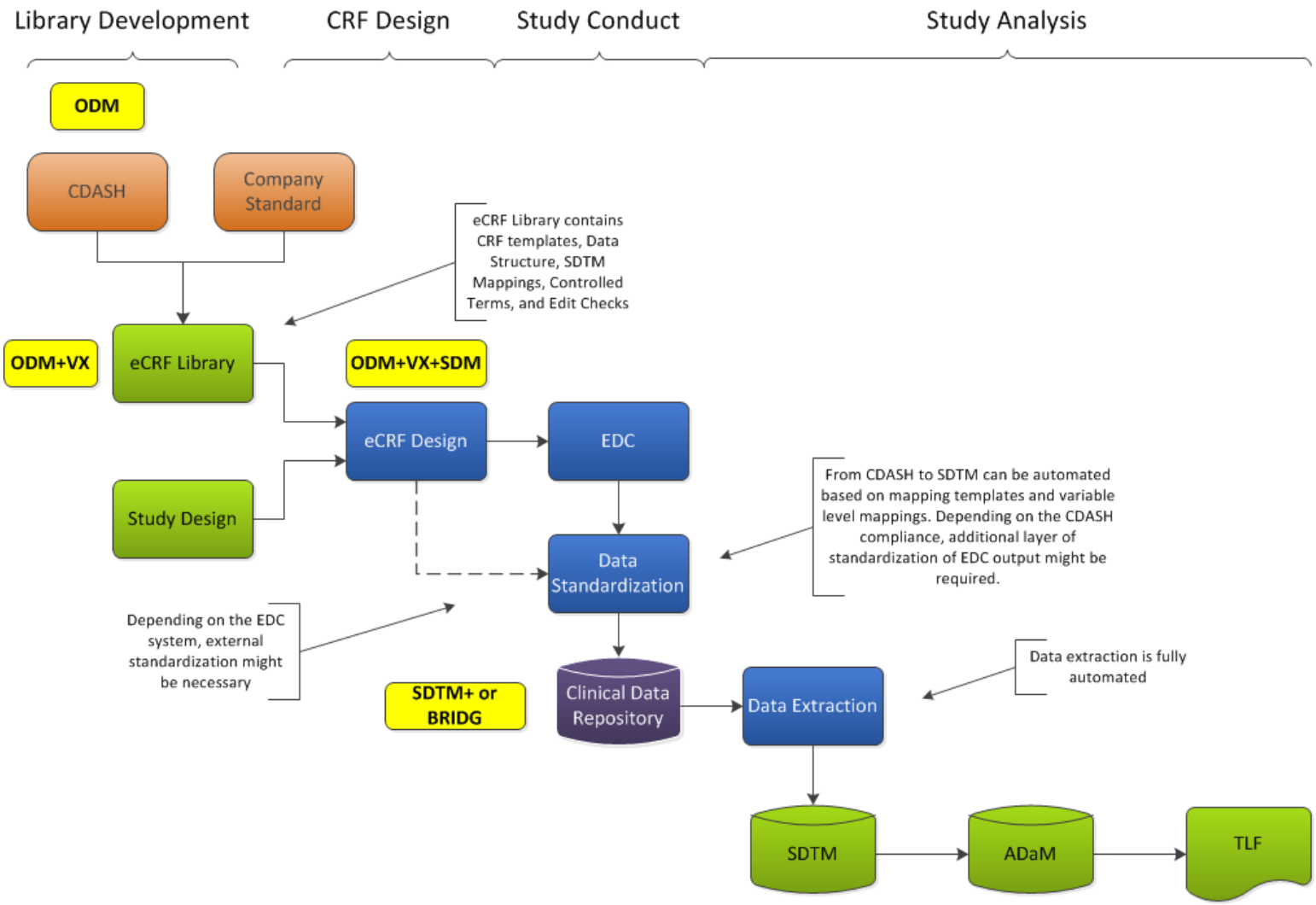
MH Number	Diagnosis/Procedure	Start Date (DD/MMM/YYYY)	End Date (DD/MMM/YYYY)	Ongoing
☐☐		☐☐-☐☐☐-☐☐☐☐	☐☐-☐☐☐-☐☐☐☐	☐



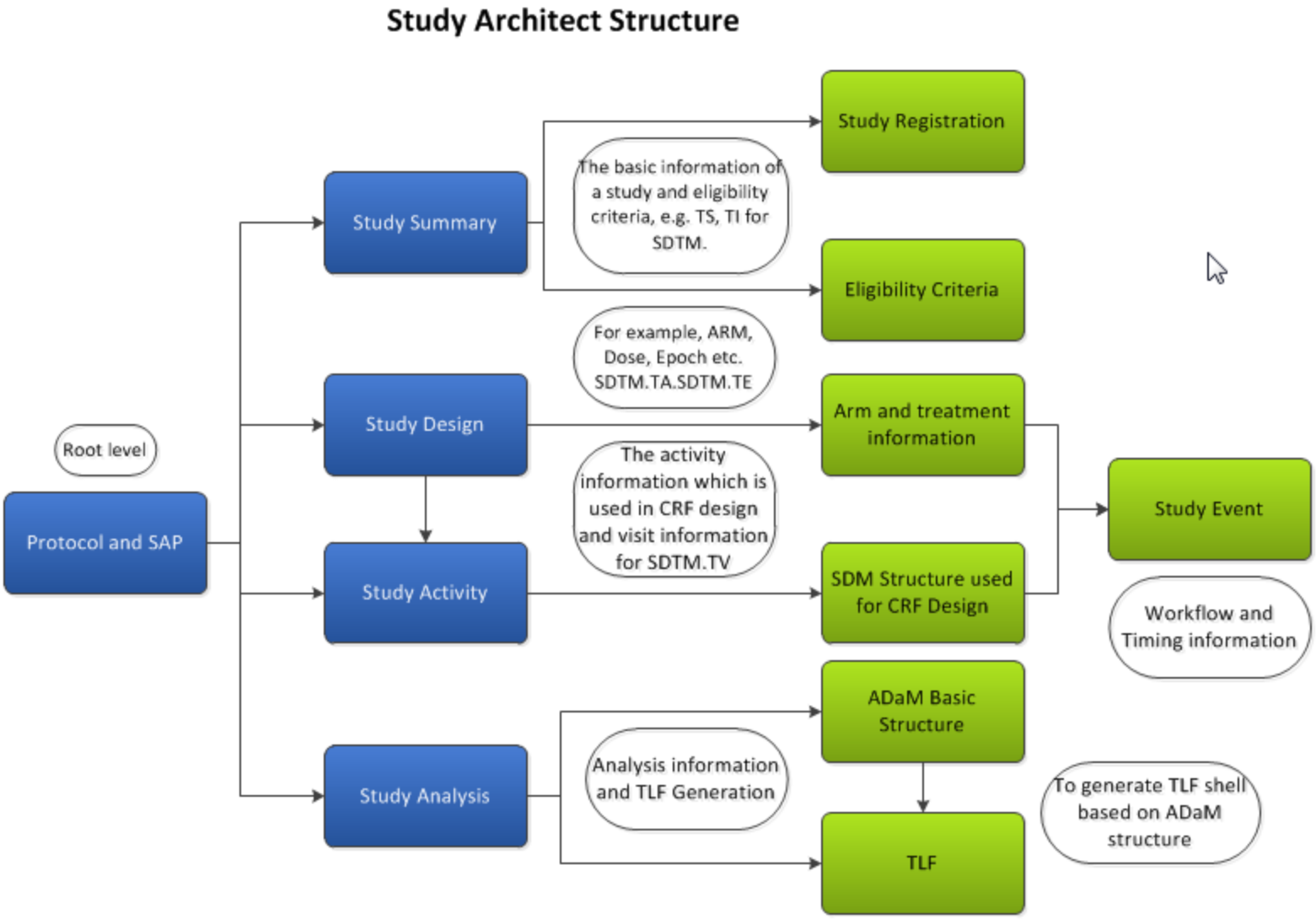
TLFs

- Common Endpoints
- Reporting Templates
- BRIDG or ADaM Mappings

Flow from Data Collection to Study Analysis



Step 3 – Study Design using Study Architect



STEP 3 : STUDY DEVELOPMENT - PROTOCOL

1.0 PROTOCOL SYNOPSIS

Study Title:	A Phase 1, Open-Label, Two-Period, Crossover Study to Evaluate the Relative Bioavailability of a New Tablet Formulation Versus the Current Capsule Formulat PPI-668 Administered in Healthy Adult Volunteers
Protocol Number:	PPI-668-103
Study Phase:	Phase 1
Study Objectives:	**Primary Objective:** 1. To compare the relative bioavail of PPI-668 when administered as a sin oral dose in a tablet formulation vs. in a capsule formulation **Secondary Objective:** 1. To evaluate the safety and tolerability of single oral doses of PPI-668 in health subjects
Study Design and Duration of Treatment:	This will be an open-label, single-center, single-dose, crossover study. A total of 24 healthy subjects will be enrolled. On Day 1, under a 1:1 randomization scheme eligible volunteers, 12 subjects will receive a single oral dose of PPI-668 in the fo the current capsule formulation and 12 subjects will receive a single oral dose of t test tablet formulation of PPI-668, under fasted conditions. After a one week wash period, on Day 8, subjects will receive under fasted conditions the alternate PPI-6 formulation to the one they received on Day 1. Study medication will be adminis approximately 0800 hours on Days 1 and 8 after an overnight fast of at least 10 ho Screening will take place within 28 days before Study Day 1. Eligible subjects wil

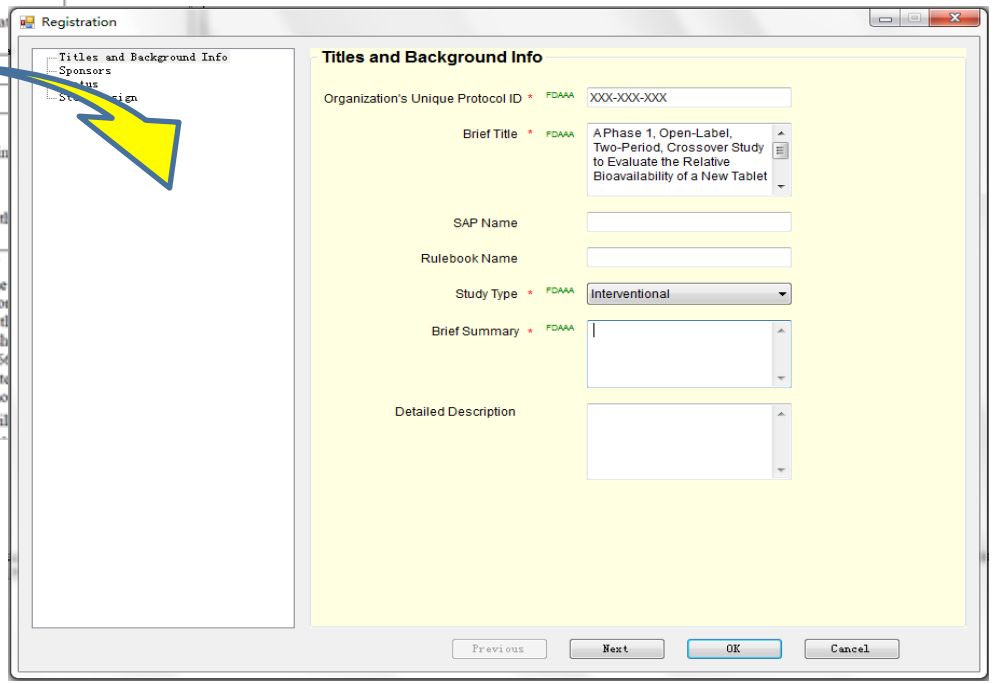


Table 7-1 Schedule of Study Procedures

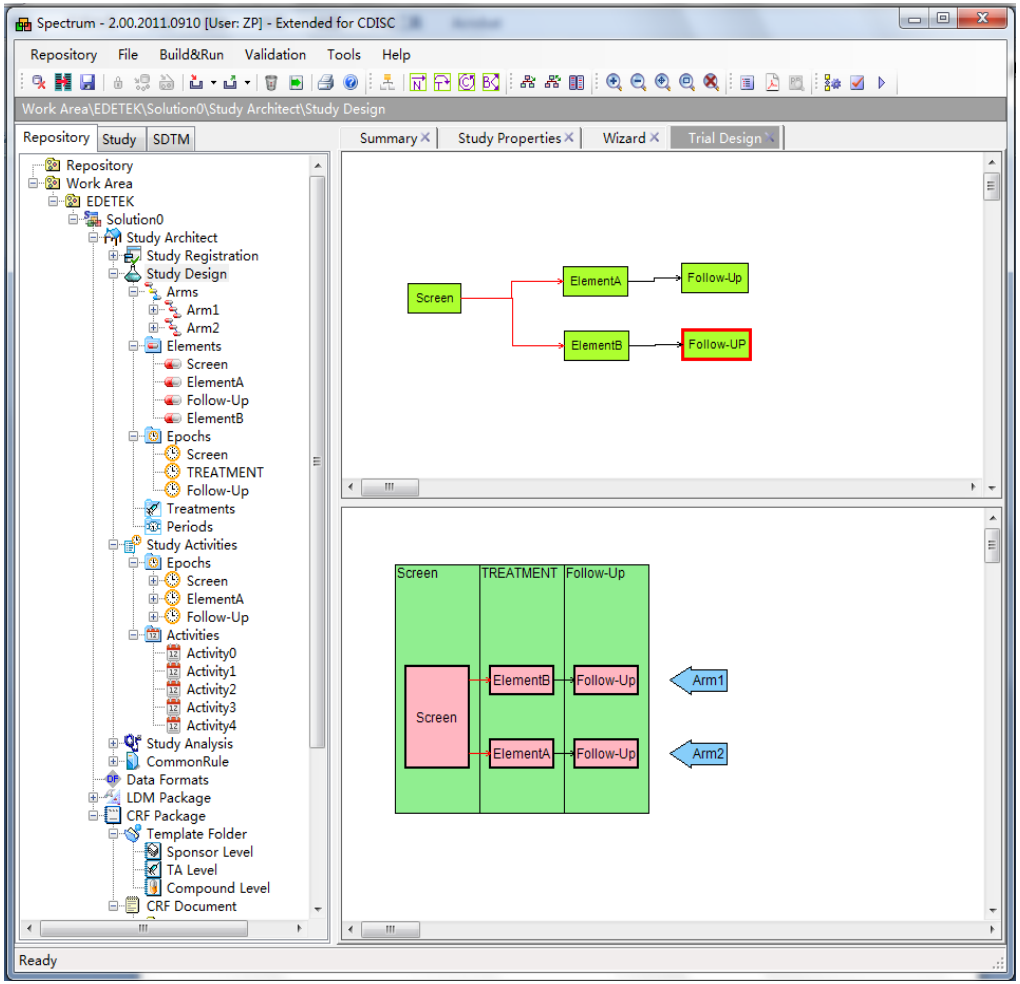
	Screening[a]	Day -1	Day 1	Day 2	Day 3	Wash-out	Day 7	Day 8	Day 9	Day 10	Day 15 ± 3 days Follow-up
Written Informed Consent	X										
Medical History	X										
Complete Physical Examination	X										X
Brief Physical Examination		X					X				
Vital Signs[b]	X	X			X		X			X	X
Height	X										
Weight	X	X					X				
Hematology[c]	X	X					X				X
Chemistry[d]	X	X					X				X
Urinalysis	X	X					X				X
Urine Drug and Breath Alcohol Screening	X	X					X				
Pregnancy	X	X					X				X

Study Protocol Versions are developed, and are associated with Defined Activities

Step 3 : Study Design – Data Collection

- Define Arms
- Define Epochs
- Define Segments
- Define Cells
- Define Study Events
- Define Workflow

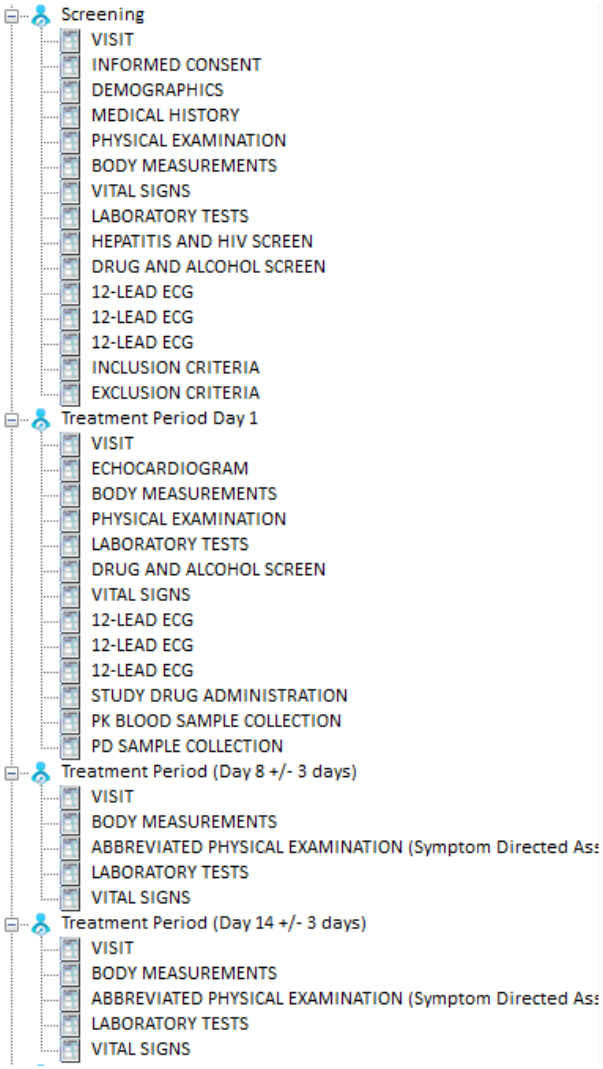
Study Design Model (SDM)



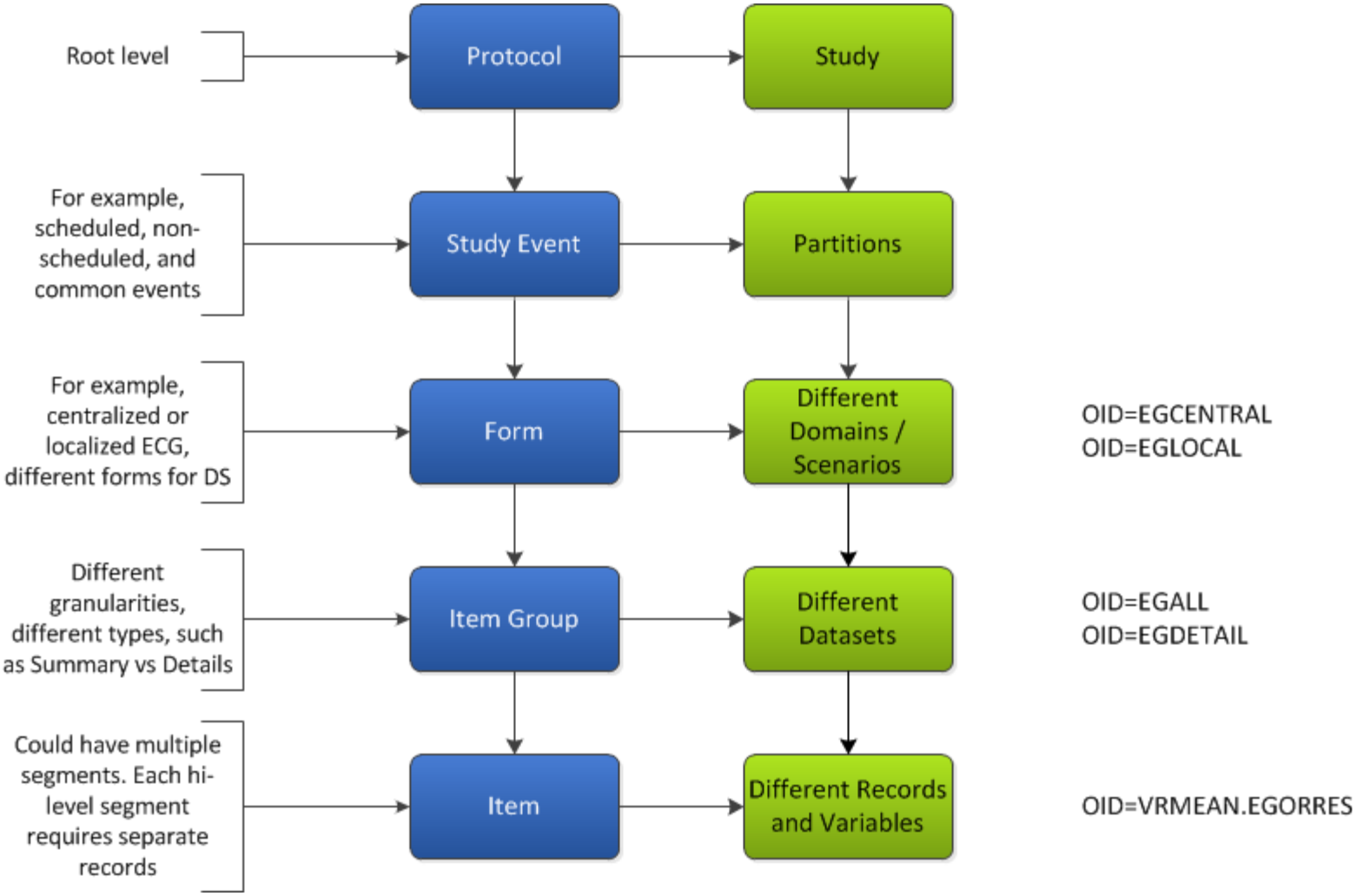
STUDY ACTIVITY FROM SA TO CRF

Epoch	Screening						
Activity\Event	Screening	Treatment Period Day 1	Treatment Period (Day 8 +/- 3 days	Treatment Period (Day 8 +/- 3 days	Treatment Period (Day 14 +/- 3	Treatment Period (Day 21 +/-3	Treatment Period (Day 28+/- 3 day
INFORMED CONSENT	X						
DEMOGRAPHICS	X						
MEDICAL HISTORY	X						
PHYSICAL EXAMINATION	X	X					
BODY MEASUREMENTS	X	X	X	X	X	X	X
VITAL SIGNS	X	X	X	X	X	X	X
LABORATORY TESTS	X	X	X	X	X	X	
HEPATITIS AND HIV SCREEN	X						
DRUG AND ALCOHOL SCREEN	X	X					X
12-LEAD ECG	X	X					
INCLUSION CRITERIA	X						
EXCLUSION CRITERIA	X						
ECHOCARDIOGRAM		X					
STUDY DRUG ADMINISTRATION		X					
PK BLOOD SAMPLE COLLECTION		X					X
ABBREVIATED PHYSICAL EXAMINATION (Symptom Directed Assessment)		X	X	X	X	X	X
PD SAMPLE COLLECTION		X					X
STUDY COMPLETION							
ADVERSE EVENTS							
CONCOMITANT MEDICATIONS							

CRF Design Structure from SA



STANDARDIZED CDASH OUTPUT STRUCTURE



CDASH OID Importance

- **Map to unique scientific concepts (logical model element)**
- **Item Group (IG) defines datasets**
  - Develop standard naming convention to cover different scenarios
  - Each name should represent unique set of contents
- **Item defines variable names and hierarchies**
  - Leveraging multiple segment names
    - Category, Sub-category, Test Code, Variable Name
  - Use CDISC Controlled Terminology to convey semantic information

DM CRF Design

DEMOGRAPHICS

OID=F.DM

OID=IG.DM

Date of Birth ☐☐ - ☐☐☐ - ☐☐☐☐ (Day - Month - Year)

OID=I.BRTHDAT

Sex

○ Male

○ Female

OID=I.SEX

Race (Check all that apply)

☐ White

☐ Black/African-American

☐ Asian

☐ Native Hawaiian/Other Pacific Islander

☐ American Indian/Alaska Native

☐ Other Specify

OID=I.RACE

OID=I.RACEOTH

Ethnicity

○ Hispanic

○ Not Hispanic

OID=I.ETHNIC

CDASH DM Dataset Obtained

BRTHDAT	ETHNIC	RACE1	RACE2	RACE3	RACE4	RACE5	RACE6	RACEOTH	SEX
23-MAR-1968	2					1			M
27-DEC-1982	2					1			M
15-DEC-1987	2			1					M
06-JAN-1976	2			1					M
14-AUG-1990	2					1			M
13-JAN-1986	1					1			M
20-SEP-1994	2			1					M
25-JUL-1976	2			1					F

EG CRF Design

12-LEAD ECG OID=F.EG

In triplicate after 5 minutes of resting in the supine position.

OID=IG.EGPERF

☐ Not Done OID=I.EGPERF		
OID=IG.EGDETL Date of ECG	__ __ - __ __ __ - __ __ __ __ (Day - Month - Year)	OID=I.EGDAT
Actual Time	__ __ : __ __ (Hour : Minute)	OID=I.EGTIM
Ventricular Rate	__ __ __ (beats/min)	OID=I.VRMEAN.EGORRES
PR Interval	__ __ __ (msec)	OID=I.VRMEAN.EGORRES
RR Interval	__ __ __ __ (msec)	OID=I.RRMEAN.EGORRES
QRS Interval	__ __ __ (msec)	OID=I.QRSDUR.EGORRES
QT Interval	__ __ __ (msec)	OID=I.QTC.EGORRES
QTcF Interval	__ __ __ (msec)	OID=I.QTCF.EGORRES

CDASH EG Dataset Obtained

EGDAT	EGORRES	EGORRESN	EGORRESU	EGTEST	EGTIM
28-MAR-2014		165	(msec)	PR Interval	0918
28-MAR-2014		985	(msec)	RR Interval	0918
28-MAR-2014		90	(msec)	QRS Interval	0918
28-MAR-2014		410	(msec)	QT Interval	0918
28-MAR-2014		412	(msec)	QTcF Interval	0918
28-MAR-2014	1	.			0918
28-MAR-2014		60	(beats/min)	Ventricular Ra	0918
28-MAR-2014		178	(msec)	PR Interval	0920
28-MAR-2014		966	(msec)	RR Interval	0920
28-MAR-2014		88	(msec)	QRS Interval	0920
28-MAR-2014		411	(msec)	QT Interval	0920
28-MAR-2014		415	(msec)	QTcF Interval	0920
28-MAR-2014	1	.			0920
28-MAR-2014		62	(beats/min)	Ventricular Ra	0920
28-MAR-2014		170	(msec)	PR Interval	0922
28-MAR-2014		979	(msec)	RR Interval	0922
28-MAR-2014		86	(msec)	QRS Interval	0922
28-MAR-2014		407	(msec)	QT Interval	0922
28-MAR-2014		409	(msec)	QTcF Interval	0922
28-MAR-2014	1	.			0922
28-MAR-2014		61	(beats/min)	Ventricular Ra	0922

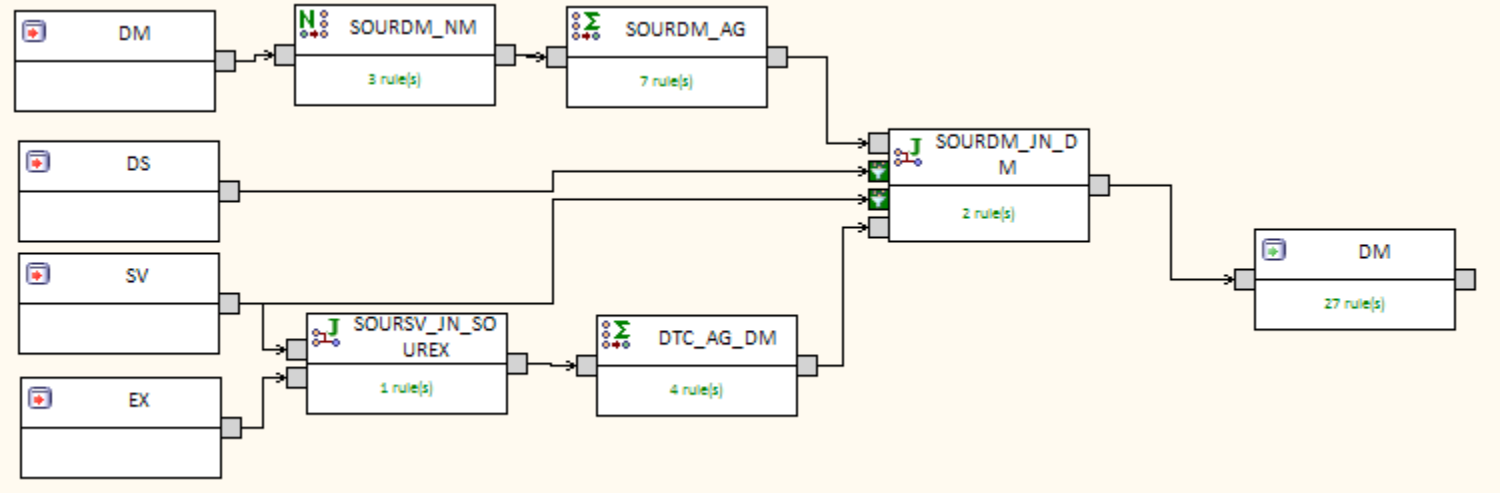
Auto-SDTM Creation

- **Match IG-based EDC Output with Partition-Level Templates**
  - Scenario based templates
  - Auto-matching based on available IG-based outputs
  - Auto-Pruning of extra template variables
- **Match Domain-Level Templates**
  - All partitions are merged together
  - Merged results are matched against domain-level templates
  - Finalization stage is applied based on domain class
    - Baseline calculation
    - --SEQ assignment
    - SUPPQUAL creation

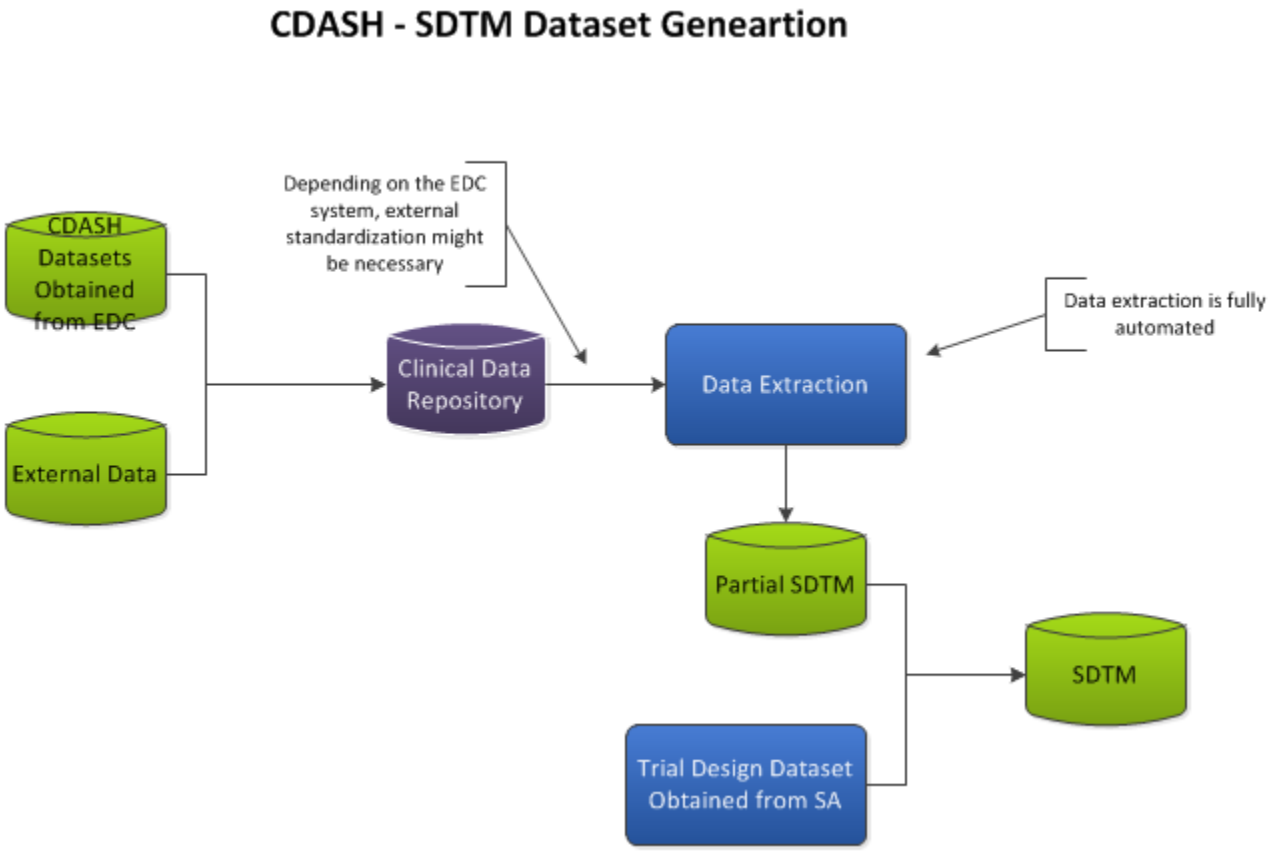
Variables of Protocol and Derived Origins

- **Non-standard derived variables have to be defined based on SAP or mapping standards.**
  - Baseline Definition
    - Default: Last non-missing observation before dosing
    - Average of repeated measurements
  - Reference Dates
    - First Dose/Last Dose Dates
  - Treatment Emergent AE
  - Last Participation Date
  - Unit Systems for Lab Standardization
- **Variables of Protocol Origin have to be defined**
  - Trial Design
  - Other protocol specified information not available elsewhere
- **These are specified at study design**

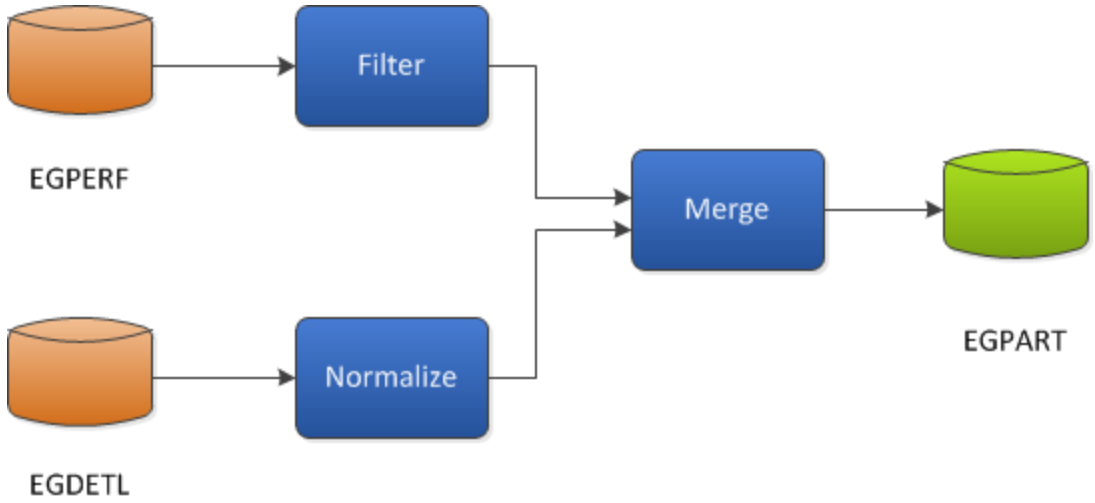
CREATING SDTM/ADAM TEMPLATES



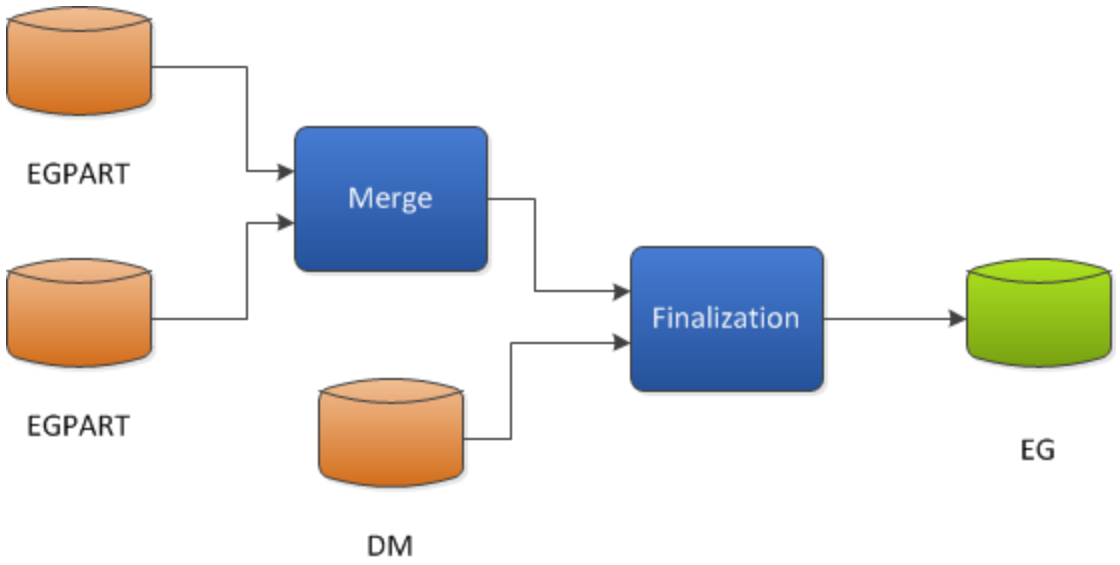
SDTM Generation



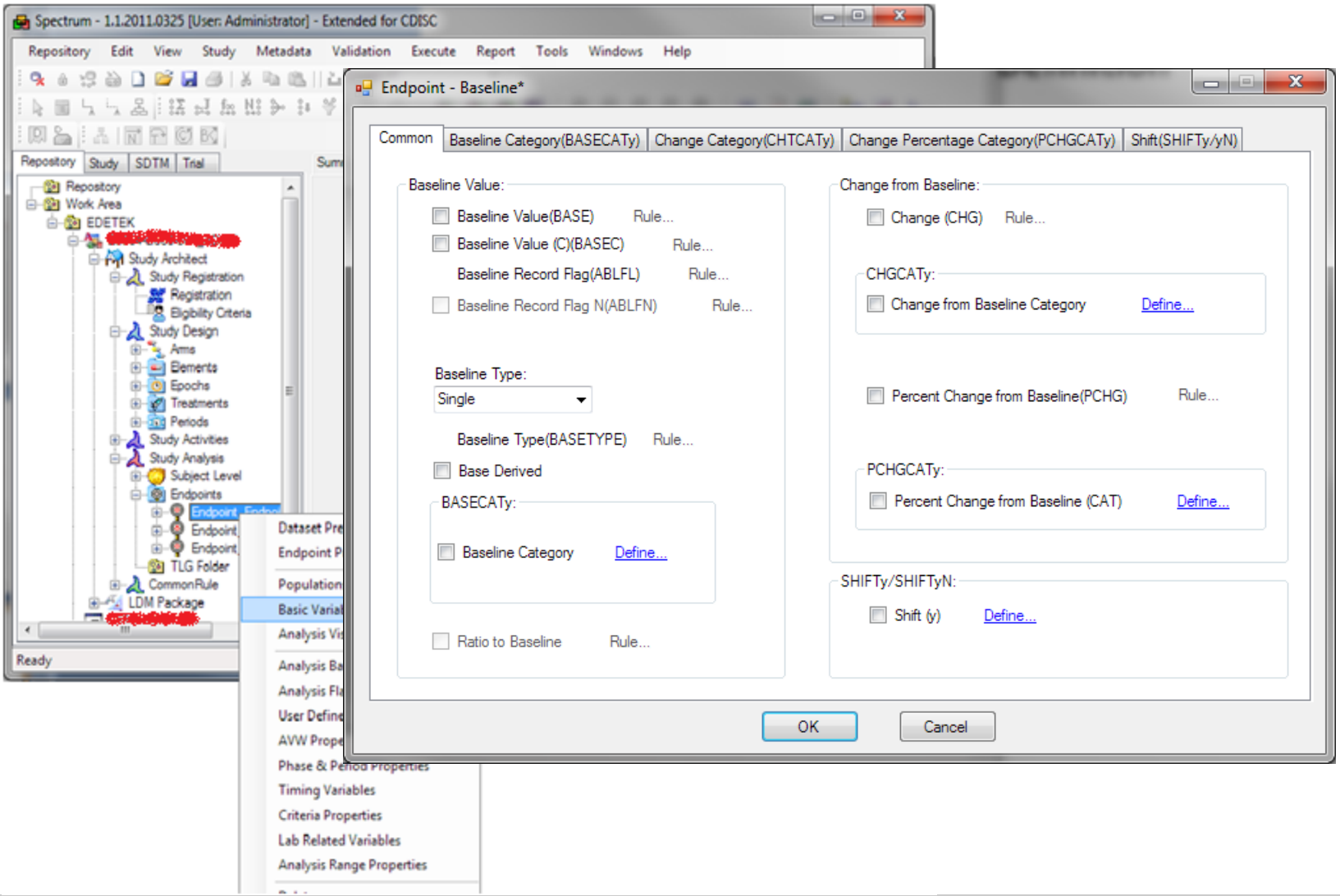
Partition Level Templates



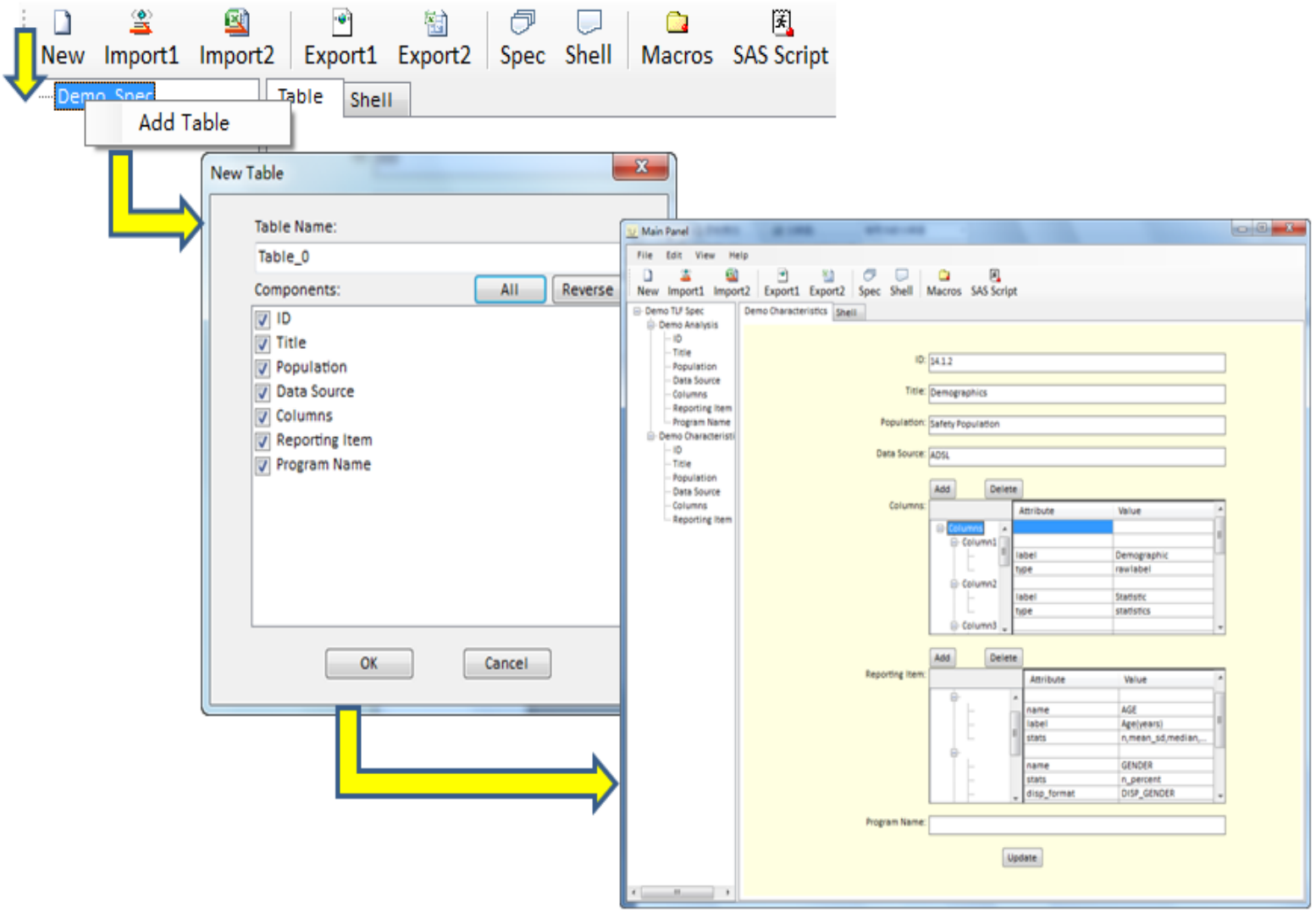
Domain Level Templates



STEP 3 : STUDY DESIGN - ANALYSIS



STUDY 3 : STUDY DESIGN - REPORTING



TLF SHELL IN SA

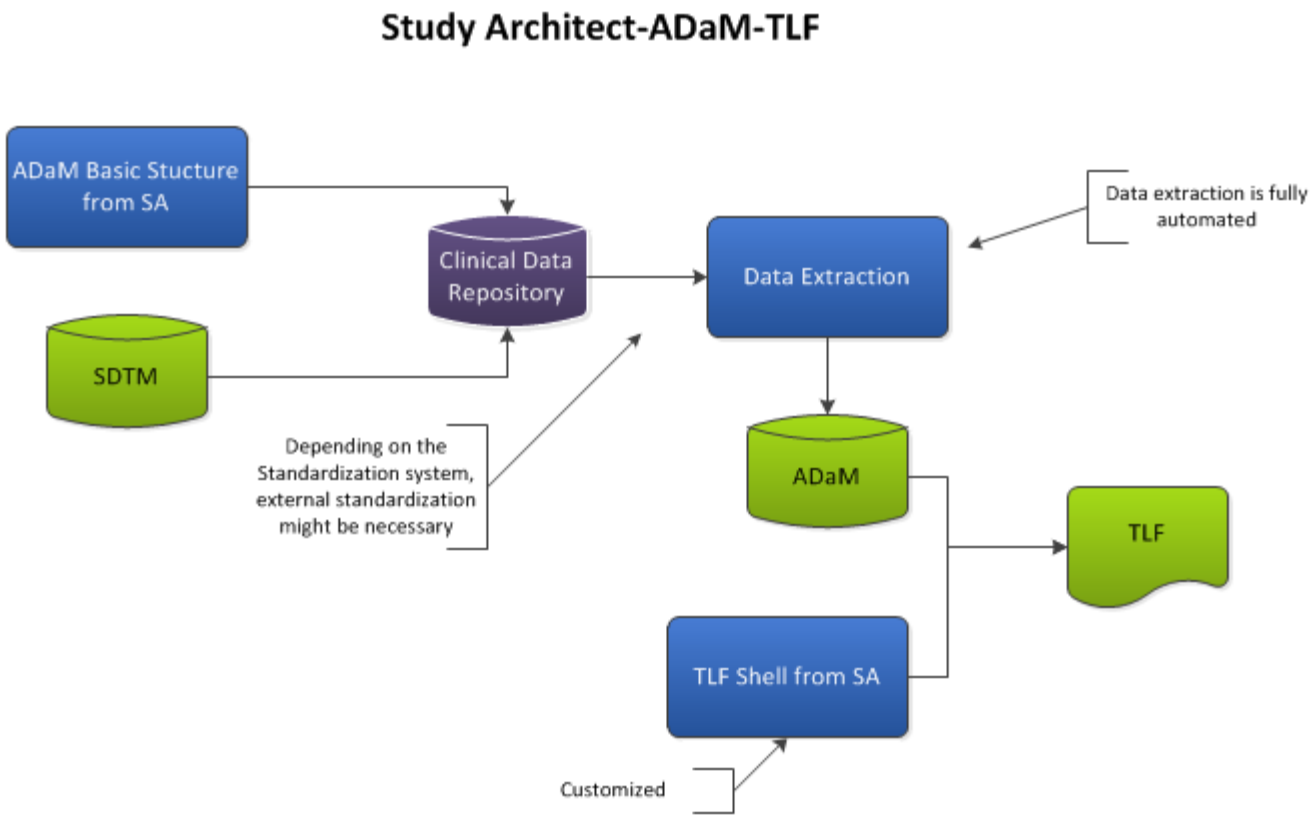
Table 14.1.2

Demographics

Safety Population

Demographic	GROUP A	GROUP B	GROUP C
Age (years)			
n	xxx	xxx	xxx
Mean (Sd)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
Min, Max	xx, xx	xx, xx	xx, xx
Median	xx	xx	xx
Gender			
Female	xx(x.xx)	xx(x.xx)	xx(x.xx)
Male	xx(x.xx)	xx(x.xx)	xx(x.xx)
Race			
American Indian or A	xx(x.xx)	xx(x.xx)	xx(x.xx)
Black	xx(x.xx)	xx(x.xx)	xx(x.xx)
White	xx(x.xx)	xx(x.xx)	xx(x.xx)
Asian	xx(x.xx)	xx(x.xx)	xx(x.xx)
Native Hawaiian or o	xx(x.xx)	xx(x.xx)	xx(x.xx)

AUTOMATED ADAM & TLF CREATION



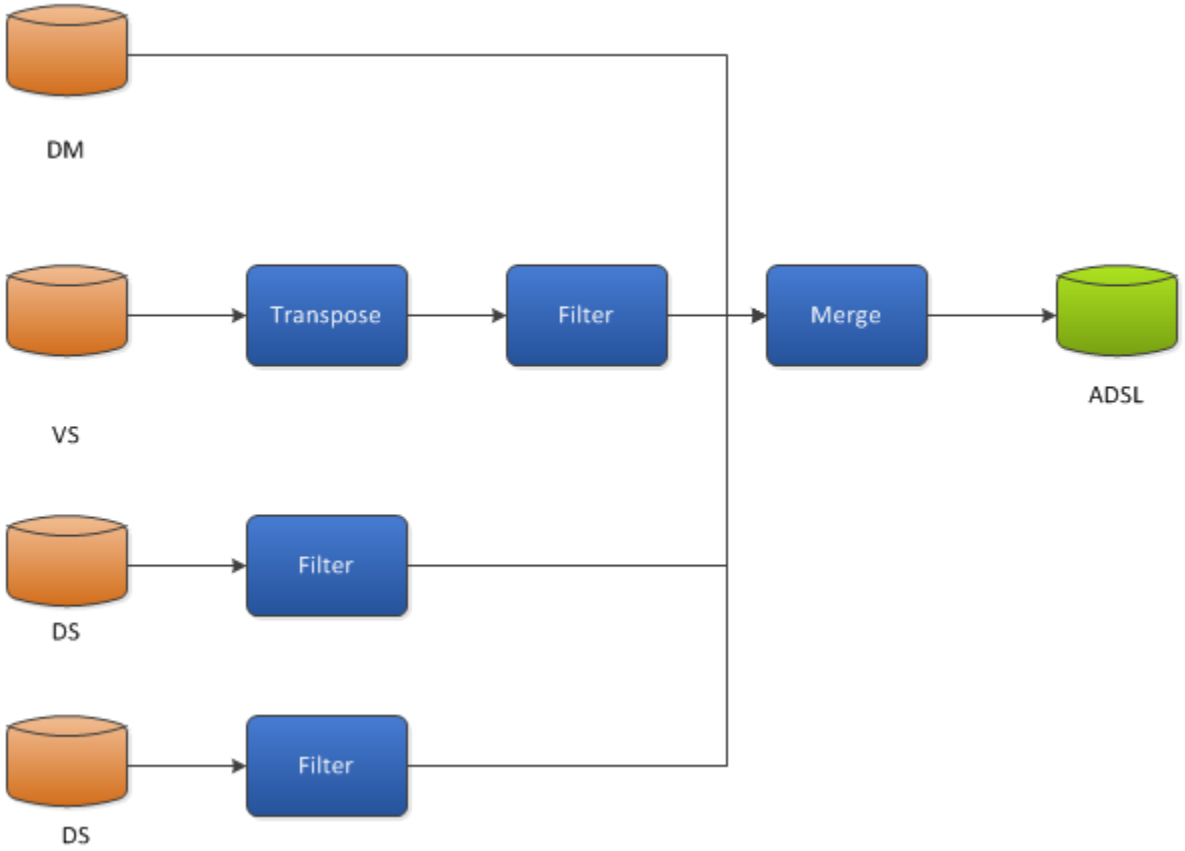
After the clinical data repository which contains rule template as the SDTM repository in page 14-16, ADaM datasets are generated finally.

Using ADaM Structure to Check SDTM Completeness

- **ADaM structure is generated to support TLF creation**
  - Analysis parameter variables
  - Analysis group variables
  - Analysis population variables
  - Analysis visit windowing variables
  - Analysis flag variables
  - Analysis enabling variables
- **Traceability to SDTM is established through ADaM templates**
  - Derivations
  - Macros – with metadata wrapper
- **Traceability to data collection is established through SDTM templates**

Data Flow in Repository

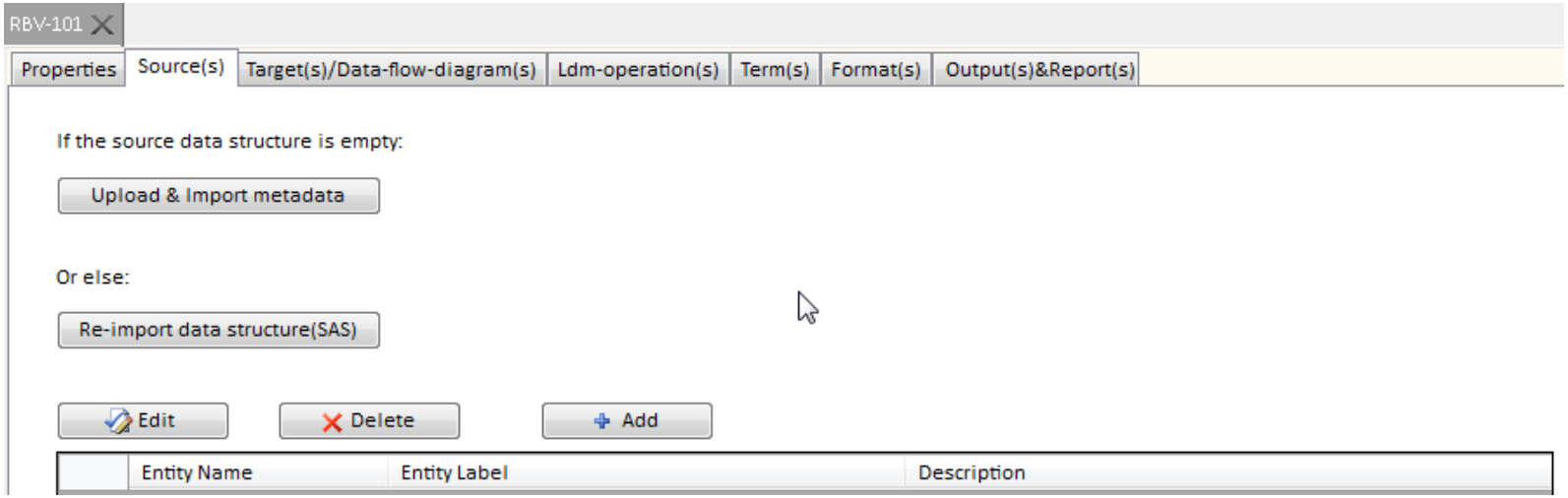
- **ADSL**



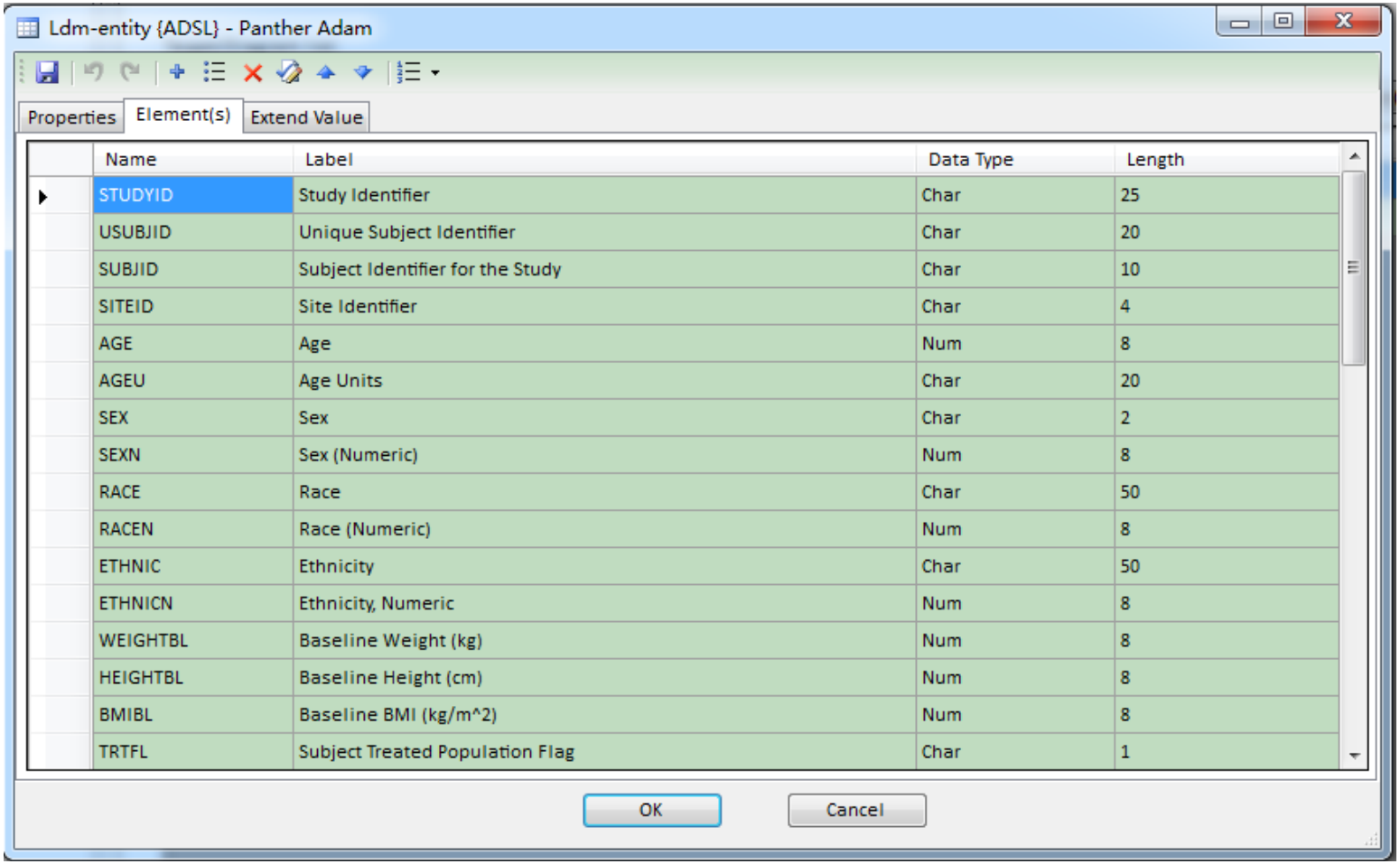
ADaM Mapping

- **ADaM data obtained from Combining SA ADaM Structure and Repository**
  - Create different templates firstly. Templates are grouped by dataset.
  - Import SA ADaM structure.
  - Select one or more template for one dataset by turns.
  - Deal with promptions during template reference.
  - Process variables and dataset without definition in template.
  - Generate ADaM data finally.

IMPORT SDTM DATA

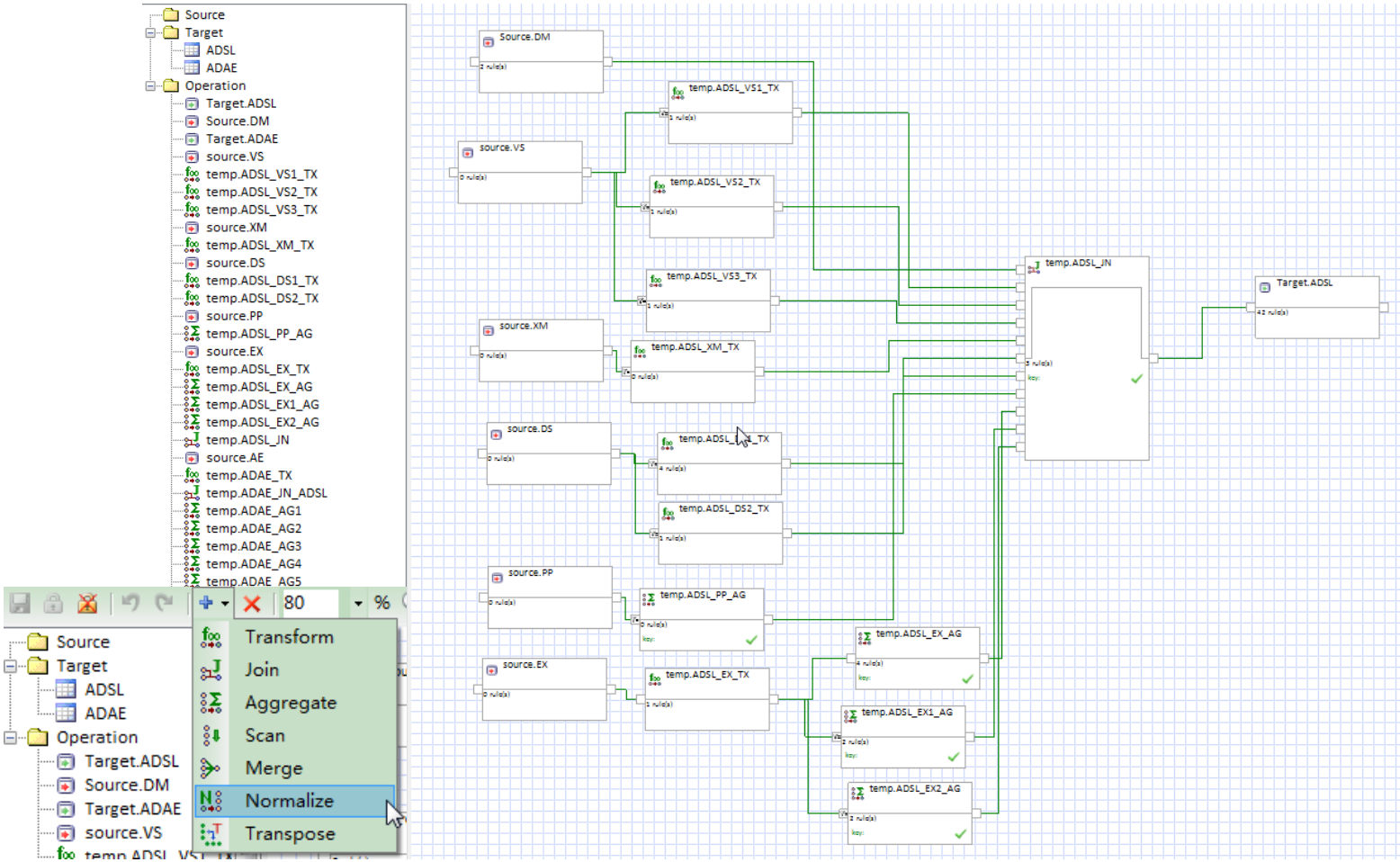


Data Structure From SA

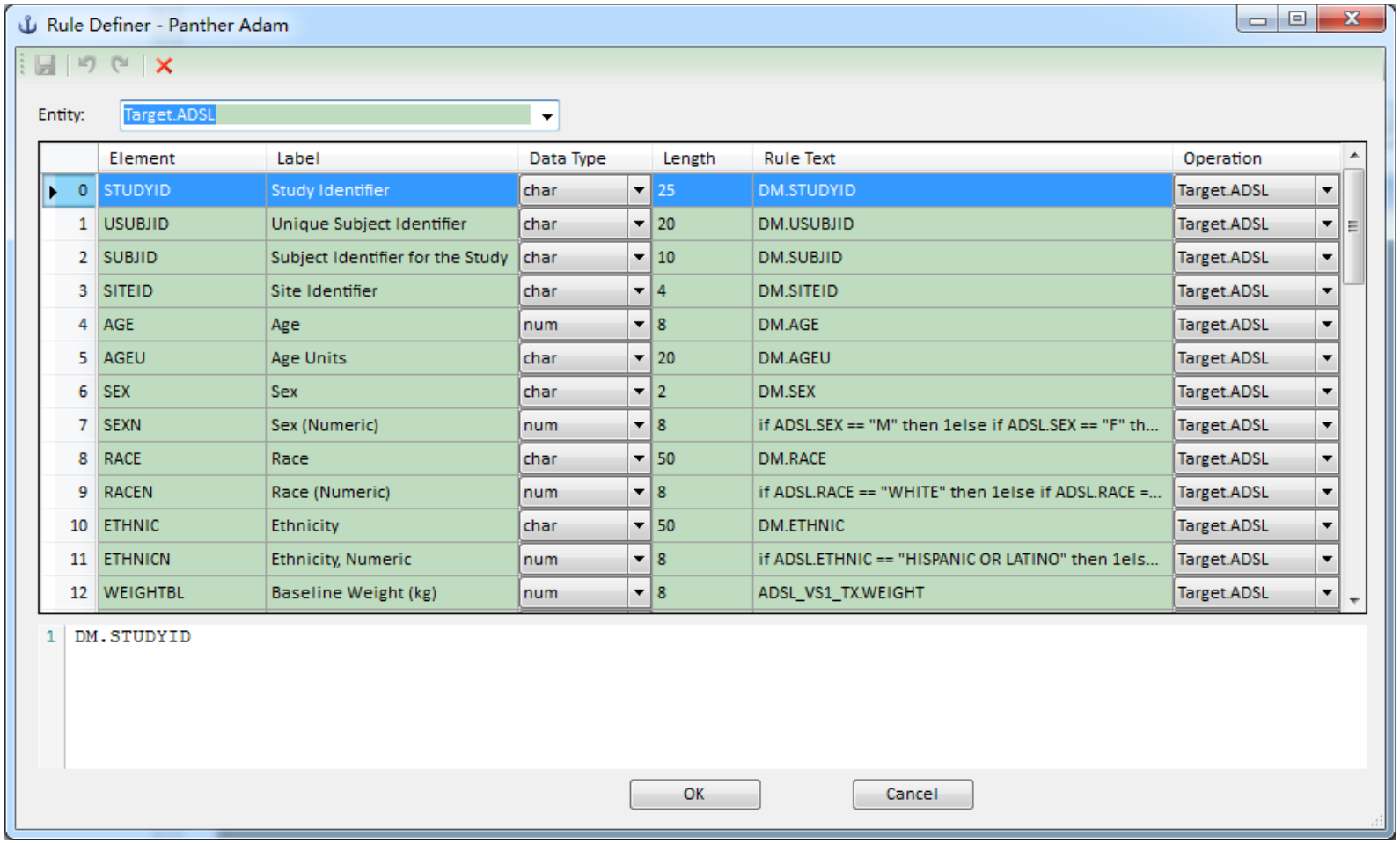


Name	Label	Data Type	Length
STUDYID	Study Identifier	Char	25
USUBJID	Unique Subject Identifier	Char	20
SUBJID	Subject Identifier for the Study	Char	10
SITEID	Site Identifier	Char	4
AGE	Age	Num	8
AGEU	Age Units	Char	20
SEX	Sex	Char	2
SEXN	Sex (Numeric)	Num	8
RACE	Race	Char	50
RACEN	Race (Numeric)	Num	8
ETHNIC	Ethnicity	Char	50
ETHNICN	Ethnicity, Numeric	Num	8
WEIGHTBL	Baseline Weight (kg)	Num	8
HEIGHTBL	Baseline Height (cm)	Num	8
BMIBL	Baseline BMI (kg/m^2)	Num	8
TRTFL	Subject Treated Population Flag	Char	1

Data Flow Diagram



Conversion Rule



TLF Report Obtained

Table 14.1.2
Demographics
Safety Population

Demographic	Statistic	KD025 (N=8)	Total (N=8)
Age (years)	N	8	8
	Mean (SD)	31.0 (8.88)	31.0 (8.88)
	Median	29.5	29.5
	Min, Max	19.0, 46.0	19.0, 46.0
Gender			
Male	n (%)	7 (87.50%)	7 (87.50%)
Female	n (%)	1 (12.50%)	1 (12.50%)
Race			
American Indian or Alaska Native	n (%)	0 (0.00%)	0 (0.00%)
Asian	n (%)	0 (0.00%)	0 (0.00%)
Black	n (%)	4 (50.00%)	4 (50.00%)
Native Hawaiian or other Pacific Islander	n (%)	0 (0.00%)	0 (0.00%)
White	n (%)	4 (50.00%)	4 (50.00%)
Other	n (%)	0 (0.00%)	0 (0.00%)
Ethnicity			
Hispanic or Latino	n (%)	1 (12.50%)	1 (12.50%)
Not Hispanic or Latino	n (%)	7 (87.50%)	7 (87.50%)
Weight (kg)	N	8	8
	Mean (SD)	78.8 (12.84)	78.8 (12.84)
	Median	80.6	80.6
	Min, Max	58.2, 96.4	58.2, 96.4